



Universidad de Extremadura

Escuela Politécnica

MÁSTER EN DIRECCIÓN DE EMPRESAS TIC

Trabajo de Fin de Máster

“Business Intelligence applied to tourism”

Manuel Domingo Ramos Barro

**Septiembre, 2015**





Universidad de Extremadura

Escuela Politécnica

**MÁSTER EN DIRECCIÓN DE EMPRESAS TIC**

Trabajo de Fin de Máster

**“Business Intelligence applied to tourism”**

Autor/es: Manuel Domingo Ramos Barro  
Fdo:

Director/ es: Pablo García Rodríguez y  
José Manuel Hernández Mogollón  
Fdo:

**Tribunal Calificador:**  
Presidente: Manuel Barrena García

Fdo:

Secretario: Andrés Caro Lindo

Fdo:

Vocal: María Luisa Durán Martín Merás

Fdo

CALIFICACIÓN:

FECHA:





*Dedico este proyecto a toda mi familia, a mi novia, a mis amigos, a todos los que siempre estáis cuando os necesito, a los que nunca estáis demasiado ocupados, a los que me habéis hecho como soy, porque sin vosotros nada de esto tiene sentido.*

*Gracias de corazón.*



## Agradecimientos

Este proyecto me ha llevado más de un año de duro trabajo y no es posible dar las gracias en unas pocas líneas a todas las personas que, sin tener obligación de hacerlo, me han ayudado durante este tiempo.

En primer lugar, por supuesto, dar las gracias a mis tutores, Pablo García Rodríguez y José Manuel Hernández Mogollón, por los consejos y la orientación que me han dado a lo largo de este proyecto.

Como instituciones, quiero dar las gracias INSA, que me dio la oportunidad de trabajar y conocer el mundo de Business Intelligence, y a EscapadaRural.com por proporcionarme los datos necesarios para poder realizar el estudio que compone este proyecto.

Muy especialmente, quiero dar las gracias a Alberto Cabanillas Rodríguez, Business Analytics Manager de INSA, por ayudarme a entender a este monstruo que es el Business Intelligence, a Sergio Vasco Hernández, gran compañero en INSA y mejor amigo, que siempre supo encontrar una solución a todo donde no la había y, sobretodo, a mi novia, Anabel Ledo Pérez, porque además de la paciencia que ha tenido conmigo y con este proyecto, es quién me ha enseñado el camino para realizar el estudio de este proyecto.

No me puedo olvidar de mi familia que, siempre ha estado a mi lado, de mis amigos que siempre me han animado, de mis compañeros que tanto me han enseñado y de todos aquellos que sin tener por qué han dedicado un solo minuto de su tiempo a ayudarme cuando más perdido estaba.

A todos vosotros y por todos vosotros, hoy culmino una etapa de mi vida.



## INSA



Ingeniería del Software Avanzado, S. A. (INSA), es una empresa española fundada en 1991 por IBM España y Catalana Occidente. El objetivo inicial era aprovechar la experiencia del equipo de desarrollo interno de la aseguradora sobre plataformas IBM para comercializar soluciones conjuntas para los sectores económicos (banca, seguros,...) y de la administración pública. Su objeto social es el diseño, desarrollo, gestión y explotación de sistemas de información basados en la tecnología de la información, así como la comercialización de productos y servicios informáticos. Presta también servicios de consultoría, y tiene unos 2.000 empleados. Cuenta con certificado ISO 9001.

Está inscrita en el Registro Mercantil de Madrid, Tomo 3238, Libro 0, Folio 78, Sección 8, Hoja M-55112, y su sede social se encuentra en la Avenida de Burgos 8A. Edificio Bronce de Madrid. Cuenta con delegaciones en Barcelona, Valencia, Guecho, Sevilla y Lisboa y tiene centro de desarrollo tecnológico en Cáceres, Salamanca, Almería, Reus y Ourense. Es parte del grupo mercantil IBM España.

Ofrece servicios de desarrollo y mantenimiento de Software, ERP, CRM, desarrollos sobre la plataforma IBM iSeries, e-Business, Business Analytics, y servicios bancarios y tributarios.



## ESCAPADA RURAL



EscapadaRural es una empresa española con sede en Barcelona fundada en el año 2007. La empresa nació como un proyecto de tiempo libre de sus dos directores de marketing y a día de hoy cuenta con una plantilla de 25 trabajadores y presencia en varios países del litoral mediterráneo europeo. Actualmente se ha situado como líder en número de casas rurales promocionadas en España con casi 14.000 y superando ostensiblemente el medio millón de usuarios.

Defienden su compromiso con la calidad en los servicios ofrecidos, la cercanía al cliente, la honestidad en la comunicación, la colaboración con los agentes del sector y una lucha por mejorar constantemente.

La empresa posee un observatorio de turismo rural, se trata de una iniciativa de investigación conjunta desarrollada por EscapadaRural.com, la EUHT CETT-UB y Netquest que genera conocimiento y aporta información de valor sobre el sector del turismo rural en España, tanto en su ámbito de la oferta (“propietarios”) como en el de la demanda (“viajeros”). A nivel de cifras, este Big Data dispone de un tamaño de muestra de 10.219 viajeros y un total de 2.275 propietarios, con un nivel de confianza del 95% y un error muestral del 1.0% para viajeros y del 1.9% para los propietarios.





## Resumen

Este proyecto versa sobre la materia del Business Intelligence, un conjunto de herramientas que buscan servir de soporte a la hora de tomar decisiones. Se analizan cuáles son sus ventajas y cómo puede ayudar al desarrollo de un negocio u organización ayudándoles a aprovechar sus recursos y capacidades y a comprender la situación que posee en el mundo exterior a fin de lograr ventaja competitiva.

Este proceso se realizará a través del estudio, análisis y predicción de la disciplina estadística conocida como "Series cronológicas" o temporales. Se expondrán los fundamentos, características y métodos de análisis de éstas desde un punto de vista puramente teórico tratando de no caer en explicaciones y exposiciones que excedan los límites de este trabajo.

Los conceptos expuestos se pondrán en práctica mediante el estudio de un caso de aplicación práctica que consistirá en el análisis de las pernoctaciones mensuales en establecimientos catalogados como "establecimiento rural" a nivel nacional por provincias desde enero de dos mil diez hasta mayo de dos mil quince. El objetivo consistirá en extraer aquellas que presenten mayor interés de inversión de las cuáles se analizará el perfil del visitante que recibe a fin de concretar el tipo de oferta que se debería poner en juego para cada territorio.

El estudio se realizará utilizando varias herramientas software de manejo y coste asequible a fin de tratar de demostrar si es posible implantar un sistema de *Business Intelligence* alejándose de las grandes suites comerciales. De estas herramientas se compararán tanto sus resultados como la experiencia de trabajo con cada una a fin de ofrecer una valoración sobre su idoneidad para este propósito. Finalmente se expondrán las conclusiones y posibles líneas futuras que pudiesen ser desarrolladas a raíz de este trabajo.

## Abstract

This project concerns the field of Business Intelligence, a set of tools that seek to provide support when making decisions. We analyze what are its advantages and how it can help the development of a business or organization helping them to leverage their resources and capabilities and to understand the situation that has in the outside world in order to achieve competitive advantage.

This will be done through study, analysis and prediction of statistics or temporary discipline known as "Time series". The fundamentals, characteristics and methods of analysis of these will be presented from a purely theoretical point of view, trying not to fall into explanations and exposures exceeding the limits of this work.

In order to implement the concepts presented a case of practical application will be made. This study will consist in analyzing monthly overnight in places classified as "rural setting" for provinces nationwide since January two thousand ten until May two thousand and fifteen establishments. The objective will be extract the provinces involving an increased investment interest. On these provinces, the visitor profile will be analyzed to set the type of tourism that should to offer each province.

The study will be carried out using various affordable software tools in order to try to test whether it is possible to implement a Business Intelligence system away from large commercial suites. Results obtained from these tools will be compared as well as the experience of working with each to provide an assessment of their suitability for this purpose. Finally, conclusions and possible future lines that could be developed as a result of this work will be presented.



## Índice

<b>Agradecimientos .....</b>	<b>4</b>
<b>INSA.....</b>	<b>6</b>
<b>ESCAPADA RURAL.....</b>	<b>8</b>
<b>Resumen      Abstract.....</b>	<b>10</b>
<b>Figuras .....</b>	<b>16</b>
<b>Tablas .....</b>	<b>20</b>
<b>1.-Introduction .....</b>	<b>22</b>
1.1.- Contextualization.....	22
1.2.- Relationship among Business Intelligence, ERP and CRM.....	22
1.3. - State of art. ....	23
1.4 Report of motivations and goals.....	27
<b>2.- Business Intelligence (BI) .....</b>	<b>30</b>
2.1.- ¿Qué es Business Intelligence? .....	30
2.2.- ¿De qué se compone BI? .....	31
2.2.1.- Data Sources. ....	33
2.2.2.- Data Analyst.....	34
2.2.2.1.- Data Warehouse .....	34
2.2.2.2. - Data Exploration–Data Mining. ....	36
2.2.3.-Business Analytics.....	38
2.2.4.- Decision Making.....	45
2.3.- Conclusión .....	45
2.4.- Business Intelligence vs Big Data .....	46
<b>3.-Series Temporales .....</b>	<b>48</b>
3.1.- ¿Qué son las series temporales?.....	48
3.2.- ¿De qué se componen las series temporales?.....	49
3.2.- ¿Cómo se clasifican las series temporales? .....	51
3.3.-Análisis de las Series Temporales. ....	54
3.3.1.- Análisis preliminar de una serie temporal.....	55
3.3.2.- Métodos Clásicos de estudio. ....	59
3.3.4- Modelos ARIMA.....	70

<b>4.- Caso de aplicación práctica. Presentación.....</b>	<b>74</b>
<b>4.1.- Turismo Rural en España. ....</b>	<b>74</b>
<b>4.2.- Presentación de las herramientas de trabajo.....</b>	<b>77</b>
4.2.1.- SPSS.....	77
4.2.2.- Microsoft Excel. ....	78
4.2.3.- Weka. ....	79
4.2.4.- Tableau ....	80
4.2.5.- R Programming .....	81
4.2.6.- ForecastBI .....	82
<b>4.3.- Aplicaciones comerciales.....</b>	<b>83</b>
4.3.1.- SAP Business suite. ....	83
4.3.2.- IBM Cognos Business Intelligence. ....	85
4.3.3.- Pentaho.....	87
<b>5.- Caso práctico. Realización. ....</b>	<b>90</b>
<b>5.2.- Herramienta de desarrollo propio. Forecast BI. ....</b>	<b>93</b>
<b>5.2.- Weka. ....</b>	<b>97</b>
<b>5.3.- Excel ....</b>	<b>101</b>
<b>5.4.- SPSS .....</b>	<b>105</b>
<b>5.5.- R programming .....</b>	<b>109</b>
<b>5.6.- Tableau .....</b>	<b>113</b>
<b>5.7.- Resumen agrupado del estudio. ....</b>	<b>117</b>
<b>5.8.- Provincia número 1. Cantabria .....</b>	<b>119</b>
<b>5.9.- Provincia Número 2. Madrid .....</b>	<b>122</b>
<b>5.10.- Provincia número 3. Baleares .....</b>	<b>125</b>
<b>5.11.- Provincia Número 4. Cáceres .....</b>	<b>128</b>
<b>5.12 Valoración del estudio realizado.....</b>	<b>131</b>
<b>5.13.- Comparativa de las Herramientas utilizadas. ....</b>	<b>133</b>
<b>6.- Debate. Business Intelligence, justificación de la inversión. ....</b>	<b>136</b>
<b>7.- Conclusions and Future Aims .....</b>	<b>138</b>
<b>7.1.- Job Conclusions. ....</b>	<b>138</b>
<b>7.2.- Future Aims.....</b>	<b>140</b>

<b>Anexos .....</b>	<b>142</b>
<b>Anexo I Diagramas de la aplicación Forecast BI.....</b>	<b>143</b>
AI1.- Diagrama de Clases. ....	143
AI2.- Diagrama de casos de uso. ....	144
<b>A II. - Study of time series using software tools. ....</b>	<b>145</b>
AI11.- Study of time series using SPSS. ....	145
AI12.- Time series study using Excel. ....	157
AI13. - Time Series Study using R.....	170
AI14.- Time Series study using Tableau .....	176
AI15.- Time series study using WEKA.....	185
<b>AIII.- Datos de estudio. Series temporales de pernoctaciones en hoteles rurales por provincias.....</b>	<b>190</b>
Álava.....	190
Albacete.....	190
Alicante .....	190
Almería .....	190
Asturias.....	191
Ávila .....	191
Badajoz .....	191
Baleares.....	191
Barcelona.....	192
Bizkaia .....	192
Burgos .....	192
Cáceres .....	192
Cádiz.....	193
Cantabria .....	193
Castellón.....	193
Ciudad Real.....	193
Córdoba .....	194
Coruña .....	194
Cuenca .....	194

<b>Guipúzcoa .....</b>	<b>194</b>
<b>Girona.....</b>	<b>195</b>
<b>Granada .....</b>	<b>195</b>
<b>Guadalajara .....</b>	<b>195</b>
<b>Huelva .....</b>	<b>195</b>
<b>Huesca .....</b>	<b>196</b>
<b>Jaén .....</b>	<b>196</b>
<b>La Rioja .....</b>	<b>196</b>
<b>Las Palmas .....</b>	<b>196</b>
<b>León.....</b>	<b>197</b>
<b>Lleida .....</b>	<b>197</b>
<b>Lugo.....</b>	<b>197</b>
<b>Madrid.....</b>	<b>197</b>
<b>Málaga.....</b>	<b>198</b>
<b>Murcia .....</b>	<b>198</b>
<b>Navarra.....</b>	<b>198</b>
<b>Ourense .....</b>	<b>198</b>
<b>Palencia .....</b>	<b>199</b>
<b>Pontevedra .....</b>	<b>199</b>
<b>Salamanca .....</b>	<b>199</b>
<b>Santa Cruz.....</b>	<b>199</b>
<b>Segovia .....</b>	<b>200</b>
<b>Sevilla .....</b>	<b>200</b>
<b>Soria .....</b>	<b>200</b>
<b>Tarragona .....</b>	<b>200</b>
<b>Teruel .....</b>	<b>201</b>
<b>Toledo.....</b>	<b>201</b>
<b>Valencia .....</b>	<b>201</b>
<b>Valladolid.....</b>	<b>201</b>
<b>Zamora .....</b>	<b>202</b>
<b>Zaragoza .....</b>	<b>202</b>
<b>Bibliografía.....</b>	<b>204</b>

## Figuras

Figura 1 BI: Datos - Información - Conocimiento.....	30
Figura 2 BI: Decisión - Acción - Información .....	30
Figura 3.- Pirámide Business Intelligence .....	32
Figura 4.- BI vs BA.....	38
Figura 5 Volumen del flujo anual del río Nilo en Asuán 1871 - 1970 (nilo.wf1).....	49
Figura 6.- Paro en España 97-02 .....	50
Figura 7.- Tendencia Paro en España 97 - 02 .....	50
Figura 8 .- Estacionalidad del Paro en España 97 - 02 .....	51
Figura 9 .- Componente Aleatoria del Paro en España 97 - 02 .....	51
Figura 10.- Concentración CO2 97-10.....	52
Figura 11.- Ruido Blanco.....	53
Figura 12.- Serie No Estacionaria.....	54
Figura 13.- Pasajero aerolínea 49 - 60. ....	55
Figura 14.- Nivel de la serie aerolínea 49-60. ....	56
Figura 15.- Diagrama de cajas por mes de la aerolínea. ....	56
Figura 16.- Representación de dos series con distintos esquema, aditivo o multiplicativo.....	57
Figura 17.- Gráfico de dispersión por nivel de la serie de la aerolínea. ....	57
Figura 18.- Cambio de tendencia en la serie del paro en USA. ....	58
Figura 19.- Histograma y diagrama de cajas del paro femenino en USA entre 1961 y 1985. ....	58
Figura 20.- Volumen de flete por aviones en USA entre 69-80. ....	59
Figura 21.- Muestra de la representación de una serie dividida. ....	59
Figura 22.- Media móvil centrada de una serie con información diaria (periodo 7).....	60
Figura 23.- Posibles medias móviles centradas en una serie con información trimestral. ....	60
Figura 24.- Diagrama de secuencia y cajas de las ventas de cava en España 62-69. ....	61
Figura 25.- Diagrama de cajas y gráfico de dispersión por nivel de la serie de venta de cava. ....	62
Figura 26.- Gráfico de secuencia de la serie original y la serie desestacionalizada.....	62
Figura 27.- Tipos de rectas de ajuste de tendencia. ....	63
Figura 28.- Gráfico de secuencia de la componente tendencia-ciclo de la serie de venta de cava en España 62-69. ....	63
Figura 29.- Gráfico de la componente TC con los ajustes lineal, logarítmico y S. ....	64
Figura 30.- Gráfico de secuencia de la serie cava original junto con las predicciones para un año. ....	64
Figura 31.- Gráfico de secuencia de la serie venta de Combustible en España 90-06 con la recta de ajuste lineal.....	65
Figura 32.- Gráfico de secuencia de la serie de venta de Combustible con el ajuste efectuado considerando únicamente el dato anterior. ....	66
Figura 33.- Gráfico de secuencia de la serie de venta de Combustible con el ajuste efectuado considerando un modelo de medias móviles anteriores de amplitud 5. ....	67
Figura 34.- Gráfico de secuencia de la serie Combustible con el ajuste efectuado considerando un modelo exponencial simple (serie suavizada) y la previsión a varios meses vista. ....	68
Figura 35.- Gráfico de secuencia de la serie suavizado Holt con el ajuste efectuado considerando un modelo exponencial doble (serie suavizada) y la precisión a varios meses vista. ....	69
Figura 36.- Gráfico de la serie Papel con el ajuste efectuado considerando un modelo de Holt-Winters aditivo y la previsión a un año vista. ....	70

Figura 37.- Distribución del turismo rural por CC.AA. en España año 2013 .....	74
Figura 38.- Pernoctaciones en España turismo rural 01-13 .....	75
Figura 39.- Pernoctaciones por provincias 01/10 - 05/15.....	91
Figura 40.- FBI antes de la realización del estudio.....	93
Figura 41.- FBI después de la realización del estudio. ....	94
Figura 42.- WEKA antes de la realización del estudio .....	97
Figura 43.- WEKA después de la realización del estudio. ....	98
Figura 44.- Estado de Excel antes de realizar el estudio. ....	101
Figura 45.- Estado Excel después de realizar el estudio .....	102
Figura 46.- Estado SPSS antes de estudio. ....	105
Figura 47.- Estado SPSS después de la realización del estudio. ....	106
Figura 48.- Estado R antes de la realización del estudio.....	109
Figura 49.- Estado R después de la realización del estudio. ....	110
Figura 50.- Estado de Tableau antes del estudio .....	113
Figura 51.- Estado Tableau después del estudio. ....	114
Figura 52.- Provincias de mayor interés de inversión.....	118
Figura 53.- Distribución de pernoctaciones en provincia 1 por meses. ....	119
Figura 54.- Acompañantes provincia 1. ....	119
Figura 55.- Actividades realizadas provincia 1.....	120
Figura 56.- Reclamo provincia 1. ....	121
Figura 57.- Distribución de pernoctaciones en provincia 2 por meses. ....	122
Figura 58.- Acompañantes provincia 2. ....	122
Figura 59.- Actividades provincia 2.....	123
Figura 60.- Reclamo provincia 2. ....	124
Figura 61.- Distribución de pernoctaciones en provincia 3 por meses .....	125
Figura 62.- Acompañantes provincia 3. ....	125
Figura 63.- Actividades provincia 3.....	126
Figura 64.- Reclamo provincia 3. ....	127
Figura 65.- Distribución de pernoctaciones en provincia 4 por meses .....	128
Figura 66.- Acompañantes provincia 4. ....	128
Figura 67.- Actividades realizadas provincia 4.....	129
Figura 68.- Reclamo provincia 4. ....	130
Figura 69.- Gráfica del coste de suites BI para empresas pequeñas, medianas y grandes. ....	136
Figura 70.- Valor vs Tiempo.....	137
Figura 71.- Valor = Información / Tiempo .....	137
Figura 72.- SPSS Image 1 .....	145
Figura 73.- SPSS image 2 .....	146
Figura 74.- SPSS image 3 .....	146
Figura 75.- SPSS image 4 .....	146
Figura 76.- SPSS image 5 .....	147
Figura 77.- SPSS Image 6 .....	148
Figura 78.- SPSS Image 7 .....	148
Figura 79.- SPSS Image 8 .....	149
Figura 80.- SPSS Image 9. ....	149
Figura 81.- SPSS Image 10 .....	150



Figura 82.- SPSS Image 11 .....	151
Figura 83.- SPSS image 12 .....	152
Figura 84.- SPSS image 13 .....	152
Figura 85.- SPSS Image 14 .....	153
Figura 86.- SPSS Image 15 .....	154
Figura 87.- SPSS Image 16 .....	155
Figura 88.- SPSS Image 17 .....	155
Figura 89.- SPSS Image 18 .....	155
Figura 90.- SPSS Image 19 .....	156
Figura 91.- SPSS Image 20 .....	156
Figura 92.- Excel Image 1.....	157
Figura 93.- Excel Image 2.....	157
Figura 94.- Excel Image 3.....	158
Figura 95.- Excel Image 4.....	158
Figura 96.- Excel Image 5.....	159
Figura 97.- Excel Image 6.....	159
Figura 98.- Excel Image 7.....	159
Figura 99.- Excel Image 8.....	160
Figura 100.- Excel Image 9.....	160
Figura 101.- Excel Image 10.....	161
Figura 102.- Excel Image 11.....	162
Figura 103.- Excel Image 12.....	162
Figura 104.- Excel Image 13.....	163
Figura 105.- Excel Image 14.....	163
Figura 106.- Excel Image 15.....	163
Figura 107.- Excel Image 16.....	164
Figura 108.- Excel Image 17.....	164
Figura 109.- Excel Image 18.....	165
Figura 110.- Excel Image 19.....	165
Figura 111.- Excel Image 20.....	166
Figura 112.- Excel Image 21.....	167
Figura 113.- Excel Image 22.....	168
Figura 114.- Excel Image 23.....	169
Figura 115.- R Image 1.....	170
Figura 116.- R Image 2.....	170
Figura 117.- R Image 3.....	172
Figura 118.- R Image 4.....	172
Figura 119.- R Image 5.....	173
Figura 120.- R Image 6.....	173
Figura 121.- R Image 7.....	174
Figura 122.- R Image 8.....	174
Figura 123.- R Image 9.....	175
Figura 124.- R Image 10.....	175
Figura 125.- Tableau Image 1 .....	176
Figura 126.- Tableau Image 2 .....	176

Figura 127.- Tableau Image 3 .....	177
Figura 128.- Tableau Image 4 .....	177
Figura 129.- Tableau Image 5 .....	177
Figura 130.- Tableau Image 6 .....	178
Figura 131.- Tableau Image 7 .....	178
Figura 132.- Tableau Image 8 .....	179
Figura 133.- Tableau Image 9 .....	179
Figura 134.- Tableau Image 10 .....	180
Figura 135.- Tableau Image 11 .....	180
Figura 136.- Tableau Image 12 .....	181
Figura 137.- Tableau Image 13 .....	181
Figura 138.- Tableau Image 14 .....	182
Figura 139.- Tableau Image 15 .....	183
Figura 140.- Tableau Image 16 .....	183
Figura 141.- Tableau Image 17 .....	184
Figura 142.- Weka Image 1 .....	185
Figura 143.- Weka Image 2 .....	185
Figura 144.- Weka Image 3 .....	186
Figura 145.- Weka Image 4 .....	186
Figura 146.- Weka Image 5 .....	187
Figura 147.- Weka Image 6 .....	188
Figura 148.- Weka Image 7 .....	188

## Tablas

Tabla 1.- Serie temporal pernотaciones en alojamientos rurales en Badajoz 01/10 - 05/15. ....	91
Tabla 2.- Serie temporal pernотaciones en alojamientos rurales en Cáceres 01/10 - 05/15. ....	91
Tabla 3.- Relación del número de establecimientos y sus plazas estimadas de las provincias seleccionadas. .	92
Tabla 4.- Resumen resultados del estudio FBI. ....	94
Tabla 5.- Resultados de estudio FBI. ....	95
Tabla 6.- Resumen resultados del estudio WEKA. ....	98
Tabla 7.- Resultados de estudio WEKA. ....	99
Tabla 8.- Resumen resultados del estudio Excel. ....	102
Tabla 9.- Resultados de estudio Excel. ....	103
Tabla 10.- Resumen resultados del estudio SPSS. ....	106
Tabla 11.- Resultados estudio SPSS. ....	107
Tabla 12.- Resumen de resultados del estudio R. ....	110
Tabla 13.- Resultados del estudio R. ....	111
Tabla 14.- Resumen resultados del estudio Tableau. ....	114
Tabla 15.- Resultado estudio Tableau. ....	115
Tabla 16.- Resumen del estudio. ....	117
Tabla 17.- Pernотaciones esperadas en provincias de mayor interés de inversión. ....	118
Tabla 18.- Resumen del estudio realizado con las aplicaciones. ....	133



## 1.-Introduction

The aim of this first chapter is present the concepts that will be discussed throughout this document. A vision over ERP, CRM and BI will be offered. That vision will express the relationship and differences among Resource Planning Systems (ERP), Control Systems Customer Relationship (CRM) and Business Intelligence solutions (BI). This will be complemented with presentation of the state of art of the latter. Finally the report scope and objectives of the project will be presented.

### 1.1.- Contextualization.

Over the last few decades it is going through a period of extreme competitiveness. In this battle it is so important to win such as not to lose. This explains the constant and comprehensive control that a business make over human resources, material resources and financial resources, which has led to the strong presence of resource control systems.

But cannot win a competition only controlling what a company has, it is necessary achieve the goal. Getting customers faster and well prepared, understand their likes, needs and, in short, be better than rivals and strengthen market position.

Nevertheless, all this effort is not enough. It is necessary to achieve strategic advantage over rivals. In this context, to give companies new weapons in this eternal battle, have born Business Intelligence solutions.

The study of own resources, the strange movements, the study of customer transactions and the environment in general produce huge amounts of data. These data can be measured, segmented and related to convert them into useful information that will yield knowledge about the business. All this knowledge will allow observe, understand and even predict futures behaviors by supporting the decisions to be one step ahead of rivals.

This is the Business Intelligence and about it talks this project.

### 1.2.- Relationship among Business Intelligence, ERP and CRM.

Systems ERP, CRM and BI solutions share many common features, but there is a tendency to equate them despite their profound differences. In order to provide a correct approach to the concept of BI in this section the areas of these concepts are discussed, in what they are based, how they act and what to expect from each one.

First, it is necessary understand exactly the concepts of ERP and CRM. Although these systems can work together (and in fact usually do) the used philosophy is profoundly different (MySoftware, 2011). CRM systems (Customer Relationships Managing) are based on interactions with clients, services and sales activities and marketing of products in short these systems manage all aspects of the relationship between business and customer.

Systems Enterprise Resource Planning (ERP) focuses on production management, accounts and supply chain. ERP Systems ensures a constant flow of information among management, accounting, human resources and production.

Both systems are based on the analysis of data produced by business allowing knowing the state in which company is. So their data stores are oriented transaction (load, delete and update), also called OLTP (On-Line Transfer Processing) and its persistence is limited.

BI systems and solutions can be defined as a set of applications and technologies that are responsible for collecting, storing, analyzing and providing access to data in order to provide support when making decisions (Nelson, 2007).

BI systems aim to provide historical, current and predictive view of the business. To carry out this objective, data stores have an ETL structure (Extraction, Transformation and Load), also called OLAP (On-Line Analytical Processing). Also they have a long persistence of data to analyze trends and developments followed by the business (Martinez, 2014).

Nowadays, companies and businesses produce a huge amount of data, generally produced by the ERP or CRM systems. These data contains valuable information regarding the business, which is used by BI systems to provide solid support at the time of making decisions.

Therefore, in summary, although three systems are based on the use of data produced by the business, ERP and CRM systems are aimed at offering a vision of the present business, managing the business aspects or relations with the customers respectively, while BI systems take and analyze data generated by the company (usually by ERP and CRM systems) to provide historical, current and predictive view of the business in order to be a solid support decision making.

### 1.3. - State of art.

Previously it mentioned the notorious and exponential increase of data produced by the economic activity of a company. Not only the volume has increased, relations among enterprises, competitiveness and the need for communication with customers have increased too (UOP, 2014). In 2009 the amount of information generated by companies was estimated at about 0.8 petabytes, while for 2020 it is predicted that this number could increase by 5,000% (40 petabytes) (Mearian, 2011). This enormous amount of data raises many questions, mainly relating to security, availability, accessibility and quality.

**-Security:** The vast amount of data generated favors leaks of information in case of neglecting the control over access to them.

**-Availability:** Information systems need to move a large amount of data, which may cause the access time to data slows down significantly.

**-Accessibility:** If the indexing and storage system is not efficient, it produces that users must invest a significant effort and time to locate the desired data.

**-Quality:** Many data cause there also lots of "junk" or no useful data that are counterproductive and must be separated from the valid data.

It is necessary remember that the process of BI is to turn data into information and turn it into knowledge. Therefore, if the basis of the process (data) does not meet business expectations, the end result will be significant and adversely affected.

This trend towards the growth of data is given by the increased use of devices and applications in both businesses such as private purposes (Chichilla, 2011). Specifically in the business world it is to bet:

**- The increase in products and applications that generate data:** The use of ERP, CRM applications, social networking presence, analysis of web applications...

-**Digitizing knowledge:** Invoices or even the books nowadays are not printed on paper.

-**Combine data horizontally and vertically:** collaboration is enhanced at an internal level as well as synergies with external fields, leading to an exponential increase in the volume of data.

-**Market data:** At present there are often agreements between companies in order to use third-party's data through contracts protection and use thereof.

However, this large amount of data generated in the business field, is the fountain where BI applications drink. Is called "latency Business" to space of time between an event in the field of business happens until a decision is made. The higher it is, the smaller the end value of the business (Business and Technology, 2013). This latency is subdivided into three sections:

-**Data latency:** The time that elapses between a business event is recorded and data are retrieved.

-**Latency analysis:** is identified with the time period ranging from a data is recorded to obtain information containing.

-**Latency of the decision:** defines how many time passes since information is obtained until the decision is made.

BI tools and applications work especially at latency analysis, obtaining information produced from data and helping to make decisions faster and more efficiently. This ability to reduce the response time in business gives these tools great strength and power, making the market, competitiveness and efficiency grow steadily.

Nowadays, most (if not all) large global companies are adopting BI systems in order to generate competitive advantages over its rivals. Following are some big success stories of BI.

- **Xerox increases the number of parking spaces in Spain.**

The Xerox' analysis processes promote and enhance the number of parking projects in cities of Spain and around the world. This company offers software and hardware solutions that enable analysis of traffic conditions in real time, allowing access to the availability of parking and modification of rates according to demand, improving driving experience and allowing concessionaires to have a single point of control and benefiting both the environment and tourism of big cities (Xerox, 2015).

- **Banco Mare Nostrum (BMN) is managed by iPad.**

BMN is a leading Spanish financial institution with a staff of 4,700 employees and 800 branches. In September 2014 BMN hired Microstrategy to develop a system of interactive controls. The system will analyze and forecast trends and lines of real-time business quickly and efficiently from users' iPad (BMN, 2015).

- **Cognos for Nestle.**

The Swiss-based company, a leader in food distribution, has more than 280,000 distributors worldwide. Nestlé implemented the SAP Business Warehouse system for data storage. Subsequently, the decision to implement the Cognos system as a suite for the proper management of the business was taken. This solution enables those responsible for decision making at all levels monitor, analyze and get a better performance,

making possible to make decisions that directly influence the overall success of the business. This makes possible meet the diverse information needs of the many communities of users that are part of the group (MarketingDirecto, 2007).

- **Open source for analyzing credit risk.**

Equifax is a company whose headquarters is in Atlanta (USA), with presence in 17 countries and is a global leader in credit risk analysis. Its division in Iberian serves more than 1,400 companies so that the management and analysis of data is crucial. In order to ensure the service and consumer confidence, was introduced an ETL system (extract, transform and load) based on Open Source (Pentaho) which allows monitoring, analysis and data integration without license fee. This solution has allowed Equifax Iberia placed in a leading position in its industry, reducing costs and ensuring the quality of their product (stratebi, 2013).

Previous cases show that the BI market is booming although it's still in what could be considered "implementation phase". Technology's evolution is very large, in 2008 its turnover was estimated at about 9,000 million dollars while in 2013 it was estimated at about 14,400 million (Gartner, 2009).

This growth is also reflected in the workplace. There is currently a high demand for new profiles related exclusively with BI, such as would be Architect BI business solutions, BI Analyst, Strategist BI or data hygienist.

It's in the sphere of big business where a higher penetration of BI is observed. In general, BI solutions are big and expensive, require a perfect and complete adaptation to the business and are difficult to learn and use. Even large multinationals recognize that their use becomes very complex, however admit that the returns at the management level of information and human capital compensate for these difficulties.

However, in the field of SMEs things change. These companies have fewer human and financial resources, so their view of BI tools is different. Specifically, the main difficulties encountered can be summarized in:

- These applications are expensive to implement because it requires software, hardware and professional experts' time.
- It's complex because BI systems are designed for large companies.
- Requires time dedicated by the company.

These difficulties expose the main deficiency of BI today: usability. Despite all its power and benefits they can bring to a company, if a BI solution is not easy to use, it will not be implemented properly or it will not draw full advantage, and the efforts engaged to its development and implementation will be in vain (Ajuntament de Barcelona, 2012).

So far there has been an exposition of the concept of BI system, which in general can be understood as a set of great tools and programs installed on servers looking to offer the best possible solutions to the user. However, like everything in the ICT world, BI systems can't afford to stay stuck in this position and is struggling to change at the same rate needs of users change. As trends of this change in BI systems can be mentioned the following:

## **MOBILE BI**

It represents a unique trend in BI, however is giving birth too many sub-trends affecting the adoption and general use. Advances in technology can satisfy the needs of BI users in terms of interaction and usability.



The main advantage offered by these solutions is mobility by allowing users to create dashboards with the same features as desktop applications. Although is expected to be established as a standard, today this technology has some limitations (Hagerty, 2012).

- Many applications allow users to view data without the possibility to modify.
- These applications are consumption tools, no analysis.
- Lack of interactivity does not allow users to explore data.
- Many companies have more than one BI application requiring users to have more than one application increasing its cost.

### **LOCATION INTELLIGENCE (LI)**

This concept uses methods developed in the domain of BI in combination with geographic information systems allowing for a greater understanding of the analysis of location data. LI generates information from labeling and storage of raw data with its temporary location to turn it into actionable information. Combining BI with GIS generates some interesting values:

- It allows companies to better understand the external features and how they affect their procedures.
- Sets an analysis of the geographic and temporal dimensions of business location to understand in more detail the influence of occurred events.

LI premises are promising, but the cost of storage and processing needs have relegated it at small solutions. However, it's beginning to capture data through sensors to put commercially available. As main areas of activity could be cited health care, insurance fraud or disaster response (Milton, 2011).

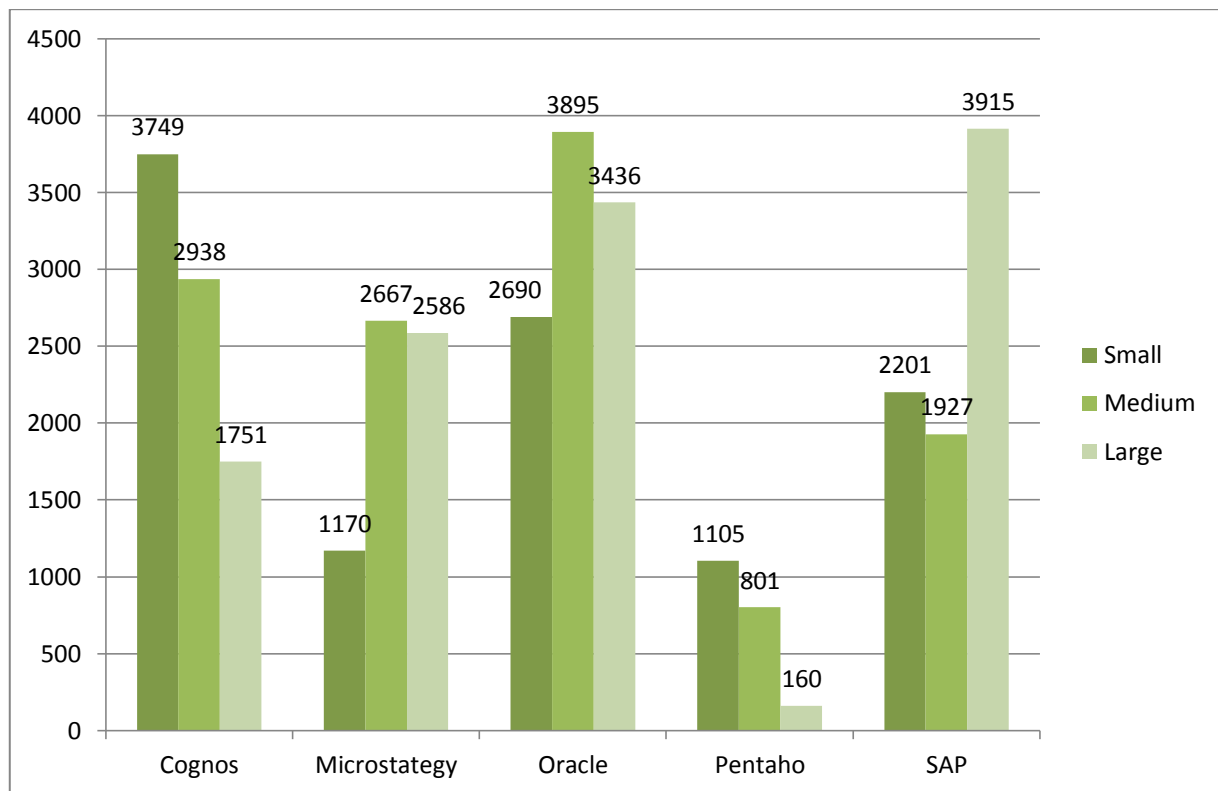
### **Software as a Service (SaaS), Cloud BI**

This is the concept of BI systems as Cloud Computing or Cloud BI. In this model, applications are hosted on servers where users access via secure connections. The main difference with the traditional solutions is the use by a pay-per-use license instead of an annual license, and allows an easier installation than traditional desktop applications. However, Cloud BI must be carefully weighed whether an organization is reluctant to analyze their data outside its borders, regardless of the security guarantees if a high degree of customization is required or if the data have tendency to change quickly (Skyrius, 2012).

Therefore, in summary, the BI industry today is booming and full health. It has affected the way of doing business and gaining competitive advantage, customer and competition relationships and even the emergence of new job profiles. Although most of the benefits are intangible and is complex to calculate Business Intelligence return on investment (ROI), is evident that these tools provide an increase worker productivity and better management of the company. This fact in economic terms means a reduction of costs between 5% and 15%. However, these tools still have a long way to go, especially in terms to usability and access that small businesses could have this technology.

### 1.4 Report of motivations and goals.

As discussed throughout this document, this project is based on Business Intelligence, its uses, viability and need and situation at business level. As it was said at the end of the previous section, at present, the main weakness of this technology is based on its usability, its cost and its learning time. Especially with regard to economic cost, it is interesting to show the following chart in which, for orientation, can appreciate the acquisition and maintenance cost of the large commercial suites (Third Nature, 2010).



As can be seen in this graph traditional BI solutions represent an almost prohibitive cost for small companies, therefore the objective raised by this work is answer the following questions:

- Are there BI tools outside large commercial suites?
- Is it possible obtaining effective BI solutions at an affordable cost?

If the answer to the above questions is affirmative:

- What costs have such applications?
- Really these applications can offer BI solutions?
- Which is better?

And as last question, benefits justify the investment required by BI tools?

Throughout this project, data, knowledge and everything related to Business Intelligence technology will be presented. As the main point of the project, a case of practical application will be developed. This case will try to obtain knowledge about the current situation of rural tourism by provinces nationwide. The objectives

of this practical case application will be presented in more detail in the corresponding chapter, but in essence, they are the followings:

- What are the provinces with a higher investment potential in rural tourism?
- What is the visitor profile of those provinces?
- What product should offer each of these provinces?

Based on this, will be presented conclusions which will try answer questions raised before.



## 2.- Business Intelligence (BI)

Una vez sentadas las bases del proyecto, llega el momento de entrar en los detalles del trabajo. En el presente capítulo se pretenden exponer con la máxima claridad posible todos los conceptos relativos al *Business Intelligence*. Se pretenden presentar las partes de las que se compone, el campo que abarca cada una, aclarándose diferencias y relaciones entre algunas de ellas, entre otros aspectos de interés.

### 2.1.- ¿Qué es Business Intelligence?

En pocas palabras, *Business Intelligence* se puede definir como el proceso de transformar los datos en información y la información en conocimiento.

Más en profundidad, se puede considerar que Business Intelligence es el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLTP / OLAP, alertas...) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio.



Figura 1 BI: Datos - Información - Conocimiento

*Business Intelligence* actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas del negocio: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, etc...

Los sistemas y soluciones BI, al contrario de los sistemas operacionales (ERP, CRM,...), están orientados a la consulta y divulgación de datos. Por este motivo, en un sistema *Data Warehouse*, los datos se encuentran sin normalizar para apoyar consultas de alto nivel. En este sentido, los sistemas BI se apoyan en procesos ETL (extracción, transformación y carga). Estos procesos tienen que traducir de uno o varios sistemas operacionales normalizados e independientes a un único sistema desnormalizado, cuyos datos estén completamente integrados (Sinnexus, 2012).

Esta filosofía permite:

- Observar** el entorno y el comportamiento del negocio.
- Comprender** el estado del negocio.
- Predecir** qué tendencia puede seguir el negocio.
- Determinar** qué líneas de trabajo emprender.
- Decidir** qué estrategia seguir.



Figura 2 BI: Decisión - Acción - Información

## 2.2.- ¿De qué se compone BI?

Una vez expuesto y definido el concepto de *Business Intelligence*, es necesario tratar las partes que lo componen. El objetivo de este apartado se basa en ofrecer una visión de las capas que componen un sistema BI incluyendo los conceptos, herramientas y tecnologías que engloban y las relaciones existentes entre las mismas.

En primer lugar, es necesario aclarar que BI es una metodología relativamente nueva. Ello implica que no exista una definición estricta y clara de las partes que la componen. De hecho existe una cierta polémica con el ámbito y la limitación de las mismas e incluso con el concepto de BI.

Más adelante se tratarán los conceptos con mayor detalle. Sin embargo, la falta definición llega al punto de que unas definiciones llegan a contradecir otras.

Existen teorías que postulan que el ámbito de los *Data Marts* se encuentran fuera del ámbito del *Data Warehouse* (tinko, 2012) mientras que otras consideran que aquéllos son parte de éste (Standen, 2008).

Incluso existe una mayor polémica en la relación entre *Business Intelligence* (BI) y *Business Analytics* (BA). Existen teorías que consideran que BI engloba a BA (Rouse, 2014), otras postulan que no existe diferencia entre ambos, siendo un problema de nomenclatura (Elliot T. , 2011) mientras que por otro lado también existe la opinión de que el trabajo de ambos está relacionado pero son distintos (Business Analytics, 2013).

Visto lo visto, no es fácil sintetizar una estructuración clara del concepto de Business Intelligence. Así lo expresa la empresa Webmining Consultores (WebMining Consultores, 2012)

*“Después de haber trabajado en este negocio por más de 10 años, podemos decir con confianza que todo el mundo tiene una noción diferente de cualquier término relacionado con analítica.”*

No son éstas las únicas discrepancias, por así decirlo, que existen en torno al BI. También es difícil situar dentro de su organización elementos como los infocubos o el *Data Mining*. Ello hace que sea delicado y complejo establecer el ámbito y las relaciones a las que responden los distintos elementos que componen esta tecnología. A pesar de todo ello, se tratará de establecer de la forma más precisa y correcta posible utilizando como base trabajos de divulgación y opiniones como podrían ser las aportaciones de (Kotorov, 2014), (Biskup, 2014), (Elliot T., 2014), (Tech-faq, 2011) o (annalect, 2013) entre otros. Con esta base, se ha construido la siguiente pirámide, que pretende ser un esquema de referencia que permita entender la organización de los sistemas y soluciones BI.

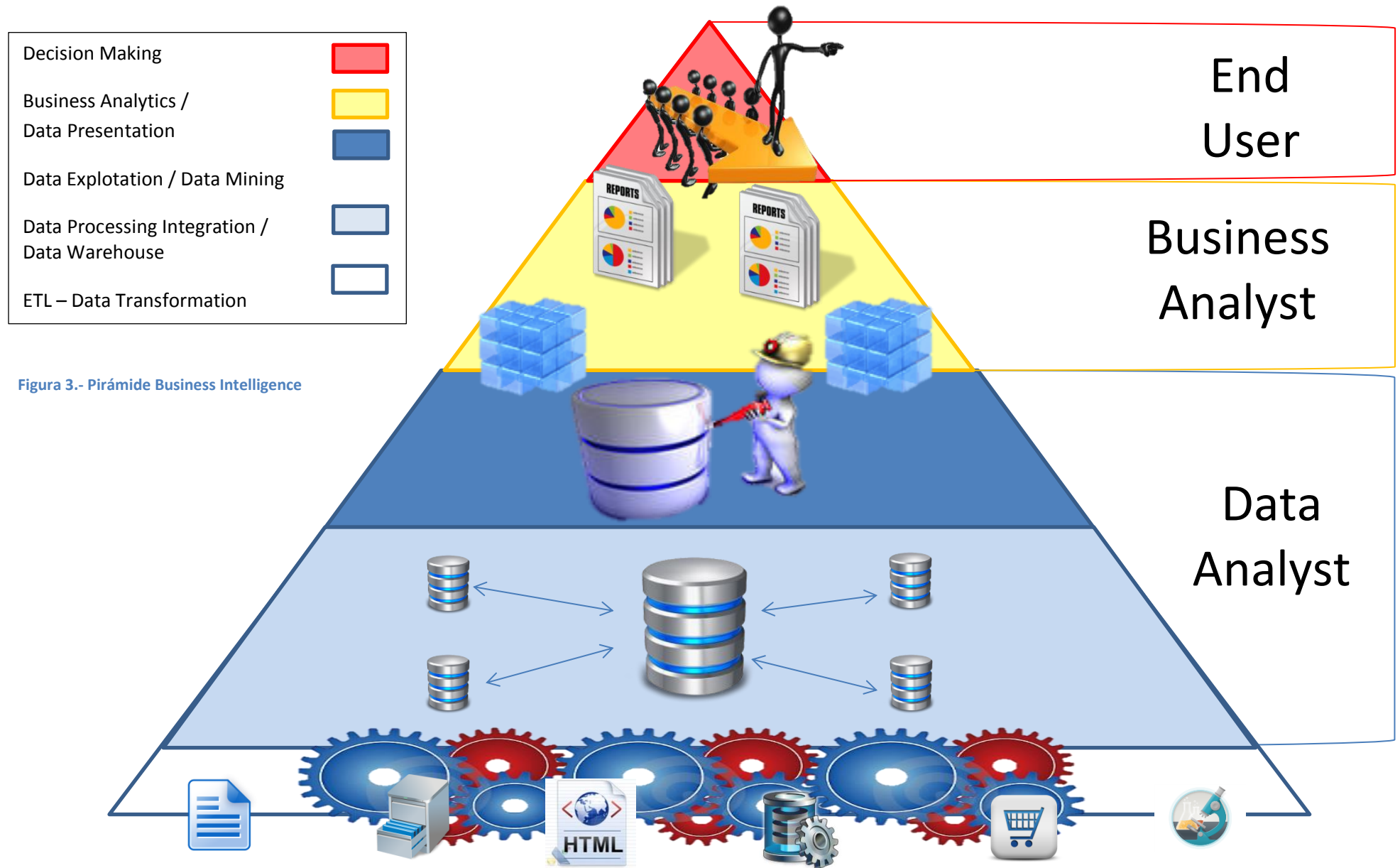


Figura 3.- Pirámide Business Intelligence

Como se puede ver en el diagrama, la composición de la estructura del BI se puede dividir en tres bloques principales:

**-Data Analyst:** En esencia es la parte donde se almacenan y transforman los datos. Se subdivide a su vez dos escalones:

-Data Warehouse.

-Data Exploration-Data Mining.

**-Business Analyst:** Este segmento se centra en el análisis y la presentación de los datos para el usuario final.

**-Making Decision:** El principal elemento de esta sección es el usuario final, el encargado de tomar las decisiones fundamentadas en la información ofrecida por el sistema completo.

A los que se deben añadir los datos de entrada.

A continuación se ofrece una visión más completa y detallada de los estamentos de los que se compone un sistema BI.

### 2.2.1.- Data Sources.

Sin duda alguna, el combustible y la fuerza de cualquier sistema *Business Intelligence* son los datos (annalect, 2013). Con esta afirmación, lo que se trata de expresar es que el origen de la potencia y de todas las capacidades que puede ofrecer un sistema BI se basa en los datos de los que se nutre. Por ello es importante tanto el número como la calidad de los mismos (punto 1.2). Como ya se ha dicho, estos datos pueden proceder de muy diversas fuentes pues es fácil que una empresa disponga de una amplia variedad de puntos a lo largo de su sistema en los que se produzcan.

A la hora de establecer un sistema BI, es muy importante saber identificar correctamente los puntos de los que obtener datos de calidad y verdadera utilidad para el sistema (Hartzen, 2010). Este proceso es vital, pues es el punto de origen a partir del cual se desarrollará toda la implementación del sistema. Es importante entender cómo funcionan los flujos de información del sistema. Es posible que la misma información se esté compartiendo en varios puntos y por tanto se podría incurrir en la inclusión de información redundante en el sistema. Así mismo también es necesario conocer las posibles relaciones existentes entre datos pues esto repercutirá en la eficiencia de la solución implementada.

De entre los distintos puntos de generación de datos pueden resaltarse los siguientes:

**-Bases de datos:** Este es el origen estrella de los datos, es donde se encuentra la mayor parte de los datos del día a día de un negocio. El volumen y la fuerza de los datos que se localizan en las bases de datos es enorme, especialmente si se cuenta con sistemas ERP/CRM.

**-Flat Files:** Se entienden éstos como ficheros sin información relacionada generados directamente por la actividad comercial. Suelen ser ficheros de texto plano con un registro por línea.

**-Información externa:** Información generada por terceros, bien pública o privada que permite conocer el estado del mundo exterior a la empresa. En esta categoría se pueden englobar desde bases de datos compradas a empresas hasta trabajos de divulgación científica.



Como se puede ver, datos pueden tener procedencias, intereses y valores muy distintos, lo que puede convertir el sistema en un caótico compendio de registros sin relación. Por ello es importante que se establezca una normalización que regule el aspecto que deben tener los datos para figurar en el sistema. Esta parte de la función del siguiente estamento de un sistema BI, data Analyst.

### 2.2.2.- Data Analyst.

Como resumen, este escalón del BI se encarga del almacenaje y transformación de los datos con el objetivo de extraer la información contenida en los mismos. Como se comentaba en el anterior punto, antes de entrar a formar parte del sistema BI, los datos de sufrir un proceso de normalización que garantice que son aptos para el objetivo perseguido. Es en este punto del sistema donde se lleva a cabo dicha normalización.

Este proceso se conoce con el nombre de ETL (Extracción, Transformación y Carga) y en esencia se compone de los siguientes subprocesos (Exelia, 2014):

- Extracción:** Obtención de los datos desde sus fuentes de origen.
- Mezcla:** Se agrupan los datos de similar contenido.
- Borrado** de datos inconsistentes: Los datos que no son útiles al sistema son desechados.
- Normalización:** Es necesario adaptar la naturaleza de los datos a la esperada por el sistema.
- Carga:** Los datos entran a formar parte del sistema.

Una vez realizado el proceso de adecuación de los datos, estos son cargados en el *Data Warehouse*. A partir de este punto, es posible dividir el escalón del DA en dos estamentos, *Data Warehouse* (almacenaje puro) y *Data Exploration* y *Data Mining* (extracción y explotación), que serán expuestos a continuación.

#### 2.2.2.1.- Data Warehouse

A la hora de implantar un sistema de Business Intelligence una de las primeras y más cruciales decisiones concierne al sistema de Data Warehouse a implementar. En esencia, el concepto de DW se podría definir como un lugar donde almacenar datos para su posterior consulta, análisis y propósitos de seguridad (Tech-faq, 2011). Estrictamente, esta definición podría englobar a todo sistema de almacenamiento de datos. Sin embargo la principal diferencia estriba en la persistencia de los datos.

Los sistemas tradicionales de Bases de Datos, también llamados OLTP (On-Line Transaction Processing) se caracterizan por la volatilidad así como por la capacidad para realizar un gran número de operaciones rápidas (inserciones, borrados y actualizaciones) destinadas al procesado, mantenimiento e integración de datos. En estos sistemas se almacenan datos actuales y detallados y se utiliza un sistema transaccional (rainmakerworks, 2012).

Por su parte, los sistemas Data Warehouse conocidos como OLAP (On-Line Analytical Processing) son sistemas orientados a persistencia, lo que quiere decir que los datos almacenados no tienden a ser eliminados y pueden remontarse a años atrás. Las consultas realizadas tienden a ser muy complejas y suelen implicar agregación (rainmakerworks, 2012). Este tipo de consultas suelen requerir mucho tiempo en su ejecución, especialmente en la carga de datos desde los sistemas de origen. Por este motivo, este proceso se realiza generalmente en las llamadas “horas valle” que son las franjas de tiempo de menor utilización del sistema que usualmente suele corresponder con las horas de madrugada.

Con los sistemas DW se busca alcanzar los siguientes objetivos:

**-Centralización:** Localizar datos procedentes de una gran variedad de puntos del sistema en repositorio único.

**-Fiabilidad:** Los datos cargados son, por definición, consistentes pues proceden del propio sistema. Sin embargo, las herramientas ETL deben asegurar además la calidad de los mismos, evitando además que los datos irrelevantes formen parte del sistema

**-Homogeneidad:** Los datos residentes en un sistema DW proceden de distintos puntos y eventos, que sin embargo deben ser categorizados utilizando una denominación y una métrica consistente que evite equívocos a la hora de la interpretación de la información.

**-Eficiencia:** A pesar de la ingente cantidad de datos que se deben mover en estos sistemas es imprescindible que se asegure una respuesta del sistema que evite la ralentización del sistema.

Usualmente, se consideraba que la única forma de mantener un sistema DW consistía en implementarlo sobre bases de datos no relacionales que premien el rendimiento y la velocidad de respuesta frente a la coherencia de los datos. Sin embargo, el avance y desarrollo de los DW ha permitido implementarlo sobre sistema RDBMS utilizando sistemas complejos que a la postre se han generalizado (Sinnexus, 2012). Estos sistemas son evoluciones de OLAP y se conocen como ROLAP, MOLAP y HOLAP que se definen a continuación:

**-ROLAP:** Esta fue la primera de estas tecnologías en desarrollarse y su nombre significa literalmente *Relational OLAP*. Estos sistemas utilizan una arquitectura de tres niveles. El primero es un sistema de bases de datos relacional que maneja los requerimientos de almacenamiento y el más alto de los niveles es una capa de presentación OLAP tradicional. Entre ambas capas se sitúa el motor ROLAP capaz de convertir las consultas dinámicamente a SQL, realizar integración de datos y generar los índices que permitan aumentar la eficiencia de los tiempos de acceso a consultas.

**-MOLAP:** Esta tecnología apuesta por la implementación del sistema sobre bases de datos Multidimensionales (Multidimensional OLAP). Este sistema utiliza una arquitectura de dos niveles: un sistema de base de datos multidimensional y un motor analítico. La información procedente de los sistemas operacionales se carga mediante rutinas por lotes. Una vez cargado el dato en la Base de Datos Multidimensional (MDBD) se generan los índices y los algoritmos de tablas hash para mejorar los tiempos de acceso a datos. El nivel de aplicación es el responsable de la ejecución de los procesos OLAP.

**-HOLAP:** Un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

En cualquier caso independientemente de la arquitectura implementada, un sistema DW depende de la eficiencia y la optimización de los tiempos de acceso. Ello obliga a la implementación de soluciones internas que permitan atender a esta necesidad. Entre estas soluciones, el principal y más potente mecanismo son los llamados *Data Mart*.

Dentro de toda organización existen comunidades o sub-grupos más pequeños con necesidades y capacidades específicas. Estas necesidades específicas se reflejan también en los sistemas de información, concretamente en lo referente al sistema DW. A este respecto, nacen los Data Marts para cubrir las necesidades de estas comunidades específicas.

Técnicamente, se puede definir un Data Mart como un subconjunto de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades (Oracle, 2007).

Otra de las grandes discordias o incongruencias existentes en torno al BI se basa en la relación entre DW y Data Mart. Ambos términos tienden a utilizarse de un modo incorrecto como sinónimos. Sin embargo, un Data Mart es un subconjunto del Data Warehouse, en un lenguaje más coloquial se podría entender que un Data Mart es un Data Warehouse en miniatura.

Una vez almacenados los datos se hace necesaria su explotación. En el siguiente apartado se exponen las técnicas y metodologías a este respecto.

#### **2.2.2.2. - Data Exploration-Data Mining.**

Esta sección del Data Analyst es especialmente difusa y complicada de delimitar. En realidad ejerce de frontera entre el almacenamiento de los datos y la obtención de información útil para el negocio. Incluso la propia literatura parece incapaz de exponer un consenso. Así es posible encontrar autores que lo identifican exclusivamente con el Data Mining (Steinbach, 2005), quién lo identifica únicamente con Data Exploration (Jansen, 2014) e incluso quién lo identifica con términos distintos (Lixto, 2005) aunque de un modo más minoritario. De esta forma se establece otra de las frecuentes polémicas relacionadas con la terminología del BI.

En esencia, la misión que se cumple en este segmento de la pirámide del BI, consiste en la extracción de datos del Data Warehouse. Como ya se ha comentado previamente, la información almacenada en el DW debe haber sido seleccionada, clasificada y normalizada, por tanto toda información alojada en éste se considera útil y correcta. Sin embargo, su número puede (y debe) seguir siendo ingente y puede resultar poco rentable según el objetivo que se persiga por cada usuario y momento.

Los dos términos antes expuestos, Data Mining y Data Exploring, son los que generalmente, se consideran los abanderados de este segmento. Su base es similar, sin embargo presentan profundas diferencias. Data Mining presenta un aspecto de automatismo, en cambio Data Exploring tiene un objetivo manual. Los procesos automáticos presentan una gran optimización mientras que los procesos manuales permiten una mayor flexibilidad y adaptabilidad.

#### **Data Mining**

Representa una metodología cuya función consiste en el análisis y la obtención de datos de interés dentro de grandes conjuntos de datos. Los procesos DM son algoritmos automáticos avanzados en los que se incluyen inteligencia artificial y aprendizaje de máquina. El gran volumen de datos a manejar hace que sea necesario disponer de automatización para que las herramientas sean fiables, sin embargo es posible utilizar estas técnicas con BBDD de cualquier tamaño (Datamology, 2010).

Existe gran cantidad de técnicas de DM entre las que cabe destacar:

**Vecino más cercano.** Esta técnica de trabajo se centra en obtener de una serie de valores almacenados en la base de datos que presentan un determinado valor de predicción el valor “más cercano” no almacenado que responda a ese valor de predicción. El objetivo es identificar nuevos patrones sin que exista una coincidencia exacta con patrones o casos almacenados (thearling, 2011).

Un sencillo ejemplo de esta técnica sería observar en los habitantes de un vecindario. Si un vecino del vecindario tiene unos ingresos de cien mil euros, es fácil que la haya más vecinos con ingresos similares. Sin embargo si un vecino obtiene unos ingresos de 20.000 es poco probable que haya muchos vecinos que alcancen los 100.000. El ejemplo es muy sencillo y obvia muchos factores a tener en cuenta, como por ejemplo el nivel cultural o el número de hijos.

**Clustering.** Este método consiste en agrupar registros semejantes para ofrecer una visión de alto nivel sobre el estado de la base de datos. Estos procedimientos buscan características similares que permitan establecer conjuntos asignándoles las llamadas “etiquetas de grupos”. Generalmente son métodos muy robustos que ofrecen adaptabilidad a cambios en el sistema (thearling, 2011).

Un ejemplo sencillo de clustering, sería el modo en que se almacenan productos en un supermercado. Entre los muchos productos que se venden existen productos perecederos, no perecederos, tecnológicos... Es importante tener en cuenta que los productos perecederos deben permanecer refrigerados para conservar sus propiedades, algunos incluso congelados. Los productos tecnológicos deben estar alejados de las zonas de humedad y los productos textiles situarse de modo que no dañen sus tejidos.

**Árboles.** Los árboles de decisión son sistemas predictivos organizados partiendo desde un punto único. Cada rama del árbol se corresponde con una característica de clasificación y cada hoja se corresponde con una partición del conjunto de datos. Un árbol de decisión realiza un test a medida que se recorre desde la raíz (padre) hasta las hojas (hijos) para alcanzar así una decisión (thearling, 2011).

Un ejemplo de árbol de decisión podría ser el agrupamiento de los clientes potenciales de una determinada marca de ropa deportiva para conocer si sería beneficioso instalarse en un lugar determinado. En primer lugar se separaría la población por segmento, asumiendo que se busca un rango de entre 15 y 55 años de edad. De ellos se realizaría una nueva división para obtener aquellos que realicen deporte habitualmente. Una vez extraído este segmento, se dividirían aquellos que practiquen el deporte especializado de la marca (*running* por ejemplo). Se pueden seguir estableciendo divisiones hasta lograr concretar el público objetivo como podrían ser el sexo o el poder adquisitivo.

**Redes Neuronales.** También llamadas redes neuronales artificiales son grafos, sistemas de organización en los que los datos representan nodos y existen varios caminos que los permiten relacionarse unos con otros e incluso conectar dos nodos que no tienen relación directa entre sí. Estos sistemas permiten conocer el grado de conectividad que presentan dos agrupaciones de datos a fin de conocer si pueden resultar beneficiosos para el negocio (thearling, 2011).

Como ejemplo de estas redes podrían exponerse las redes sociales (cuyo nombre no es casual). En estas redes es posible establecer relaciones con personas con las que no existe relación previa a partir de personas entre las que sí existe una relación.

## Data Exploration

Como ya se ha expresado, Data Exploration es una metodología que emplea técnicas manuales con el fin de recorrer conjuntos de datos para obtener aquellos que permitan realizar análisis. Su naturaleza manual está orientada a la utilización en conjuntos de datos pequeños aunque pueden utilizarse con conjuntos de gran tamaño (Datamology, 2010).

Data Exploration se fundamenta en la metodología estadística del análisis exploratorio de datos. Éste es el análisis al que se someten las muestras recogidas en cualquier campo científico con la idea de obtener los datos necesarios para apoyar o rechazar una hipótesis o simplemente mostrar el estado en que se encuentra una población de datos.

Un ejemplo de aplicación podría ser la realización de un estudio para conocer el índice de sobrepeso presente en una sociedad. Para ello se realizaría un estudio estadístico en el que se obtendrían por géneros y por edad una serie de segmento de los que se obtendrían la altura y el peso medio. Con esos datos generados se calcularía el IMC medio de cada segmento obteniéndose los segmentos de edad con mayor índice de sobrepeso.

Dentro de este segmento se podrían incluir también los Infocubos por ser un elemento de extracción de datos. No obstante, debido a que su uso fundamental se realiza en el análisis de negocio, se definirán en el siguiente escalón de la pirámide

A lo largo del proceso de BI se han obtenido y limpiado los datos para su posterior almacenamiento en el DW. Las técnicas de *Data Extraction* han permitido al sistema obtener los datos más relevantes que se hallaban en el DW. Estos datos, que por sí mismos son muy valiosos, son lo que permitirá el funcionamiento del siguiente segmento de la estructura BI. A continuación los datos se procesarán y se obtendrá la información que será presentada al usuario final con el fin de generar conocimiento y poder tomar decisiones fundamentadas.

### 2.2.3.-Business Analytics.

Si se considera el Business Intelligence como una herramienta poderosa, en este segmento es sin duda la fuente de su potencial. En este segmento se genera la información necesaria que será presentada al usuario para la toma de decisiones que permitan obtener ventaja competitiva. Los procesos BA utilizan la información extraída de los datos procesados. Como se ha dicho previamente, los datos son el combustible del BI en general y de estas herramientas en particular. Es por tanto en esta parte del sistema donde el proceso de refinado, normalizado y extracción de los datos tendrá especial repercusión. La calidad de las decisiones tomadas será consecuencia directa de la calidad de los procesos utilizados hasta el momento.

De entre las muchas discusiones o faltas de acuerdos relativas a la estructura del BI, la relación entre BI y BA es sin duda la mayor de todas, como bien ilustra la Figura 4 que acompaña a este texto. Es un tema ampliamente discutido en cualquier foro de debate empresarial y las opiniones al respecto no pueden ser más dispares.



Figura 4.- BI vs BA

Es fácil encontrar citas de autores que consideran que BA y BI son dos ciencias separadas pero relacionadas,

quién considera que BA es un subconjunto de BI y también quién considera lo contrario. Para ofrecer una visión de las distintas opiniones al respecto, a continuación se exponen las consideraciones de siete expertos en el mundo empresarial.

Para Pat Roche, vicepresidente de ingeniería de Magnitud Software (USA), son dos conceptos distintos, y su diferencia radica en qué persigue cada uno:

*“Business Intelligence es necesario para que funcione el negocio mientras Business Analytics busca cambiarlo.*

*BI se centra en crear una operatividad eficiente a través del acceso a datos en tiempo real para que cada persona pueda crear desempeñar su trabajo con la mayor eficacia posible. BI también incluye el análisis de los datos históricos desde múltiples fuentes y la identificación y resolución de problemas para una toma de decisiones informada.*

*Por su parte BA se refiere a la explotación de datos históricos desde muchas fuentes de origen a través del análisis estadístico, cuantitativo, Data Mining, modelado predictivo... para identificar las tendencias y comprender la información que puede impulsar el cambio empresarial apoyado en decisiones satisfactorias.” (Roche, 2014)*

Por tanto, Roche postula que BI se centra en la comprensión del estado concreto de la empresa, para tener una visión global de la misma, identificándolo con la función que cumplen herramientas software como ERP/CRM, mientras BA estaría basado en el aprovechamiento de los recursos del sistema para la toma de decisiones que cambien el negocio.

En una línea semejante se expresa Mark van Rijmenam, CEO de BigData-Startups (Países Bajos):

*“La diferencia se encuentra en que BI mira por el retrovisor, desde hace un minuto hasta años atrás, mientras BA mira al frente siempre. De esta forma, BI puede decir que ha pasado mientras BA puede predecir qué pasará. BI toma decisiones sobre la base de los resultados anteriores, mientras BA permite analizar el negocio para alcanzar y entender lo que podría llegar a suceder” (Rijmenam, 2014)*

Por tanto Rijmenam, entiende que tanto BI como BA se basan en el pasado del negocio, pero entiende que BA puede sustentar decisiones a un nivel muy superior al de BI.

Una línea distinta de pensamiento toma François Ajenstat, Director Senior del Departamento de Producto de Tableau (Canadá)

*“Business Intelligence se ha centrado siempre en informes altamente formateados creados por unas pocas personas que posteriormente son distribuidos al resto de un departamento. La tendencia de BA es proporcionar a los trabajadores que tengan preguntas sobre el sistema las respuestas correctas en cada momento, permitiendo que se conviertan en sus propios analistas.” (Ajenstat, 2014)*

Para Ajenstat, la diferencia entre BA y BI no se basa en que puede hacer cada uno por la empresa, sino en cómo mejora la capacidad de trabajo de los trabajadores, convirtiendo BA en una herramienta cercana y flexible para conocer aspectos internos del negocio.

Por otra parte, Dipak Bhudia, arquitecto jefe de productos de Clear Analytics (India), entiende que existe una diferencia entre BI y BA, pero que en realidad es casi una diferencia semántica.

*“BI es esencialmente un sustantivo, ya que es un término general que engloba la adquisición, la persistencia, el almacenamiento, el análisis y reporte de ideas junto con todo lo que rodea a estos conceptos. Business Analytics es más bien un verbo, el acto*

*de descubrir ideas utilizando cualquier herramienta o servicio disponible.” (Bhudia, 2014)*

También hay voces que abogan por que, en realidad, BI y BA son mutuamente dependientes como Tim Biskup, Director del CRM de Progressive Business Publications (USA).

*“Existen muchas opiniones y visiones conflictivas en la relación entre BI y BA, y muchas de éstas están en realidad basadas en tendencias de marketing y en cómo se están intentando vender.  
(...)”*

*Cualquier tipo de análisis es una forma de usar la inteligencia con colectiva como una forma inteligente de tomar decisiones perspicaces<sup>1</sup>. En este sentido, definiría BI como la colección de métodos y herramientas usadas para analizar y consumir inteligencia de forma inteligente como el fin de tomar decisiones perspicaces sobre el negocio.*

*Análisis sin inteligencia no puede hacerse, es basura. Coger inteligencia sin hacer análisis es una pérdida de tiempo e inteligencia que no está basada en inteligencia no es inteligente.” (Biskup, 2014)*

El análisis es parte de la inteligencia y la inteligencia es parte del análisis, la una no puede persistir sin la otra. En esta misma línea, dando un paso más allá, se expresan autores como Rado Kotorov, jefe de la oficina de innovación de Information Builders (Rusia).

*“BI utilizada para referirse a la capacidad de la plataforma para acceder a los datos, gestionar metadatos, herramientas de desarrollo para de informes, cuadros de mandos, la edición, la programación y las capacidades de distribución. Analytics se refiere a los métodos de análisis de la información (regresión, redes, descriptivos, redes neuronales...) o las herramientas para llevar a cabo dichos métodos.*

*Por lo tanto, el análisis es un subconjunto de las capacidades de la organización. Las soluciones actuales de BI incluyen herramientas de análisis cada vez más complejas.” (Kotorov, 2014)*

Kotorov, entiende que el análisis es una parte inseparable del BI y por tanto BA sólo puede ser la respuesta a la evolución de los procesos de BI, es decir, que BA es la parte fundamental de BI donde se ven los avances tecnológicos más importantes.

Como se ve, la controvertida relación entre BI y BA da mucho que hablar. Al final, la opinión con la más de acuerdo se puede estar, es la de Timo Elliot, Evangelist Innovation de SAP (USA).

*“¿Cuál es la diferencia entre BI y BA? La respuesta correcta es que cada uno tiene una opinión, pero nadie sabe realmente cual es la verdad, y no hay que preocuparse por ello. Confidencialmente, puedo decir que todo el mundo tiene una noción diferente de cualquier término relacionado con el significado de análisis. Por ejemplo, cuando SAP habla de BA en lugar de BI, está tratando de indicar que BA es un paraguas que incluye los términos DW, BI, gestión de la información empresarial, gestión del desarrollo empresarial, análisis de aplicaciones... es decir, BA incluye a BI. Por otra parte, empresas como SAS usan BA para indicar algún nivel de*

---

<sup>1</sup> En la edición utiliza la palabra “smart” para diferenciarla de “intelligence”, imposible de traducir literalmente en castellano sin incurrir en redundancia.



*conocimiento vertical u horizontal ligado al análisis estadístico o predictivo de los datos, es decir BI incluye BA.” (Elliot T. , 2014)*

Alcanzar un consenso en lo tocante a la relación entre BI y BA es casi utópico a día de hoy. Es, sin duda, consecuencia del hecho de ser una tecnología relativamente nueva y emergente. Como se ha visto en las opiniones antes expuestas, las divergencias responden a las concepciones tecnológicas, intereses económicos o incluso a la misma semántica de los propios conceptos.

Sin embargo, tal como expresaba Biskup, la unión entre inteligencia y análisis es indisoluble, la una carece de sentido sin la otra. Partiendo de esa base, se puede entender que es imposible, o más bien ilógico, mantener soluciones BI sin un módulo de análisis del negocio, independientemente del punto de vista que se emplee, tal como razona Kotorov. Por tanto, la causa de la controversia podría ser la misma novedad del término y la dificultad para adaptarlo a las evoluciones metodológicas y de tecnología, tal como expresaban Bhudia y Ajenstat, BA es una nueva forma de proceder con el mismo concepto de BI.

En resumen, partiendo de la opinión de Biskup de que A e I no se pueden separar, la concepción de Kotorov de que BI necesita al análisis para subsistir y siguiendo las pautas de Bhudia y Ajenstat de que en BA es una nueva forma de proceder pero siguiendo las pautas de BI, se puede colegir que, la concepción postulada por Roche y Rijmenam de BA respondería en realidad al propio BI incorporando los avances en los procedimientos de análisis del negocio. Por tanto, cuando Roche y Rijmenam hablan de BA, en realidad se podrían estar refiriendo al concepto de BI 2.0 (stratebi, 2013), siendo ésta la ideología en la que se ha basado este proyecto.

Y sin embargo, al final Timo Elliot tiene razón, cada cual tiene su opinión sobre BA y BI y tampoco es imprescindible preocuparse por ello. Lo importante es que cada negocio disponga de la mejor solución que le permita tomar decisiones con la mayor certidumbre y fiabilidad para maximizar sus resultados y mejorar su rendimiento.

Una vez expuesta la controvertida relación y su repercusión entre BI y BA, es momento de entrar en materia en lo exclusivamente relativo a BA. Como ya se ha expresado, *Business Analytics* es la parte de la estructura de BI donde se genera la información que permitirá tomar decisiones y esta información se fundamenta en los datos que se hallan en el DW. Estos datos son lo que permite al BA funcionar, y una de las principales herramientas para su obtención son los llamados *Cubos OLAP* o *infocubos*.

## INFOCUBOS

Previamente se introdujo el concepto de Infocubo, como herramienta de extracción de datos, sin embargo debido a que son utilizados principalmente en este segmento, se ha preferido incluirlos en el análisis de negocio. Ya se habló antes de la arquitectura OLAP, que se puede identificar con la estructura de un DW (consultas muy grandes sobre grandes cantidades de datos orientados a persistencia). Los infocubos se pueden entender como arrays multidimensionales o como un conjunto de tablas relacionales organizadas en un esquema de estrella compuestos por una tabla de hechos<sup>2</sup> y varias de dimensiones<sup>3</sup>, que contienen información necesaria para el análisis del negocio desde varias perspectivas (área geográfica, canales de ventas, personal, etc.) (SAP, 2014).

---

<sup>2</sup> Es la tabla central de un esquema dimensional (en estrella o en copo de nieve) y contiene los valores de las medidas de negocio o dicho de otra forma los indicadores de negocio.

<sup>3</sup> Son elementos que contienen atributos (o campos) que se utilizan para restringir y agrupar los datos almacenados en una tabla de hechos cuando se realizan consultas sobre dichos datos en un entorno de DW o Data Mart.



Los infocubos se rellenan con datos procedentes uno o más “*infosources*” o “*infoproviders*” (orígenes de datos) permitiendo, a su vez, servir como orígenes de datos para análisis o propósitos informativos. Los datos son almacenados físicamente en los Infocubos (no se enlazan con las bases de datos de origen) y se componen de una serie de unidades de información o infoObjects que se encuentran organizados en forma de esquema de estrella.

Los infocubos son generalmente estáticos, sin embargo también existen los llamados “Real-Time InfoCube” que tienen la capacidad de ofrecer datos en tiempo real que permiten la escritura desde múltiples lugares, incluso directamente desde bases de datos, aumentando la capacidad de maniobra de los usuarios finales y la consistencia de las decisiones tomadas.

La información proporcionada por los infocubos, permite al *Business Analytics* responder a las siguientes preguntas:

- ¿Por qué ocurrió este evento?
- ¿Volverá a suceder?
- ¿Qué ocurrirá si cambiamos una variable del negocio?
- ¿Qué pueden mostrarnos los datos que no se nos haya ocurrido preguntar?

Son preguntas que van más allá del ¿qué?, ¿cómo? y ¿cuánto? Son preguntas destinadas a generar conocimiento y a dar una vuelta de tuerca al negocio que permita maximizar las posibilidades del mismo. Para llegar a responder a estas preguntas se ha puesto al servicio del BA toda la infraestructura definida hasta el momento. Para aprovechar toda esta estructura existe una gran cantidad de técnicas y métodos que pueden ser agrupados a un alto nivel en tres tipos determinados.

**-Análisis Descriptivo:** Se basa en modelo de agregación de datos y de data mining para permitir una visión histórica de la empresa que permita entender qué ha sucedido.

**-Análisis Predictivo:** Utiliza modelos estadísticos y técnicas de predicción para comprender que podría ocurrir en caso de tomar una decisión.

**-Análisis Prescriptivo:** Se usan algoritmos de optimización y simulación que permitan tomar la mejor decisión posible.

A continuación se exponen en detalle los tres tipos de análisis.

### **Análisis Descriptivo**

Es tipo de análisis se basa en la estadística descriptiva, es decir, permite describir y concretar una gran conjunto de datos con el fin de hacerlo entendible para los humanos. Este análisis describe el pasado lo que significa que se refiere a un evento que ocurrió en el pasado y cuyos datos se hallen almacenados en el sistema DW. La principal utilidad del análisis descriptivo es la compresión y el conocimiento de eventos para obtener una visión de cómo podrían afectar a eventos futuros.

En general, la estadística utilizada dentro de una empresa o en la vida real se sitúa dentro de este grupo, los datos son contados, organizados, filtrados... La estadística descriptiva es útil a la hora de mostrar, por

ejemplo, stock de un producto, la inversión media anual, el dinero medio ingresado por cliente... (Major, 2014)

Un ejemplo de este tipo de análisis sería el tratar de comprender cómo una determinada campaña afecta a los resultados de un negocio. Para ello sería necesario, en primer lugar extraer toda la información relativa a los periodos de esa campaña (es de prever que se haya realizado más de una vez). Una vez obtenido los datos se debe extraer el gasto medio en obtención de stock junto con hechos más puntuales como la inversión en publicidad, salarios de posible personal extraordinario, amén de otros posibles gastos indirectos o fijos como suministros. Por otro lado se deben obtener los ingresos generados por la acción directa de esas campañas. Una vez obtenida toda esta información, se enfrentarán los datos y se situarán en el tiempo obteniendo el conocimiento de la productividad real de dicha campaña.

### **Análisis Predictivo.**

Esta categoría de análisis tiene su raíz en la capacidad de “predecir” que podría ocurrir. La idea consiste en comprender el futuro del negocio basándose en la información procesable de la base de datos, siendo importante recordar que ningún algoritmo estadístico permite predecir el futuro, sino realizar estimaciones basadas en probabilidades.

El método consiste en rellenar los datos obtenidos de sistemas ERP/CRM, recursos humanos, etcétera, con conjeturas a fin de descubrir patrones que permitan capturar relaciones entre distintos conjuntos de datos. La potencia y versatilidad del análisis predictivo permite utilizarlo en cualquier sistema de la empresa, desde estimar el comportamiento de los clientes hasta identificar tendencias en las actividades de venta. (Major, 2014)

Una aplicación común para el análisis predictivo es la producción de puntuaciones de crédito (*credit score*). Estas puntuaciones son utilizadas por los servicios financieros para determinar la probabilidad de que los clientes hagan pagos crediticios a tiempo. Los usos comerciales típicos incluyen, la comprensión de cómo las ventas podrían cerrar a finales de año, la predicción de qué artículos se compran en conjunto o la previsión de los niveles de inventario basados en una gran cantidad de variables.

### **Análisis Prescriptivo.**

Este análisis es el más avanzado de los tres pues permite a los usuarios prescribir (establecer) un número de diferentes acciones posibles con el fin de guiarlos hasta una solución satisfactoria. El análisis prescriptivo trata de cuantificar el efecto de las decisiones futuras con el fin de asesorar sobre los posibles resultados antes de que las decisiones se lleven realmente a cabo. En sus mejores versiones, este tipo de análisis permite conocer no sólo el qué va a pasar, sino también por qué ocurrirá junto con recomendaciones para obtener ventaja de las predicciones.

Estos análisis van más allá de los análisis descriptivos y predictivos mediante la recomendación de uno o más cursos de acción posibles. Esencialmente predicen múltiples futuros y permiten a las empresas evaluar una serie de posibles resultados en base a sus posibles acciones. Se utiliza una combinación de técnicas y herramientas como reglas de negocio, algoritmos de aprendizaje automático y procedimientos de modelado computacional. Estas técnicas se aplican contra la entrada de muchos conjuntos de datos diferentes, incluyendo datos históricos, transaccionales, datos en tiempo real y almacenamiento en Big Data.

Como mayor contra, estos análisis son relativamente complejos de administrar y la mayoría de las empresas aún no los han incorporado a su actividad diaria. Sin embargo, implementados correctamente pueden tener un gran impacto en las grandes decisiones de la empresa permitiendo optimizar la producción, la organización o el inventario de la cadena de suministros mejorando la experiencia del cliente (Major, 2014).

Un ejemplo de la aplicación de los análisis prescriptivos se encuentra en la industria energética. Ésta es. Sin duda, la industria más grande y prolífica del mundo. Las decisiones relacionadas con la exploración y explotación de petróleo y gas natural generan enormes cantidades de datos, además de repercutir en prácticamente todo el mundo. Utilizando estos datos capturados se pueden generar modelos e imágenes de la estructura de la Tierra en capas de varios kilómetros bajo la superficie para descubrir pozos, tasas de flujos de aceite, temperaturas de yacimientos, presiones... Los procesos de análisis prescriptivos permiten tomar decisiones sobre cómo y dónde perforar con el fin de optimizar la recuperación y minimizar tanto el coste como la huella ambiental.

Una de las premisas del BI es ayudar en la gestión del negocio reduciendo el coste, económico y temporal, de las tareas. En este sentido tan importante como la generación de información es la interpretabilidad de la misma. Debido a ello, existen las llamadas Técnicas de Visualización que permiten a los usuarios finales (humanos) comprender la información generada por el procesamiento de los datos.

### **Técnicas de visualización**

En esencia, las técnicas de visualización se pueden entender como herramientas de informes, éstas son una exploración interactiva y una presentación visual de los datos devueltos por los procesos de BA, alojados en DW o incluso en un infocubo. De entre las múltiples técnicas para la visualización se expondrán las siguientes:

**Key Performance Indicator (KPI):** Los indicadores de clave de rendimiento o KPI son informes que permiten conocer la medida del nivel del desempeño de un proceso. Los KPI's son diseñados con el fin de entender cómo progresa un aspecto del negocio y pueden ser utilizados por distintos sectores de la organización. Este tipo de informe se utiliza en BI para reflejar el estado actual de un negocio y definir una línea de acción futura.

**BI Dashboard:** Son una herramienta de visualización de datos que muestra el estado actual que arrojan los KPI's para una empresa. Éstas herramientas consolidan y ordenan los números, las métricas y en ocasiones los cuadros de mando de rendimiento en una sola pantalla. Pueden ser adaptados para métricas específicas de roles y visualización dirigidas a un solo punto de vista o departamento. Las características esenciales de un BI Dashboard incluyen una interfaz personalizable y la capacidad de presentar datos en tiempo real desde múltiples fuentes.

**Balanced ScoreBoard:** se trata de un enfoque integral que analiza el desempeño general de una organización desde cuatro puntos de vista, análisis financiero (costos y retornos de una operación), análisis de clientes (satisfacción y fidelización), análisis interno (producción e innovación junto con la medición del desempeño) y el análisis del crecimiento (satisfacción de los empleados y el rendimiento de la retención del sistema de información).

Los procesos y técnicas de segmento BA han permitido transformar la ingente cantidad procedente del DW en información útil y comprensible por el usuario. Es ahora el momento de convertir toda esta información en conocimiento que permita tomar decisiones.

#### 2.2.4.- Decision Making

La toma de decisiones consiste en el proceso de seleccionar una concreta de entre un grupo de opciones disponibles. Cuando se intenta tomar una decisión, una persona debe ser capaz de valorar los aspectos positivos y negativos de cada alternativa, siendo capaz de anticiparse a las posibles consecuencias de cada opción eligiendo la que más se ajuste a la situación concreta en la que se encuentre (Business Dictionary, 2014).

Como se ve en la anterior definición, la toma de decisiones se fundamenta en el campo en el que BI ejerce su influencia, realizando la labor de soporte de decisiones y ofreciendo todo su potencial que, explotado apropiadamente, permite adoptar estrategias tácticas y operativas para mejorar el rendimiento del negocio (Evelson, 2008). Sin embargo, como se ha expuesto varias veces a lo largo de este documento, la toma de decisiones en BI depende exclusivamente de la calidad de los datos en las que esté cimentada, que sean fiables, actualizados y estén disponibles cuando se necesiten.

En este segmento de la arquitectura BI no existe software como tal ni soluciones implementadas de relevancia, pues todo el procesamiento de los datos para obtener la información necesaria y consistente para tomar decisiones se ha realizado ya a lo largo de la “pirámide” BI. Por tanto, la pieza más importante del segmento, por no decir la única, es el usuario final. En consecuencia a esto, el principal elemento a tener en cuenta aquí, es el factor humano (Skirus, 2012).

#### FACTOR HUMANO

Sin duda alguna, son los usuarios finales la parte fundamental de la estructura BI, desde la elección de la tecnología, el desarrollo o la provisión de datos afectan en gran medida a los resultados arrojados. Cuando se considera el factor humano, la diferencia entre la toma de decisiones reactiva y proactiva, se puede identificar con el nivel de sofisticación de los métodos de BA de la solución BI implementada. De tal forma que los métodos BA más sencillos simplemente permitirían realizar un análisis descriptivo, permitiendo tomar decisiones tras la ocurrencia de un evento, mientras que los métodos más avanzados, predictivos y prescriptivos, permitirán anticiparse a las situaciones.

Se ha demostrado que la inteligencia funciona mejor cuando se organiza como un esfuerzo de grupo entre todas las partes involucradas. Coordinando conveniente este esfuerzo, estableciendo prioridades, responsabilidades y procedimientos al tiempo que se genera un intercambio horizontal eficiente de información entre los participantes se logra una optimización del proceso estableciendo una cultura de la inteligencia del negocio. Existen muchos casos de fracasos de soluciones BI debido a la reticencia de los usuarios a proveer información a los sistemas o incluso el caso contrario, la inserción de una gran cantidad de datos pero de baja calidad (UOC, 2012).

Por tanto, al igual que cualquier otra herramienta, el desempeño final de la misma depende en último término del uso que se haga de ella. Por avanzada, precisa y potente que sea ésta, los resultados obtenidos pueden quedar muy por debajo de las expectativas generadas si no se cumple con el uso especificado de la misma.

#### 2.3.- Conclusión

BI es una herramienta de soporte de decisiones avanzado y de gran potencial que puede resumirse en el proceso de convertir los datos en información y la información en conocimiento. Esta estructura recibe datos desde múltiples y muy variadas fuentes y es de la calidad de estos datos de la que depende la calidad de la

información ofrecida. BI dispone de muy avanzadas herramientas que permiten anticiparse a las circunstancias, pero su desempeño final dependerá del uso que los usuarios finales hagan de la misma tanto en la entrada como en los resultados ofrecidos.

#### **2.4.- Business Intelligence vs Big Data**

Antes de finalizar este apartado de exposición de las características y propiedades del Business Intelligence, es interesante enfrentar el concepto de BI al de Big Data. Realmente, incluso su definición hace que sea difícil de diferenciar pues se puede entender que el término Big Data se refiere a sistemas informáticos basados en la acumulación a gran escala de datos y de los procedimientos usados para identificar patrones recurrentes dentro de esos datos (Gartner, 2014).

La diferencia por tanto es más sutil que el propio propósito de ambos conceptos. BI se centra en la estructuración de información empresarial útil y relevante para la toma de decisiones corporativas permitiendo generar sistemas de previsión pero siempre basándose en información existente y conocida.

Big data por su parte trata información obtenida desde el interior del propio sistema o desde entornos dispares como pueden ser las redes sociales independientemente de la utilidad o incluso la confiabilidad de los datos pero tratando de localizar patrones y tendencias dentro de los mismos y con éstos sustentar la toma de decisiones.

La diferencia por tanto por se encuentra en la procedencia en sí de los datos y cómo son éstos tratados por cada uno de los paradigmas de soporte de decisión. Sin embargo, esta línea de por sí difusa y tenue presenta tendencia a desaparecer. En la actualidad, los principales proveedores de soluciones BI están apostando por la inclusión de productos para facilitar la extracción y visualización de información de Big Data, es la transición normal que deben llevar los grandes proveedores de soluciones BI: Microsoft con Insights (HD Insight + Hadoop), Oracle Big Data, Pentaho Big Data Analytics, QlikView... Por tanto la tendencia natural de BI y de Big Data será la de unirse a fin de ofrecer a las organizaciones o negocios la mejor solución posible que actúe como soporte en su toma de decisiones (García González, 2014).



### 3.-Series Temporales

Si en este documento el Business Intelligence es el “*Qué*”, sin duda las series temporales y su análisis son el “*Cómo*”. En este capítulo se ofrecerá una visión de todo lo concerniente a las series temporales, desde su definición hasta su análisis, procurando en la medida de lo posible evitar entrar en un nivel de detalle muy profundo y emplear complejas definiciones matemáticas pues quedan muy lejos de los objetivos de este proyecto.

#### 3.1.- ¿Qué son las series temporales?

También llamadas Series Cronológicas, son conjuntos de observaciones sobre un hecho determinado recolectadas a lo largo del tiempo. Es fácil encontrar ejemplos de las mismas en muy diversos campos, desde la economía, los precios diarios de ventas de determinados productos, las exportaciones totales, los beneficios trimestrales... pasando por la meteorología, temperaturas, precipitaciones, presión atmosférica... o la demografía, nacimientos, defunciones, índice de población anual... (García, 2009).

Las observaciones de las series temporales se pueden realizar sobre una sola característica, lo que se conoce como serie variante o escalar, o sobre varias características, series multivariantes o vectorial, de una unidad observable en diferentes momentos (Mauricio, 2013).

Matemáticamente se pueden expresar de la siguiente forma:

Serie variante:

$y_1, y_2, \dots, y_N; (y_t)_{t=1}^N; (y_t : t = 1, \dots, N)$  Donde  $y_t$  es la observación nº  $t$  ( $1 \leq t \leq N$ ) de la serie y  $N$  es el número de observaciones de que consta la serie completa (el tamaño o la longitud de la serie). Las  $N$  observaciones  $y_1, y_2, \dots, y_N$  pueden recogerse en un vector columna  $\mathbf{y} \equiv [y_1, y_2, \dots, y_N]'$  de orden  $N \times 1$ .

Series Multivariantes:

$y_1, y_2, \dots, y_N; (y_t)_{t=1}^N; (y_t : t = 1, \dots, N)$  Donde  $\mathbf{y} \equiv [y_{t1}, y_{t2}, \dots, y_{tN}]'$  ( $M \geq 2$ ) es la observación nº  $t$  ( $1 \leq t \leq N$ ) de las serie y  $N$  es el número de observaciones de que consta la serie completa. Las  $N$  observaciones  $y_1, y_2, \dots, y_N$  pueden recogerse en una matriz  $\mathbf{Y}$  de orden  $N \times M$ .

$$\mathbf{Y} \equiv \begin{bmatrix} \mathbf{y}'_1 \\ \mathbf{y}'_2 \\ \vdots \\ \mathbf{y}'_N \end{bmatrix} \equiv \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1M} \\ y_{21} & y_{22} & \cdots & y_{2M} \\ \vdots & \vdots & & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NM} \end{bmatrix}$$

Donde  $y_{tj}$  es la observación nº  $t$  ( $1 \leq t \leq N$ ) sobre la característica o variable nº  $j$  ( $1 \leq j \leq M$ ) que es la misma en todo momento  $j$ .

### 3.2.- ¿De qué se componen las series temporales?

La representación de las series temporales se realiza situando en el eje horizontal la escala del tiempo y en el eje vertical el valor de las observaciones correspondientes. Es habitual observar que los datos aparentemente fluctúan a lo largo del tiempo siguiendo algún patrón interno, formalmente, este hecho responde al concepto de proceso estocástico<sup>4</sup>. Un ejemplo sería el de la Figura 5:

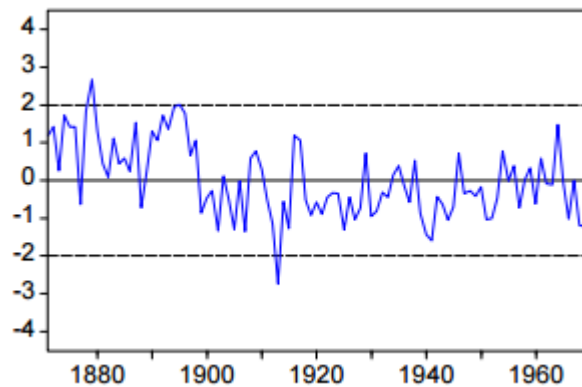


Figura 5 Volumen del flujo anual del río Nilo en Asuán 1871 - 1970 (nilo.wf1)

En este ejemplo se pueden observar las principales componentes de variación de datos. Estas componentes se identifican generalmente como:

**-Componente estacional:** también conocida como efecto estacional de una serie temporal se puede definir como una conducta repetitiva a lo largo del tiempo, provocada generalmente por factores que se presentan de forma periódica (trimestralmente, mensualmente, semanalmente...) y que influyen en el comportamiento de la serie. Así pues, por ejemplo, la lectura del consumo mensual de electricidad presenta picos altos de consumo en los meses de verano e invierno (debido a la climatización) reduciéndose en los meses de primavera y otoño. Normalmente se identifica la periodicidad del efecto estacional con una duración igual o inferior a un año.

**-Componente cíclica:** o cambio cíclico se identifican como conductas repetitivas sin periodo fijo, debido a algunas otras causas físicas o bien la periodicidad de las mismas es superior a un año, debido generalmente a las fluctuaciones de la actividad económica y social.

**-Componente de tendencia:** o simplemente tendencia, recoge la variación de la serie a lo largo del tiempo. Se presenta la tendencia cuando hay una variación significativa en el valor medio de la serie conforme vamos añadiendo datos a la misma. Existen series económicas que presentan efectos cíclicos con un periodo de

<sup>4</sup> Secuencia de variables aleatorias, ordenadas y equidistantes cronológicamente, referidas a una características (proceso univariante o escalar) o a varias (proceso multivariantes o vectorial) de una unidad observable en diferentes momentos. Como restricciones es este tipo de procesos, se acepta que deben ser:

-Estacionarios: las relaciones existentes entre las variables aleatorias que lo componen deben ser estables en el tiempo.

-Ergódico: se utiliza la función de correlación, es decir, que conforme  $k$  se hace más grande, la autocorrelación se hace más pequeña. Esto significa que lo que ocurre hoy, conforme va pasando el tiempo va teniendo menos importancia.



medio siglo o más incluso, lo que hace que si existen pocos datos de tales series, los efectos del ciclo se puedan confundir con la tendencia. Debido a este hecho es habitual agrupar ambas componentes en la llamada componente tendencia-ciclo.

**-Componente Irregular:** o variaciones irregulares que corresponden con todas aquellas variaciones no explicadas por ninguna de las anteriores componentes.

Para ilustrar la anterior exposición, se ofrece el siguiente ejemplo (Figura 6). En él, se observa la evolución del paro en España entre enero de 1997 y diciembre de 2002 (INE, 2002).

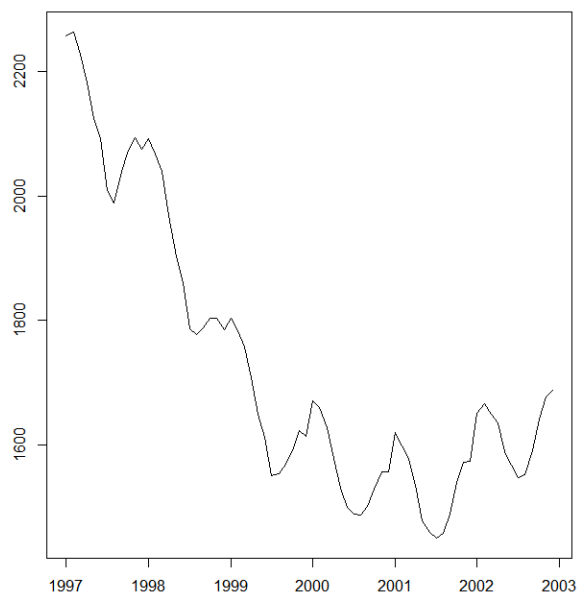


Figura 6.- Paro en España 97-02

Descomponiendo la serie (Figura 7), se obtiene la tendencia junto con la componente cíclica.

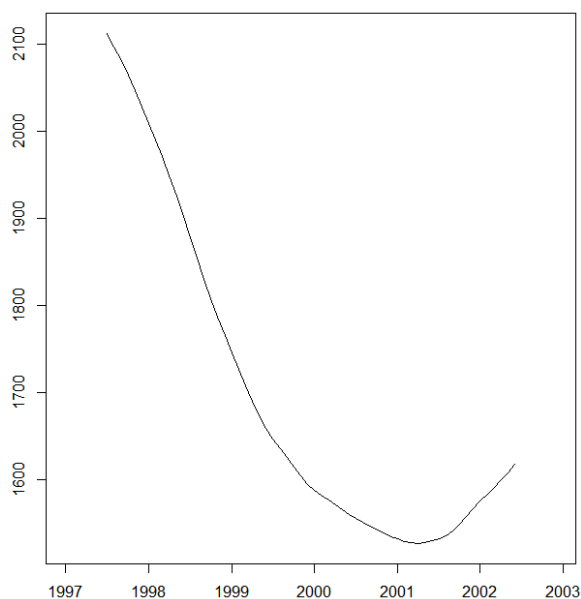


Figura 7.- Tendencia Paro en España 97 - 02

Del mismo modo, es posible obtener la estacionalidad de la misma (Figura 8).

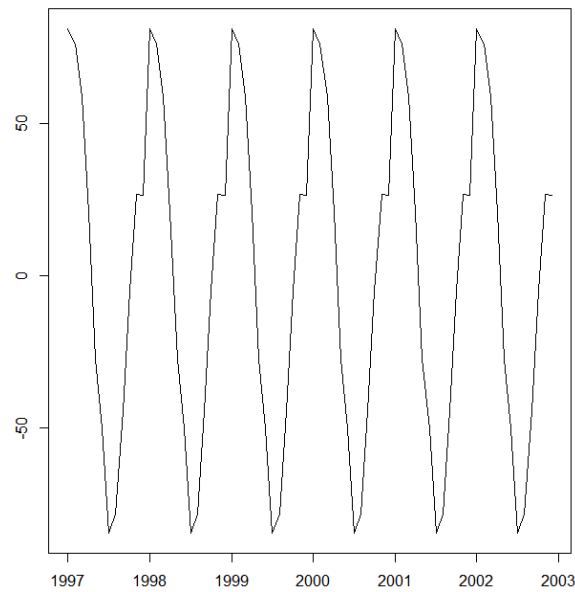


Figura 8 .- Estacionalidad del Paro en España 97 - 02

Por último se muestra la componente aleatoria (Figura 9).

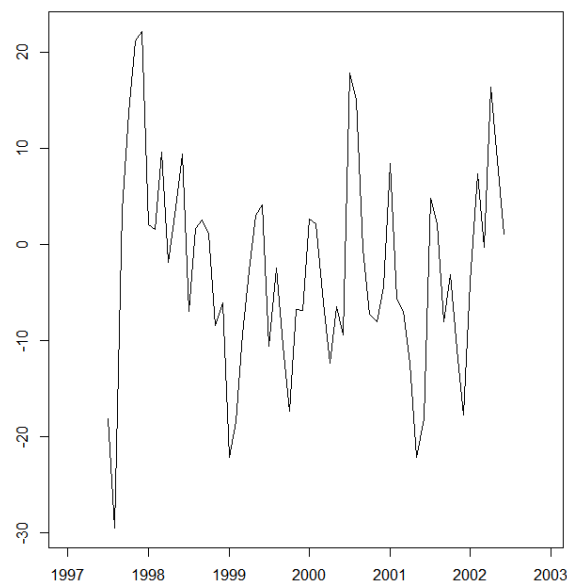


Figura 9 .- Componente Aleatoria del Paro en España 97 - 02

### 3.2.- ¿Cómo se clasifican las series temporales?

Las series temporales se pueden clasificar atendiendo a varios criterios como la capacidad de observación en el tiempo, el número de variables que se ven afectadas o el modo en que varían a lo largo del tiempo. Atendiendo a la observación de la muestras, las series temporales podrían clasificarse en:

**Discretas:** Son las series cuyas variables únicamente pueden ser observadas en instantes aislados del tiempo.

**Continuas:** Hacen referencia a las series que pueden ser observadas en cualquier instante de tiempo.

Es importante resaltar, que esta clasificación está más relacionada con la capacidad de observación de los datos que con la propia naturaleza de las variables. Es por tanto habitual disponer de series discretas obtenidas al observar variables (tanto discretas como continuas) en intervalos marcados por horas, días, meses, etc. Por tanto, las series discretas pueden surgir de varias maneras:

**-Muestral:** Dada una serie de tipo continua es posibles construir una serie de tipo discreta asociada a ella, tomando los valores en intervalos de tiempo de igual longitud. Un ejemplo de serie temporal de tipo continua es la temperatura en un lugar determinado. Teóricamente, se podría observar la temperatura en cualquier instante, sin embargo, si es observada cada hora se obtendría una serie temporal discreta.

**-Agregada o acumulada:** Este tipo de series se obtiene cuando se contabiliza el número de veces que ocurre un determinado suceso en intervalos de tiempo igualmente espaciados. Las series observadas pueden ser, por ejemplo, las dadas por las lluvias recogidas anualmente, los accidentes de tráfico mensuales o el número de pasajeros mensuales en las líneas aéreas. En cualquiera de estos casos lo que se observa en cada periodo considerado (años, meses, etc.) es una agregación o acumulación de sucesos discretos (lluvias torrenciales, accidentes de tráfico, número de pasajeros)

**-Inherentes o discretas.** Todas las que se obtienen al observar realmente los datos en momentos discretos, por ejemplo, el salario mensual.

Por otro lado, si se observa el número de variables involucradas en una serie temporal las series podrían agruparse en:

**Series Univariantes:** son series que se sustentan en una única variable (escalar) observada en el tiempo. Aunque este tipo de series se representan generalmente como un único conjunto de datos, el tiempo es, de hecho, una variable implícita. Si los datos son equiespaciados, la variable del tiempo se encuentra implícitamente dada, sin embargo es esta variable la que se utiliza para representar gráficamente la serie. Como ejemplo de series temporales se puede citar la concentración mensual de  $CO_2$  (Figura 10).

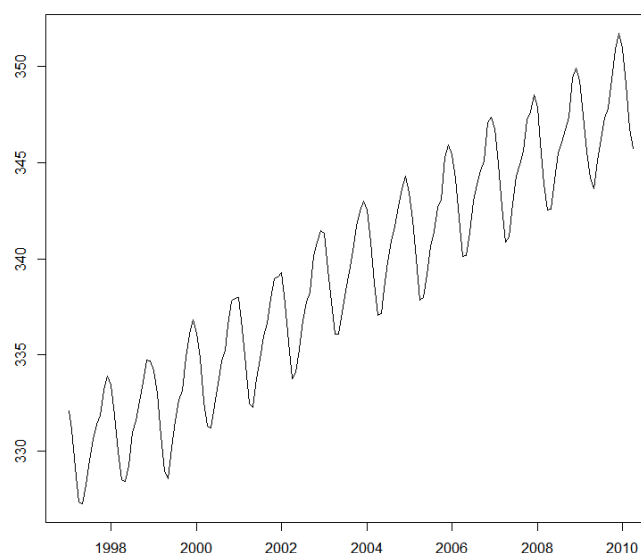


Figura 10.- Concentración CO2 97-10

**Serie Multivariantes:** Este tipo de series, se fundamentan en la observación de varias variables interrelacionadas (vectorial). En esencia, este tipo de series se basan en modelos que persiguen explicar una variable, generalmente aleatoria, a través de otras. El estudio de estas series se realiza mediante el empleo de métodos que buscan la simplificación de la misma. Estos modelos se pueden agrupar en diversos tipos:

-**Métodos de dependencia:** se utiliza generalmente para tratar de explicar cómo una variable puede ser prevista conociendo otra e intentar predecir el comportamiento de ciertas variables a partir de otras.

-**Métodos de interdependencia:** clasifica una muestra de entidades (individuos o variables) en un número pequeño de grupos de forma que las observaciones pertenecientes a un grupo sean muy similares entre sí y muy disimilares del resto

-**Métodos Estructurales:** analizan las relaciones existentes entre un grupo de variables representadas por sistemas de ecuaciones simultáneas en las que se suponen que algunas de ellas (denominadas constructos) se miden con error a partir de otras variables observables denominadas indicadores.

Atendiendo a las características que presenten algunas de sus propiedades, como por ejemplo la media o la varianza, se puede establecer otra categorización. De este modo, se pueden encontrar.

**Serie Estacionaria:** Son aquellas en las que la media y la varianza se mantienen estables a lo largo del tiempo en torno a los mismos valores, lo que se conoce como equilibrio estadístico. Como ejemplo clásico de este tipo de serie se podría citar el ruido blanco (Figura 11).

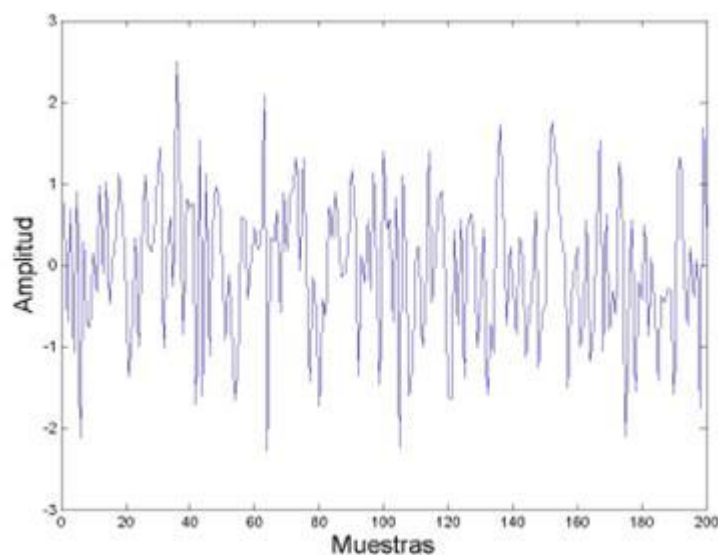


Figura 11.- Ruido Blanco

**Serie No Estacionaria:** Por tanto, en el otro extremo, se encuentran las series cuyas propiedades tienden a variar con el tiempo. Pueden por tanto aparecer cambios en la media o la varianza, apareciendo tendencias o incluso efectos estacionales, lo que son comportamientos parecidos en distintos periodos de tiempo (Figura 12).

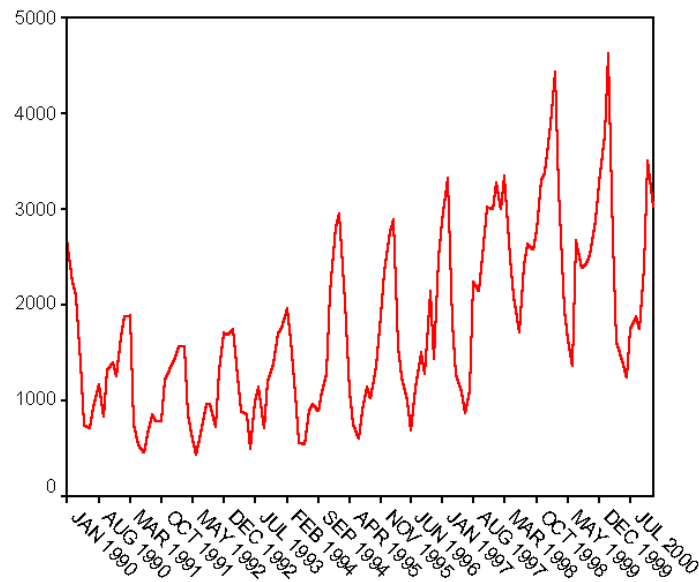


Figura 12.- Serie No Estacionaria

Por último, las series se pueden agrupar dependiendo de la relación de sus componentes, de esta forma pueden presentar un modelo:

-**Aditivo.**  $x_t = T_t + E_t + C_t + R_t$

-**Multiplicativo.**  $x_t = T_t \cdot E_t \cdot C_t \cdot R_t$

-**Mixto.**  $x_t = T_t + E_t \cdot C_t \cdot R_t$   
 $x_t = T_t + E_t \cdot R_t$   
 $x_t = T_t \cdot E_t \cdot C_t + R_t$

Esta clasificación se expondrá en mayor profundidad en el siguiente apartado, que trata del análisis de las series temporales.

### 3.3.-Análisis de las Series Temporales.

Intrínsecamente, las series temporales contienen información que debidamente estudiada puede permitir conocer los valores futuros que pueda presentar. El estudio de las series temporales permite comprender un conjunto de técnicas destinadas a alcanzar este objetivo mediante la generación de un modelo que explique el comportamiento evolutivo del fenómeno observado a lo largo del tiempo. Este análisis pretende extraer las regularidades que se observan en el comportamiento pasado del mecanismo que la genera.

El estudio de una serie temporal puede responder a distintas motivaciones y, en atención a esto, el modelo de estudio puede variar, dependiendo igualmente de las preguntas que se pretenda responder. Si el objeto de estudio comprende únicamente una variable, el objetivo suele ser construir un modelo para explicar la estructura (descripción) y prever la evolución (predicción) de la variable de interés. Si, por el contrario, se observa simultáneamente un grupo de variables, el objetivo que se persigue consiste generalmente en analizar las posibles relaciones entre las variables observadas así como su evolución conjunta. Este documento se centra en el primer caso, quedando el segundo por tanto fuera de los límites cubiertos por el mismo (Esparza Catalán, 2001).

### 3.3.1.- Análisis preliminar de una serie temporal.

Esta fase constituye en el primer paso a seguir a la hora de estudiar una serie temporal, permitiendo detectar las características más importantes de una serie, su tendencia, la existencia de ciclos, presencia de valores atípicos... Como primer paso, es interesante conformar un gráfico de secuencia dando la posibilidad de obtener la información anteriormente citada de un modo visual. Como ejemplo, se presenta el volumen de pasajeros de una compañía aérea estadounidense entre 1949 y 1960 (Figura 13).

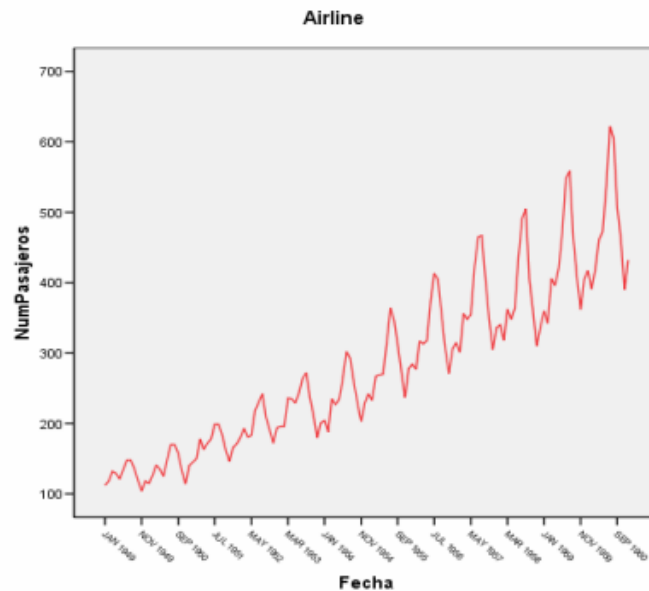


Figura 13.- Pasajero aerolínea 49 - 60.

En adelante, se expondrán algunas de las características más importantes que pueda presentar una serie.

#### Tendencia y nivel de una serie.

El nivel de una serie es una medida local de tendencia central, como por ejemplo la media, de cada periodo de tiempo considerado. Una vez establecido el nivel de una serie, se debe estudiar si la medida de tendencia centra elegida tiene valores similares en los periodos de tiempo establecidos, siendo necesario observar la tendencia en los mismos, es decir, si presenta una dirección constante de cambio de nivel. En las serie de la compañía aérea presentada anteriormente, se observa que el número de pasajeros va aumentando progresivamente, de manera que presenta tendencia creciente.

El nivel de una serie puede ser:

**-Estable:** Siempre se encuentra entorno a unos valores semejantes.

**-Inestable:** En este caso los valores cambian, sin embargo pueden presentar cambios sin tendencia, es decir que los valores no siguen un patrón determinado, o presentar una tendencia siguiendo una línea de crecimiento o decrecimiento.

**-Circunstancial:** El fenómeno que explica la serie suele presentar un nivel estable acompañado de determinados comportamientos estacionarios.

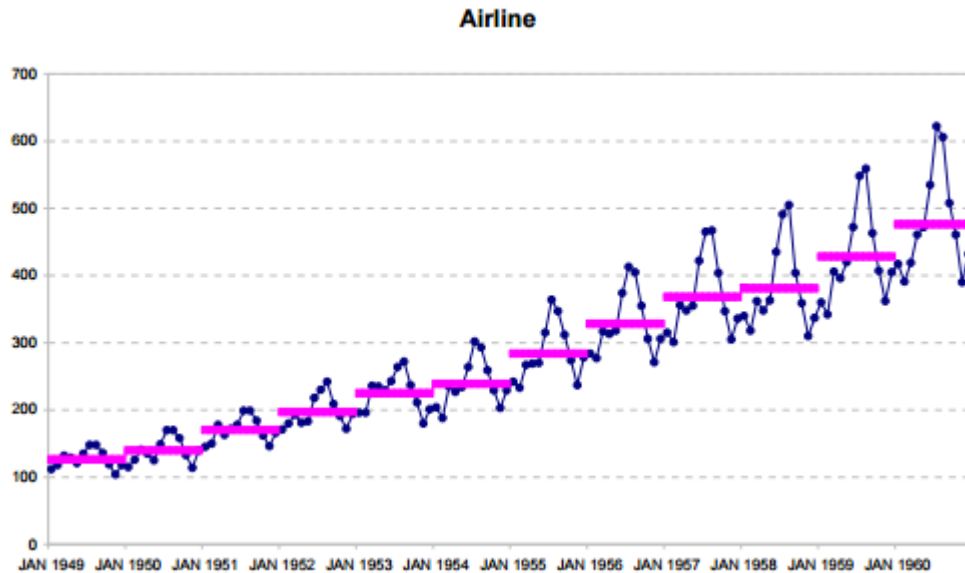


Figura 14.- Nivel de la serie aerolínea 49-60.

### Estacionalidad

Tal y como se comentó cuando se presentaban las componentes que contiene una serie, la estacionalidad o ciclicidad, se identifica con un patrón sistemático que se repite periódicamente. Este comportamiento es fácilmente detectable en un gráfico de secuencia, tal como se observaba en el ejemplo de la línea aérea presentado anteriormente. Sin embargo es más efectivo para este objetivo, representar la serie mediante un diagrama de cajas (caja y bigotes) o box-plot, realizando una división por meses. De esta forma se observa la presencia en la serie de un pico en verano y otro menos pronunciado en la época navideña.

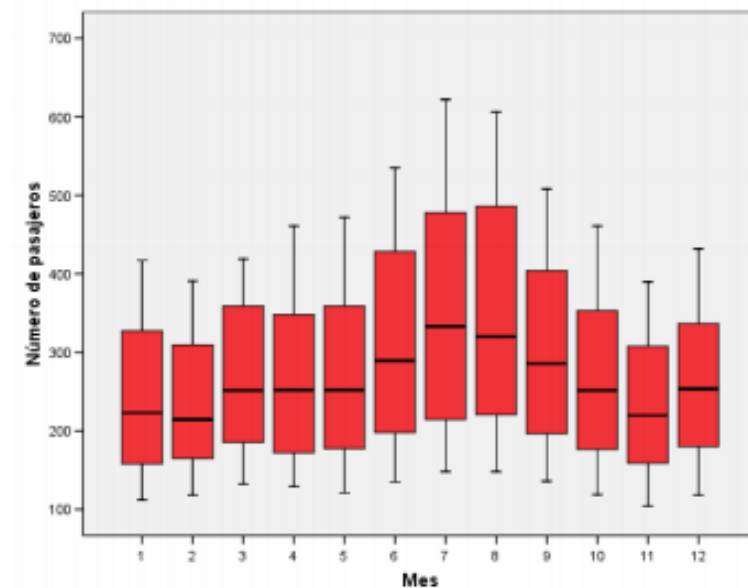


Figura 15.- Diagrama de cajas por mes de la aerolínea.

### Dependencia entre variabilidad y nivel

Otra de las cuestiones a determinar cuándo se realiza el análisis preliminar de una serie es si existe dependencia entre su variabilidad y su nivel. Si la variabilidad de una serie no depende del nivel significa que las componentes de la serie se combinan de forma aditiva, es decir, el incremento debido a la estacionalidad siempre es el mismo aunque exista tendencia creciente o decreciente. Si la variabilidad y el nivel dependen entre sí los elementos de la serie se combinan de forma multiplicativa, por lo que el incremento debido a la estacionalidad aumenta o disminuye conforme la tendencia crece o decrece. Esta diferenciación se expresa mediante un diagrama de dispersión en el que se representa el logaritmo neperiano de la mediana (medida de tendencia central) frente a l logaritmo neperiano de la diferencia entre los percentiles 80 y 20 (medida de la dispersión) de cada uno de los periodos considerados en la serie.

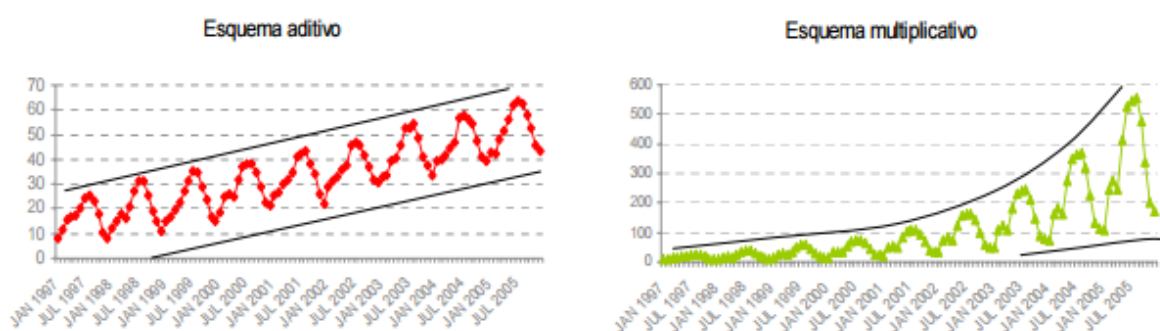


Figura 16.- Representación de dos series con distintos esquema, aditivo o multiplicativo.

Considerando el ejemplo anterior, se puede observar en el gráfico de dispersión una importante dependencia entre la variabilidad y el nivel, ya que los puntos del diagrama pueden ajustarse bastante a una recta, por tanto se presenta un esquema multiplicativo. Se puede también observar que la variabilidad crecer con el nivel, es decir, que cuando mayor es el nivel correspondiente un año de la serie mayor es su variabilidad.

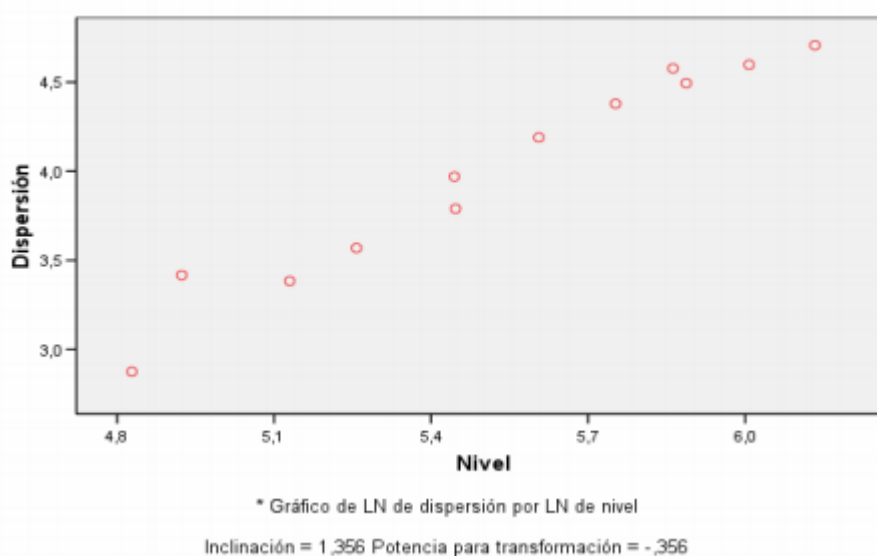


Figura 17.- Gráfico de dispersión por nivel de la serie de la aerolínea.



## Comportamientos anómalos

Como último aspecto a tener en cuenta a la hora de realizar el análisis preliminar de una serie temporal es el de los comportamientos extraños. Si una serie temporal tiene valores perdidos o valores raros, no es posible ignorarlos. Los comportamientos anómalos pueden englobarse en tres posibles tipos:

-**Cambios de tendencia:** En estas series se observa un cambio de tendencia creciente a decreciente o viceversa. Un ejemplo sería el paro entre 1982 y 1997 en USA (Figura 18).

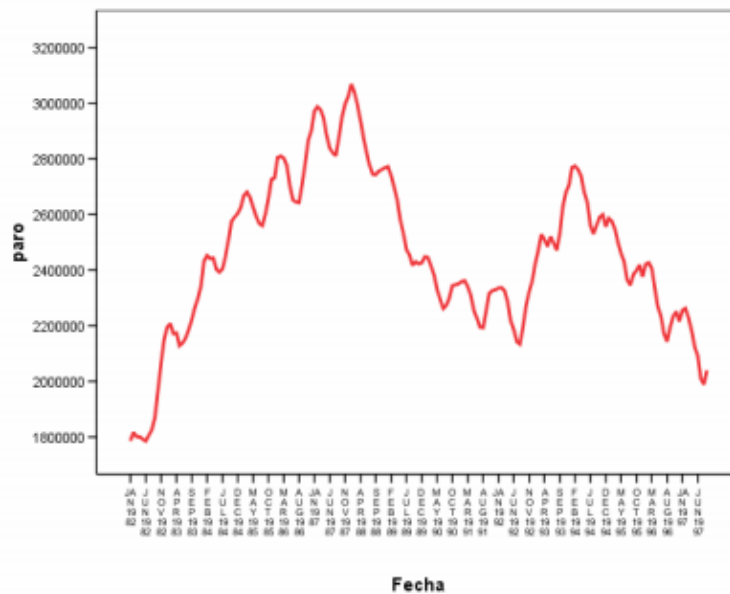


Figura 18.- Cambio de tendencia en la serie del paro en USA.

-Cambios bruscos de nivel: Anomalía en la que se observan subidas o bajadas bruscas del nivel en una serie. Como ejemplo se presenta tanto el histograma como el diagrama de cajas del paro femenino entre 1961 y 1985 en USA (Figura 19).

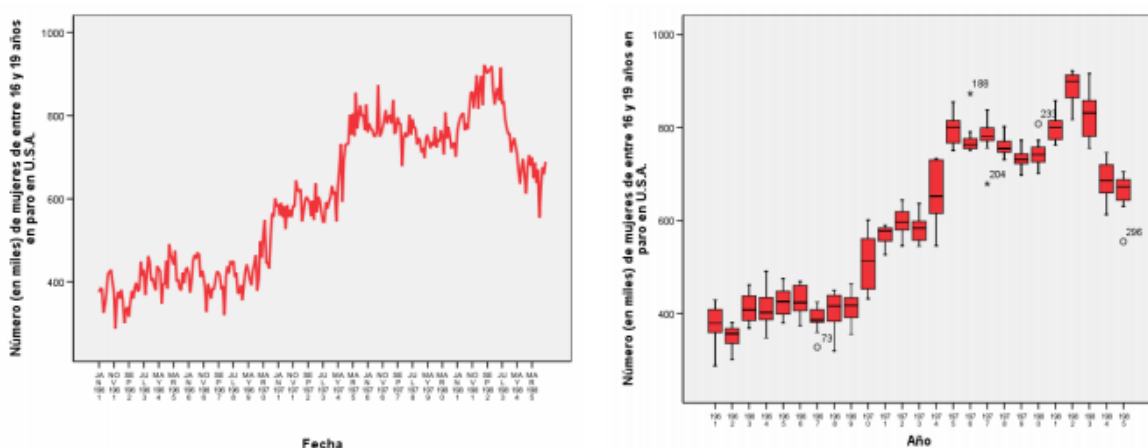


Figura 19.- Histograma y diagrama de cajas del paro femenino en USA entre 1961 y 1985.

-**Valores raros:** También llamados outliers, representan valores anormalmente bajos o altos estando fuera del nivel habitual de la serie o del entorno de puntos. Ante estos valores, es necesario

determinar si se trata de un error en la recogida de los datos o de causas externas que han afectado a la variable.

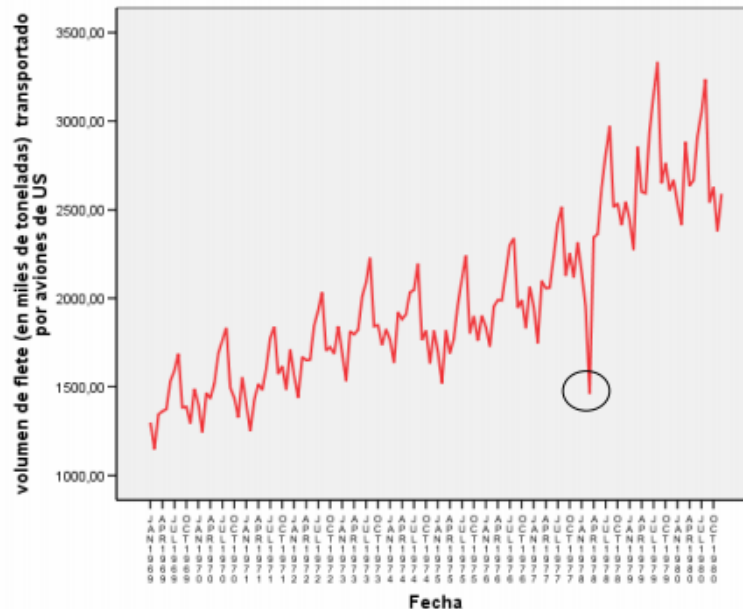


Figura 20.- Volumen de flete por aviones en USA entre 69-80.

### 3.3.2.- Métodos Clásicos de estudio.

El análisis de las series temporales considera el comportamiento de una variable en el tiempo como el resultado de la integración de cuatro las componentes de la serie (tendencia, cíclica, estacional y aleatoria) quedando por tanto expresada la serie en función de estas cuatro componentes  $X_t = f(C_t, T_t, S_t, E_t)$ . Los métodos clásicos de análisis de series no son excesivamente complejos, pero su vez, responden a preguntas menos ambiciosas. Se pueden emplear para realizar predicciones a corto plazo, pero no son efectivos ni fiables en periodos de tiempo mayores.

#### Método de descomposición estacional y ajuste de tendencia.

Los métodos de descomposición estacional buscan la división de la serie en las sub-series que representa cada una de las componentes de la misma, aunque la dificultad para separar la tendencia y el ciclo motiva que se presenten juntas. Es importante resaltar, que para aplicar este método, la serie debe presentar un comportamiento estacional.

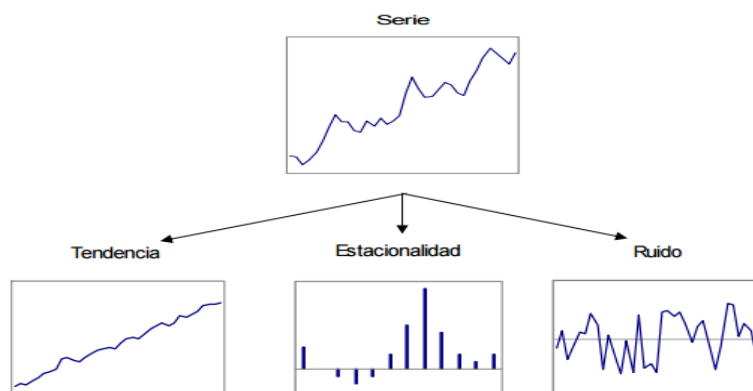


Figura 21.- Muestra de la representación de una serie dividida.

Motivados por las necesidades u objetivos de estudio o por las características intrínseca de la componentes de la serie (que la fuerza de una no permita ver otras) los métodos de descomposición permiten observar cada una de las componentes de forma aislada. Sin embargo estos métodos no realizan predicciones, para ello es necesario combinarlos con métodos de ajustes de tendencia.

Cuando se realiza una predicción, se considera la estacionalidad constante periodo a periodo y se desprecia el ruido, pues es impredecible por su propia naturaleza y tiene tendencia a cero. Por tanto el proceso consistirá en realizar un ajuste de la tendencia y posteriormente añadir la estacionalidad.

Como primer paso a la hora de descomponer una serie es determinar cómo se combinan sus componentes, es decir, si presentan un esquema multiplicativo, aditivo o mixto. Los procesos de estudio son relativamente similares, en este caso se realizará un proceso aditivo. En la sección de anexos, el análisis de series mediante herramientas software, se ha llevado a cabo siguiendo un esquema multiplicativo.

Una vez determinado el esquema que sigue la serie, se hace necesario eliminar el ruido (componente aleatoria) y la estacionalidad. El ruido se elimina sustituyendo cada observación por lo ocurrido anteriormente (media móvil anterior) y la estacionalidad realizando un proceso de media móvil centrada. Este último procedimiento suaviza cada observación tomando la media de igual número de valores anteriores y posteriores a la misma. El origen de la media móvil centrada, es decir, el número total de observaciones que generará cada media móvil centrada, habitualmente es igual al periodo de la serie. En cualquier caso deber ser tal que no incluya más observaciones de una unidad de periodo que de las demás. Es decir, si la serie tiene periodicidad semanal, en la media móvil centrada no deben estar incluido los valores correspondientes a dos lunes si no se incluyen también dos veces los del resto de los día de las semana.

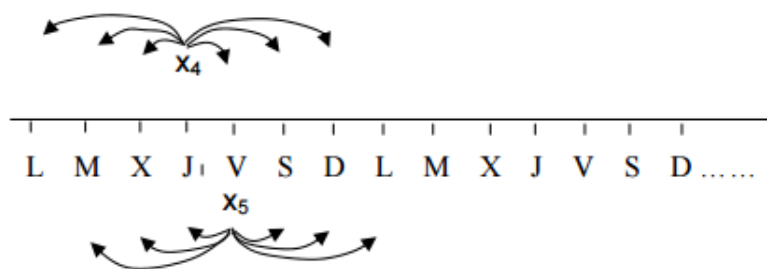


Figura 22.- Media móvil centrada de una serie con información diaria (periodo 7).

En el caso de que el periodo de la serie sea par la media móvil centrada se calcula tomando la media de las dos posibles medias móviles a considerar.

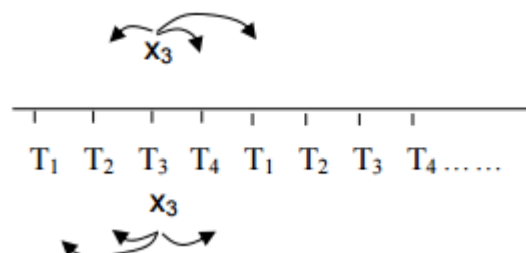


Figura 23.- Posibles medias móviles centradas en una serie con información trimestral.

Una vez eliminados la estacionalidad y el ruido se tiene una serie compuesta únicamente por la tendencia y el ciclo:  $M_t = C_t + T_t$ . Del mismo modo, se elimina la tendencia y el ciclo de la serie original para conseguir una serie integrada sólo por la estacionalidad y el ruido:  $X_t - M_t = E_t + S_t$ . Para estimar el factor estacional a partir de esta última serie se realiza la media de todas las observaciones disponibles de cada unidad de periodo (por ejemplo, cada mes de un año). De esta forma la serie del factor estacional estará formada por los valores de estas medias, es decir, será una serie que repetirá constantemente los mismos valores en cada periodo. Igualmente, se puede obtener una serie compuesta únicamente por el ruido del siguiente modo:  $X_t - M_t - E_t = S_t$ .

Como ejemplo de aplicación, se estudiarán las ventas de cava en España entre los años 1962 y 1969 (Figura 24). En primer lugar, se presenta el gráfico de secuencia y el de cajas de las ventas, donde se observa un comportamiento estacional que presenta un aumento en los últimos meses del año, cuando se acercan las fechas navideñas.

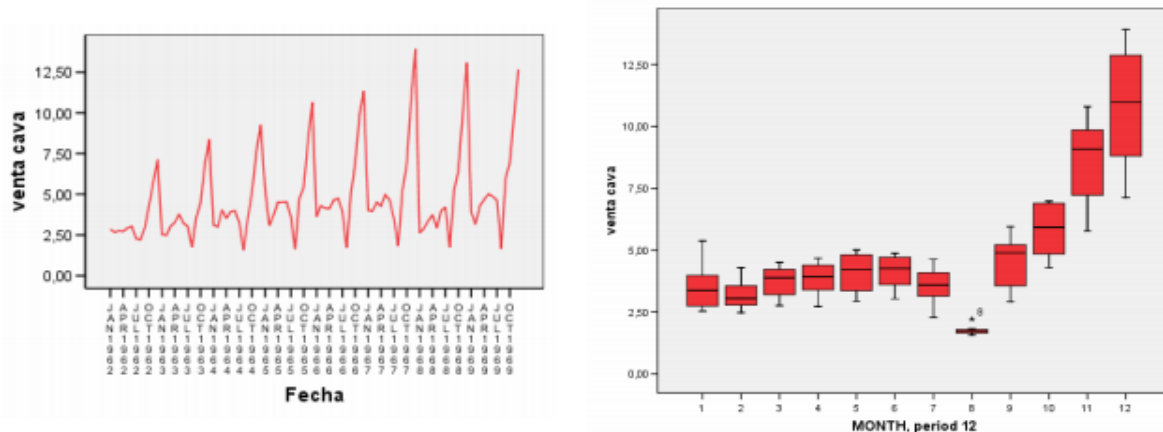


Figura 24.- Diagrama de secuencia y cajas de las ventas de cava en España 62-69.

Si observa el gráfico de dispersión que arroja la serie, se observa que los puntos se ajustan a una recta, lo que implicaría dependencia, sin embargo existen puntos que se alejan de la misma, descendiendo el valor de  $R^2$  hasta 0.494. Si se observa el diagrama de cajas por año, la variabilidad no parece depender del nivel, por tanto se trataría de puntos anómalos que no representarían el comportamiento real de la serie, por lo que es posible considerar que no existe dependencia entre nivel y variabilidad, aceptando por tanto que se trata de un esquema aditivo.

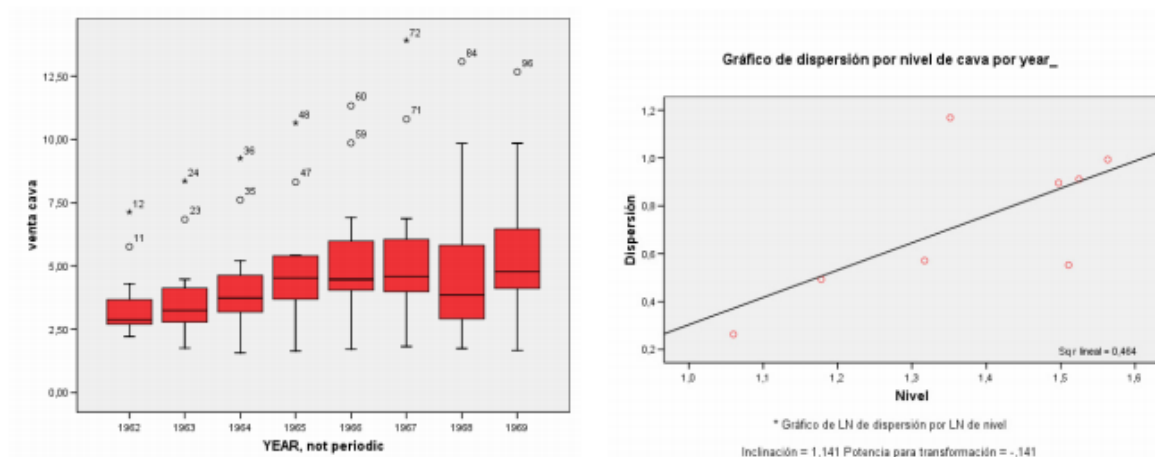


Figura 25.- Diagrama de cajas y gráfico de dispersión por nivel de la serie de venta de cava.

Tras estas consideraciones se pueden aplicar los métodos de división de la serie que se presentaron anteriormente, permitiendo contrastar la serie original con la misma serie al eliminar la estacionalidad.

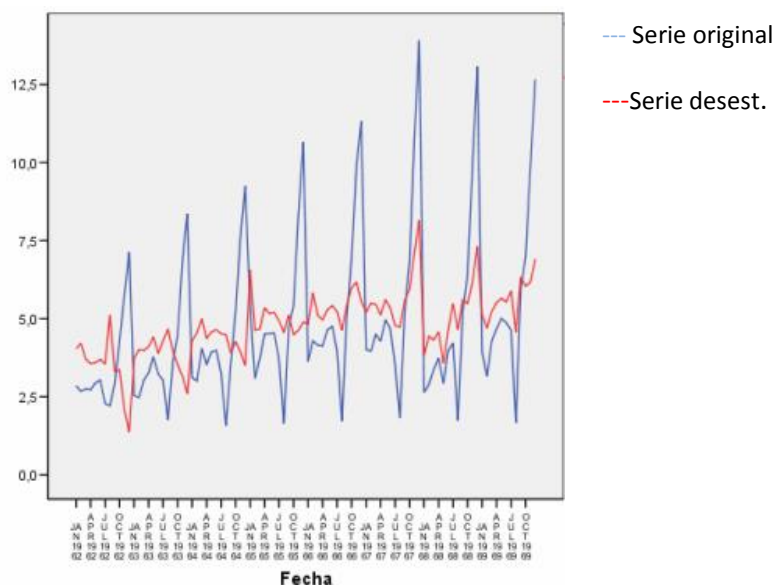


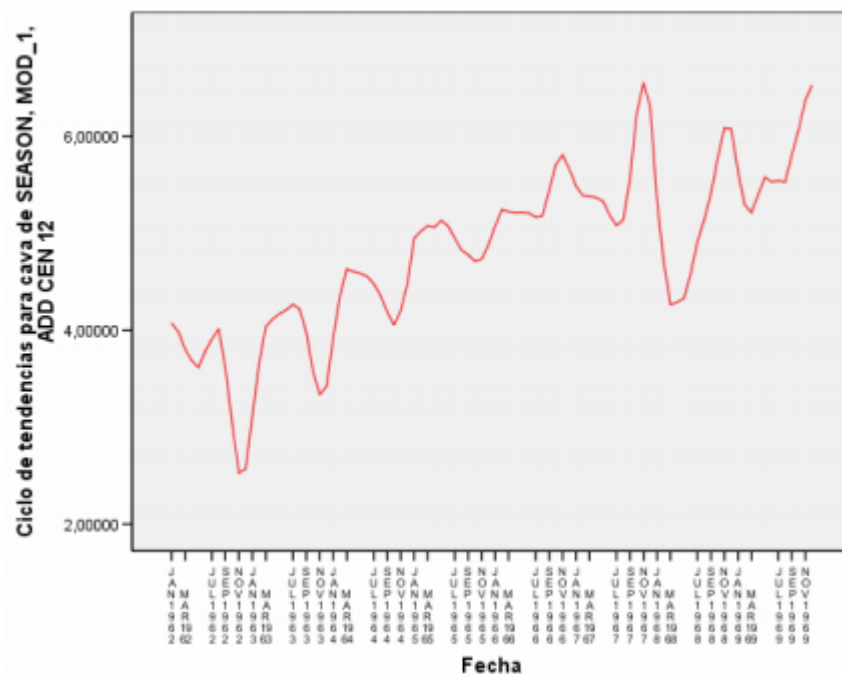
Figura 26.- Gráfico de secuencia de la serie original y la serie desestacionalizada

Una vez llegados a este punto, es muy posible que el objetivo final de estudio sea realizar una predicción, en primer lugar, es necesario ajustar la tendencia de la serie. Para ello será necesario encontrar el modelo que más se ajuste a la recta de entre los que se pueden citar:

Lineal $y = b_0 + b_1 \cdot t$	Cuadrático $y = b_0 + b_1 \cdot t + b_2 \cdot t^2$	Inverso $y = b_0 + \frac{b_1}{t}$
Logarítmico $y = b_0 + b_1 \cdot \ln t$	Cúbico $y = b_0 + b_1 \cdot t + b_2 \cdot t^2 + b_3 \cdot t^3$	Compuesto $y = b_0 \cdot b_1^t$
Potencia $y = b_0 \cdot t^{b_1}$	S $y = e^{b_0 + \frac{b_1}{t}}$	Exponencial $y = b_0 \cdot e^{b_1 \cdot t}$
Crecimiento $y = e^{b_0 + b_1 \cdot t}$	Logístico $y = \frac{1}{\frac{1}{u} + b_0 \cdot b_1^t}$	

**Figura 27.- Tipos de rectas de ajuste de tendencia.**

Teniendo en cuenta el gráfico de secuencia de la componente tendencia-ciclo (Figura 28):



**Figura 28.- Gráfico de secuencia de la componente tendencia-ciclo de la serie de venta de cava en España 62-69.**

De entre los distintos modelos, tras realizar la estimación de  $R^2$  y la de  $R^2$  corregido, se observa que, de entre las distintas opciones, existen que tienen mayores posibilidades de ajuste:

**-Lineal:**  $R^2 = 0.667$ ,  $R^2$  corregido = 0.663.

-Logarítmica:  $R^2 = 0.558$ ,  $R^2$  corregido = 0.573

-S:  $R^2 = 0.120$ ,  $R^2$  corregido= 0.110

Tras mostrar estos valores, al representar la información gráficamente, se puede considerar que el modelo de ajuste que representa mejor los valores de la tendencia de la serie es el esquema línea.

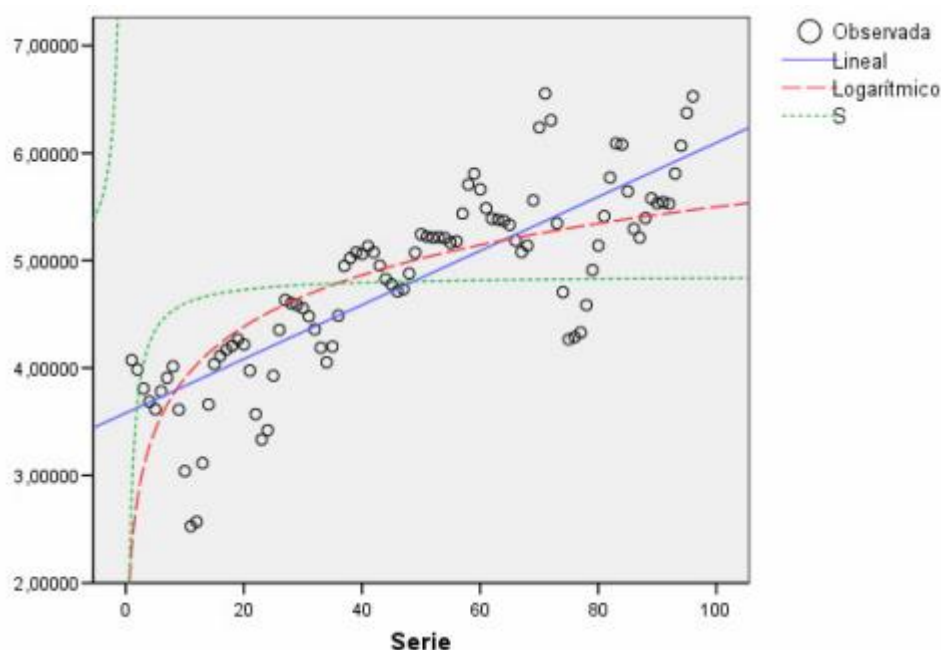


Figura 29.- Gráfico de la componente TC con los ajustes lineal, logarítmico y S.

Es necesario recordar, que la recta de ajuste obtenida se encuentra expresada en función del tiempo,  $y = b_0 + b_1 * t$ , por lo que para realizar la predicción de los siguientes intervalos de tiempo en los que se esté interesado, se debe aplicar el valor del tiempo estimado al de la recta de ajuste escogida. Una vez obtenida la estimación de la tendencia, se debe aplicar a la recta el valor de la componente estacional para cada estación de la serie  $X_{t estimada} = M_{t estimada} + E_t$  Con lo que se obtendría la siguiente estimación para los siguientes doce meses.

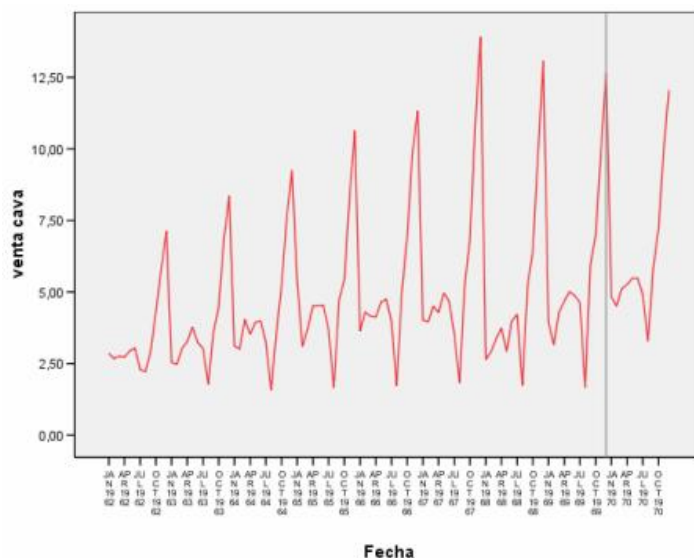


Figura 30.- Gráfico de secuencia de la serie cava original junto con las predicciones para un año.





La previsión para la siguiente de la que no se tienen datos con este método sería la propia media. Se da importancia únicamente al último de los datos de que se dispone, obviando el resto  $\hat{X}_{n+1} = X_n$ .

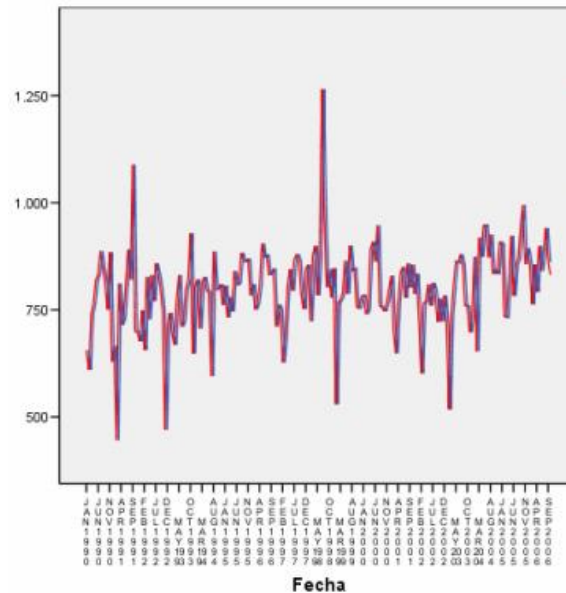


Figura 32.- Gráfico de secuencia de la serie de venta de Combustible con el ajuste efectuado considerando únicamente el dato anterior.

Como se observa en el gráfico de secuencia, el ajuste de la serie es la “sombra”, es decir, que es la misma serie pero retardada en una unidad de periodo, en este caso un mes, por tanto la previsión para el siguiente mes será el último valor de la serie.

Modelos de medias móviles.

Se basan en considerar únicamente las últimas  $k$  observaciones, de esta manera se da el mismo peso a los últimos  $k$  datos ( $1/k$ ) y cero al resto, mediante un procedimiento de medias móviles. Este procedimiento no es tan extremo como los anteriores, y al sustituir cada dato por una media de los  $k$  últimos la serie se suaviza y se elimina ruido, obteniendo el patrón subyacente de la misma. Cuantas más observaciones relevantes ( $k$ ) tomemos al aplicar este tipo de ajuste más se suaviza la serie.

$$\hat{X}_{n+1} = \frac{\sum_{i=1}^k X_{n-i+1}}{k}$$

Una vez aplicado el modelo, se obtiene una nueva serie derivada de la anterior en la que se encuentra la serie ajustada. Hay que tener en cuenta que esta nueva variable no dispone de valores en los  $k$  primeros registros, puesto que en esos casos no existe  $k$  observaciones para calcular la media.

La previsión del siguiente mes para el que no se dispone de datos con este método será la media de los últimos  $k$  valores de la serie.

## Modelos de suavizado exponencial simple

$$\hat{X}_{n+1} = \alpha * X_n + (1 - \alpha) * \hat{X}_n \text{ donde } 0 \leq \alpha \leq$$

67

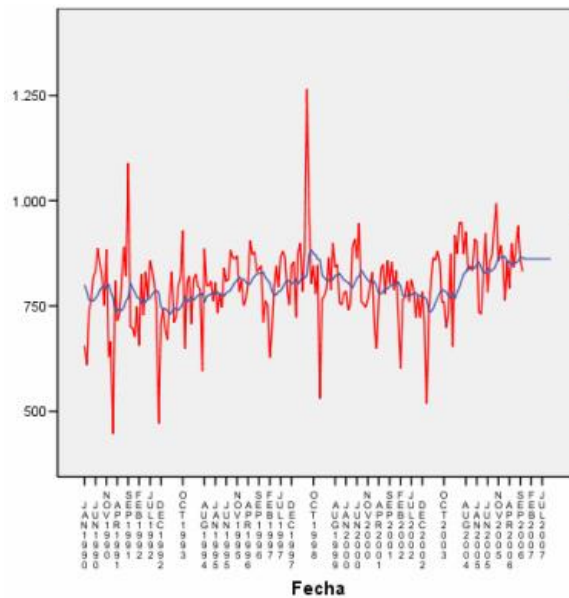


Figura 34.- Gráfico de secuencia de la serie Combustible con el ajuste efectuado considerando un modelo exponencial simple (serie suavizada) y la previsión a varios meses vista.

### Modelos aplicables a series no estacionales con tendencia.

En el caso de series con tendencia lineal (creciente o decreciente) pero sin comportamiento estacional, el modelo clásico que más se suele aplicar es el de Holt. Se trata de modelo de estimación exponencial que atenúa directamente la tendencia al obtener la diferencia entre los valores sucesivos (de la atenuación exponencial), para pronosticar a futuro hacia  $n$  periodos. Permite reducir el efecto de la aleatoriedad (usando la diferencia entre los promedios calculados en dos periodos sucesivos). Se actualiza la estimación de la demanda o tendencia a pronosticar. Se evita un pronóstico con una reacción retrasada al crecimiento.

El procesamiento del método de Holt ofrece resultados análogos a los obtenidos en el caso del suavizado exponencial simple, con la diferencia de que en lugar de un valor inicial se obtienen dos, y el resultado proporciona dos parámetros de suavizado.

Serie	Alpha (Nivel)	Sumas de los errores cuadráticos	gl error
Gasolina	,11000	1723447,19	201

Sin embargo, al contrario de los modelos de suavizado exponencial simple, en el de Holt las previsiones conforman una recta. Como ejemplo de aplicación se presenta en un gráfico de secuencia una serie original (rojo), el ajuste y las predicciones logradas mediante el método de Holt (azul) (Figura 35).

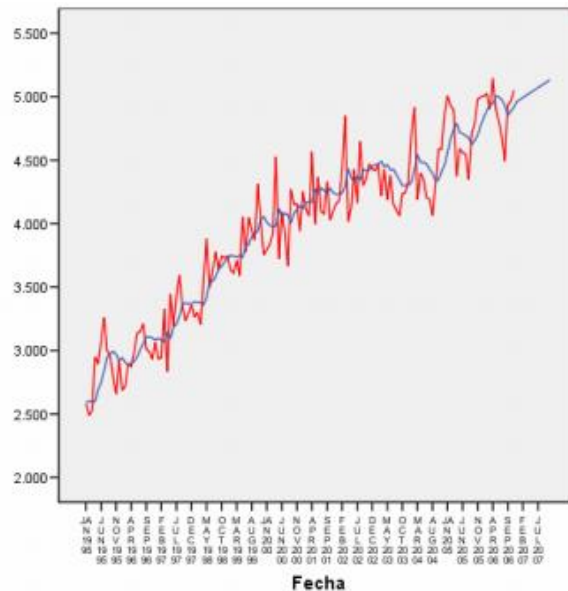


Figura 35.- Gráfico de secuencia de la serie suavizada Holt con el ajuste efectuado considerando un modelo exponencial doble (serie suavizada) y la precisión a varios meses vista.

### Modelos aplicables a series con tendencia y estacionalidad

Como tercer caso a presentar, se tratan series con tendencia lineal (creciente o decreciente) y comportamiento estacional, el modelo clásico que se aplica es el de Holt-Winters. Es una extensión del modelo de Holt, visto en el apartado anterior, que considera estacionalidad. La tendencia y la estacionalidad se pueden combinar de diferentes maneras, pero las que se consideran más frecuentes son la combinación aditiva y multiplicativa, vistas en el Análisis preliminar. A diferencia de otras técnicas, el modelo este puede adaptarse fácilmente a cambios y tendencias, así como a patrones estacionales. En comparación con otras técnicas como ARIMA (que se detallará en el siguiente apartado) el tiempo y esfuerzo empleado en calcular el pronóstico es considerablemente más rápido, lo que le confiere una considerable versatilidad haciéndolo adaptable a cualquier negocio.

Como ejemplo de aplicación, se presenta la siguiente serie, en la que se presenta la información trimestral sobre producción de papel, estacional y con tendencia creciente. Tras realizar un análisis preliminar, se concluye que no existe dependencia entre la variabilidad y el nivel, de manera que se considera que la tendencia y la estacionalidad se combinan de forma aditiva. Al igual que ocurría con el método de Holt puro, Holt-Winters arroja un parámetro de ajuste más con lo que la predicción forma una curva.

En el siguiente gráfico de secuencia, se muestra la serie de producción de papel trimestral antes mencionada (rojo) junto con la resultante del suavizado y la predicción (azul) (Figura 36).

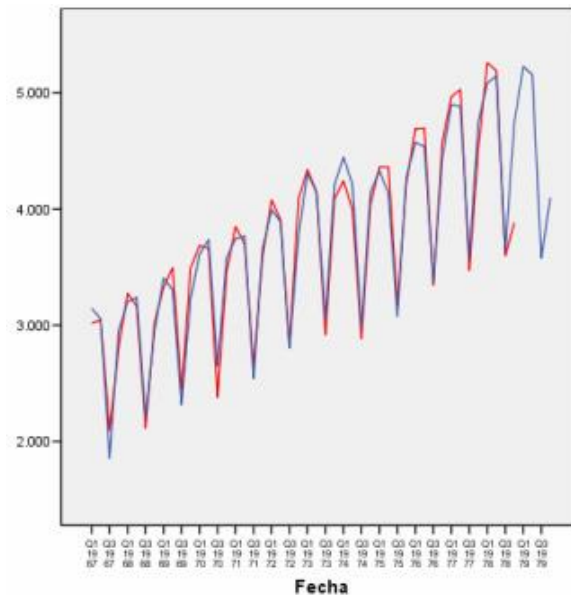


Figura 36.- Gráfico de la serie Papel con el ajuste efectuado considerando un modelo de Holt-Winters aditivo y la previsión a un año vista.

La información presentada a lo largo del presente apartado permite conocer el concepto y funcionamiento de los métodos clásicos de análisis de series temporales, sin embargo no se ha tenido en cuenta de dónde salen los datos de las mismas o cuál es el mecanismo que las genera.

### 3.3.4- Modelos ARIMA.

A la hora de estudiar una serie temporal empleando los modelos ARIMA se está suponiendo que la serie puede estar generada por un proceso estocástico, definido en el apartado de presentación de las series temporales.

El objetivo perseguido al analizar series temporales es estimar el proceso estocástico que la genera y para ello, según hemos visto hasta ahora, debemos partir del supuesto de que dicho proceso estocástico es estacionario y ergódico, como ya se expuso en la definición de estos procesos.

Los métodos ARIMA en realidad se pueden agrupar dependiendo del acrónimo que compone su nombre:

AR (AutoRegresivo)

I (Integrado)

MA (Medias Móviles)

#### Modelo de medias móviles (MA)

Un proceso de medias móviles de orden  $q$  que es un proceso en el que la variable  $X_t$ , se obtiene como un promedio de variables de ruido blanco ( $a_i$ ), siendo los  $\theta_i$  sus coeficientes de ponderación.

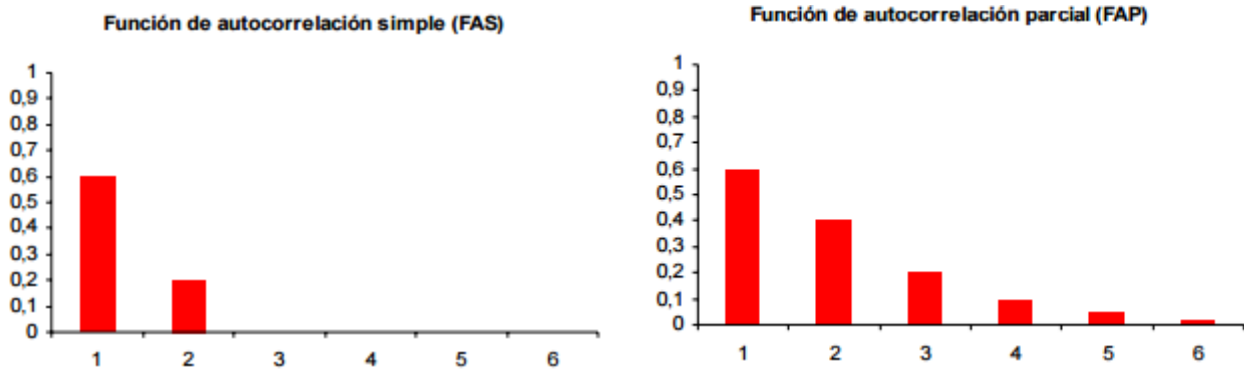
Todos los procesos de medias móviles son procesos estacionarios.

Existe  $2q$  procesos de medias móviles de orden  $q$  que poseen la misma función de autocorrelación, pero solo uno de ellos es invertible. De esta manera si sólo se consideran procesos invertibles la función de autocorrelación determina unívocamente un proceso.

-La función de autocorrelación simple de un modelo de medias móviles se corta (se hace cero) en el orden del modelo ( $q$ ) y caracteriza los procesos de medias móviles.

-La función de autocorrelación parcial de un modelo de medias móviles no se corta, tiende a cero rápidamente, exponencialmente.

Como ejemplo de funciones de autocorrelación simple de un modelo de medias móviles de orden dos:



Hay que tener en cuenta que los coeficientes de correlación de ambas funciones pueden tomar valores positivos o negativos.

### Modelos Autorregresivos (AR)

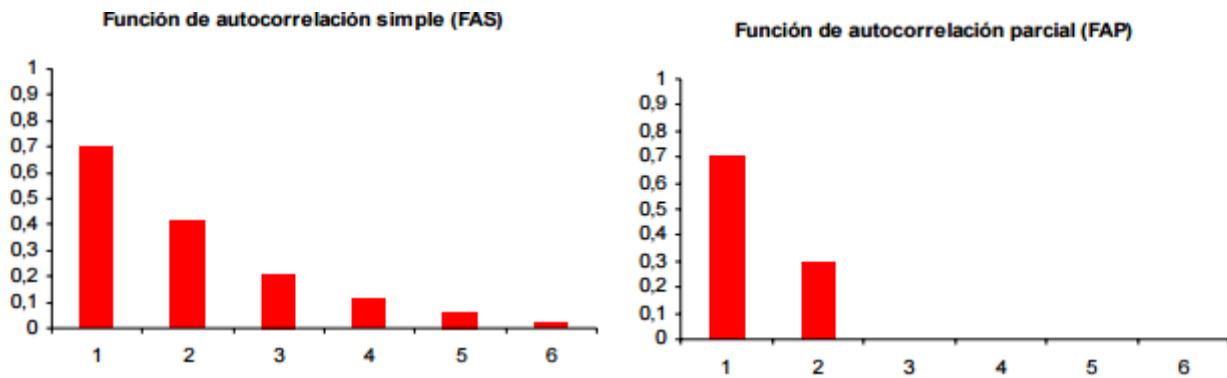
Un proceso autorregresivo de orden  $p$  es un proceso en el que la variable  $X_t$  se obtiene efectuando una regresión sobre valores pasados de las mismas.

Todos los procesos autorregresivos son procesos invertibles, por tanto su modo de trabajo es el opuesto al de las medias móviles.

-La función de autocorrelación simple de un modelo autorregresivo no se corta, tiende a cero de forma exponencial, rápidamente.

-La función de autocorrelación parcial de un modelo autorregresivo se corta (se hace cero) en el orden del modelo ( $p$ ).

Los siguientes gráficos muestran funciones de autocorrelación de modelo autorregresivo de orden dos:



Al igual que el caso anterior, ambas funciones pueden tomar valores positivos y negativos.

### Modelos autorregresivos de medias móviles ARMA

A pesar de que el método de trabajo de los métodos AR y MA sea opuesto, ambos métodos se pueden combinar. Los métodos ARMA son, por tanto, la suma de un proceso autorregresivo de orden  $p$  y uno de medias móviles de orden  $q$ .

Las funciones de autocorrelación simple y parcial de un proceso ARMA también son la suma de las correspondientes de un proceso MA y otro AR. De esta forma ninguna de las dos se corta.

-La función de autocorrelación simple de un proceso ARMA en cuanto supera el orden de la parte MA,  $q$ , se comporta como si sólo hubiera parte AR.

Función de autocorrelación parcial se comporta como si sólo hubiera parte MA en cuanto se supera el orden de la parte AR,  $p$ .

Estas características de las funciones de autocorrelación dificultan la identificación del proceso a través de las mismas, de manera que no es posible identificar  $p$  y  $q$  al mismo tiempo. En un primer paso, se debe identificar uno de los procesos, independientemente de cuál de ellos sea, y en el segundo paso el otro.

### Modelos no estacionarios ARIMA

Por definición, al igual que existen series estacionarias, existen también series no estacionarias. Por tanto los modelos vistos hasta ahora no se pueden utilizar para este tipo de series.

Sin embargo es posible transformar las series no estacionarias para que verifiquen los supuestos que necesitamos. En primer lugar, se analiza la dependencia entre variabilidad y nivel. Si existe dependencia de este tipo la serie no es estacionaria en varianza, de manera que se debe transformar. Cuando la serie de trabajo es no estacionaria de nivel (serie con tendencia) se consigue que la serie sea estacionaria diferenciándola, es decir, integrarla. No obstante, es posible que, tras diferenciar una serie, esta siga siendo no estacionaria, en cuyo caso es posible que necesite una nueva diferencia. De esta manera, la notación que se emplea con estos modelos es ARIMA ( $p,d,q$ ) donde  $d$  es el número de diferencias que se efectúan sobre la serie original antes de ajustar un ARMA.

Como conclusión del capítulo, se puede decir que el corazón del BI se encuentra en el análisis de las series temporales. Estas permiten almacenar las observaciones recogidas en el desarrollo del proceso de negocio,

que por definición es estacionario, pues las variables de estudio mantienen relación en el tiempo, y ergódico, las observaciones más influyentes en el proceso de decisión son las más actuales.

De esta forma, los métodos de estudios de las series temporales permiten realizar predicciones de posibles valores futuros que den sustento a las decisiones que se pretenden tomar.



## 4.- Caso de aplicación práctica. Presentación

Los conceptos expuestos a lo largo de este trabajo se pondrán en práctica en un caso de aplicación real. En este capítulo realizará una contextualización del entorno sobre el que se realizará el estudio, el sector turístico rural en España. Del mismo modo se expondrán las herramientas mediante las cuales se realizará dicho estudio lo cual servirá para establecer una comparativa entre las mismas. Para finalizar el capítulo se expondrán una serie de grandes sistemas de gestión empresarial cuyas capacidades y requerimientos exceden los límites de este trabajo.

### 4.1.- Turismo Rural en España.

España es líder mundial en turismo, tal y como se demuestran en los indicadores de actividad turística. Desde hace décadas este país es destino de un gran número de viajeros tanto nacionales como extranjeros. Sin embargo tradicionalmente se identifica este turismo con los productos asociados al “*sol y playa*” cuyo momento de disfrute se concentra en los grandes periodos vacacionales y en las zonas costeras (minetur, 2014).

El turismo rural como tal se encuentra menos desarrollado aunque su implantación está siendo fuerte especialmente en zonas del interior y el Norte peninsular (Figura 37).

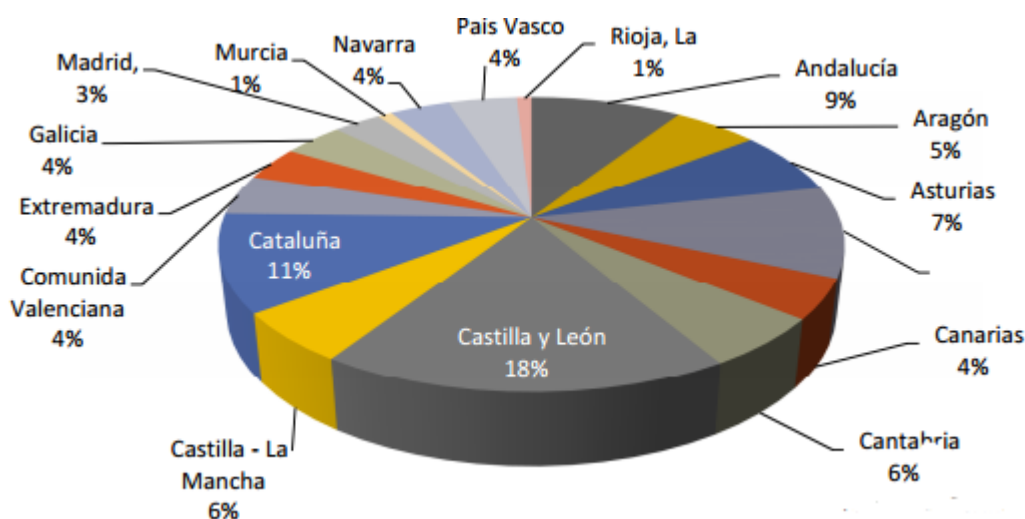


Figura 37.- Distribución del turismo rural por CC.AA. en España año 2013

Se entiende por turismo rural a la actividad turística que se realiza en un espacio rural, habitualmente en pequeñas localidades o fuera del casco urbano en localidades de mayor tamaño. Las instalaciones suelen ser antiguas viviendas de gran tamaño o fincas de explotación primaria rehabilitadas, reformadas o adaptadas y su gerencia suele hacerse de modo familiar o por medio de pequeñas empresas. Entre las modalidades de turismo rural más demandadas se pueden citar:

El agroturismo.

El turismo ecológico.

El enoturismo.

Como actividad económica, el turismo rural presenta puntos a favor que permiten el desarrollo de una zona:

- Genera riqueza y dinamiza áreas geográficas con menor flujo turístico o económico y a la integración del territorio.

- Diversifica las actividades económicas de determinadas zonas y permite preservar las tradiciones del territorio sobre el que se fundamenta.

Además los indicadores de la actividad turística también se ven beneficiados:

- Distribuye los flujos de demanda a lo largo de todo el año al explotar principalmente la modalidad de turismo de “*fin de semana*” y de estancias tipo “*fly&drive*” para el turismo internacional.

- Adapta la oferta turística a las exigencias del nuevo entorno turístico internacional que demanda productos turísticos experienciales.

Como se ha dicho previamente, por lo general el turismo en España se ha asociado con el modelo “*sol y playa*” al contrario que en el resto del continente europeo, especialmente en la zona centro y norte donde por orografía y clima se ha explotado este tipo de turismo desde el comienzo del fenómeno del turismo de masas ofertando elementos como el contacto con la naturaleza, la calidad material y humana o la gastronomía y patrimonio locales. Por su parte en España el turismo rural vivió una fase crecimiento exponencial entre finales del siglo XX y principios del XXI que supuso una rápida expansión de la oferta al ser percibido como una solución atractiva para dinamización de determinadas zonas rurales (minetur, 2014).

Sin embargo, esta relativa juventud del turismo rural en España provoca que el sector presente una serie de carencias y debilidades entre las que se podrían citar:

- Ausencia de planificación estratégica.

- Falta de visión integral de destino turístico rural.

- Difícil acceso del turismo exterior de algunas localizaciones.

- Necesidad de implantar estándares de calidad que faciliten su comercialización en mercados internacionales.

Sin embargo el turismo rural se encuentra en un periodo de crecimiento, a pesar de que se observe un descenso en los últimos años, principalmente provocado por la crisis económica (Figura 38).

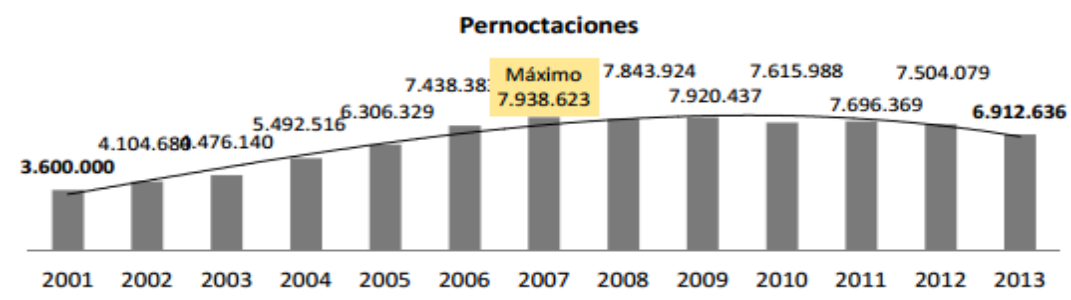


Figura 38.- Pernoctaciones en España turismo rural 01-13

Sin embargo representa un muy pequeño porcentaje del total de pernoctaciones en España que en 2013 llegó al total de 285 millones (INE, 2014) representando un 2.45% del total.

En comparación con la oferta turística que se presenta en Europa, el turismo rural español se haya en un periodo de desarrollo y necesidad de madurez. A continuación se presentan las características de cuatro destinos turísticos europeos por su experiencia y proximidad a España (Universidad Pablo de Olavide, 2009).

**Francia.** En el país galo existe una fuerte categorización, normalización y homogenización de los establecimientos turísticos que son sometidos a controles de calidad periódicamente. Los establecimientos se encuentran especializados en diferentes tipos de usuarios existiendo aquellos que están especializados en niños, otros en actividades deportivas y recreativas para practicar en la naturaleza, establecimientos que se encuentran en parques nacionales, zonas de nieve, viñedos... A todo esto hay que añadir una buena relación calidad precio y un importante énfasis en la tradición cultural y gastronómica del lugar.

**Portugal.** Con una menor distinción tipológica que el caso anterior (o incluso que España) y una oferta dirigida principalmente a extranjeros, el país luso presenta una oferta turística menor pero muy segmentada y de calidad.

**Reino Unido.** Se trata de un país pionero en turismo rural que ofrece productos consolidados y maduros. Como en el caso de Francia presenta una gran estandarización y compromiso con la calidad dirigiéndose a muy distintos segmentos de público que van desde las casas de huéspedes centenarias hasta nuevos conceptos como las cabañas o refugios con servicios básicos que en ocasiones ni siquiera cuentan con electricidad.

Por tanto, como reto para alcanzar el grado de madurez que presentan países como los anteriores, el sector turístico rural en España tiene que afrontar las siguientes metas (minetur, 2014):

- Articular el producto.

- Armonizar la clasificación de alojamientos.

- Mejorar la comercialización vía web.

- Desarrollar iniciativas de turismo experiencial.

- Posicionar España como destino de turismo rural en mercados internacionales con potencial de generación de pernoctaciones.

- Formación y profesionalización.

Por supuesto, el principal problema del turismo rural en España, al contrario de lo que sucede en otros países europeos (como los anteriores) es que el atractivo del turismo costero estival es muy alto, acaparando la mayor parte de los visitantes. Sin embargo, como se dijo previamente, el turismo rural en España tiene un gran potencial debido a la riqueza cultural y gastronómica del país y debe ser considerado como un motor económico para zonas rurales del centro del país con menor desarrollo empresarial.

## 4.2.- Presentación de las herramientas de trabajo

A lo largo del presente apartado, se presentarán las herramientas y soluciones software utilizadas para este proyecto. Se han utilizado herramientas que permitan una utilización sencilla y asequible a pequeña escala tratando de cubrir la mayor cantidad de casos posible, software libre y propietario, aplicaciones estadísticas y generalistas, software comercial y autogenerado...

### 4.2.1.- SPSS

SPSS es un programa estadístico informático creado originalmente al final de la década de los sesenta muy usado en las ciencias sociales y las empresas de investigación de mercado. Originalmente SPSS fue creado como el acrónimo de Statistical Package for the Social Sciences aunque también se ha referido como "Statistical Product and Service Solutions". En la actualidad el programa es propiedad de la empresa estadounidense IBM (IBM, 2012).



Es uno de los programas estadísticos más conocidos teniendo en cuenta su capacidad para trabajar con grandes bases de datos y un sencillo interface para la mayoría de los análisis. En la versión 12 de SPSS se podían realizar análisis con 2 millones de registros y 250.000 variables. El programa consiste en un módulo base y módulos anexos que se han ido actualizando constantemente con nuevos procedimientos estadísticos. Cada uno de estos módulos se compra por separado, de entre estos módulos destacan:

- Modelos de Regresión,
- Modelos Avanzados (Reducción de datos, Clasificación, Pruebas no paramétricas)
- Tablas.
- Análisis conjunto.
- Mapas.

#### 4.2.2.- Microsoft Excel.

Se trata de una aplicación de hojas de cálculo desarrollado por Microsoft integrada dentro de la suite Microsoft Office que puede ser utilizada en diversos sistemas operativos. Sus características de cálculo, herramientas para presentar gráficos, tablas y la posibilidad de insertar programación en lenguaje *Visual Basic* la hace destacar como uno de los principales y más extendidos software de gestión empresarial a pequeña, e incluso mayor, escala del mundo (Microsoft, 2014).



En la actualidad, la capacidad de cómputo de Excel permite manejar hasta un total de 16.386 columnas y hasta 1.048.576 filas por hoja de cálculo de las que puede llegar a tener hasta 1.024 lo que en términos prácticos lo hace prácticamente inagotable. A nivel de usuario, Excel presenta las mismas características que cualquier software de hojas de cálculo, utilizando una parrilla de celdas ordenadas en filas numeradas y columnas nombradas mediante letras de forma que se puede organizar la manipulación de datos a modo de operaciones aritméticas.

Como gran baza a favor, la amplia difusión de la herramienta permite que exista una gran comunidad en Internet que facilita un mayor aprovechamiento de la herramienta para conseguir un trabajo más óptimo, a esto hay que añadir la relación que la mayoría de los programas y soluciones software permiten establecer con Excel mediante exportación directa de datos. A lo anterior hay que añadir que los principales programas de análisis de datos en incluso SGBD permiten establecer conexión con Excel para explotar los datos obtenidos mediante análisis.

#### 4.2.3.- Weka.

Esta herramienta es una solución software orientada al aprendizaje automático y la minería de datos escrito en java y distribuido bajo licencia CNU-GPL. El acrónimo WEKA significa “Entorno para análisis del conocimiento de la Universidad de Waikato” haciendo referencia a la universidad neozelandesa en la que fue desarrollado (Waikato, 2013).



El paquete Weka contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades. A nivel técnico, weka ofrece procesamiento de datos, clustering, clasificación, regresión, visualización y selección. Todas las técnicas de Weka se fundamentan en la asunción de que los datos están disponibles en un fichero plano o una relación, en la que cada registro de datos está descrito por un número fijo de atributos. Weka también proporciona acceso a bases de datos vía SQL gracias a la conexión JDBC (Java Database Connectivity) y puede procesar el resultado devuelto por una consulta hecha a la base de datos. No puede realizar minería de datos multi-relacional, pero existen aplicaciones que pueden convertir una colección de tablas relacionadas de una base de datos en una única tabla que ya puede ser procesada con Weka.

La flexibilidad y capacidad de adaptación de WEKA permite que se integre dentro de suites de mayor tamaño, como el caso de Pentaho, que se comentará más adelante.

#### 4.2.4.- Tableau

Es una herramienta desarrollada por la compañía americana del mismo nombre con sede en Seattle destinada a la visualización de datos y al BI. La empresa ofrece cinco productos según las necesidades de cada usuario Tableau Desktop, Tableau Server, Tableau Online, Tableau Reader y Tableau Public. Tableau Public y Tableau Reader son herramientas de uso gratuito mientras que Server y Desktop presentan un periodo de prueba de 14 días después de los cuales es necesario abonar la licencia que en el caso de Desktop está disponible tanto en versión profesional como en edición personal de bajo coste. Por su parte Tableau Online es una herramienta en Cloud Computing disponible mediante suscripción anual (tableau, 2014).



Tableau permite trabajar con prácticamente cualquier fuente de información, desde bases de datos hasta ficheros planos lo que confiere una enorme adaptabilidad y facilidad de uso. Entre las características de Tableau se pueden citar:

- Permite gestión de datos en tiempo real logrando seguir los datos minuto a minuto agilizando la toma de decisiones.
- La lectura de datos desde varias fuentes referida previamente se complementa con la posibilidad de combinar los datos procedentes de cada una de las mismas para trabajar con todas ellas.
- Interfaz de uso intuitiva y sencilla que permite un aprendizaje rápido.
- Creación de cuadros de mandos informativos e interactivos permitiendo combinar varias vistas al mismo tiempo.
- Creación de entornos públicos para compartir informes de forma que estén disponibles para todos los implicados en el negocio aumentando la ubicuidad de la información.

#### 4.2.5.- R Programming

Se trata de un lenguaje de programación y un entorno software orientado al análisis estadístico y gráfico. R es ampliamente utilizado tanto por estadísticos como por técnicos de data mining para el desarrollo de soluciones ad-hoc. Realmente R es una implementación del lenguaje de programación S combinado con semántica inspirada en Scheme. Fue desarrollado por la Universidad de Auckland en Nueva Zelanda y actualmente mantenido por el *R Development Core Team*. El entorno de R es un proyecto distribuido mediante GNU escrito en C, Fortran y R.



A nivel estadístico implementa una enorme cantidad y variedad de técnicas estadísticas y gráficas incluyendo modelado lineal y no lineal, test clásicos estadísticos, análisis de series temporales, clasificación o clustering. R es fácilmente extensible a través de funciones y extensiones junto con una enorme y activa comunidad que desarrolla gran cantidad de paquetes y funcionalidades. Otra fortaleza de R es la producción de gráficos, con la cual es posible producir y publicar gráficos de gran calidad incluyendo símbolos matemáticos.

A nivel de programación R es un lenguaje interpretado que debido a su herencia con S es fuertemente orientado a objetos. Presenta una interfaz de scripting que, combinado con el gran repositorio de funciones de la comunidad de R, permite su utilización tanto a bajo como a alto nivel facilitando además que lo puedan utilizar usuarios sin necesidad de profundos conocimientos tanto en estadística como en programación.

Como ya se ha comentado, la mayor ventaja de esta herramienta radica en su repositorio de paquetes que permiten especialización de procesos a alto nivel, importación y exportación de capacidades, etc. Estos paquetes son desarrollados principalmente en R, lo que permite a los usuarios seguirlos e incluso, si se permite, modificarlos, aunque también pueden codificarse en Java, C, C++ o Fortran.



#### 4.2.6.- ForecastBI

Esta es una herramienta especialmente desarrollada para este proyecto orientada específicamente al análisis de series temporales permitiendo el estudio de una gran cantidad de series al mismo tiempo e incluso establecer relaciones entre las mismas. La aplicación se basa en un core en código R, lenguaje en el que se deben escribir las funciones de relación entre las series, y una interfaz gráfica desarrollada en Java.



El motivo de desarrollar una aplicación propia es el de obtener una solución completamente personalizada que responda exactamente a los problemas del negocio. Las capacidades y complejidad de la herramienta desarrollada dependerán completamente tanto de las necesidades del negocio como de las capacidades y conocimientos del desarrollador.

Como punto negativo de esta solución cabe destacar que su especificidad la hace tremendamente limitada, pues únicamente responde a las necesidades específicas del negocio en el momento de la construcción y cualquier requerimiento posterior exigirá un nuevo desarrollo. Por otra parte, esta misma característica la hace difícilmente adaptable a otros negocios. Además, es necesario poseer conocimiento de desarrollo lo que, en caso de no tener, puede provocar que se externalice su desarrollo a otra organización con los problemas que ello puede acarrear.

### 4.3.- Aplicaciones comerciales.

En el apartado anterior se han presentado una serie de programas y soluciones capaces de cubrir de un modo más o menos solvente las necesidades de organizaciones y negocios de tamaño pequeño o medio. Sin embargo para empresas de gran tamaño se necesita una capacidad de cómputo, almacenamiento y distribución que queda muy lejos de las anteriormente citadas. En orden a resolver estos problemas existen grandes “suites” o conjuntos de programas de entre las que se citarán tres de los principales ejemplos.

#### 4.3.1.- SAP Business suite.

Es un conjunto de programas desarrollados por la empresa SAP basada en una plataforma de tecnología llamada netWeaver<sup>5</sup> que permiten a las organizaciones ejecutar y optimizar distintos departamentos como los sistemas de ventas, finanzas, operaciones bancarias, compras, fabricación, inventarios y relaciones con los clientes. Ofrece la posibilidad de realizar procesos específicos de la empresa o crear módulos independientes para funcionar con otro software de SAP o de otros proveedores. La suite puede soportar sistemas operativos, bases de datos, aplicaciones y componentes de hardware de casi cualquier proveedor (Rouse, <http://searchsap.techtarget.com/>, 2014).



SAP Business Suite se encuentra dividido en cinco módulos:

- SAP CRM
- SAP ERP
- SAP PLM
- SAP SCM
- SAP SRM

**SAP CRM:** Se encarga de interactuar con todos los temas relacionados con el cliente ya sea ventas, marketing o servicios. No sólo realiza operaciones a corto plazo, como reducir costes, sino que también adquiere capacidades que permite llevarlas a cabo a largo plazo. Sus características son las siguientes:

- Soporta todos los procesos que se llevan a cabo con el cliente como suministro, facturación y la contabilidad de deudores.
- Proporciona conocimientos de los clientes a toda la empresa.

---

<sup>5</sup> Es conocida como una aplicación orientada a servicios y a la integración que provee al usuario de un vínculo entre lenguajes y aplicaciones. Está construido usando estándares abiertos de la industria por lo que es sencillo negociar transacciones de información con desarrollos de Microsoft .NET, Sun Java EE, e IBM WebSphere.

-Permite obtener unos resultados inmediatamente, a la vez que establece las pautas para conseguir unos objetivos a medio/largo plazo.

Las áreas que abarca CRM son: marketing, ventas, servicio, aplicaciones analíticas, soporte de aplicaciones a domicilio, E-commerce, operaciones y gestión de centros de atención al cliente y gestión de canales.

**SAP ERP:** Da soporte a las funciones esenciales de los procesos y operaciones de la empresa. A su vez, se subdivide en:

-Finanzas: permite cumplir con los estándares de generación de informes financieros, mejorar el flujo de caja y gestionar los riesgos financieros.

-Gestión del capital humano: optimiza los procesos de selección y motivación de los empleados.

-Operaciones: se mejoran las operaciones para reducir costes, aumentar ingresos, maximizar la rentabilidad y la atención al cliente.

Las áreas que abarca SAP ERP son: análisis empresarial, contabilidad financiera e interna, gestión del capital humano, gestión de operaciones, gestión de servicios corporativos y autoservicios.

**SAP PLM:** Las funciones de este módulo más importantes son crear y suministrar productos y optimizar los procesos de desarrollo de los productos y sistemas para acelerar su introducción en el mercado. Este módulo comprende áreas como gestión del ciclo de vida de la información, gestión de programas y proyectos, colaboración en el proceso completo, gestión de calidad, gestión del ciclo completo de los activos, y medio ambiente, salud y seguridad.

**SAP SCM:** Permite diseñar, construir y poner en marcha la cadena de suministro. Las funciones más importantes que ofrece son reducir los costes a la hora de distribuir el producto, aumentar los ingresos por la venta de estos y la reducción de costes, y mejorar el servicio a los clientes.

**SAP SRM:** Ofrece funciones tales como el análisis de gastos, abastecimiento, contratos operativos, pedidos, facturas y gestión de proveedores. Permite reducir costes a la hora de comprar materiales, elegir aprovisionamientos y colaboración entre la empresa y pequeños comercios.

#### 4.3.2.- IBM Cognos Business Intelligence.

Se trata de una suite de IBM que proporciona un conjunto de herramientas orientadas a reporting, análisis, scoreboarding, y monitoring de eventos y métricas. La aplicación consiste en una gran cantidad de componentes que proporcionan distintas informaciones y requerimientos sobre el negocio al usuario (IBM, 2014).



Como era de esperar, IBM Cognos está dividida en gran cantidad de componentes que se citan a continuación:

**Cognos Connection:** Es en realidad un portal web accesible y ubicuo provistos con todas la funcionalidades propias de la suite. Gracias a este portal, se pueden compartir los informes, las tablas de resultados, programación o cualquier otro tipo de información entre distintos tipos de usuario autorizado para verlo.

**Query studio:** Permite la realización de consultas y la generación de informes rápidos para el usuario con el objetivo de responder a necesidades básicas. La capa de presentación puede ser personalizada y los datos pueden ser filtrados, ordenados y formateados e incluso ofrece generación de diagramas.

**Report studio:** Esta herramienta es utilizada para crear informes de gestión que pueden incluir gráficos o mapas utilizando tanto bases de datos relacionales como multidimensionales.

**Analysis studio:** Este módulo permite a los usuarios realizar análisis de una gran cantidad de fuentes de datos y obtener información sobre eventos y acciones. El análisis multidimensional permite identificar tendencias y comprender las anomalías y desviaciones, las cuales son no obvias en otro tipo de informe. La herramienta permite la generación KPI's, la inclusión de filas o columnas de la tabla e incluso funciones OLAP.

**Event studio:** Es una herramienta de notificaciones que informa al usuario sobre eventos dentro de la organización en tiempo real. De esta forma, un usuario puede crear o detectar la ocurrencia de eventos de negocio en circunstancias excepcionales basadas en el cambio de eventos específicos o condiciones de datos. Estas notificaciones son presentadas en formas muy distintas según las preferencias del usuario mediante correo electrónico, publicándolo en un portal o lanzando informes.

Además de estas cinco herramientas básicas, IBM Cognos dispone también de una serie de componentes adiciones:

**Go! Office:** Este componente permite a los usuarios trabajar con contenido de IBM Cognos en el entorno de Microsoft Office. Esta herramienta proporciona acceso al todas las funciones de generación de informes de IBM Cognos permitiendo a los usuarios utilizar los informes predefinidos o crear nuevos contenidos con

Query studio, Analysis Studio o Report Studio. Es posible importar los informes a la herramienta de Excel y publicar los resultados mediante Cognos Connection.

**Go! Search:** En Cognos Connection se realizan búsquedas basadas en texto completo en el interior de análisis, informes, cuadros de mando, etc. Utilizando como base el cuerpo de los documentos, los títulos, encabezamientos, etc. La búsqueda mediante IBM Cognos Go! Search se realiza de modo semejante al de los grandes motores de búsqueda tipo Google permitiendo una búsqueda optimizada y obteniendo mejores resultados.

**Go! Dashboard:** Este componente permite crear informes y cuadros de mandos interactivos tanto de información interna como externa para proporcionar la información concreta a cada usuario individual. Los cuadros de mandos pueden presentar informes de objetos presentados mediante Cognos Viewer. Las partes de los informes, como listas, tablas cruzadas y diagramas son presentados mediante *portlet interactivos*. Los contenidos pueden ser mostrados u ocultos dinámicamente mediante el uso de *combos* y *checkboxes*. Además, las direcciones web y los links pueden ser presentados dentro del propio informe.

#### 4.3.3.- Pentaho

Al contrario que las anteriores suites, Pentaho es un conjunto de programas distribuidos mediante licencia de software libre lo que lo hace extraordinariamente portable reduciendo los costes pero no por ello la calidad. Existen dos versiones del mismo, una edición llamada *Community Edition* gratuita y distribuida mediante licencia GNU GPL y otra versión denominada *Enterprise Edition* que se debe conseguir mediante una suscripción de pago que incluye servicio de soporte y mejora anual de productos. Incluye herramientas integradas para generar informes, minería de datos o ETL entre otros. En esencia es una plataforma de BI orientada a la solución y centrada en procesos. Las soluciones que Pentaho pretende ofrecer se componen fundamentalmente de una infraestructura de herramientas de análisis e informe integrado con un motor de flujo de trabajo de procesos de negocio.



Como en los casos anteriores, la suite de Pentaho está compuesta una serie de componentes que a continuación se citan:

**Pentaho Analysis Services:** cuyo nombre código Modrian es un servidor OLAP (procesamiento analítico en línea) escrito en Java. Es compatible con el MDX (expresiones multidimensionales) y el lenguaje de consulta XML para el Análisis y especificaciones de la interfaz olap4j.

**Pentaho Reporting:** Consiste en un motor de presentación, capaz de generar informes programáticos sobre la base de un archivo de definición XML. Sobre esta solución se han desarrollado muchas herramientas, por ejemplo informes, diseñadores de interfaz gráfica de usuario, y asistentes tipo wizard. Un uso notable de esta herramienta es el Generador de informes para OpenOffice.org

**Pentaho Data Mining:** Es una envoltura alrededor del proyecto Weka. Es una suite de software que usa estrategias de aprendizaje de máquina, aprendizaje automático y minería de datos. Cuenta con series de clasificación, de regresión, de reglas de asociación, y de algoritmos de clustering, para así apoyar las tareas de análisis predictivo.

**Pentaho Dashboard:** Es una plataforma integrada para proporcionar información sobre sus datos, donde se pueden ver informes, gráficos interactivos y los cubos creados con las herramientas Pentaho Report Designer.

**Pentaho para Apache Hadoop:** Es un conector de bajo nivel para facilitar el acceso a MUY grandes volúmenes manejados en el proyecto Apache Hadoop, la Suite de Pentaho BI para Hadoop permite abordar los mayores desafíos que experimentan los usuarios de Hadoop -, sobre su empinada curva de aprendizaje técnico, la falta de personal técnico cualificado y la falta de disponibilidad de las aplicaciones de desarrollo y despliegue para llevar a cabo la integración de datos e inteligencia de negocios con Hadoop.

**Pentaho Data Access Wizard:** Este plug-in es soportado por todos los servidores y permite al usuario crear nuevas fuentes de datos a partir de bases de datos externas o incluso de archivos CSV. Durante los pasos de creación de una fuente de datos el usuario puede incluso generar modelos de datos describiendo como las

columnas o campos se relación creando jerarquías de relaciones. El modelo resultante es utilizado por Modrian o cualquier otro analizador para crear nuevas consultas que consuman la nueva fuente creada.

**Pentaho Mobile:** Este componente permite explotar las capacidades del análisis OLAP en cualquier dispositivo permitiendo obtener informes en tiempo real en cualquier lugar. Mobile también agrega características para bookmarking contenidos favoritos para un fácil acceso y el concepto de la apertura de varias piezas de contenido en pestañas.

**Pentaho for Big Data:** Pentaho para Big Data es una herramienta de integración de datos basada en Pentaho Data Integration. Se permite la ejecución de trabajos de ETL dentro y fuera de los entornos de datos grandes, como Hadoop o distribuciones como Amazon, Cloudera, EMC Greenplum, MapR y Hortonworks. También es compatible con fuentes de datos NoSQL como MongoDB y HBase.

**Pentaho Desing Studio:** El servidor de Pentaho BA soporta scripts XML especiales llamados xactions para implementar la lógica de negocio y otras formas de automatización en la plataforma. Design Studio es una versión modificada del entorno de desarrollo de Eclipse con plug-in diseñado para comprender los componentes apoyados por xaction scripts. Los Xactions son muy poderosos, y útiles, pero a veces resultan difíciles de solucionar debido a la forma de bajo nivel en que interactúan con las partes del servidor de BA. Los desarrolladores están empezando a utilizar archivos de transformación Pentaho Data Integration para llevar a cabo las tareas de automatización y la lógica de negocio. Las transformaciones se pueden ejecutar directamente por el servidor de BA y depurar visualmente en Pentaho Data Integration (PDI) y están ganando rápidamente a favor de la comunidad sobre xactions. Se trata de un pequeño salto de imaginar transformaciones PDI eventualmente reemplazará xactions completo.

**Pentaho Data Integration:** Cuyo nombre código es Kettle, consiste en un núcleo de integración de datos (ETL) y una interfaz gráfica que permite a los usuarios definir los Jobs de integración y transformación de datos. Este componente soporta despliegue tanto en ordenadores autónomos como en nubes o cluster.





## 5.- Caso práctico. Realización.

En el capítulo previo se ha dado una introducción al caso de aplicación práctica que se va realizar en este proyecto. En este se llevará a cabo el estudio de las series temporales referidas a las pernoctaciones de alojamientos rurales a nivel de provincias desde enero de dos mil diez hasta mayo de dos mil quince. Todos los datos han sido extraídos del Instituto Nacional de Estadística (INE).

En primer lugar, es importante notar que los datos referidos recogen información referente a todos los alojamientos de turismo rural según están definidos en las distintas normativas legales de cada comunidad autónoma. Se consideran, en general, alojamientos rurales, a aquellos establecimientos o viviendas destinadas al alojamiento turístico, con o sin otros servicios complementarios y que estén inscritos en el correspondiente registro de alojamientos turísticos de cada comunidad autónoma con independencia de las actividades llevadas a cabo por los huéspedes (ine, 2015).

La metodología de estudio empleada consistirá, en primer lugar, en establecer un gráfico de barras que muestre el acumulado de pernoctaciones por provincias a lo largo del periodo de estudio referido, desde enero de dos mil diez hasta mayo de dos mil quince, ambos inclusive.

Con dicho gráfico ya establecido, será posible obtener una impresión de la relevancia del turismo rural en cada una de las provincias del país. A fin de establecer una discriminación que permita conocer la información relativa a aquellas provincias con mayor potencial de inversión, se valorarán únicamente las provincias que a lo largo de estos cinco años y medio hayan superado un valor umbral, un millón de pernoctaciones totales.

Una vez obtenidas las series de interés, será el momento de realizar el estudio propiamente dicho del potencial de inversión utilizando las herramientas escogidas para dicho fin. Por este orden, las aplicaciones serán:

- Herramienta de desarrollo propio. Forecast BI.
- Weka.
- Excel
- SPSS.
- R programming.
- Tableau.

Para el estudio se han descartado las de SAP Lumira y PSPP por no permitir completar el proceso de estimación de series temporales.

El estudio consistirá en la estimación de los siguientes doce meses a la última estación seleccionada, mayo de dos mil quince, hasta mayo de dos mil dieciséis. Una vez que se obtengan estos datos se realizará una suma del acumulado de la serie estimada y en último término se relacionará este agregado con el número total de establecimientos de cada territorio.

El tercer paso consistirá en seleccionar aquellas provincias que presenten una mejor relación de estimación con respecto al número de establecimientos valorando la media arrojada por todas aplicaciones. De las provincias seleccionadas se realizará un estudio del perfil de huésped según los datos alojados en el Big Data del Observatorio de Turismo Rural alojado en el portal [turismorural.com](http://turismorural.com).

Finalmente, con las aplicaciones empleadas se establecerá una comparativa con el objetivo de valorar las fortalezas y debilidades de cada una de las herramientas en base a la experiencia del trabajo llevado con cada una de las mismas.

Como ejemplo de las series de estudio se presentan los datos referidos a las provincias extremeñas de Badajoz (Tabla 1) y Cáceres (Tabla 2). El resto de las series se encuentran en el Anexo III.

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	1007	1576	1593	2863	1788	1895	1492	3119	1172	2795	778	1766
2011	713	1131	1466	2414	1254	1576	2028	3598	1676	1736	1146	2111
2012	1392	1318	1902	4195	1571	1366	2637	3784	2486	3249	2562	3586
2013	1690	2108	3342	2226	2376	2328	3691	4101	2810	2510	2970	3692
2014	1320	2460	2209	4987	3240	2937	3844	6875	3368	3643	2407	4336
2015	1792	2467	3353	4520	3780							

Tabla 1.- Serie temporal pernотaciones en alojamientos rurales en Badajoz 01/10 - 05/15.

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	8320	11303	22983	33818	12901	12980	19720	30682	13648	25709	11384	13765
2011	7627	11884	19101	36534	13329	14750	23608	38779	16668	18942	14124	16505
2012	9009	11789	17538	37751	11325	13656	19082	38095	11920	17482	14877	16785
2013	5285	8274	24548	19080	15993	14934	19122	37305	16312	14605	16920	19921
2014	7890	10996	16522	30542	19605	17131	23370	50541	17031	17764	14285	22444
2015	9152	13717	22525	30271	20914							

Tabla 2.- Serie temporal pernотaciones en alojamientos rurales en Cáceres 01/10 - 05/15.

En el gráfico 1 se muestra el acumulado de todas las provincias durante los últimos cinco años. Como se puede apreciar en el gráfico, la mayor parte del turismo rural en España se concentra en la mitad Norte peninsular, especialmente cerca de zonas montañosas aunque también se aprecia cerca de la costa.

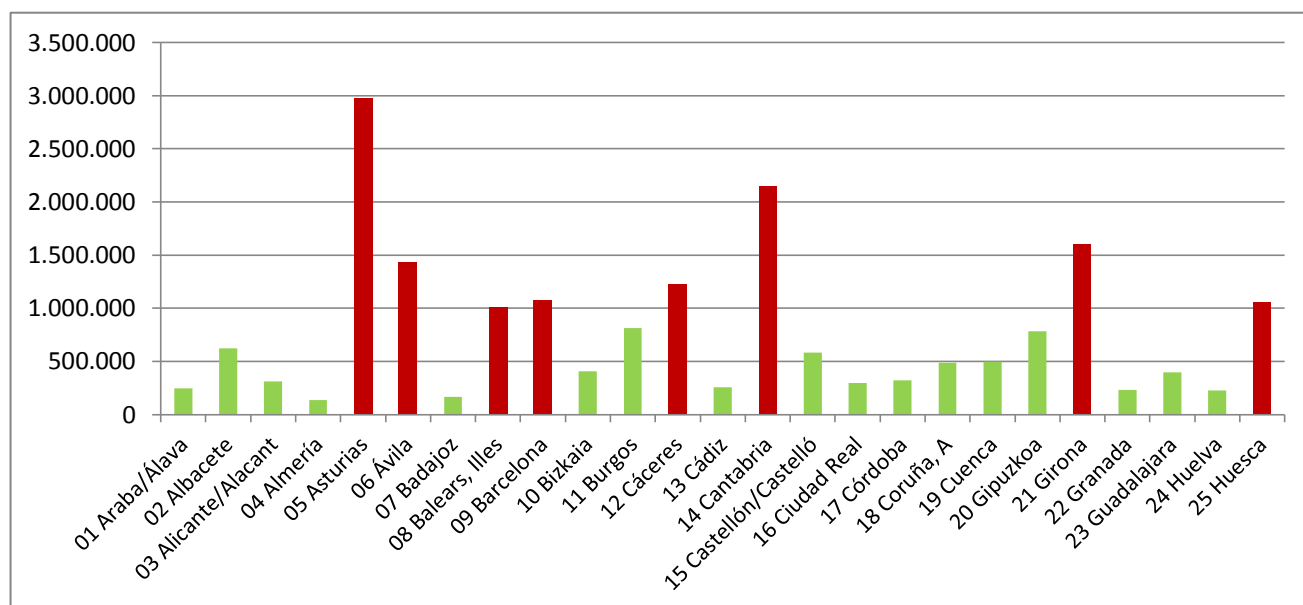


Figura 39.- Pernотaciones por provincias 01/10 - 05/15

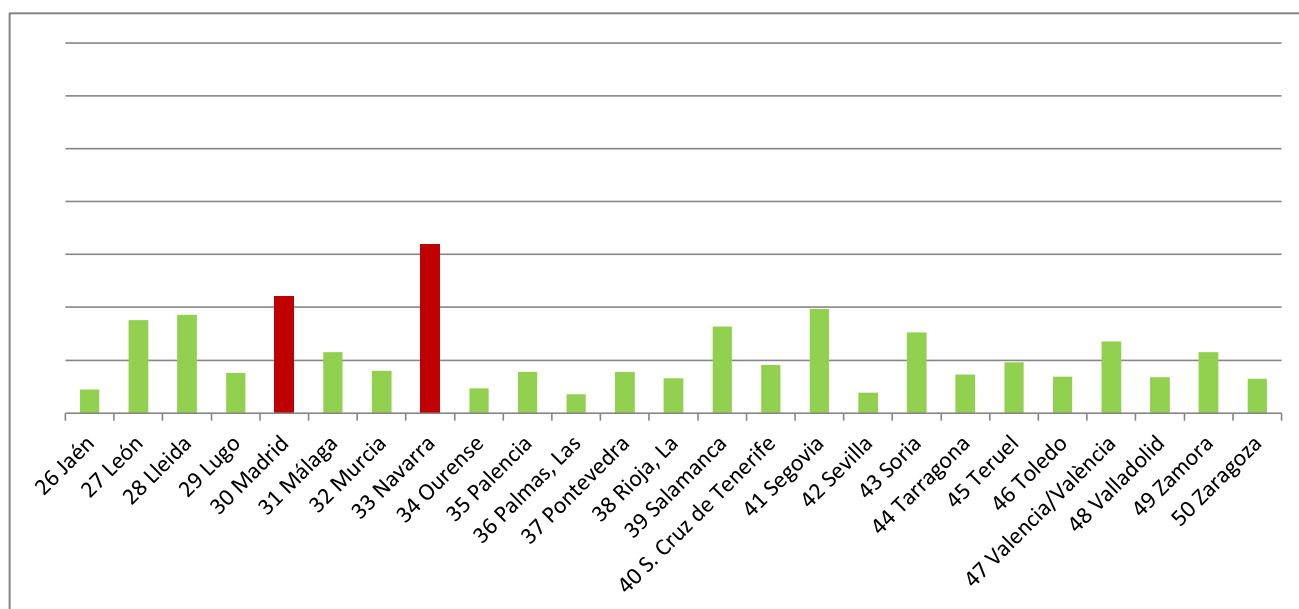
Para la realización del estudio se ha establecido una discriminación sobre aquellas provincias que han sobrepasado el millón de pernoctaciones en el periodo de estudio resaltándolas en rojo para hacer más sencilla su diferenciación. De esta forma se realizará el estudio sobre las provincias de Asturias, Ávila, Baleares, Barcelona, Cáceres, Cantabria, Girona, Huesca, Madrid y Navarra.

Para completar el estudio se relacionarán los datos pronosticados de las provincias seleccionadas con el número de establecimientos existentes en dichas provincias a mayo de 2015 como se ve en la tabla 3 (INE, 2015).

Provincia	Número de establecimiento	Plazas totales estimadas
Asturias	1316	12397
Ávila	791	5944
Baleares	304	6002
Barcelona	492	4245
Cáceres	443	5423
Cantabria	381	6884
Girona	708	5928
Huesca	632	4723
Madrid	234	3427
Navarra	657	4086

Tabla 3.- Relación del número de establecimientos y sus plazas estimadas de las provincias seleccionadas.

A continuación se detallará el estudio realizado sobre los datos obtenidos con cada una de las aplicaciones seleccionadas para el mismo.



## 5.2.- Herramienta de desarrollo propio. Forecast BI.

Este es el caso de la herramienta desarrollada en propiedad por la organización o negocio. Como ya se dijo, la principal ventaja que presenta esta opción es cubrir concisamente las necesidades específicas de la organización. En el caso de esta herramienta se perseguía obtener previsiones para series temporales, pudiendo establecer acciones de cálculo sobre cada una o relaciones entre las distintas series y que finalmente el resultado sea volcado en formato de informe a un archivo pdf.

En esta captura (Figura 39) se puede apreciar el estado inicial de la aplicación previa a la realización del estudio, poniendo como ejemplo a la provincia de Cáceres.

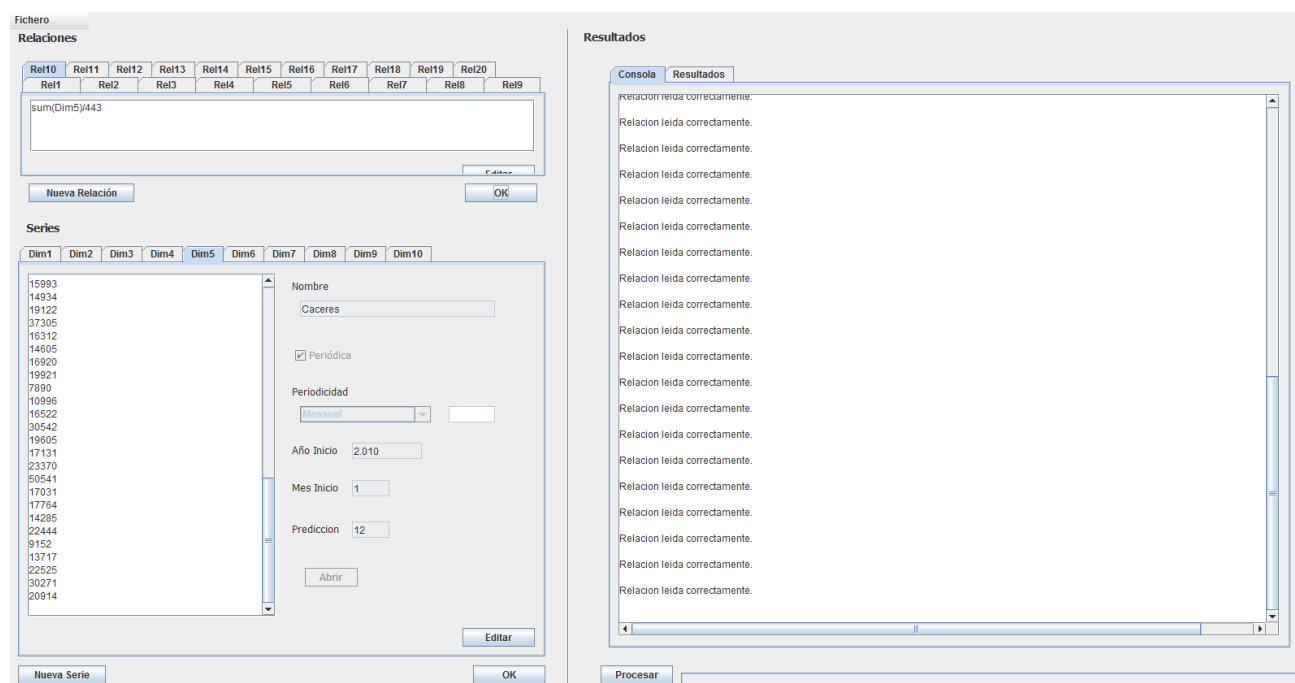


Figura 40.- FBI antes de la realización del estudio.

Aprovechando las capacidades de la aplicación se ha realizado la lectura de las series temporales además de establecer la suma acumulada de la predicción de cada una de las series y finalmente relacionando dicha suma con el número de establecimientos de cada una de las provincias seleccionadas.

Tal como se dijo en el capítulo de presentación, las relaciones y operaciones definidas sobre las relaciones se escriben en lenguaje R debido al hecho de que es en este lenguaje en el que está definido el motor de cálculos de la aplicación. Ello hace que sea necesario que el usuario tenga conocimientos en esta tecnología, aunque no se precisa que sean muy profundos.

La siguiente captura (Figura 40) muestra la estimación de la serie referida a la provincia de Cáceres para los siguientes doce meses a la última estación de la serie, es decir, el periodo comprendido entre junio de dos mil quince y mayo de dos mil dieciséis.

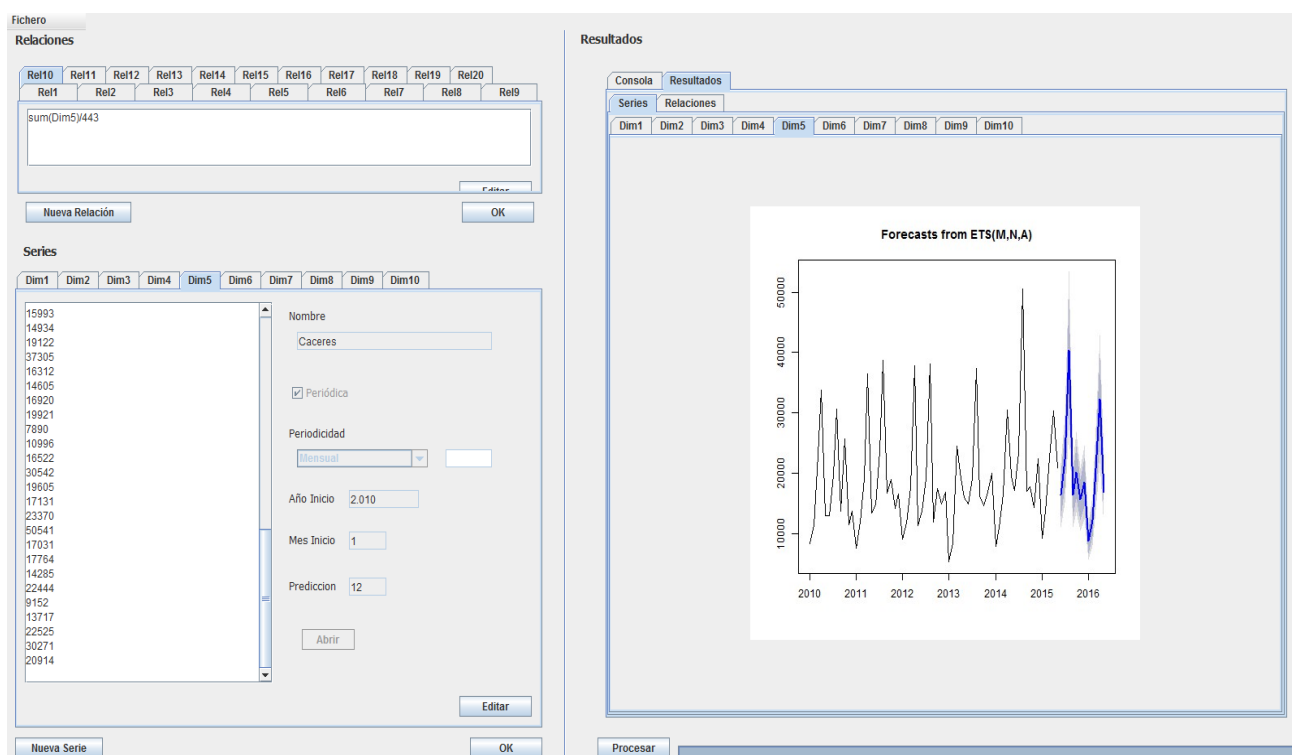


Figura 41.- FBI después de la realización del estudio.

El estudio de las series demuestra que en las provincias de Asturias, Cantabria, Girona y Navarra se espera un mayor número total acumulado de visitantes durante los próximos doce meses. Sin embargo, si se atiende a la relación que se establece entre este número de visitantes y el número de establecimientos totales abiertos en la provincia se observa un cambio en el potencial de inversión situando a la provincia de Cantabria como la que mayor potencial de inversión presenta, seguida de Madrid, Baleares y Cáceres.

El resumen del estudio con esta herramienta sería el siguiente:

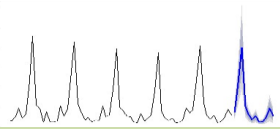


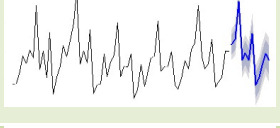
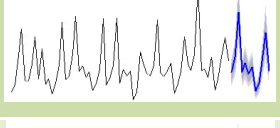
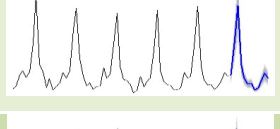
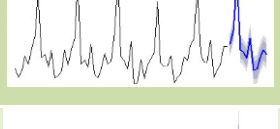
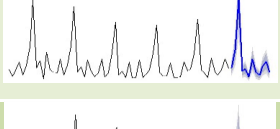
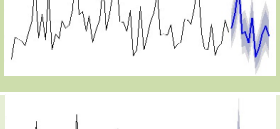
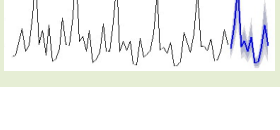
Provincia	Acumulado	Relación
Asturias	570.901	433,82
Ávila	292.231	369,45
Baleares	194.073	638,40
Barcelona	231.395	470,31
Cáceres	240.467	542,82
Cantabria	410.277	1076,84
Girona	361.908	511,17
Huesca	218.381	345,54
Madrid	193.641	827,52
Navarra	299.956	456,55

Tabla 4.- Resumen resultados del estudio FBI.

A continuación se presentan los resultados del estudio detalladamente en la tabla 5.

Provincia	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Enero
Asturias	34.270	8.836	10.537	19.889	44.821	25.118	108.972	200.518
Ávila	21.402	35.778	47.539	12.624	13.955	22.102	27.572	19.332
Baleares	30.459	38.676	2.896	4.792	9.853	18.494	24.954	3.029
Barcelona	22.094	23.839	8.367	11.105	15.099	19.027	16.655	36.526
Cáceres	16.344	22.287	8.732	12.225	20.674	32.271	16.863	40.358
Cantabria	28.499	8.247	11.393	20.454	30.856	23.905	69.726	119.698
Girona	31.238	39.318	70.323	15.864	16.533	22.724	27.982	25.275
Huesca	13.619	27.002	10.318	8.004	14.516	18.392	10.354	64.987
Madrid	16.416	19.329	25.088	10.663	12.576	15.104	16.834	14.747
Navarra	18.735	35.794	7.661	8.653	18.679	36.332	20.703	64.948

Tabla 5.- Resultados de estudio FBI

Febrero	Marzo	Abril	Mayo	Total	Número de Establec.	Relación	Gráfico
48.508	29.104	15.528	24.800	570.901	1316	433,82	
17.239	25.758	19.403	29.527	292.231	791	369,45	
33.166	19.808	4.130	3.817	194.073	304	638,40	
16.834	19.241	16.846	25.762	231.395	492	470,31	
16.380	20.197	15.578	18.559	240.467	443	542,82	
40.359	23.859	15.935	17.346	410.277	381	1076,84	
27.962	26.813	23.096	34.780	361.908	708	511,17	
11.159	12.910	7.062	20.058	218.381	632	345,54	
15.307	15.734	13.414	18.429	193.641	234	827,52	
19.917	24.528	16.550	27.455	299.956	657	456,55	

## 5.2.- Weka.

La aplicación Weka está expresamente diseñada para la minería de datos, es por ello muy apropiada a la hora de realizar predicciones de series temporales. Su interfaz es muy sencilla y su modo de trabajo no es especialmente complicado. A fin de dar a conocer el modo de realizar cálculos de series temporales con esta herramienta se ha incluido un manual de uso en el Anexo II de este documento. En este caso se precisa de la utilización de un fichero .csv para trabajar con la aplicación.

En esta captura (Figura 41) se puede apreciar el estado inicial de la aplicación previa a la realización del estudio, poniendo como ejemplo a la provincia de Cáceres.

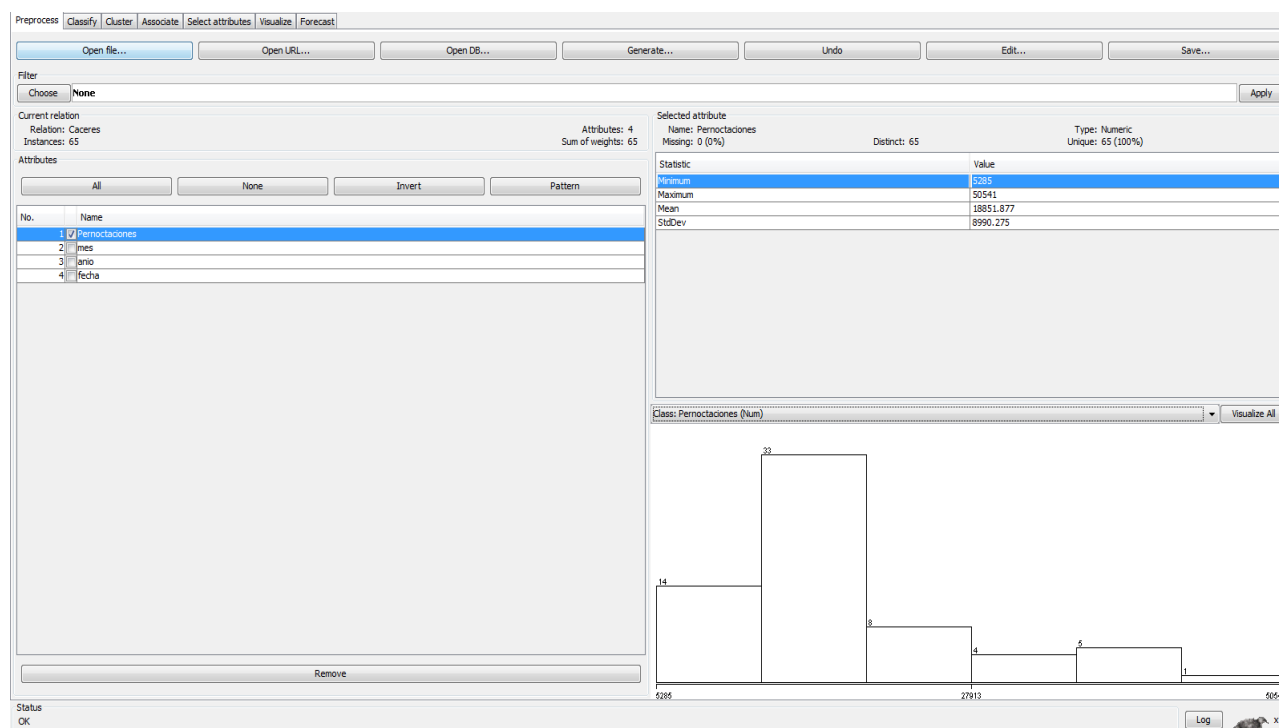


Figura 42.- WEKA antes de la realización del estudio

La interfaz de Weka es muy sencilla, como ya se ha dicho, solicitando al usuario establecer qué serie calcular, qué indicador utilizar como medida de tiempo, que en este caso se le indicará como una variable artificialmente calculada, lo que indica que no necesita que se la proporcione el usuario, además del número de estaciones a calcular y la periodicidad que presenta la serie.

En la siguiente captura se indica el estado de la aplicación después de finalizar los cálculos utilizando como ejemplo a la provincia de Cáceres para los siguientes doce meses a la última estación de la serie, es decir, el periodo comprendido entre junio de dos mil quince y mayo de dos mil dieciséis.



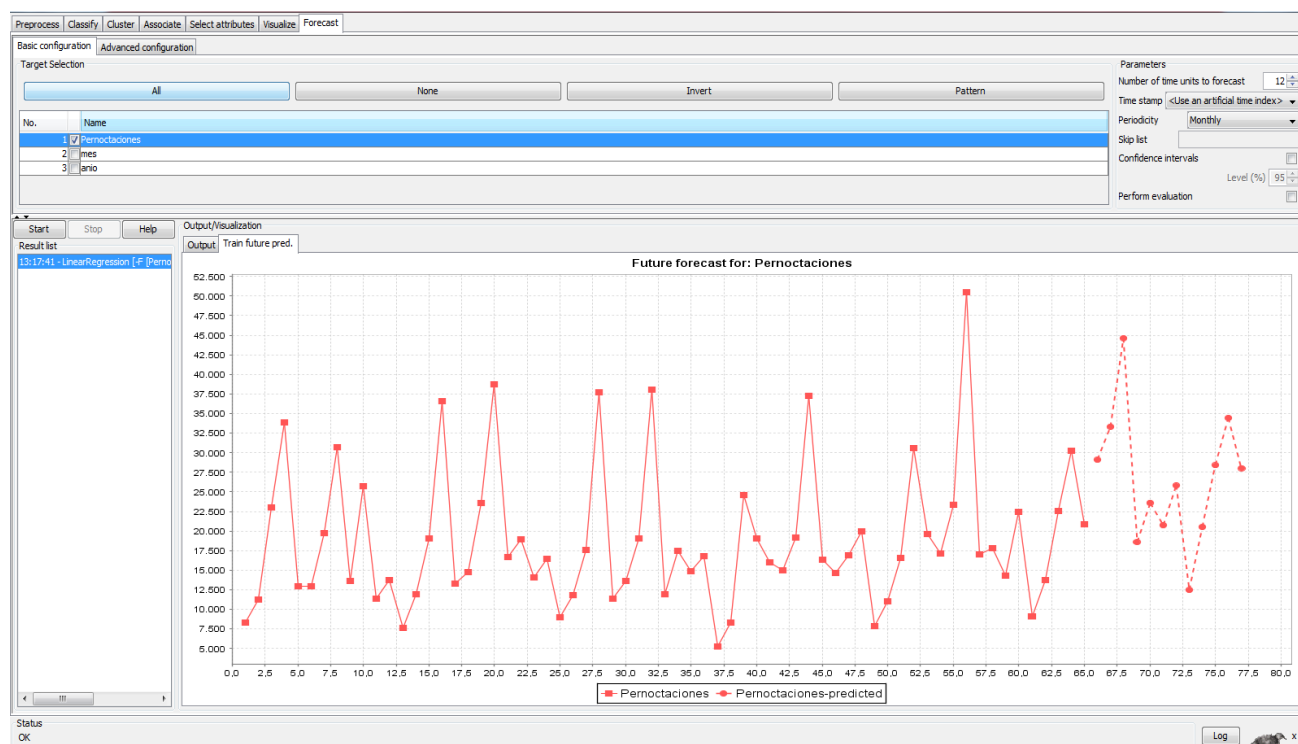


Figura 43.- WEKA después de la realización del estudio.

El estudio realizado mediante Weka arroja como resultados que las provincias de Asturias, Girona, Cantabria y Navarra recibirían un mayor número de visitantes según los cálculos estimados. Sin embargo, al aplicar la relación con el número de establecimientos existentes, son Cantabria, Ávila, Madrid y Cáceres, por este orden, las que presentarían un mayor potencial de inversión.

El resumen del estudio con esta herramienta sería el siguiente:

Provincia	Acumulado	Relación
Asturias	603.226	458,38
Ávila	323.607	409,11
Baleares	323.607	1064,50
Barcelona	275.229	559,41
Cáceres	319.672	721,61
Cantabria	435.276	1142,46
Girona	442.380	624,83
Huesca	224.161	354,68
Madrid	175.996	752,12
Navarra	330.212	502,61










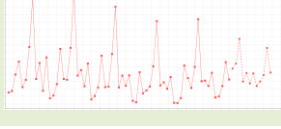
Tabla 6.- Resumen resultados del estudio WEKA.

A continuación se muestran los resultados detallados del estudio en la tabla 7.

Provincia	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Enero
Asturias	38.716	107.822	195.332	76.015	32.207	19.832	25.784	9.969
Ávila	27.093	31.343	37.073	26.212	24.664	21.865	27.141	21.735
Baleares	27.093	31.343	37.073	26.212	24.664	21.865	27.141	21.735
Barcelona	24.989	23.074	35.938	25.566	17.174	20.589	27.784	11.669
Cáceres	29.082	33.358	44.624	18.597	23.623	20.725	25.789	12.460
Cantabria	32.072	62.856	109.627	34.472	36.424	23.452	14.107	18.259
Girona	41.389	40.905	66.356	30.457	30.577	29.836	31.175	17.145
Huesca	7.226	24.650	51.152	13.028	11.700	7.821	23.764	23.832
Madrid	15.267	17.757	20.916	14.416	12.050	12.405	14.527	11.637
Navarra	28.819	32.161	48.771	20.153	25.759	18.744	25.616	17.043

Tabla 7.- Resultados de estudio WEKA

# Business Intelligence applied to tourism | 2014 -2015

Febrero	Marzo	Abril	Mayo	Total	Número de Establec.	Relación	Gráfico
10.439	18.289	39.888	28.932	603.226	1316	458,38	
20.343	23.260	31.824	31.054	323.607	791	409,11	
20.343	23.260	31.824	31.054	323.607	304	1064,50	
17.467	19.944	21.586	29.448	275.229	492	559,41	
20.536	28.392	34.451	28.034	319.672	443	721,61	
7.674	39.863	35.509	20.960	435.276	381	1142,46	
32.242	23.776	53.134	45.387	442.380	708	624,83	
3.803	10.075	36.175	10.936	224.161	632	354,68	
10.562	13.027	17.627	15.804	175.996	234	752,12	
20.139	24.326	42.622	26.058	330.212	657	502,61	

### 5.3.- Excel

La herramienta Excel es de sobra conocida pues forma parte de una de las más potentes suites ofimáticas del momento, Microsoft Office. Esta herramienta que comenzó como una simple hoja de cálculo ha ido sufriendo mejoras a lo largo de su historia y recibiendo módulos de trabajo que le han conferido una potencia y usabilidad tremenda, pues a pesar de la amplia gama de prestaciones que ofrece es sencilla de manejar y además existe una inmensa cantidad de documentación e información fácilmente accesible desde Internet.

La siguiente captura (Figura 43) muestra el estado de la aplicación antes de realizar el procesamiento, utilizando la provincia de Cáceres como ejemplo.

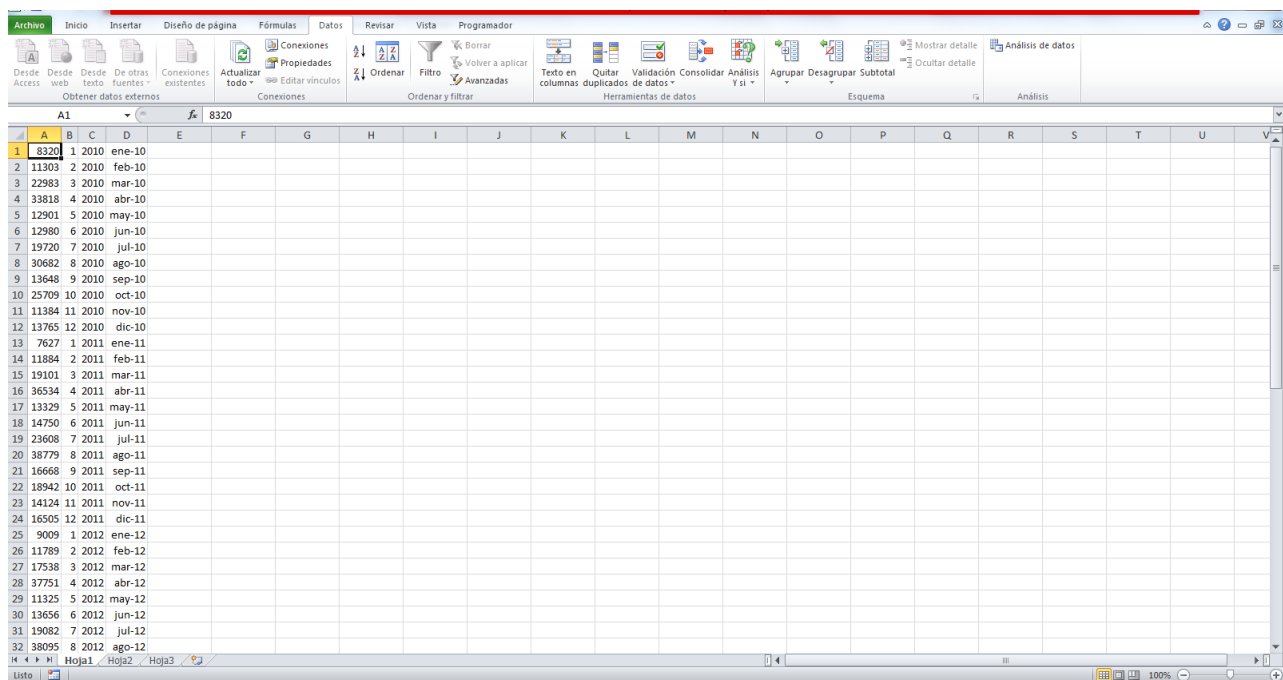


Figura 44-. Estado de Excel antes de realizar el estudio.

Excel no es un software puramente estadístico, ello puede implicar que presente carencias a la hora de realizar estimaciones sobre series temporales.

Para llevar a cabo este estudio será necesario realizar y calcular manualmente la desestacionalización de la serie para posteriormente obtener una recta de regresión que permita obtener la tendencia a partir de la cual realizar la previsión de los datos. Estas rectas son escogidas de un determinado conjunto de rectas predefinidas que posee la propia herramienta lo que puede limitar en gran medida la veracidad y validez de los datos estimados.

La siguiente captura muestra el estado de la herramienta una vez finalizados el cálculo de los datos utilizando como ejemplo la provincia de Cáceres.

## Business Intelligence applied to tourism | 2014 -2015

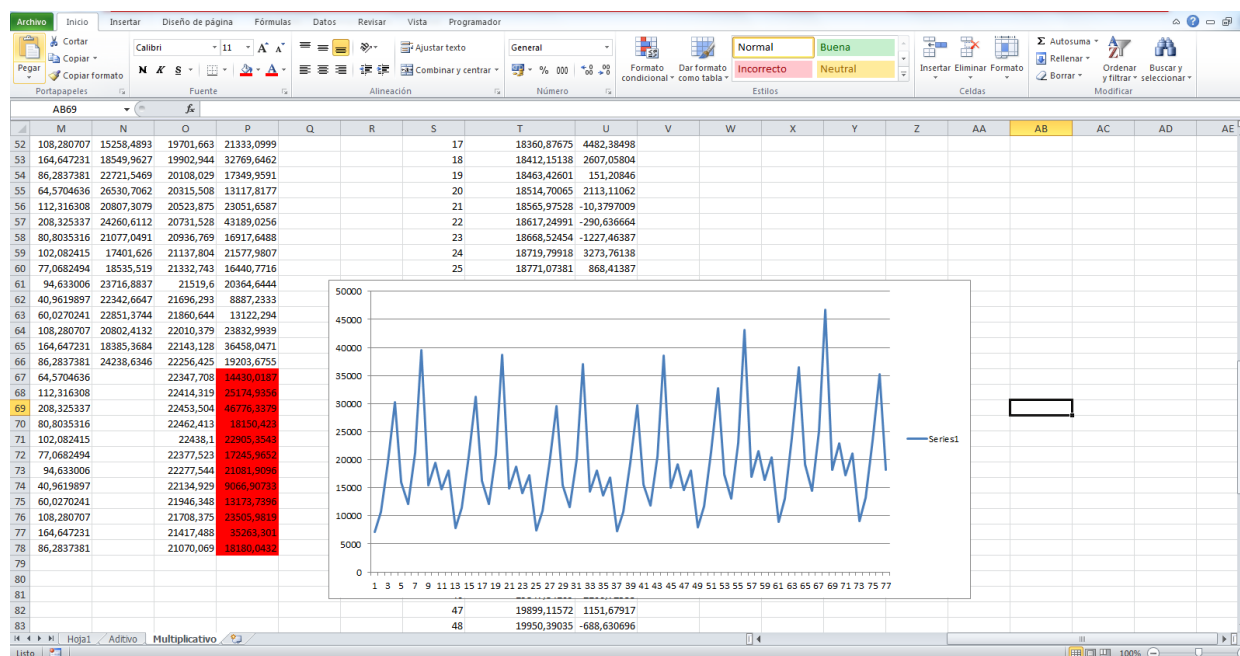


Figura 45.- Estado Excel después de realizar el estudio

Tras el estudio se desprende que se espera un número de pernoctaciones muy similar en todos los territorios otorgando una ligera ventaja a Asturias, seguida de Cantabria, Girona y Navarra. Sin embargo, al relacionar estos datos con el número de establecimientos se obtiene que Madrid es la provincia con mayor potencial de inversión en turismo rural, seguida de Baleares, Cantabria y Cáceres por este orden.

El resumen del estudio con esta herramienta sería el siguiente:

Provincia	Acumulado	Relación
Asturias	266.726	202,68
Ávila	265.578	335,75
Baleares	265.525	873,44
Barcelona	265.665	539,97
Cáceres	264.955	598,09
Cantabria	266.299	698,95
Girona	265.853	375,50
Huesca	266.045	420,96
Madrid	265.228	1133,45
Navarra	265.702	404,42

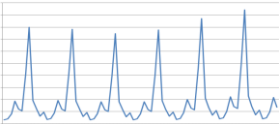


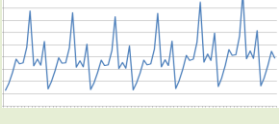



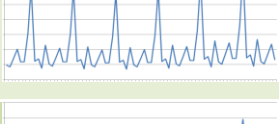


Tabla 8.- Resumen resultados del estudio Excel.

A continuación se muestran los resultados detallados del estudio.

Provincia	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Enero
Asturias	12.560	49.886	94.392	22.846	14.306	7.293	11.284	4.020
Ávila	16.037	32.541	46.611	15.851	24.216	17.985	27.316	11.037
Baleares	26.568	41.669	56.812	36.458	22.100	4.637	4.056	3.155
Barcelona	21.040	28.560	46.003	19.308	22.535	19.480	30.830	8.200
Cáceres	14.430	25.175	46.776	18.150	22.905	17.246	21.082	9.067
Cantabria	15.107	45.584	77.308	26.714	15.786	10.085	11.190	5.114
Girona	18.636	30.521	58.697	20.476	19.412	16.106	26.506	9.245
Huesca	14.120	36.040	72.577	14.604	16.417	8.932	26.760	12.172
Madrid	18.384	26.807	34.480	21.535	22.044	19.012	25.232	15.496
Navarra	12.975	32.904	62.922	17.965	22.318	14.611	25.105	6.924

Tabla 9.- Resultados de estudio Excel

Business Intelligence applied to tourism | 2014 -2015

Febrero	Marzo	Abril	Mayo	Total	Número de Establec.	Relación	Gráfico
4.916	9.936	21.690	13.597	266.726	1316	202,68	
13.290	18.396	24.911	17.389	265.578	791	335,75	
5.275	11.337	22.949	30.509	265.525	304	873,44	
11.397	16.387	22.348	19.577	265.665	492	539,97	
13.174	23.506	35.263	18.180	264.955	443	598,09	
7.607	13.202	21.269	17.333	266.299	381	698,95	
10.499	15.856	20.927	18.972	265.853	708	375,50	
10.436	16.967	23.699	13.320	266.045	632	420,96	
17.904	20.703	23.306	20.325	265.228	234	1133,45	
7.705	16.349	28.299	17.623	265.702	657	404,42	

### 5.4.- SPSS

Dentro de las herramientas y soluciones software de índole estadístico, SPSS es sin duda el más célebre de todos. Esta potente herramienta permite realizar una inmensa cantidad de cálculos y procesamiento estadísticos de un modo relativamente sencillo. Sin embargo su propia naturaleza hace que, aunque más sencillo que en otras herramientas, el usuario deba poseer importantes conocimientos tanto estadísticos como de la propia herramienta. Debido a ello, se ha incluido en el Anexo II un manual que pueda actuar como guía para el trabajo con series temporales con esta herramienta.

La siguiente captura (Figura 45) muestra el estado de la aplicación antes de realizar el procesamiento, utilizando la provincia de Cáceres como ejemplo.

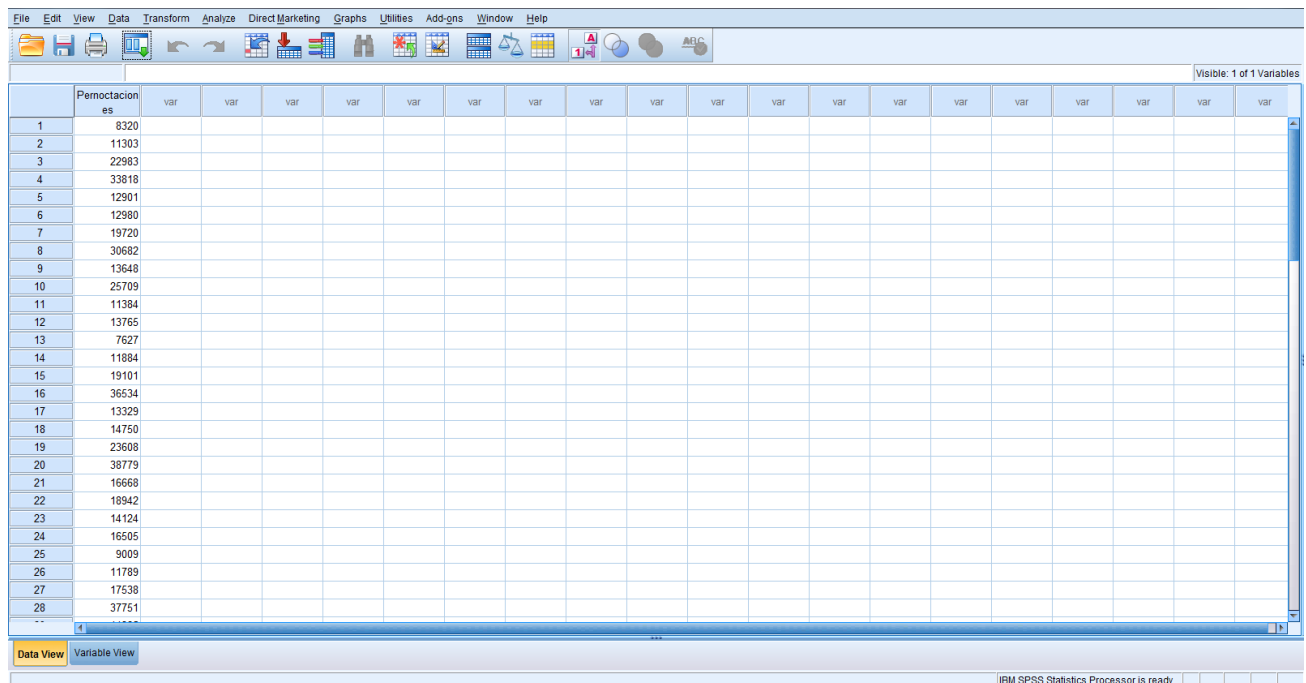


Figura 46.- Estado SPSS antes de estudio.

Como se dijo antes, SPSS es una herramienta muy poderosa aunque no precisamente trivial a la hora de trabajar con ella. La gran potencia que presenta hace que sea posible realizar cálculos que con otras herramientas serían bastante complejos agrupándolos en una misma acción. De esta forma es posible, por ejemplo, desestacionalizar una serie de forma automática, simplificando mucho el trabajo a realizar.

En la siguiente captura (Figura 46) se muestra el estado de la aplicación una vez se han realizado los cálculos para la obtención de los datos estimados, utilizando como ejemplo la provincia de Cáceres.



## Business Intelligence applied to tourism | 2014 -2015

ia	Cociente	Pernoc_MMC	ERR_1	SAS_1	SAF_1	STC_1	FIT_2	var	var	var	var	var	var	var	var
52	462,00	1,60	20564,5	,91844	17804,13306	1,71544	19385,28729	19485,28631							
53	612,00	1,23	20576,3	1,14601	23879,50897	,82100	20837,16449	19687,62833							
54	197,00	1,15	20571,6	,97532	21174,54295	,80904	21710,27916	19905,12313							
55	248,00	1,22	20729,3	,95435	21063,20160	1,10952	22070,68088	20138,21425							
56	236,00	1,35	20895,3	1,10830	24179,32936	2,09026	21816,57347	20387,34527							
57	719,00	1,04	21258,8	,99742	20859,26954	,81647	20913,21367	20652,95974							
58	159,00	1,22	21497,6	,92072	18791,68792	,94531	20409,67591	20935,50123							
59	635,00	,84	21540,9	,90922	18635,87823	,76653	20496,55246	21235,41329							
60	523,00	1,13	-	1,11383	24142,39453	,92965	21675,02078	21553,13949							
61	262,00	1,16	-	,97636	21833,90300	,41916	22362,52049	21889,12339							
62	721,00	1,25	-	1,03012	22916,75307	,59856	22246,69113	22243,80854							
63	003,00	1,36	-	1,04866	23006,80197	,97906	21939,33782	22517,63852							
64	271,00	,99	-	,80056	17646,15650	1,71544	22042,29026	23011,05687							
65	309,00	1,07	-	1,15299	25473,91229	,82100	22093,76648	23424,50717							
66	-	-	-	-	-	-	-	23858,43297							
67	-	-	-	-	-	-	-	24313,27784							
68	-	-	-	-	-	-	-	24789,48533							
69	-	-	-	-	-	-	-	25287,49900							
70	-	-	-	-	-	-	-	25807,76242							
71	-	-	-	-	-	-	-	26350,71915							
72	-	-	-	-	-	-	-	26916,81275							
73	-	-	-	-	-	-	-	27506,48678							
74	-	-	-	-	-	-	-	28120,18480							
75	-	-	-	-	-	-	-	28758,35037							
76	-	-	-	-	-	-	-	29421,42705							
77	-	-	-	-	-	-	-	30109,85841							
78	-	-	-	-	-	-	-								
79	-	-	-	-	-	-	-								

Figura 47.- Estado SPSS después de la realización del estudio.

El estudio de las series llevado a cabo arroja como resultado que las provincias de Girona, Navarra, Cantabria y Ávila serían las que registrarían un mayor número de pernoctaciones. No obstante, al poner en juego el número de establecimientos abiertos en cada territorio, serían Cantabria, Cáceres, Madrid y Girona las que demostrarían un mayor interés de inversión.

Este sería el resumen del estudio:

Provincia	Acumulado	Relación
Asturias	301.296	228,95
Ávila	386.113	488,13
Baleares	43.784	144,03
Barcelona	37.419	76,05
Cáceres	320.068	722,50
Cantabria	402.173	1055,57
Girona	488.082	689,38
Huesca	312.324	494,18
Madrid	211.626	904,38
Navarra	407.673	620,51

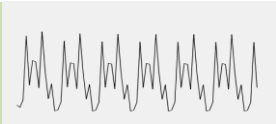

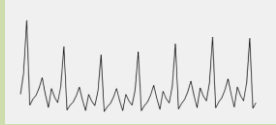
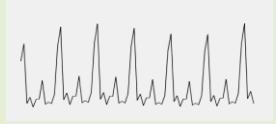

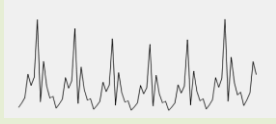
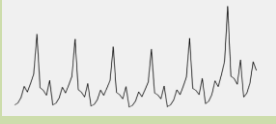
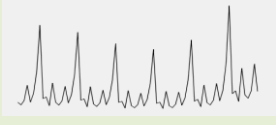

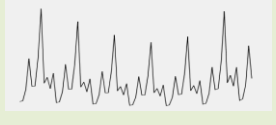
Tabla 10.- Resumen resultados del estudio SPSS

El estudio detallado se muestra a continuación.

Provincia	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Enero
Asturias	33747	33383	20442	48818	28671	15162	22720	9091
Ávila	26045	44409	62397	22081	31697	25794	40174	15887
Baleares	2880	3690	4828	3230	1910	3992	3110	2563
Barcelona	2113	2133	3745	1629	1812	1695	2535	6753
Cáceres	19302	26976	51816	20647	24396	20199	25023	11530
Cantabria	35013	86155	15076	53286	31131	20586	23050	11303
Girona	36891	48580	96352	35801	33845	28800	49905	17742
Huesca	18459	36999	77023	16145	18127	10793	33740	15245
Madrid	17055	19923	25771	16127	17131	14906	19310	12709
Navarra	21003	44401	86981	26022	32362	23191	39197	10870

Tabla 11.- Resultados estudio SPSS

Business Intelligence applied to tourism | 2014 -2015

Febrero	Marzo	Abril	Mayo	Total	Número de Establec.	Relación	Gráfico
9546	13620	44828	21267	301.296	1316	228,95	
18118	25642	45289	28580	386.113	791	488,13	
4418	9015	1785	2363	43.784	304	144,03	
8534	2075	2718	1678	37.419	492	76,05	
16832	28156	50471	24720	320.068	443	722,50	
16273	22575	49528	38197	402.173	381	1055,57	
21114	29900	48332	40821	488.082	708	689,38	
13020	18260	36731	17781	312.324	632	494,18	
14971	16452	19862	17409	211.626	234	904,38	
12139	24229	57584	29693	407.673	657	620,51	

## 5.5.- R programming

Otra de las herramientas abanderadas del ámbito estadístico, al igual que SPSS, es R. R como tal no es una aplicación sino un lenguaje de programación que permite realizar desarrollos para cubrir las necesidades de cómputo estadístico. En esencia es un lenguaje de *scripting*, lo que significa que trabaja a bajo nivel, sin embargo existe un amplio repositorio donde los usuarios pueden alojar paquetes de trabajo específico que hacen más sencillo y asequible el procesamiento de los datos según la necesidad concreta.

La siguiente captura (Figura 47) muestra el estado de la aplicación antes de realizar el procesamiento, utilizando la provincia de Cáceres como ejemplo.

```

47 3287
48 3581
49 1369
50 1248
51 2378
52 5122
53 4218
54 3517
55 4484
56 5839
57 2775
58 2961
59 2615
60 3136
61 1499
62 1402
63 2887
64 5162
65 4332
> sum(serie)
[1] 248410
> clear()
Error: no se pudo encontrar la función "clear"
> clean()
Error: no se pudo encontrar la función "clean"
> dir()
[1] "Alava_01.10-05.15.txt"      "Alicante_01.10-05.15.txt"  "Almeria_01.10-05.15.txt"  "Asturias.csv"              "Asturias_01.10-05.15.txt"
[7] "Avila.csv"                  "Badajoz_01.10-05.15.txt"   "Baleares.csv"             "Baleares_01.10-05.15.txt"  "Barcelona.csv"
[13] "Barcelona_01.10-05.15.txt"  "Bizkaia_01.10-05.15.txt"   "Caceres.csv"              "Caceres_01.10-05.15.txt"   "Cadiz_01.10-05.15.txt"
[19] "Cantabria.csv"             "Castellon_01.10-05.15.txt" "CiudadReal_01.10-05.15.txt" "Cordoba_01.10-05.15.txt"   "Coruña_01.10-05.15.txt"
[25] "Cuenca_01.10-05.15.txt"     "Gipuzkoa_01.10-05.15.txt"  "Girona.csv"               "Girona_01.10-05.15.txt"    "Granada_01.10-05.15.txt"
[31] "Huelva_01.10-05.15.txt"     "Huesca.csv"                "Jaen_01.10-05.15.txt"     "LaRioja_01.10-05.15.txt"   "LasPalmas_01.10-05.15.txt"
[37] "Leon_01.10-05.15.txt"      "Lleida_01.10-05.15.txt"    "Lugo_01.10-05.15.txt"     "Madrid.csv"                "Madrid_01.10-05.15.txt"
[43] "Murcia_01.10-05.15.txt"     "Navarra.csv"               "Ourense_01.10-05.15.txt"  "Palencia_01.10-05.15.txt"  "Pontevedra_01.10-05.15.txt"
[49] "Salamanca_01.10-05.15.txt" "SantaCruz_01.10-05.15.txt" "Segovia_01.10-05.15.txt"  "Sevilla_01.10-05.15.txt"   "Soria_01.10-05.15.txt"
[55] "Teruel_01.10-05.15.txt"     "Toledo_01.10-05.15.txt"    "Valencia_01.10-05.15.txt" "Valladolid_01.10-05.15.txt" "Zamora_01.10-05.15.txt"
[61] "Zaragoza_01.10-05.15.txt"

Error en read.table(file = file, header = header, sep = sep, quote = quote, :
objeto 'False' no encontrado
> serie<-read.csv(file="Asturias_01.10-05.15.txt", head=FALSE, sep=",")
> serieTS<-ts(serie, frequency=12,start=c(2010,1))
Error en ts(serie, frequency = 12, start = c(2010, 1)) :
unused argument (frequency = 12)
> serieTS<-ts(serie, frequency=12,start=c(2010,1))
> local(pkg <- select.list(sort(.packages(all.available = TRUE)),graphics=TRUE))
+ if(nchar(pkg)) library(pkg, character.only=TRUE))

```

Figura 48.- Estado R antes de la realización del estudio.

Cómo se ha dicho antes R cuenta con un gran repositorio en el que se aloja una gran cantidad de paquetes de trabajo. Uno de estos paquetes se llama "forecast" este paquete permite el análisis de una serie temporal de manera automática sin necesidad de realizar cálculo previos. De esta forma es sencillo lograr el objetivo de una forma relativamente fácil y confiable.

En la siguiente captura (Figura 48) se muestra el estado de la aplicación una vez se han realizado los cálculos para la obtención de los datos estimados, utilizando como ejemplo la provincia de Cáceres.

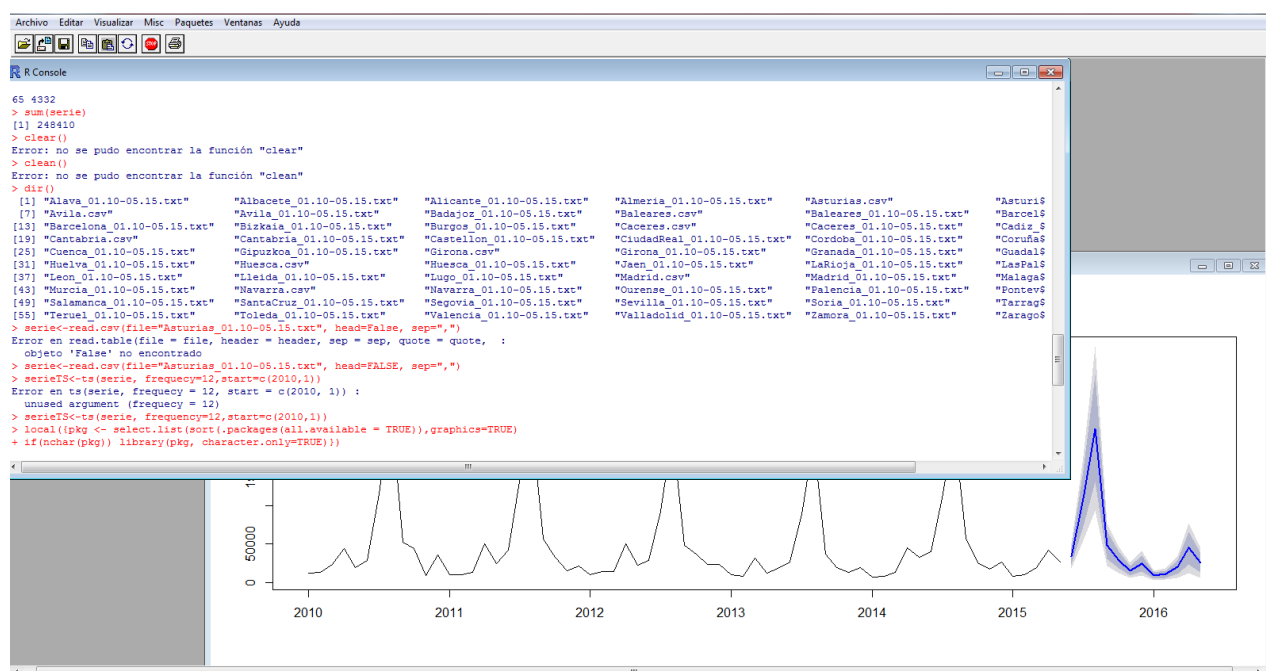


Figura 49.- Estado R después de la realización del estudio.

Tras el análisis de los datos se observa que las provincias de Asturias, Cantabria, Girona y Navarra (por este orden) esperarían un mayor número de pernoctaciones. Sin embargo, al relacionar el número de pernoctaciones estimado con el número de establecimientos abiertos en cada provincia, se desprende que Cantabria, Madrid, Baleares y Cáceres presentan un mayor potencial de inversión a la hora de establecer un nuevo local.

Este sería el resumen del estudio:

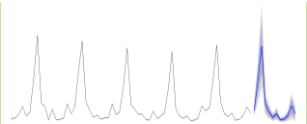









Provincia	Acumulado	Relación
Asturias	570.908	433,82
Ávila	292.231	369,45
Baleares	244.073	802,87
Barcelona	231.395	470,31
Cáceres	240.467	542,82
Cantabria	410.277	1076,84
Girona	361.908	511,17
Huesca	218.381	345,54
Madrid	193.641	827,52
Navarra	299.956	456,55

Tabla 12.- Resumen de resultados del estudio R

El estudio detallado se muestra a continuación.

Provincia	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Enero
Asturias	34271	108973	200519	48508	29104	15529	24800	8837
Ávila	21402	35778	47539	17239	25758	19403	29527	12624
Baleares	30459	38676	53029	33166	19808	4130	3817	2896
Barcelona	22094	23839	36526	16834	19241	16846	25762	8367
Cáceres	16344	22287	40358	16380	20197	15578	18559	8732
Cantabria	28499	69726	119698	40359	23859	15935	17346	8247
Girona	31238	39318	70323	27962	26813	23096	34780	15864
Huesca	13619	27002	64987	11159	12910	7062	20058	10318
Madrid	16416	19329	25088	15307	15734	13414	18429	10663
Navarra	18735	35794	64948	19917	24528	16550	27455	7661

Tabla 13.- Resultados del estudio R

Febrero	Marzo	Abril	Mayo	Total	Número de Establec.	Relación	Gráfico
10537	19889	44822	25119	570.908	1316	433,82	
13955	22102	27572	19.332	292.231	791	369,45	
4792	9853	18494	24954	244.073	304	802,87	
11105	15099	19027	16655	231.395	492	470,31	
12225	20674	32271	16863	240.467	443	542,82	
11393	20454	30856	23905	410.277	381	1076,84	
16533	22724	27982	25.275	361.908	708	511,17	
8004	14516	18392	10354	218.381	632	345,54	
12576	15104	16834	14747	193.641	234	827,52	
8653	18679	36332	20703	299.956	657	456,55	

## 5.6.- Tableau

Al contrario que las aplicaciones anteriores, Tableau no es sólo una herramienta con vertiente estadística sino una herramienta orientada al *Business Intelligence*. Por tanto, el potencial de esta solución va más allá del potencial de cálculo estadístico pues permite realizar gestiones sobre las bases de datos de una organización o negocio, generar informes, crear y mantener tablas de gestión o incluso generar mapas de datos todo ello con una interfaz amigable y sencilla que permite una experiencia satisfactoria al usuario. Como punto negativo citar el hecho de ser software de pago.

La siguiente captura (Figura 49) muestra el estado de la aplicación antes de realizar el procesamiento, utilizando la provincia de Cáceres como ejemplo.

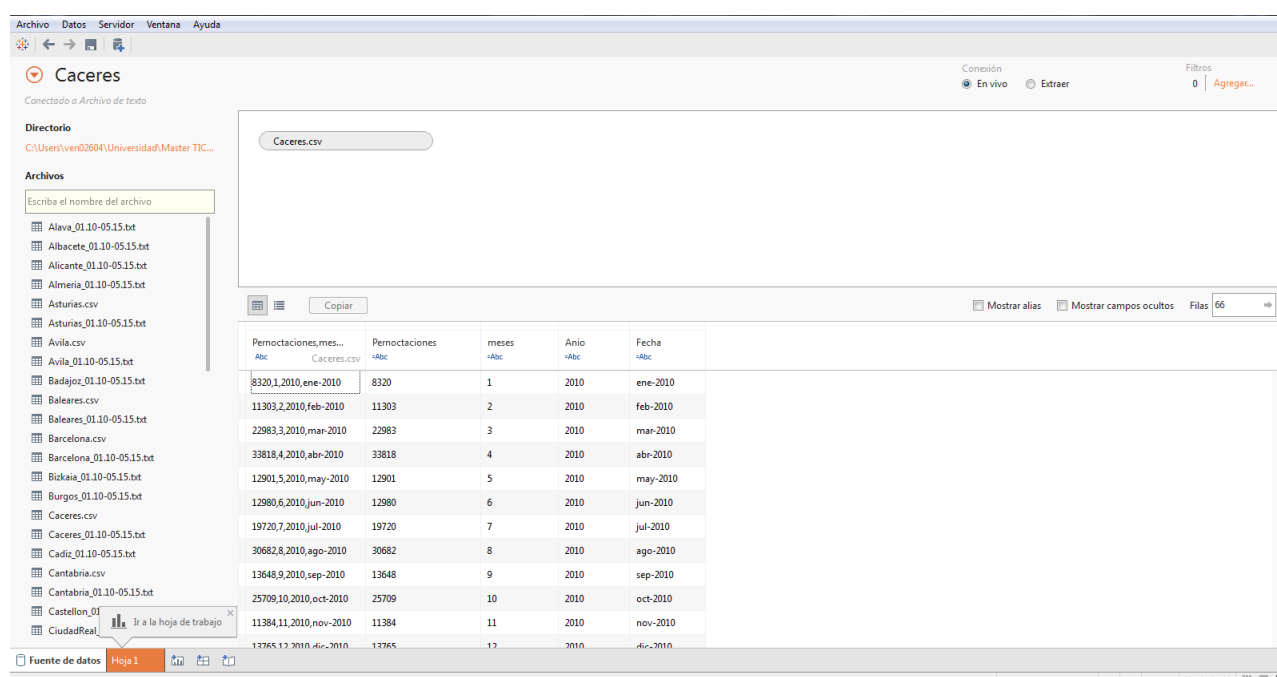


Figura 50.- Estado de Tableau antes del estudio

La sencilla e intuitiva interfaz de Tableau permite realizar el procesado de series de tiempo de una forma muy sencilla que mejoran mucho la experiencia del usuario. Además de la propia ayuda proporcionada por la plataforma existe una gran cantidad de información en Internet con la que un usuario puede resolver los problemas que se le presenten con relativa facilidad. Para ayudar en esta tarea, se ha incluido en el Anexo II un manual de ayuda para analizar series temporales con este programa.

En la siguiente captura (Figura 50) se muestra el estado de la aplicación una vez se han realizado los cálculos para la obtención de los datos estimados, utilizando como ejemplo la provincia de Cáceres.



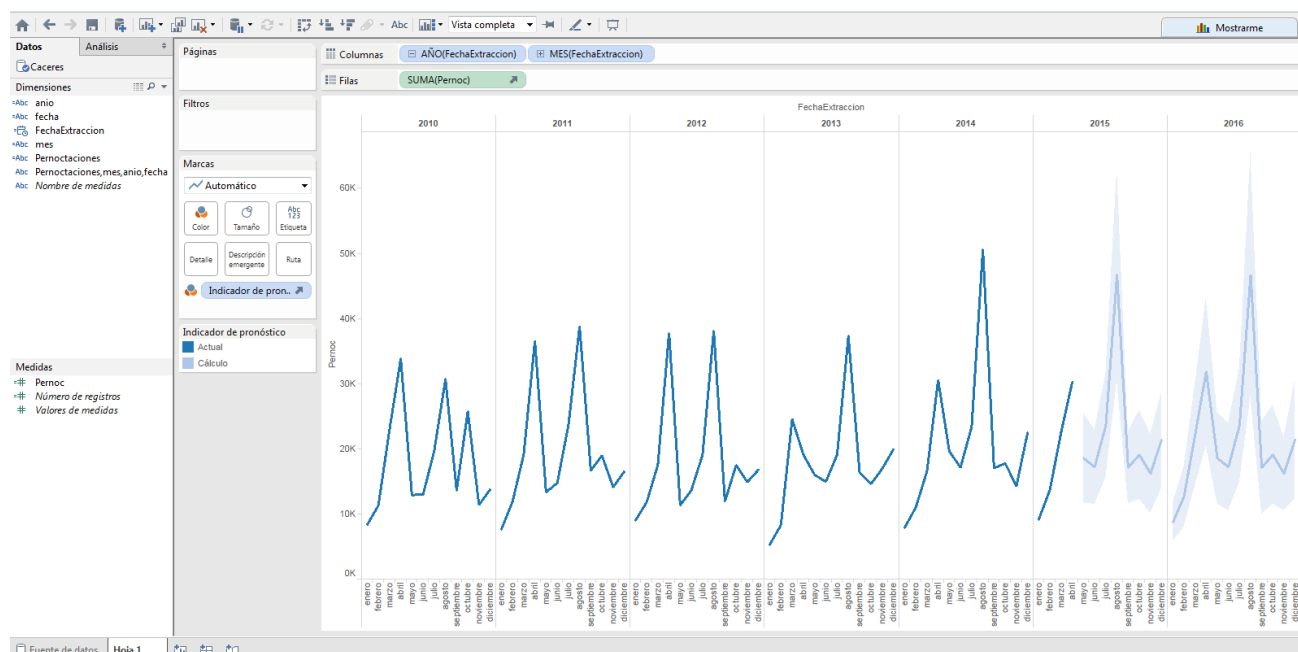


Figura 51.- Estado Tableau después del estudio.

Con la realización del estudio, se ve que las provincias de Asturias, Girona, Cantabria y Navarra obtienen una mayor previsión de pernoctaciones. Sin embargo, cuando una vez más, entra el juego el número correspondiente de establecimientos rurales, son las provincias de Cantabria, Madrid, Baleares y Girona las que tienen un mejor resultado y presenta por tanto un mayor interés a la hora de invertir en este sector.

Este sería el resumen del estudio:

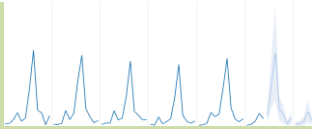
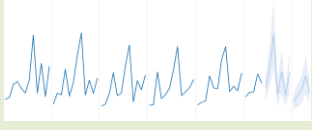
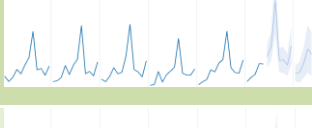
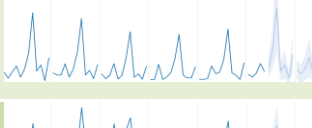

Provincia	Acumulado	Relación
Asturias	623.323	473,65
Ávila	280.103	354,11
Baleares	250.110	822,73
Barcelona	224.746	456,80
Cáceres	259.478	585,73
Cantabria	365.387	959,02
Girona	483.312	682,64
Huesca	265.169	419,57
Madrid	192.100	820,94
Navarra	296.820	451,78

Tabla 14.- Resumen resultados del estudio Tableau

El estudio completo y detallado se muestra a continuación en la tabla 15.

Provincia	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre	Enero
Asturias	35360	126225	220267	53884	38180	11869	28368	10705
Ávila	19717	31108	49766	16386	28669	15782	28398	11991
Baleares	30944	37752	51395	33935	20716	4267	3601	2957
Barcelona	21893	23317	37591	14803	19718	13752	26293	7131
Cáceres	17450	23861	47460	17338	19205	16401	21766	8833
Cantabria	27170	65221	103009	37598	23985	10393	17035	7021
Girona	37517	48599	95539	33609	34139	28736	48583	20274
Huesca	16150	32238	69076	13707	19022	7817	28049	14615
Madrid	15929	18892	24728	14984	15255	13483	18349	11027
Navarra	18183	36840	67307	19584	25808	13079	29316	9175

Tabla 15.- Resultado estudio Tableau.

Febrero	Marzo	Abril	Mayo	Total	Número de Establec.	Relación	Gráfico
11314	18100	47279	21773	623.323	1316	473,65	
15582	18783	26640	17.281	280.103	791	354,11	
5046	10490	20738	28269	250.110	304	822,73	
9818	13906	19718	16806	224.746	492	456,80	
12817	22323	32111	19912	259.478	443	585,73	
10712	18074	25944	19222	365.387	381	959,02	
20896	30319	45344	39.757	483.312	708	682,64	
11112	15818	25384	12179	265.169	632	419,57	
13085	14675	16782	14911	192.100	234	820,94	
10423	18524	32245	16336	296.820	657	451,78	

### 5.7.- Resumen agrupado del estudio.

Entre los resultados de todas las aplicaciones es frecuente y esperable que existan diferencias en los datos estimados debido a los métodos de previsión utilizados, los métodos de suavizado, la confianza otorgada o métodos más técnicos como la reserva de memoria disponible o incluso el tipo de variable de cálculo utilizada. Por este motivo, se seleccionarán las provincias con mayor interés de inversión de acuerdo a la media de la relación entre las pernoctaciones estimadas y los establecimientos abiertos otorgados por cada aplicación.

Provincia	FBI	Weka	Excel	SPSS	R	Tableau	Total
Asturias	433,82	458,38	202,68	228,95	433,82	473,65	371,88
Ávila	369,45	409,11	335,75	488,13	369,45	354,11	387,67
Baleares	638,40	1064,50	873,44	144,03	802,87	822,73	724,33
Barcelona	470,31	559,41	539,97	76,05	470,31	456,80	428,81
Cáceres	542,82	721,61	598,09	722,50	542,82	585,73	618,93
Cantabria	1076,84	1142,46	698,95	1055,57	1076,84	959,02	1001,61
Girona	511,17	624,83	375,50	689,38	511,17	682,64	565,78
Huesca	345,54	354,68	420,96	494,18	345,54	419,57	396,75
Madrid	827,52	752,12	1133,45	904,38	827,52	820,94	877,66
Navarra	456,55	502,61	404,42	620,51	456,55	451,78	482,07

Tabla 16.- Resumen del estudio.

Para expresarlo con mayor claridad, se presenta a continuación el siguiente gráfico de barras en el que figura el total promedio obtenido en la anterior tabla:

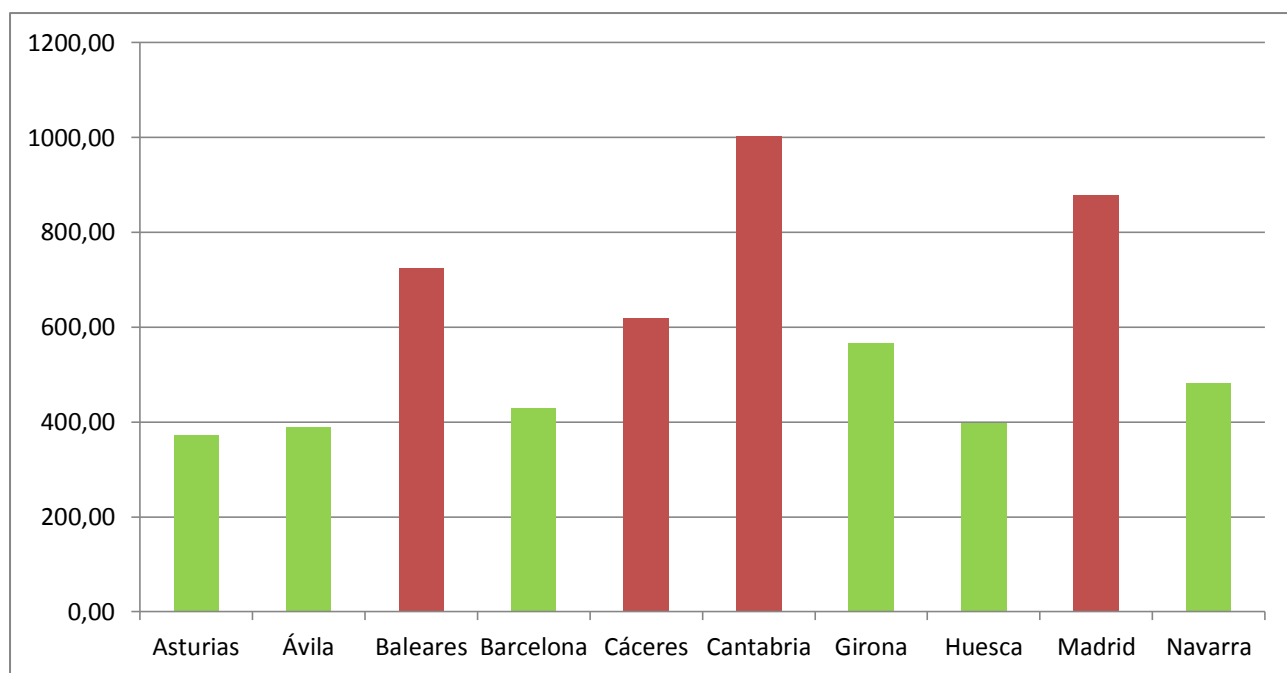


Figura 52.- Provincias de mayor interés de inversión.

Por tanto las provincias de mayor interés a la hora realizar inversión en turismo rural según el estudio realizado en base a las pernoctaciones esperadas y relacionándolas con el número total de establecimientos abiertos a mayo del año dos mil quince serían las siguientes:

Provincia	Pernoctaciones esperadas/ Establecimientos	Establecimientos 2015/05
Cantabria	1001,61	381
Madrid	877,66	234
Baleares	724,33	304
Cáceres	618,93	443

Tabla 17.- Pernoctaciones esperadas en provincias de mayor interés de inversión.

Estas provincias han sido seleccionadas utilizando como discriminación el ratio de 600 visitantes por establecimiento. A continuación se realizará un estudio del perfil del visitante mayoritario en estas provincias para conocer la oferta que debe promocionar en cada territorio.

### 5.8.- Provincia número 1. Cantabria

Atendiendo a los datos obtenidos de la serie temporal se puede ver las épocas del año donde se concentran la mayor parte de las pernoctaciones en la provincia de Cantabria.

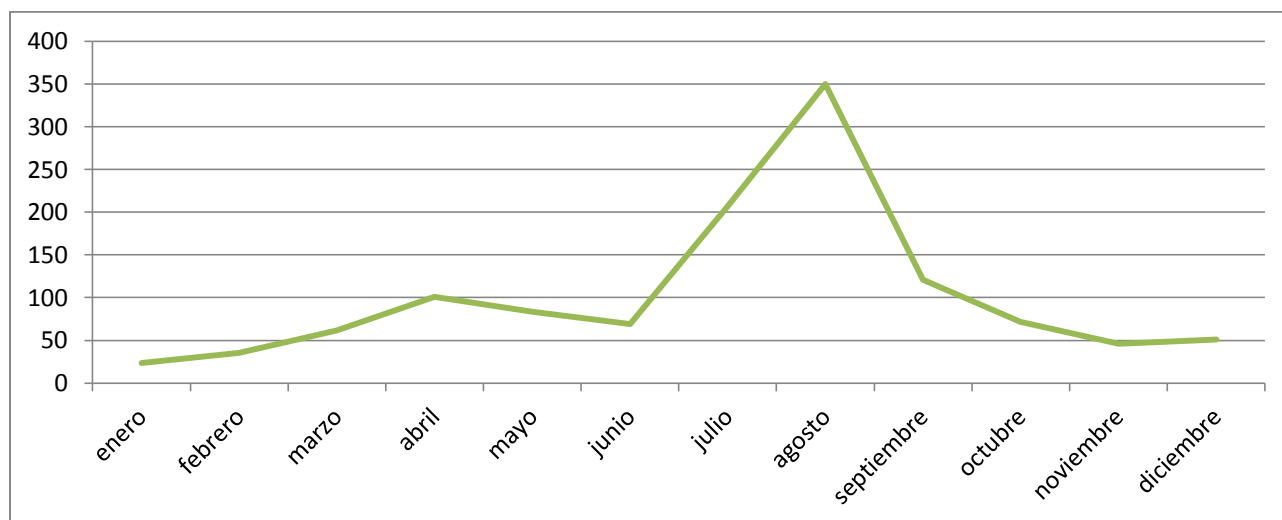


Figura 53.- Distribución de pernoctaciones en provincia 1 por meses.

Como se puede ver (Gráfico 3), en Cantabria el grueso principal de pernoctaciones se concentra en los meses de verano, principalmente en agosto. También se observan pico relevantes en primavera, principalmente en abril y a finales de año coincidiendo con las fechas navideñas (diciembre).

Para conocer debidamente el perfil del tipo de visitante de los establecimientos rurales de esta provincia se presenta el siguiente gráfico (Gráfico 4) que indica quién los acompaña en su estancia. Los datos hacen referencia a la persona a nombre de la cual realiza la reserva (EscapadaRural, 2015).

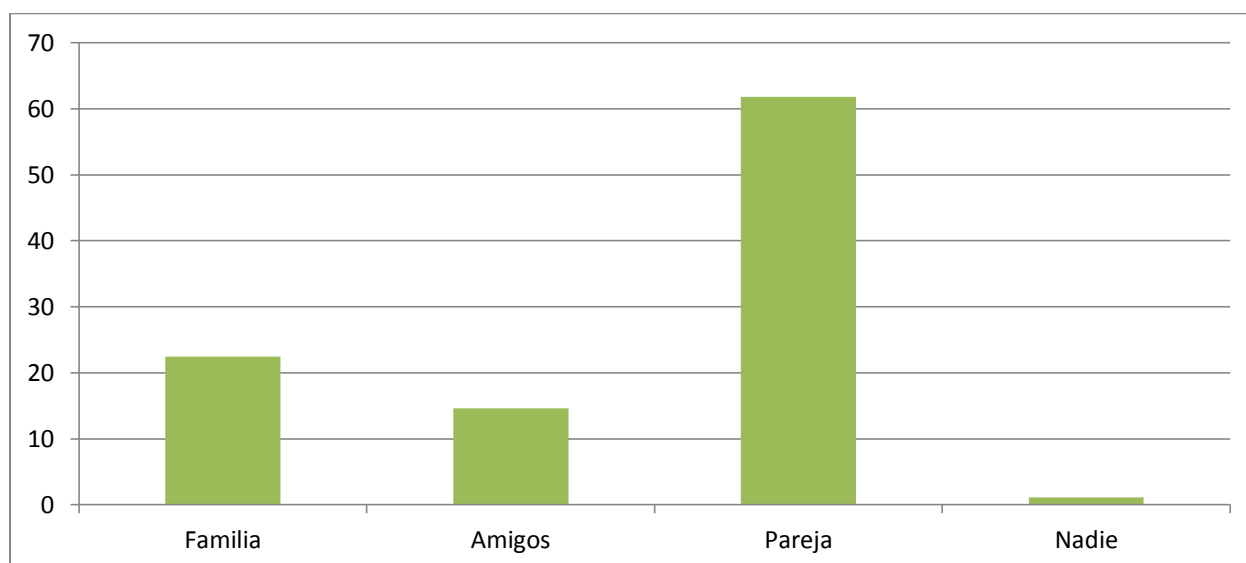


Figura 54.- Acompañantes provincia 1.

Como se puede ver la mayor parte de los usuarios se hacen acompañar de sus parejas para disfrutar de sus estancias rurales. Ello lleva relación con los datos que se desprenden del siguiente gráfico donde se pueden observar las actividades que llevan a cabo los huéspedes de dichos establecimientos.

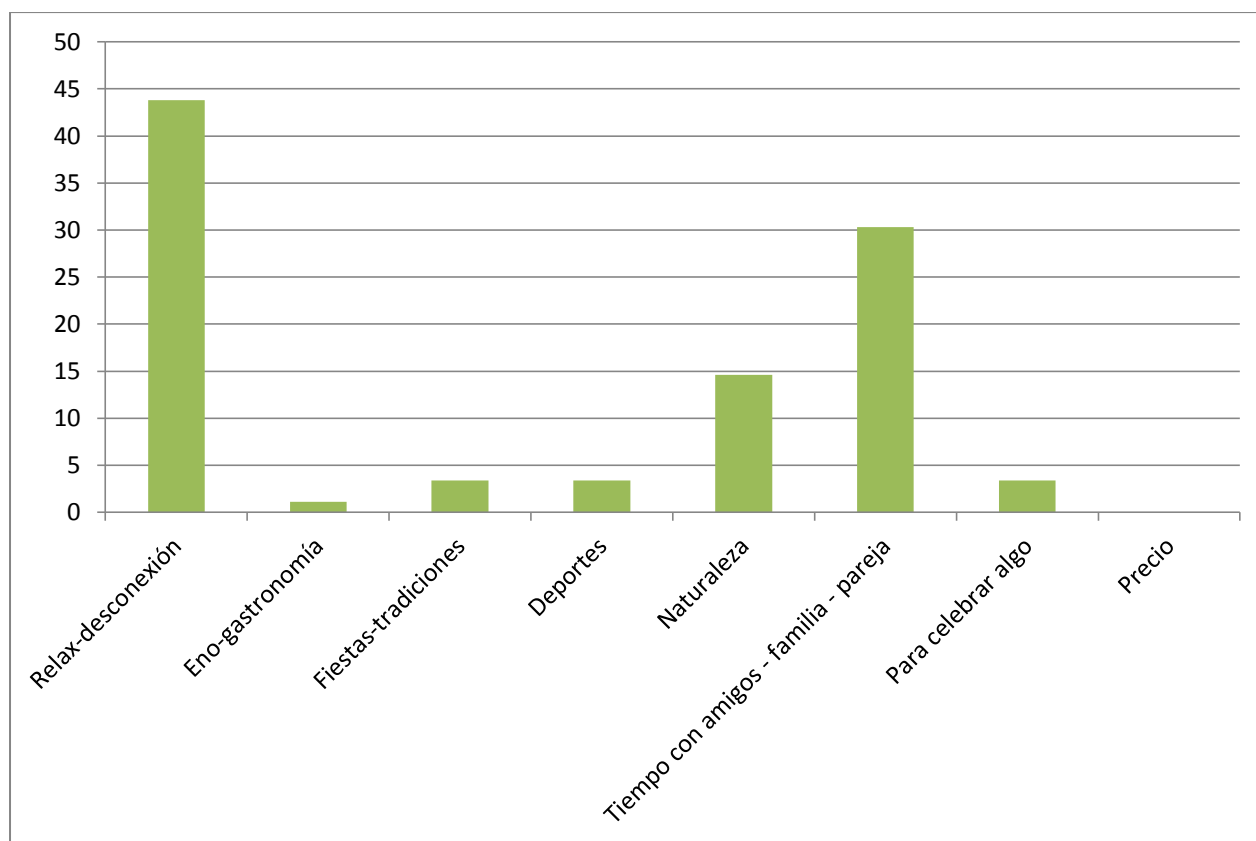


Figura 55.- Actividades realizadas provincia 1.

Como se puede ver en el gráfico anterior (Gráfico 5), gran parte de los usuarios tiene el objetivo de pasar tiempo con sus amigos, familiares o pareja, sin embargo la mayor parte de ellos tiene por objetivo descansar. Existe también un importante porcentaje de visitantes que buscan estar en contacto con la naturaleza mientras que las opciones de Eno-gastronomía, Fiestas o tradiciones, Deportes y Celebraciones tienen pocos adeptos en esta provincia. También se aprecia que los visitantes no se preocupan por el precio del alojamiento.

Por tanto, se puede colegir que el grueso de pernoctaciones en la provincia de Cantabria es realizada por huéspedes que, en su mayoría acompañados por su pareja, buscan relajarse y desconectar de su vida ordinaria concentrándose principalmente en agosto.

Con todos los datos recabados, la oferta que se debería promocionar sería algo semejante a la que aparece en la figura 51.



Figura 56.- Reclamo provincia 1.



### 5.9.- Provincia Número 2. Madrid

Atendiendo a los datos obtenidos de la serie temporal se puede ver las épocas del año donde se concentran la mayor parte de las pernoctaciones en la provincia de Madrid.

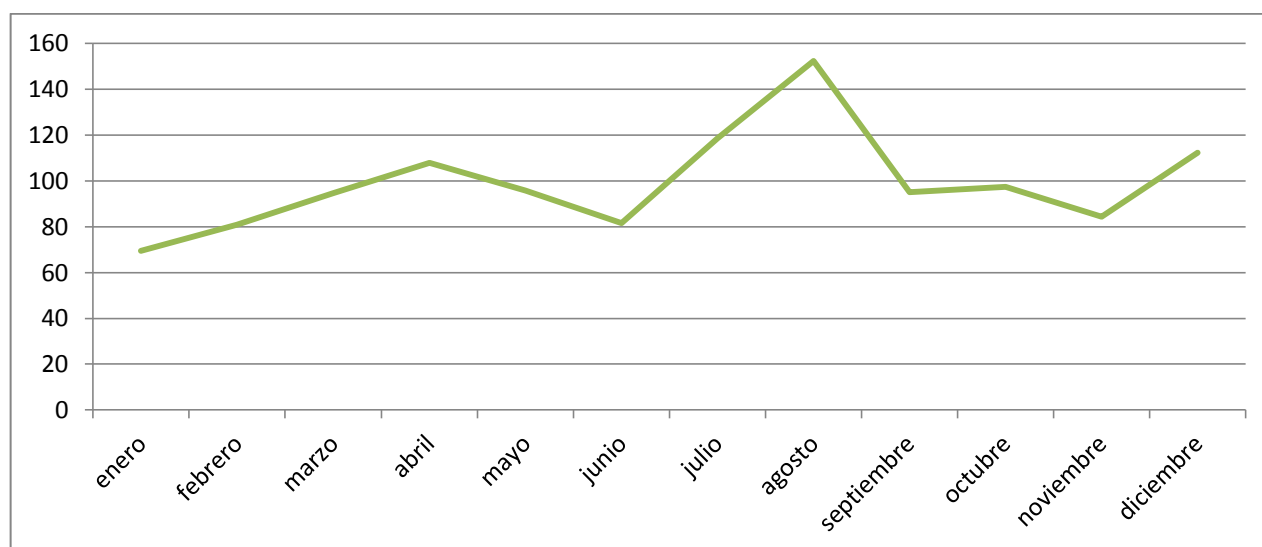


Figura 57.- Distribución de pernoctaciones en provincia 2 por meses.

Como se puede ver en el gráfico 6, en Madrid el grueso principal de pernoctaciones se concentra en los meses de verano, principalmente en agosto. También se observan pico relevantes en primavera, principalmente en abril y a finales de año coincidiendo con las fechas navideñas (diciembre). También se puede apreciar un ligero repunte de las pernoctaciones entorno al mes de octubre.

Para conocer debidamente el perfil del tipo de visitante de los establecimientos rurales en la provincia de Madrid se presenta el gráfico 7 que indica quién los acompaña en su estancia. Los datos hacen referencia a la persona a nombre de la cual realiza la reserva (EscapadaRural, 2015).

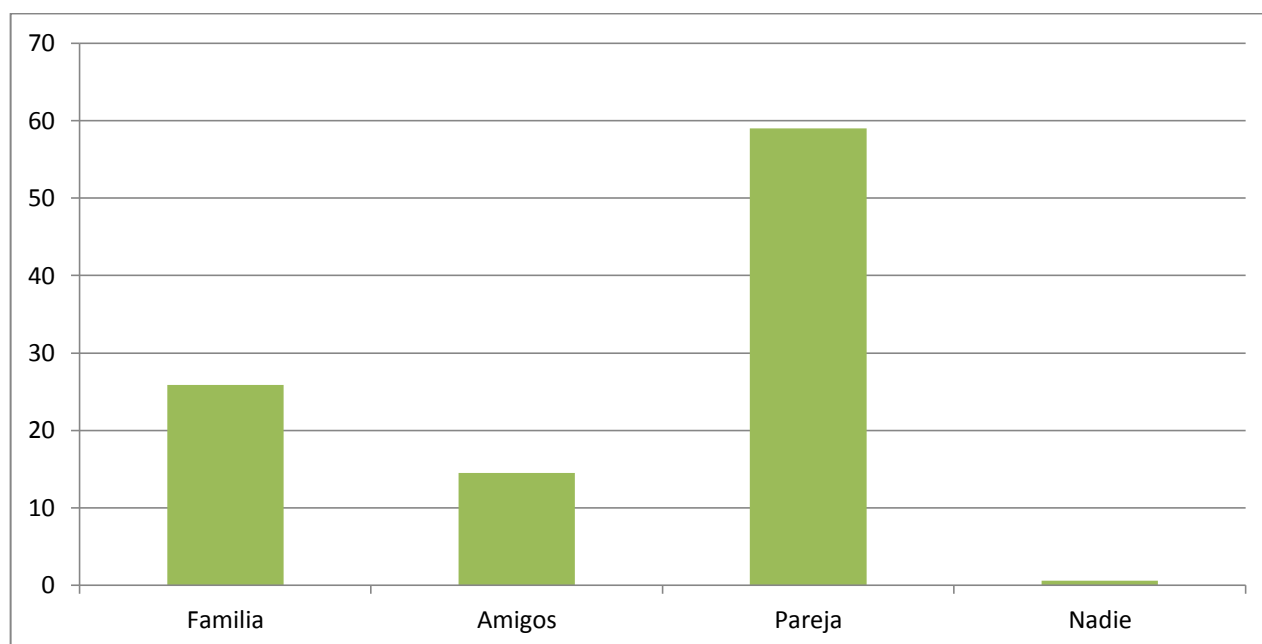


Figura 58.- Acompañantes provincia 2.

Como se puede ver la mayor parte de los usuarios se hacen acompañar de sus parejas para disfrutar de sus estancias rurales primando por encima de las escapadas familiares o con amigos y muy poca gente las realiza sin compañía.

Estos datos se refuerzan con el gráfico 8 que se muestra a continuación donde figuran las principales actividades que llevan a cabo los huéspedes de estos establecimientos en Madrid.

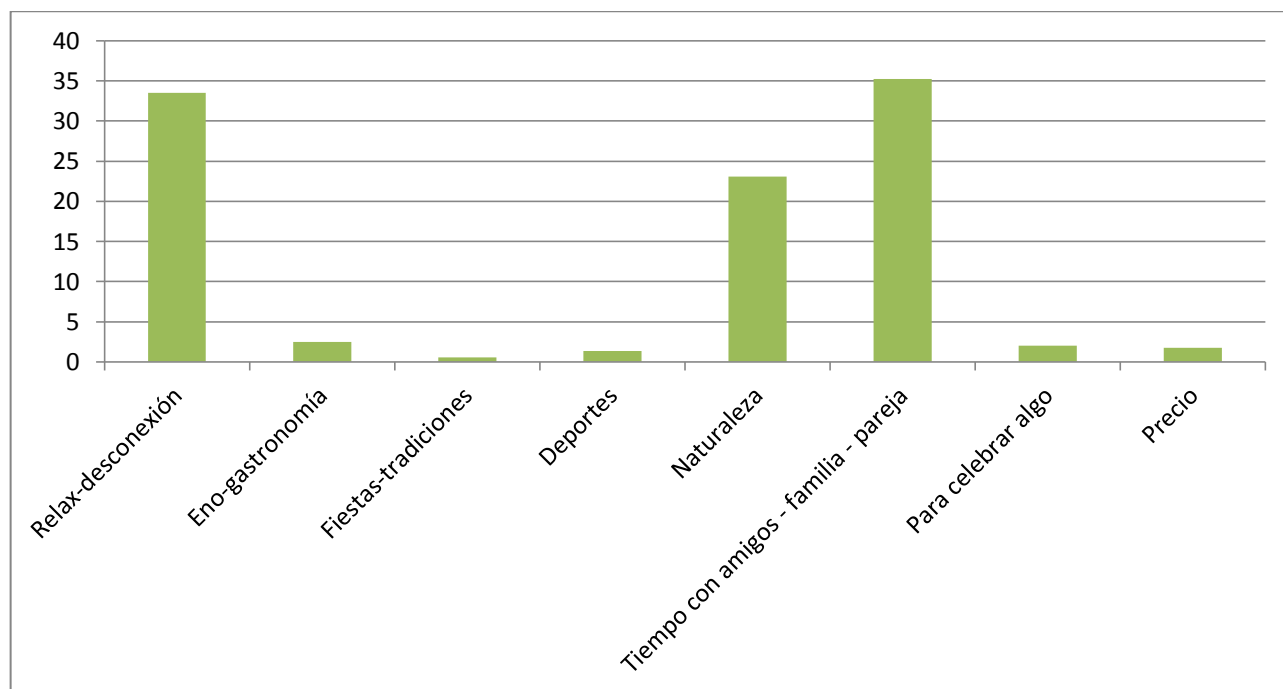


Figura 59.- Actividades provincia 2.

Como se puede ver en el gráfico 8, la mayoría de los visitantes de establecimientos rurales en la provincia de Madrid buscan pasar tiempo con sus seres queridos (pareja, amigos o familia) seguido muy de cerca por la búsqueda del relax y la desconexión teniendo también importancia la búsqueda del contacto con la naturaleza. Pocos de los visitantes buscan experiencia eno-gastronómicas o practicar deporte al igual que para celebraciones personales. Del mismo modo pocos parecen preocuparse por el precio y muy pocos tienen interés en fiestas o tradiciones.

Por tanto, se puede colegir que el grueso de pernoctaciones en la provincia de Madrid es realizado por huéspedes que principalmente buscan pasar tiempo con sus parejas encontrando su mayor concentración el mes de agosto.

Con todos los datos recabados, la oferta que se debería promocionar sería algo semejante a lo que aparece en la figura 52.



Figura 60.- Reclamo provincia 2.

### 5.10.- Provincia número 3. Baleares

Atendiendo a los datos obtenidos de la serie temporal se puede ver las épocas del año donde se concentran la mayor parte de las pernoctaciones en las Islas Baleares.

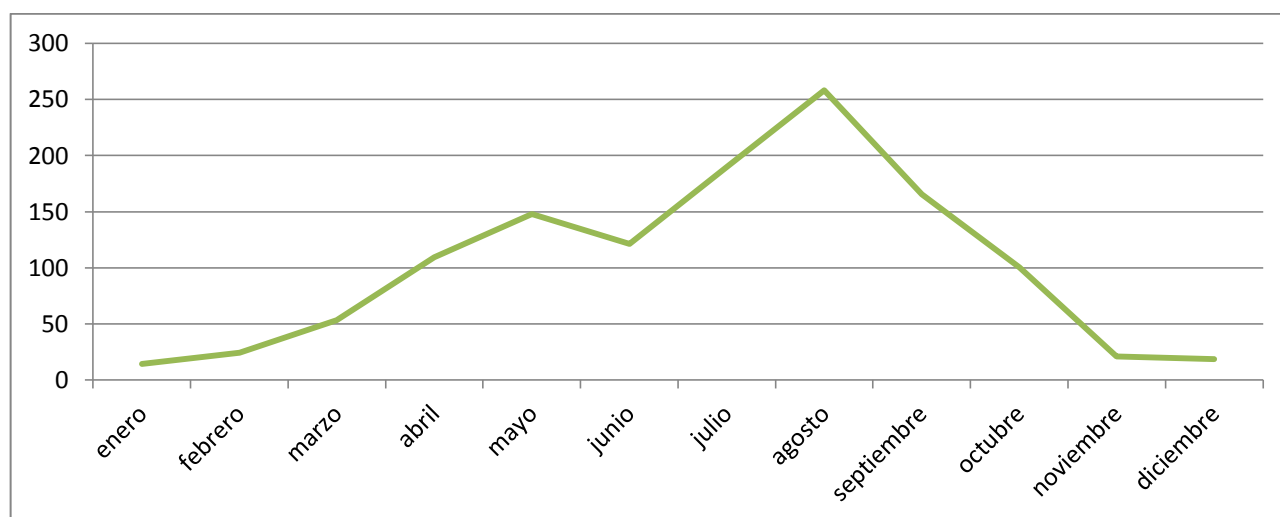


Figura 61.- Distribución de pernoctaciones en provincia 3 por meses

Tal como se refleja en el gráfico anterior (Gráfico 53), el turismo rural en las Islas Baleares se concentra en los meses de verano, especialmente agosto, aunque se aprecia un repunte en primavera (mes de mayo).

Para conocer debidamente el perfil del tipo de visitante de los establecimientos rurales en las Islas Baleares se presenta el gráfico 9 que indica quién los acompaña en su estancia. Los datos hacen referencia a la persona a nombre de la cual realiza la reserva (EscapadaRural, 2015).

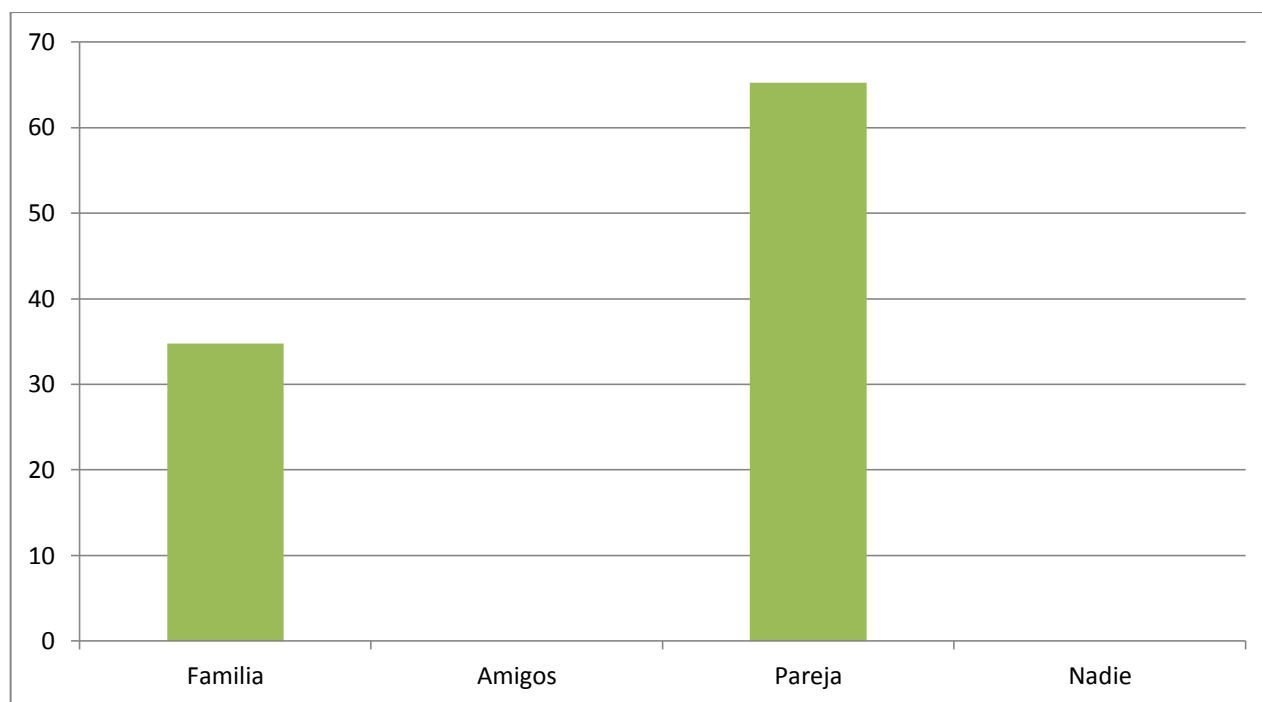


Figura 62.- Acompañantes provincia 3.

Al igual que en los anteriores casos, la mayor parte de los visitantes de alojamientos rurales en las Islas Baleares se hacen acompañar de su pareja o, en menor medida, de su familia. No se refleja que en ningún caso se acompañen de amigos u otras opciones.

Para comprender correctamente el perfil del visitante de estas islas se presenta a continuación el gráfico 10 donde se reflejan las principales actividades que llevan a cabo durante su estancia.

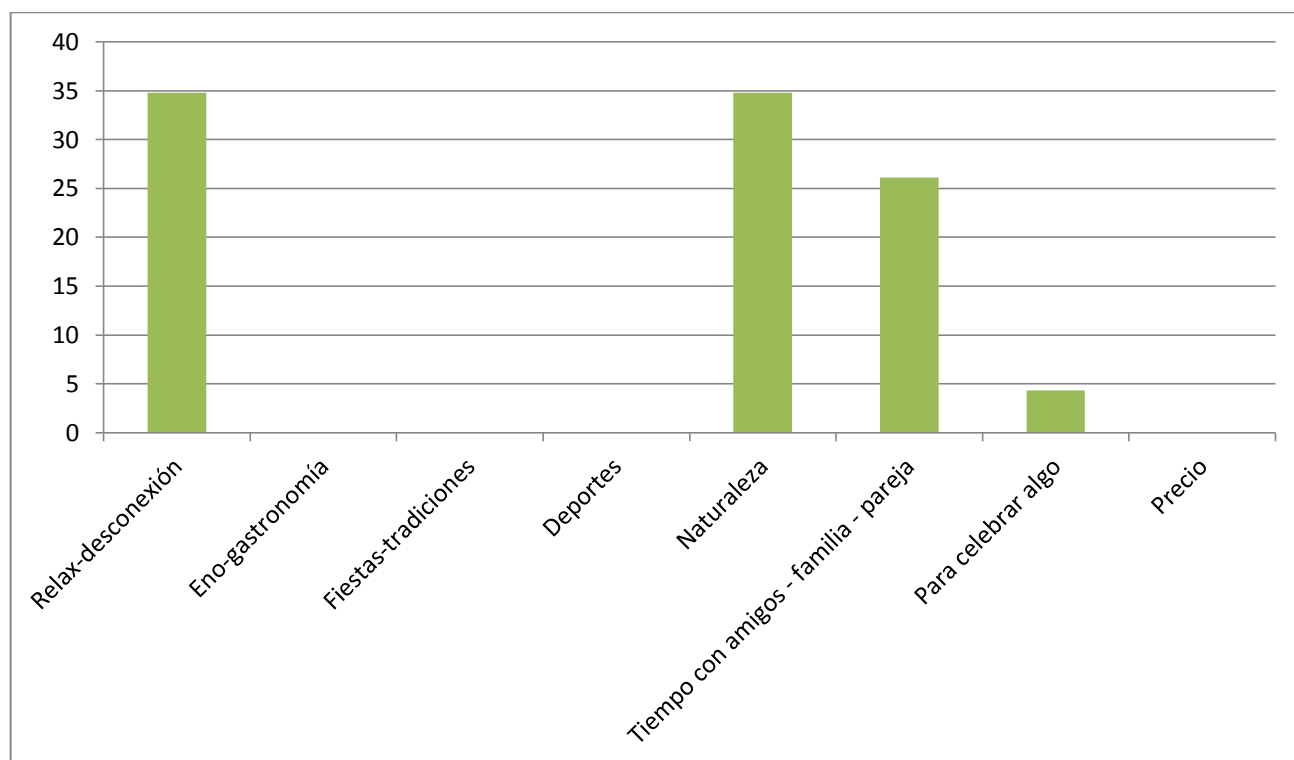


Figura 63.- Actividades provincia 3.

En el gráfico 10 se aprecia que lo que más buscan los huéspedes de los establecimientos rurales en las Islas Baleares es el contacto con la naturaleza junto con el relax y la desconexión. En menor medida, también hay bastantes personas que desean pasar tiempo con sus seres queridos y por último quienes pretenden realizar alguna celebración personal. Sin embargo no parece haber quienes desean vivir experiencias eno-gastronómicas o deportivas o conocer fiestas o tradiciones y en ningún caso parecen preocuparse por el precio.

Por tanto, se puede colegir que el grueso de pernoctaciones en las Islas Baleares son realizadas por huéspedes que, en su mayoría acompañados por su pareja, buscando tomar contacto con la naturaleza y desconectar de su vida ordinaria visitando los establecimientos principalmente en agosto.

Con todos los datos recabados, la oferta que se debería promocionar sería algo semejante a la figura 54.





Figura 64.- Reclamo provincia 3.

### 5.11.- Provincia Número 4. Cáceres

Atendiendo a los datos obtenidos de la serie temporal se puede ver las épocas del año donde se concentran la mayor parte de las pernoctaciones en la provincia de Cáceres.

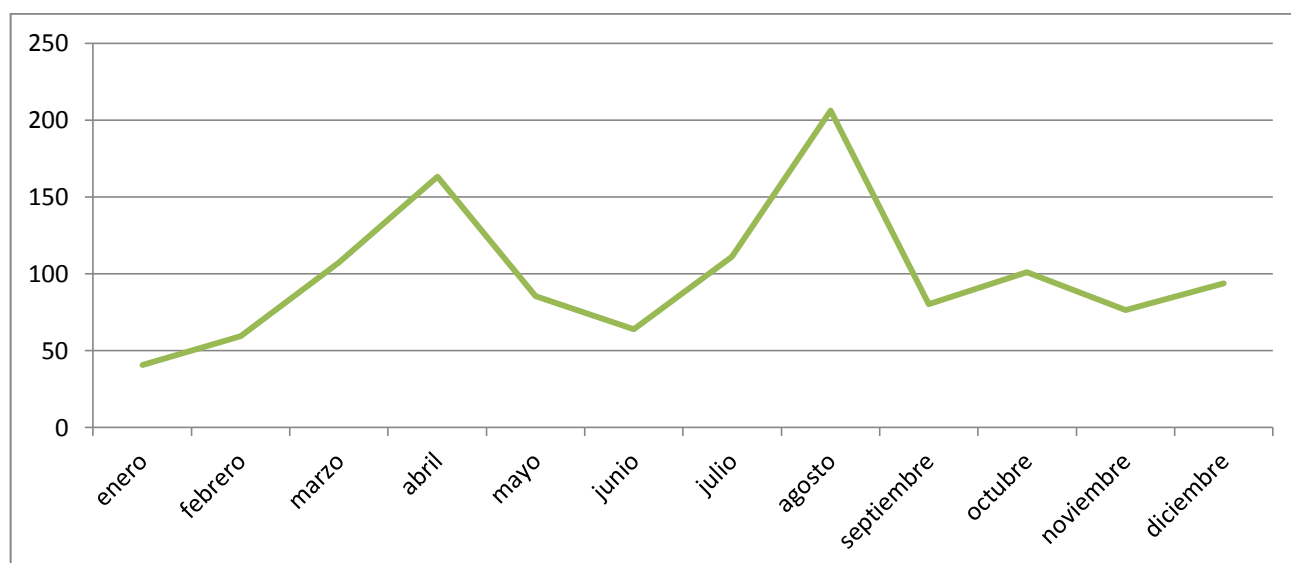


Figura 65.- Distribución de pernoctaciones en provincia 4 por meses

En la figura 55 se aprecia que la mayor parte de las pernoctaciones en la provincia de Cáceres se registran en los meses de verano, principalmente Agosto, seguido bastante de cerca por meses de primavera, especialmente abril, coincidiendo con la época del florecimiento del cerezo, lo que aporta verosimilitud al estudio. En menor medida se ve un repunte en el mes de octubre.

Para conocer debidamente el perfil del tipo de visitante de los establecimientos rurales en Cáceres se presenta el gráfico 11 que indica quién los acompaña en su estancia. Los datos hacen referencia a la persona a nombre de la cual realiza la reserva (EscapadaRural, 2015).

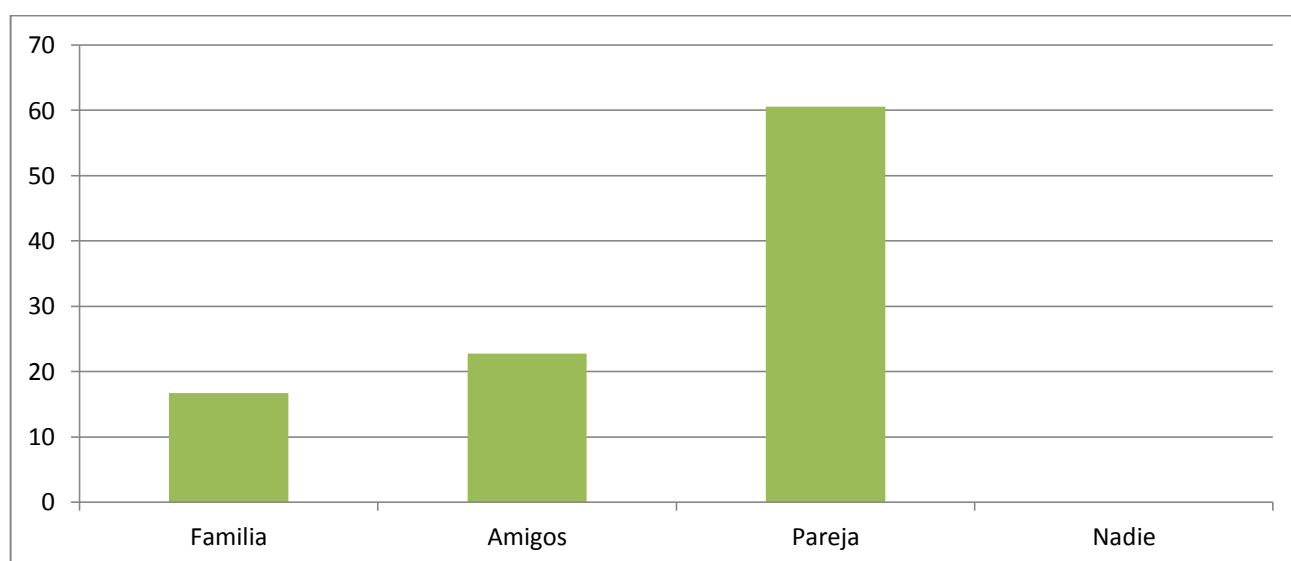


Figura 66.- Acompañantes provincia 4.

Se observa en esta representación que la mayor parte de los huéspedes de establecimientos rurales de Cáceres se hacen acompañar de sus parejas y en menor medida de amigos o familiares.

Para comprender correctamente el perfil del visitante de esta provincia se presenta a continuación el gráfico 12 donde se reflejan las principales actividades que llevan a cabo durante su estancia.

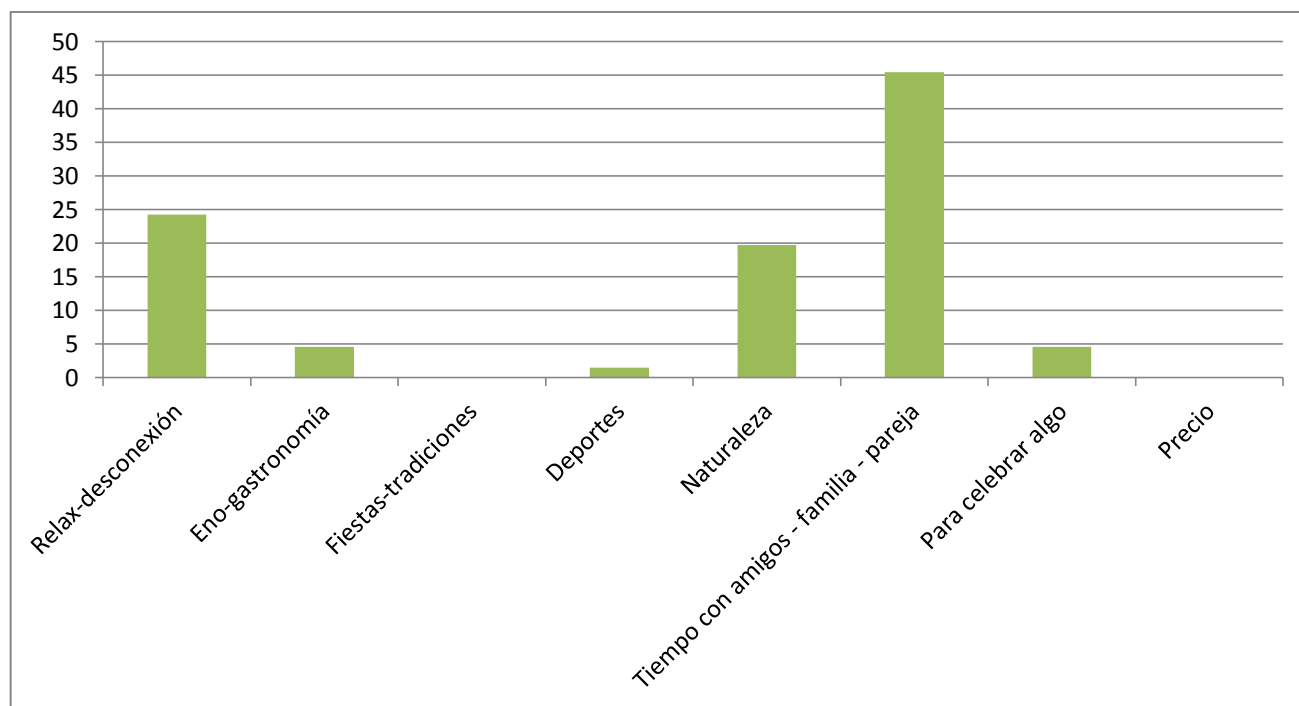


Figura 67.- Actividades realizadas provincia 4.

Con los datos que arroja este gráfico 12, se puede apreciar que la mayor parte de los usuarios de establecimientos rurales en la provincia de Cáceres persiguen pasar tiempo con sus seres queridos y en menor medida estar en contacto con la naturaleza o exclusivamente el relax. Menos son aquéllos que quieren vivir experiencias gastronómicas o festejar celebraciones personales y muy pocos los que buscan experiencias deportivas. No parece haber interés en realizar turismo rural sólo para conocer fiestas y tradiciones y del mismo modo nadie parece preocuparse por el precio.

Por tanto, a modo de síntesis, la mayor parte del turismo rural en Cáceres se concentra en agosto, aunque seguido muy de cerca por el mes de abril, y principalmente lo practican parejas que buscan pasar tiempo juntos.

Con todos los datos recabados, la oferta que se debería promocionar sería algo semejante a la figura 56.





Figura 68.- Reclamo provincia 4.

### 5.12 Valoración del estudio realizado

Tal y como se ha podido comprobar con los datos arrojados por el estudio realizado, el perfil del visitante en las cuatro provincias es similar:

- La mayor parte de los visitantes se hacen acompañar de sus parejas.
- En su mayoría, buscan como actividad principal el descanso y el relax, o quizás pasar tiempo con sus familiares.
- La mayor concentración de pernoctaciones se concentra en torno al verano especialmente en el mes de agosto.

Viendo la similitud de estos resultados no sería descabellado afirmar que el perfil de visitantes de cada provincia podría extrapolarse al global de la península. Ello significaría que existe una uniformidad en la visión que los visitantes tienen del turismo rural, el deseo de desconectar y disfrutar de la intimidad y la compañía de los seres queridos, especialmente de la pareja. Este resultado se corresponde con una visión del entorno rural de tranquilidad, calma y, en definitiva de asilamiento del estrés propio de las grandes ciudades.

Estos datos dan pie a dos posibles lecturas:

En primer lugar se podría explotar la imagen de “mar de tranquilidad” que ofrecen los entornos rurales en contraste al bullicio y estrés propio de los entornos urbanos. De esta forma se tendría un producto maduro y estable a partir del cual articular la oferta. Esto provoca que no haya una gran diferenciación entre la oferta de distintos establecimientos e incluso distintas provincias, lo que conlleva que sea muy complicado obtener diferenciación más allá del precio, lo que en el estudio se ha demostrado que no representa un factor de decisión para los visitantes.

Como segunda lectura, se observa que la mayor parte de las pernoctaciones se lleva a cabo en agosto y, en algunas provincias como es el ejemplo de Cáceres, en determinadas épocas del año en las que se producen algunos eventos de interés, como son el cerezo en flor, pero en pocos casos las festividades tradicionales son un factor de decisión. Este resultado es lógico pues al estar hablando de establecimientos que se relacionan con localidades pequeñas, es complicado que las festividades propias de las poblaciones tengan una fuerza suficiente para representar un evento que atraiga una cantidad de público considerable. Esto no significa que se deba desistir en la promoción de las festividades propias, sino que sería recomendable establecer eventos y festividades en las que se agrupe a un número de poblaciones suficiente como para resultar atractivo a los visitantes y de esta forma distribuir y aumentar el número de pernoctaciones a lo largo del año. Una vez más, el ejemplo sería el cerezo en flor de la provincia de Cáceres cuyo volumen de visitantes rivaliza con el del mes de agosto, si bien este es un evento natural, no una iniciativa ciudadana.

Por tanto, en resumen, el turismo rural es llevado a cabo mayoritariamente por parejas que buscan relajarse y desconectar en un entorno de tranquilidad y calma.

Finalmente, como propuesta para aumentar el peso del turismo rural en el global de turismo nacional, sería interesante organizar eventos que engloben a varias poblaciones de forma que unidas sumen una fuerza suficiente para atraer la atención de potenciales visitantes.



### 5.13.- Comparativa de las Herramientas utilizadas.

Aplicaciones	Manejo	Conocimientos Necesarios		Información disponible en Internet
		Tipo	Grado	
Desarrollo de aplicación a medida (FBI)	Muy complejo	Programación	Alto	Mucha
		Estadística	Alto	
		Necesidades de la Empresa	Alto	
Weka	Sencillo	Estadística	Medio	Mucha
Excel	Asequible	Ofimática	Alto	Mucha
		Estadística	Muy Alto	
		Conocimiento de la aplicación	Alto	
SPSS	Complejo	Estadística	Alto	Suficiente
		Conocimiento de la Aplicación	Muy Alto	
R programming	Asequible	Programación	Medio	Mucha
		Estadística	Medio	
Tableau	Asequible	Conocimiento de la aplicación	Medio	Insuficiente
SAP Lumira	Asequible	Conocimiento de la aplicación	Medio	Mucha
PSPP	Muy complejo	Estadística	Alto	Insuficiente
		Conocimiento de la Aplicación	Muy Alto	

Tabla 18.- Resumen del estudio realizado con las aplicaciones.

Colabora con otras aplicaciones	Aspectos positivos	Aspectos Negativos
Según se defina la herramienta	-Aplicación ajustada a las necesidades concretas de la empresa.	-Dificultad para llevarlo a cabo en la misma organización. -Cualquier cambio de necesidades exigirá desarrollo. -Difícilmente transferible a otra organización.
Sí	-Sencillo y rápido. -Software libre.	-Permite poco margen de maniobra.
sí	-Interfaz sencillo. -Permite trabajar en muy diversos campos. -Lee datos desde muchas fuentes distintas. -Existe mucha documentación en Internet.	-Tareas concretas como la predicción exigen un gran conocimiento tanto del campo de estudio como de la herramienta.
No	-Facilidad para cubrir cualquier campo de la estadística. -Simplifica cálculos complejos.	-Se necesita conocer específicamente la aplicación. -Software de pago.
Sí	-Gran repositorio de módulos. -Fácil uso. -Motor de otras aplicaciones. Libre	-No es una aplicación realmente sino un lenguaje de programación.
Sí	-Interfaz sencillo y amigable. -Gran potencia de cálculo. -Simplifica tareas complejas.	-Software de pago
Sí	-Interfaz sencillo y amigable. -Gran potencia de cálculo. -Simplifica tareas complejas.	-Software de pago
No	-Software Libre. -Válido para cálculo sencillos.	-Potencia de cálculo insuficiente. -Imposible realizar predicciones con él.



## 6.- Debate. Business Intelligence, justificación de la inversión.

A lo largo de este trabajo se ha hablado largo y tendido de los beneficios y posibilidades que presenta el paradigma del BI y cómo las aplicaciones que lo explotan permiten a las organizaciones y negocios que las utilizan obtener ventaja competitiva actuando de soporte para la toma decisiones.

Sin embargo, en este punto es interesante recordar la gráfica que se mostró en el apartado 1.4 de este documento.

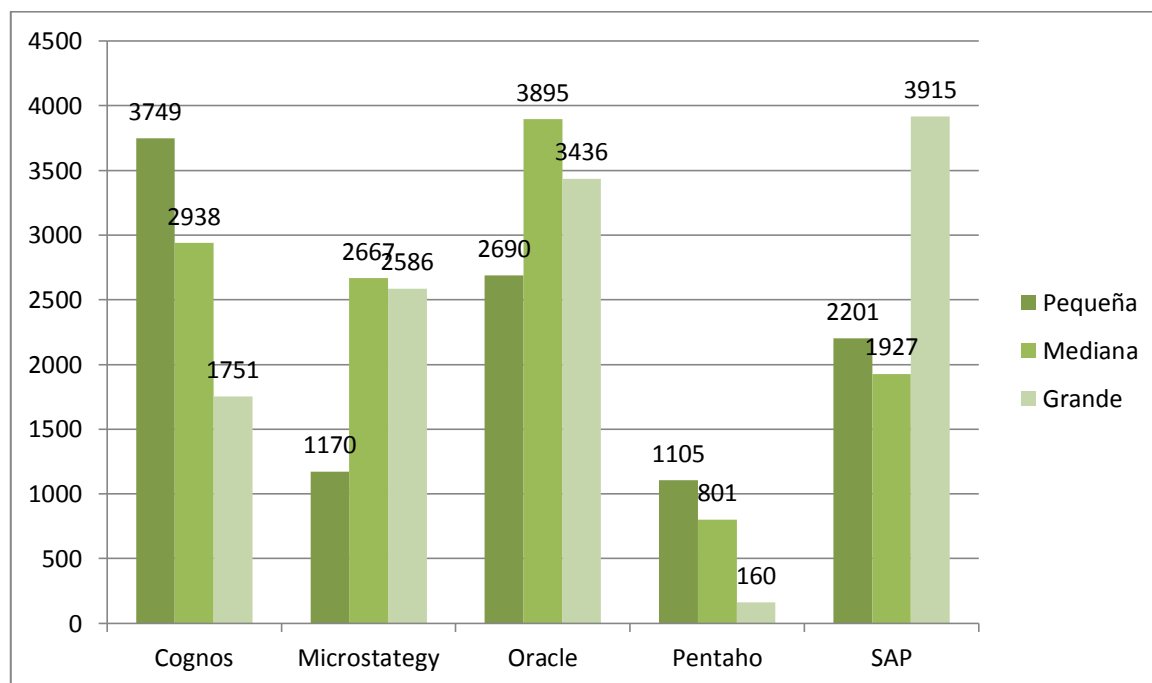


Figura 69.- Gráfica del coste de suites BI para empresas pequeñas, medianas y grandes.

En ella se expone el coste que representan las grandes suites comerciales para las pequeñas, medianas y grandes empresas por puesto. Como se puede ver, suponen un coste nada despreciable especialmente para empresas de pequeño tamaño para las que suelen quedar muy lejos de sus capacidades de adquisición.

Sobre ello es lo que trata este debate, sobre si los beneficios que reporta BI justifican sus costes. Es necesario recordar que en esta gráfica únicamente se ha presentado el coste económico, pues el coste de aprendizaje es difícilmente medible, sin embargo la existencia de cursos e incluso masters completos dedicados al aprendizaje y utilización de estas herramientas da una pista de que su manejo dista mucho de ser trivial.

Sin embargo de esta parte negativa ya se ha hablado y es la que en gran medida ha motivado este trabajo, la obtención de soluciones BI a menor coste, económico al menos. Es momento de hablar de su parte positiva.

El principal motivo que fomenta la inversión en BI es obtener un soporte confiable y válido a la hora de tomar decisiones. El valor o la corrección de una decisión tomada decrece con el tiempo que transcurre entre que ocurre un evento y esta decisión se ejecuta. Matemáticamente, se podría expresar esta afirmación como que el valor de una decisión es inversamente proporcional al tiempo que tarda en tomarse, como se expresa en el gráfico de más abajo (Gráfico 58). Con esto se pretende expresar que si una organización o



negocio no actúa con rapidez pierde ventaja respecto a sus competidores y eso, si se está hablando del mundo empresarial puede ser fatal. Esto es lo que en el primer capítulo se llamó “Latencia del Negocio”.

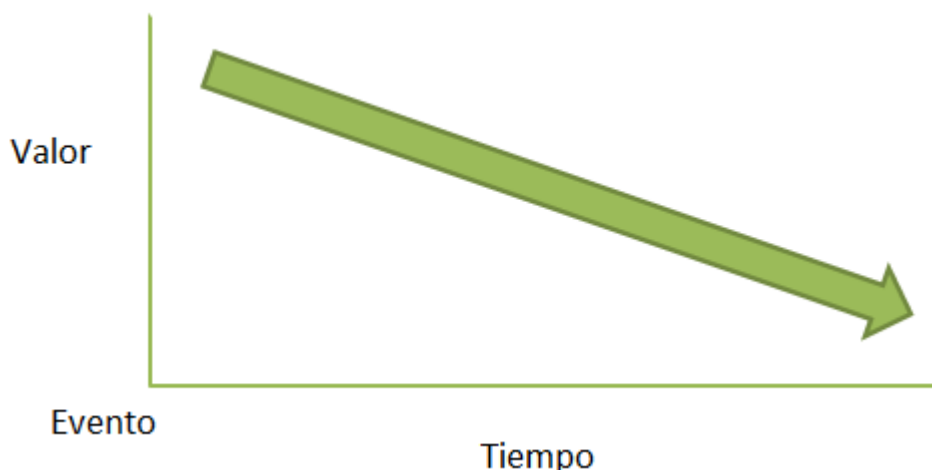


Figura 70.- Valor vs Tiempo

Las aplicaciones BI por su parte añaden una nueva variable a esta ecuación, la información. De esta forma, la toma de decisiones es más sencilla y sólida si está fundamentada sobre datos que aporten criterios para elegir el camino correcto a seguir. Matemáticamente, el valor de una decisión es directamente proporcional a la información de la que se dispone para tomarla e inversamente proporcional al tiempo que se tarda en tomar. De esta forma, el gráfico cambia para convertirse en el siguiente (Figura 59):



Figura 71.- Valor = Información / Tiempo

Los sistemas de BI no sólo deben proporcionar información necesaria para la toma de decisiones sino que esa información debe estar disponible en el menor tiempo posible para que dichas decisiones puedan ser realmente útiles.

Por tanto, como conclusión, el coste (económico y humano) que supone la adopción de un sistema BI sólo es justificable si posibilita al usuario disponer de la información precisa en el momento oportuno para que la toma de decisiones sea la correcta y permita, en último término, la obtención de ventaja sobre los competidores.



## 7.- Conclusions and Future Aims

### 7.1.- Job Conclusions.

At the beginning of this project a number of issues that motivated and justified the realization of it were raised. Throughout this document have been exposed knowledge, opinions, data, and procedures and has made a practical case for the sole purpose of responding to those questions. Once done this entire project, is possible answer them.

#### **Are there BI tools outside large commercial suites?**

Clearly there are these solutions and in this document has described some of them. An attempt was made to cover the widest possible spectrum showing, free tools, owners, general purposes, specific, self-generated ... Naturally there are many other tools that can allow cover needs here exposed. In any case, the answer to this question is yes.

#### **Is it possible obtaining effective BI solutions at an affordable cost?**

Tools presented here have been selected to present those looking for affordable and bearable costs for organizations and small scale business. Many of them are free software tools, which in itself implies economic ease, while others require payment of a license in any case cheaper than the price of a license to a commercial BI suite as shown in the following example:

One year Tableau professional license<sup>6</sup> – 362 €

Cognos professional License<sup>7</sup> – 3468 €

Therefore, these solutions even having to pay a license, present costs much more accessible than those of larger BI suites.

#### **What costs have such applications?**

Has already discussed the issue of economic cost very affordable compared to that of large BI suites. But must emphasize that these suites, despite not having a particularly easy operation (the vendors themselves offer customized courses to buyers) are designed to greatly facilitate the work when making BI processes. Therefore, the main cost that can present these tools is eminently learning.

Some of the solutions presented, such as Excel or SPSS require a deep knowledge of these applications and statistical, especially in the case of Excel. Other in turn may require knowledge of development, such as R, which however by having its repository of functions is assumable, while developing an own tool requires thorough knowledge of development that can be not found in the organization and could be necessary seek them out.

Therefore, the lower cost of these solutions has its counterpoint on the need to have deep learning costs.

---

<sup>6</sup> Tableau Desktop - Professional - First Year Maintenance - 12 month period (Per User) (Triad, 20105)

<sup>7</sup> Cognos TM1 Enterprise User (AU) - Neulizenz (Cursor Software, 2015)

### **Really these applications can offer BI solutions?**

To carry out this project, was chosen that applications used in the case of practical application to develop the main feature of BI, estimation of possible future behavior. This has been replicated successfully, and has obtained models representing possible future behavior of evolution in overnight stays by guests in rural accommodation over a year.

For these reasons, it is possible to obtain BI solutions through tools away from the main commercial solutions with valid results.

### **Which is better?**

This question is difficult to answer. Certainly applications allow achieve model to predict future behavior, but all of them have certain gaps or weaknesses that make that there is none that stands out from the rest. Tableau is the most complete application, but is chargeable, WEKA is simple and free but allows little room for work, Excel offers processing power and extra work allowing great leeway; however obtaining predictive modeling is complicated and is easy to make mistakes. R programming meanwhile is a free application and with optimum results however is essentially a programming language that makes many users could have misgivings for use it.

Thus choose a particular tool is not easy, and probably not appropriate. On the other hand, collaboration among several of these tools does seem right. Combining several of them can fill gaps presented. As an example would be possible to obtain an analysis of a time serie with WEKA or R and work and reporting with Excel. This way of working is common even among large commercial solutions, because by definition are suites, i.e. set of programs which in itself implies collaboration. It is also very frequent, almost a rule, which the suites of BI work with Excel.

Therefore, rather than choosing a tool, is recommended choose those that give better solution to the specific needs of the organization and combine them to achieve the objective pursued by it.

### **Benefits justify the investment required by BI tools?**

Actually this question was answered in the previous chapter (6. Discussion) and how was said at the time, investing in a BI tool, both economically and learning, is only justifiable if it enables the user to have the information accurately and in a timely manner to enable it to make appropriate decisions and to achieve advantage over competitors.

In this way all the questions raised at the beginning of this work have been answered. As a final thought: the study process carried out in this document is not sufficient to justify a business investment base, however, anyone thinking about a similar investment should at least conduct a process similar to this study.

### 7.2.- Future Aims.

To end this work, the possible lines of investigation that could lend continuity to this work are shown then.

First, the study process can be continued by increasing the granularity of research, carrying out a study of the state of rural tourism by municipalities, in order to obtain more accurate conclusions presented.

As a second way, over time, it is possible to repeat this same study expecting new data produced at the time and check the progress that has rural tourism at national level and the effect of measures both government and private to promote it.

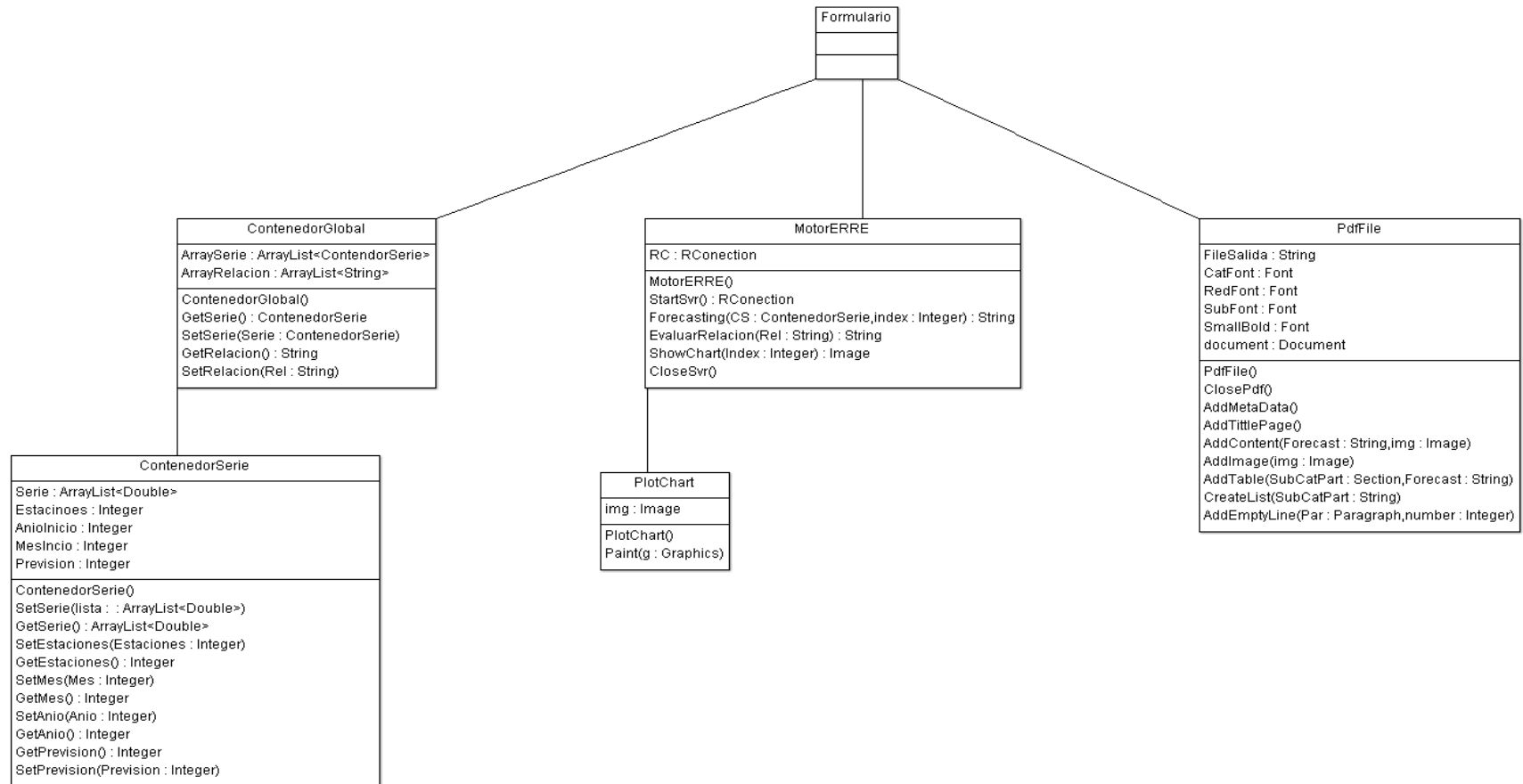
Finally, the study technique is fully transferable to other areas as long as there are sufficient data to carry out. Thus, techniques of time series analysis and data processing with the various applications described in this document allow any knowledge of the state and evolution of consumption, sales or economic indicators of any kind.



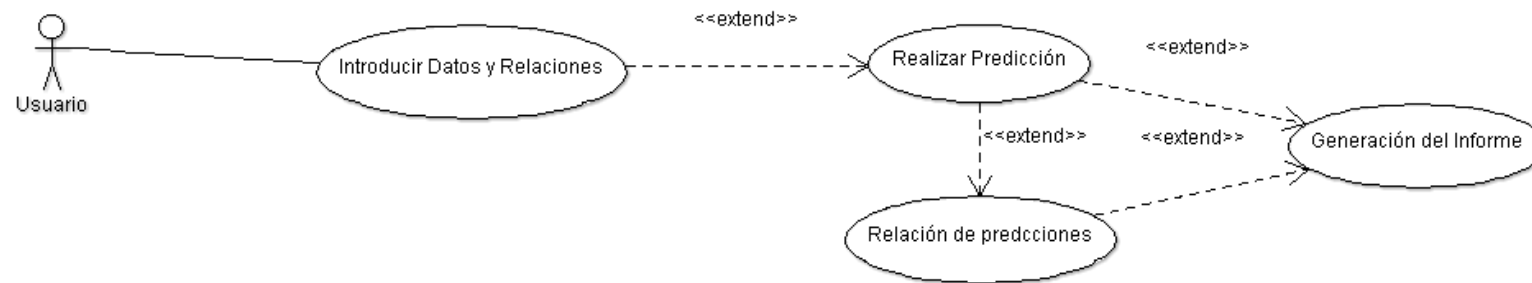
## Anexos

## Anexo I Diagramas de la aplicación Forecast BI.

### AI1.- Diagrama de Clases.



AI2.- Diagrama de casos de uso.



## A II. - Study of time series using software tools.

### AII1.- Study of time series using SPSS.

To carry on a time series study using SPSS is necessary make the following steps:

#### 1. - Introducing data into SPSS:

This point can be made into two ways:

- a) Manually
- b) Importing external files.

#### 2. - Establishing time periods which data are distributed:

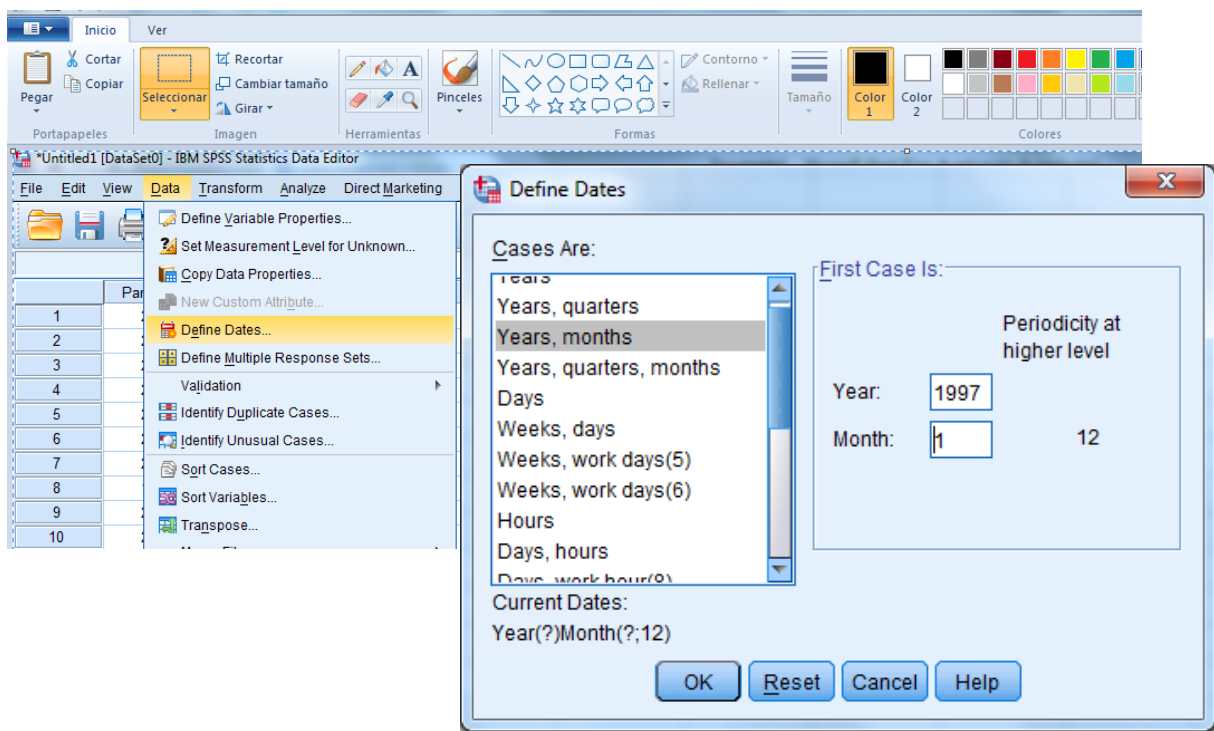


Figura 72.- SPSS Image 1

As has been said previously, is necessary insert time periods which have been samples obtained. There are many possibilities in which establish it, for example is possible select as main period an year and divide it into quarters, months, both them... even is possible establish a custom division. In the image above (Figure 72), has been selected an year (begging in 1997) divided into month (begging in January).

After that, three new columns are obtained, in this case year, month and full date:



	ParoEsp	YEAR_	MONTH_	DATE_
1	2256,50	1997	1	JAN 1997
2	2262,70	1997	2	FEB 1997
3	2227,50	1997	3	MAR 1997
4	2181,70	1997	4	APR 1997
5	2123,80	1997	5	MAY 1997
6	2091,80	1997	6	JUN 1997
7	2009,20	1997	7	JUL 1997

Figura 73.- SPSS image 2

### 3. - Generating sequence chart:

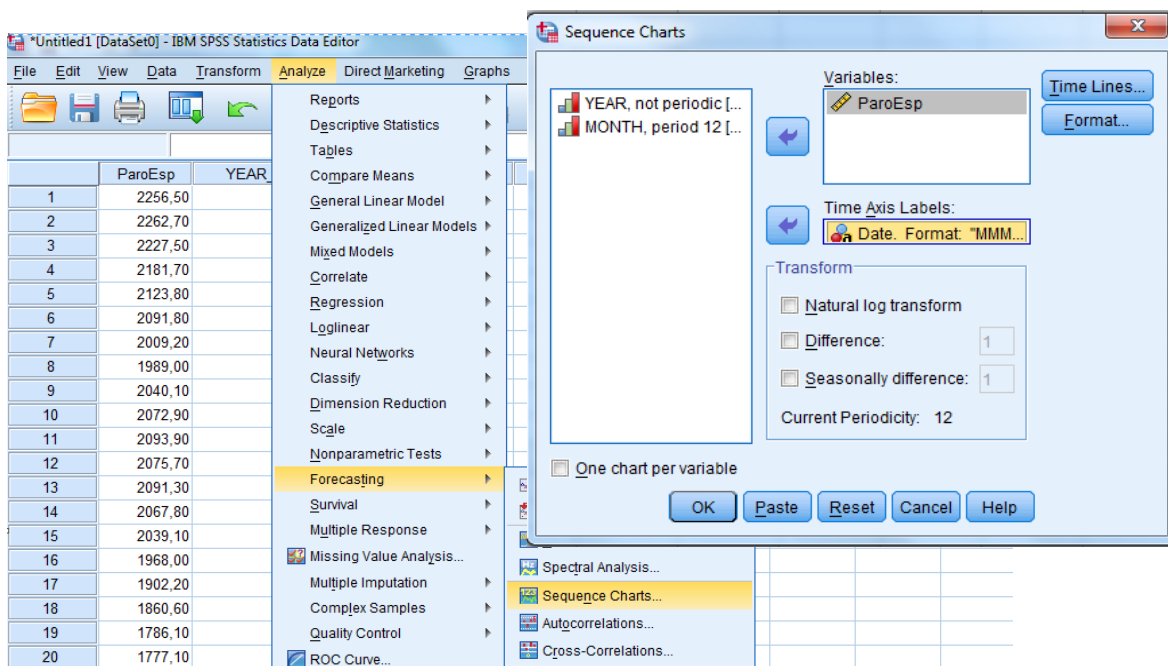


Figura 74.- SPSS image 3

To establish sequence chart is necessary select as variable column which will be analyzed, in example (Figure 74), this variable is date previously obtained.

Result is the following (Figure 75):

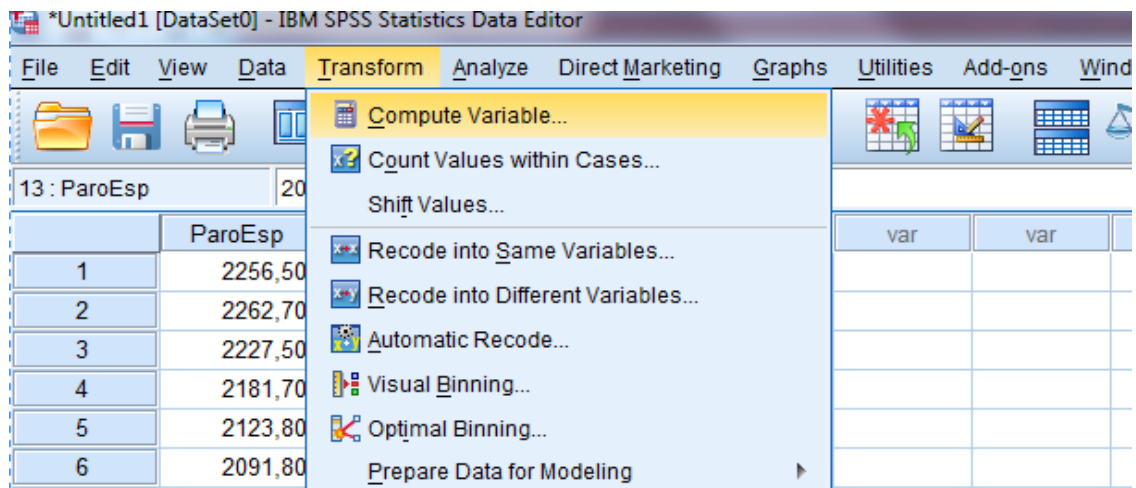


Figura 75.- SPSS image 4

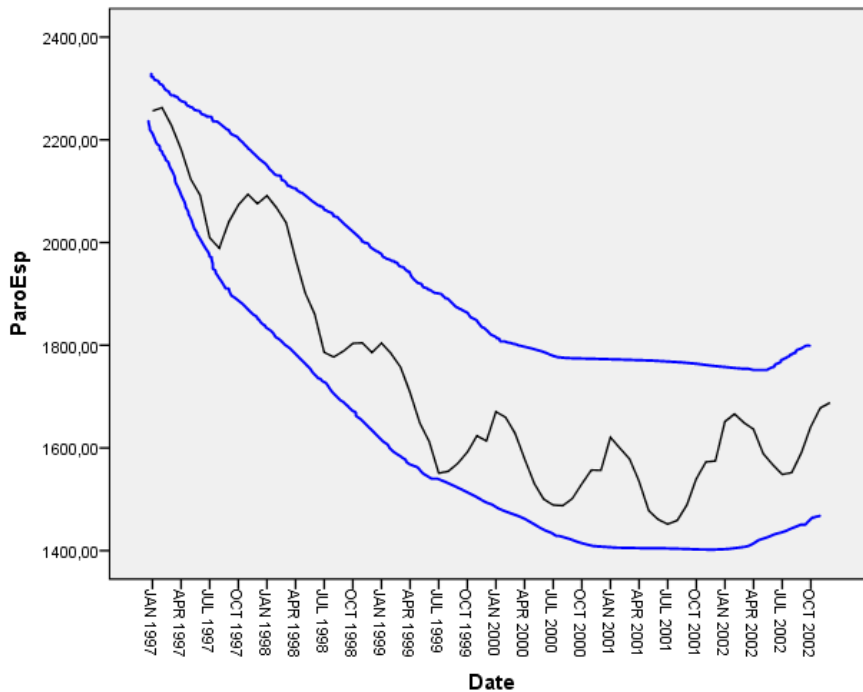


Figura 76.- SPSS image 5

In the over image (Figure 76), have been inserted two blue lines, in order to show trend observed into this simple. As it is observed, there is not a uniform trend, this vary over time. This indicates that this example has a multiplicative scheme. However, this assessment could be subjective. To stablish certainly which scheme has a sample is necessary use method variation coefficients of differences and ratios.

#### 4. - Determining scheme of sample.

Method variation coefficients is based on calculation of relationship (difference and ratios) among two samples into same station (month into this example) but belong to different periods (years into this example).

Thus, in the example, among 1997 January and 1998 January will be establish difference and ratio (always higher against fewer):

Difference: January 1998 y January 1997= 2091.3 - 2256.5 = -165.2

Ratio January 1998 y January 1997= 2091.3 / 2256.5 = 0.926

For this, two new variables (DIFFERENCE and RATIO) will be generated representing the series explained above<sup>8</sup>.

<sup>8</sup> For realize this calculations it is necessary move series until next period, for do this, it is necessary use lag function as follows LAG(PARoEsp,12) where PARoEsp is variable to study and 12 are station which it is divided.

## DIFFERENCE:

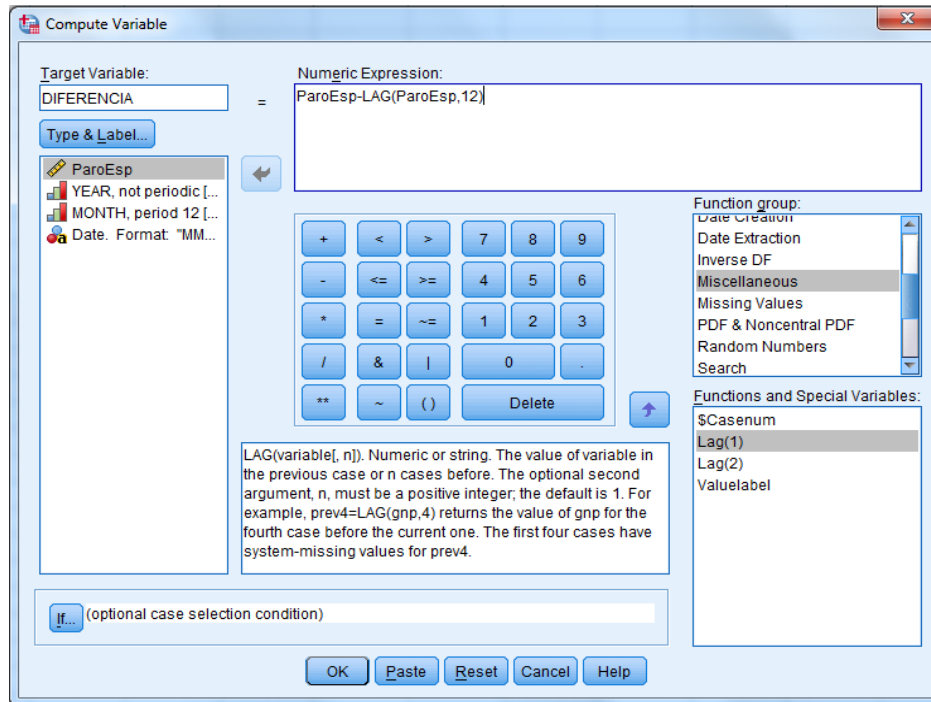


Figura 77.- SPSS Image 6

## RATIO:

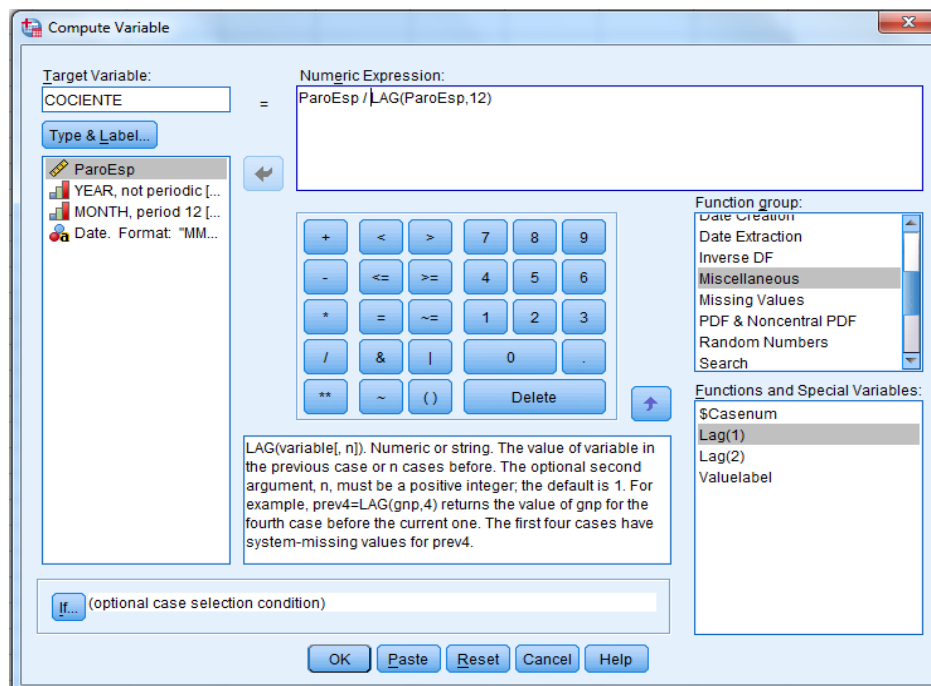


Figura 78.- SPSS Image 7

Result obtained will be two new variables with following values:

	ParoEsp	YEAR_	MONTH_	DATE_	DIFERENCIA	COCIENTE
1	2256,50	1997	1	JAN 1997		
2	2262,70	1997	2	FEB 1997		
3	2227,50	1997	3	MAR 1997		
4	2181,70	1997	4	APR 1997		
5	2123,80	1997	5	MAY 1997		
6	2091,80	1997	6	JUN 1997		
7	2009,20	1997	7	JUL 1997		
8	1989,00	1997	8	AUG 1997		
9	2040,10	1997	9	SEP 1997		
10	2072,90	1997	10	OCT 1997		
11	2093,90	1997	11	NOV 1997		
12	2075,70	1997	12	DEC 1997		
13	2091,30	1998	1	JAN 1998	-165,20	,93
14	2067,80	1998	2	FEB 1998	-194,90	,91
15	2039,10	1998	3	MAR 1998	-188,40	,92
16	1968,00	1998	4	APR 1998	-213,70	,90
17	1902,20	1998	5	MAY 1998	-221,60	,90
18	1860,60	1998	6	JUN 1998	-231,20	,89

Figura 79.- SPSS Image 8

As image above (Figure 79) shows, first period is empty, this is due to there is not a previous period.

Once obtained both series, to obtain variation coefficients, it is necessary calculate mean and standard deviation of both them. Highlighted button opens next dialog frame where is possible to establish it:

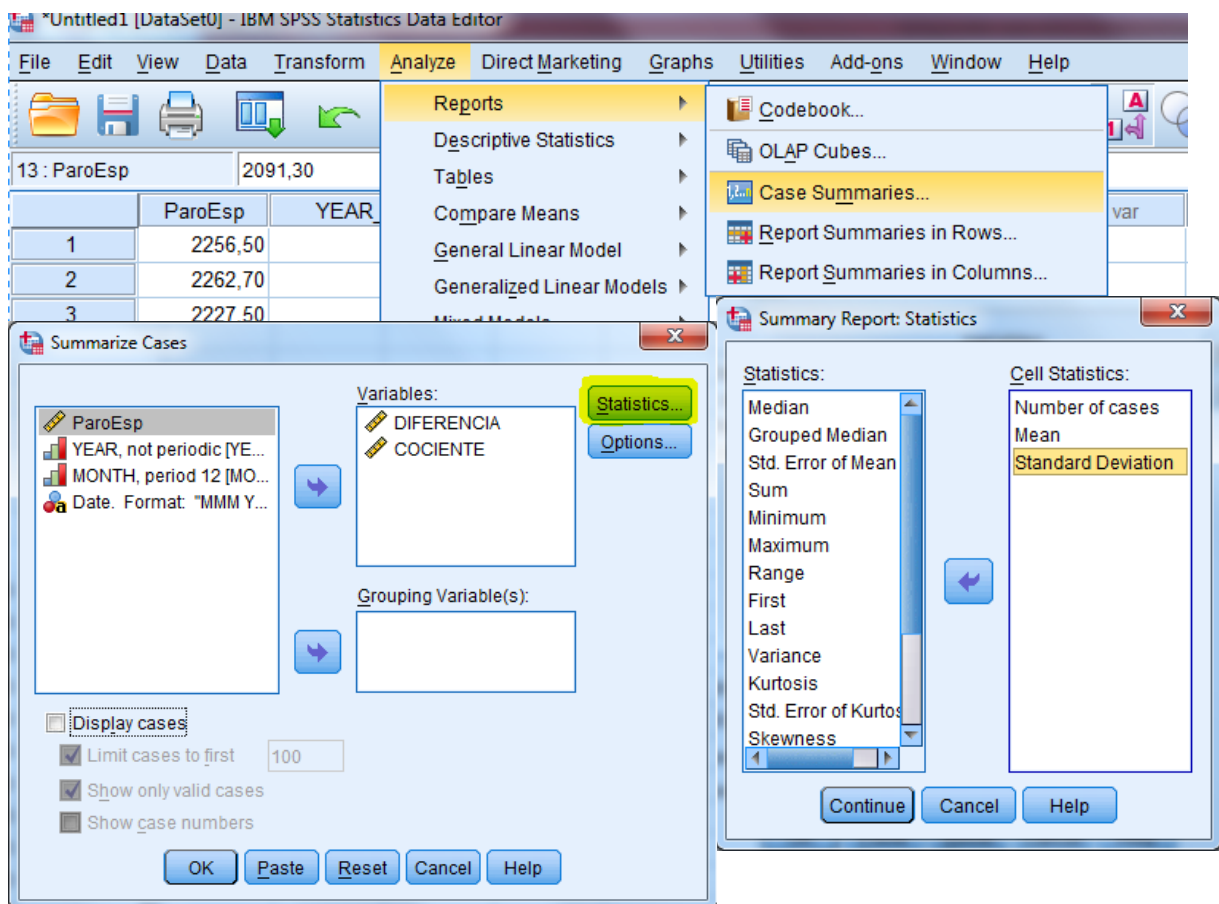


Figura 80.- SPSS Image 9.

This operation generate following results:

	DIFERENCIA	COCIENTE
N	60	60
Mean	-99,4483	,9505
Std. Deviation	129,67003	,06957

Thus, variation coefficients are:

$$CV_{\text{diferencia}} = 129.67 / -99.44 = -1.30$$

$$CV_{\text{cociente}} = 0.06957 / 0.9505 = 0.073$$

To compare both coefficients, is necessary use absolute value, without signs.

In this case  $CV_{\text{diferencia}}$  is higher than  $CV_{\text{cociente}}$  thus, this serie has a multiplicative scheme.

### 5. - Calculating centered moving average.

Thank to SPSS power, calculating centered moving avarge is extreemly simple:

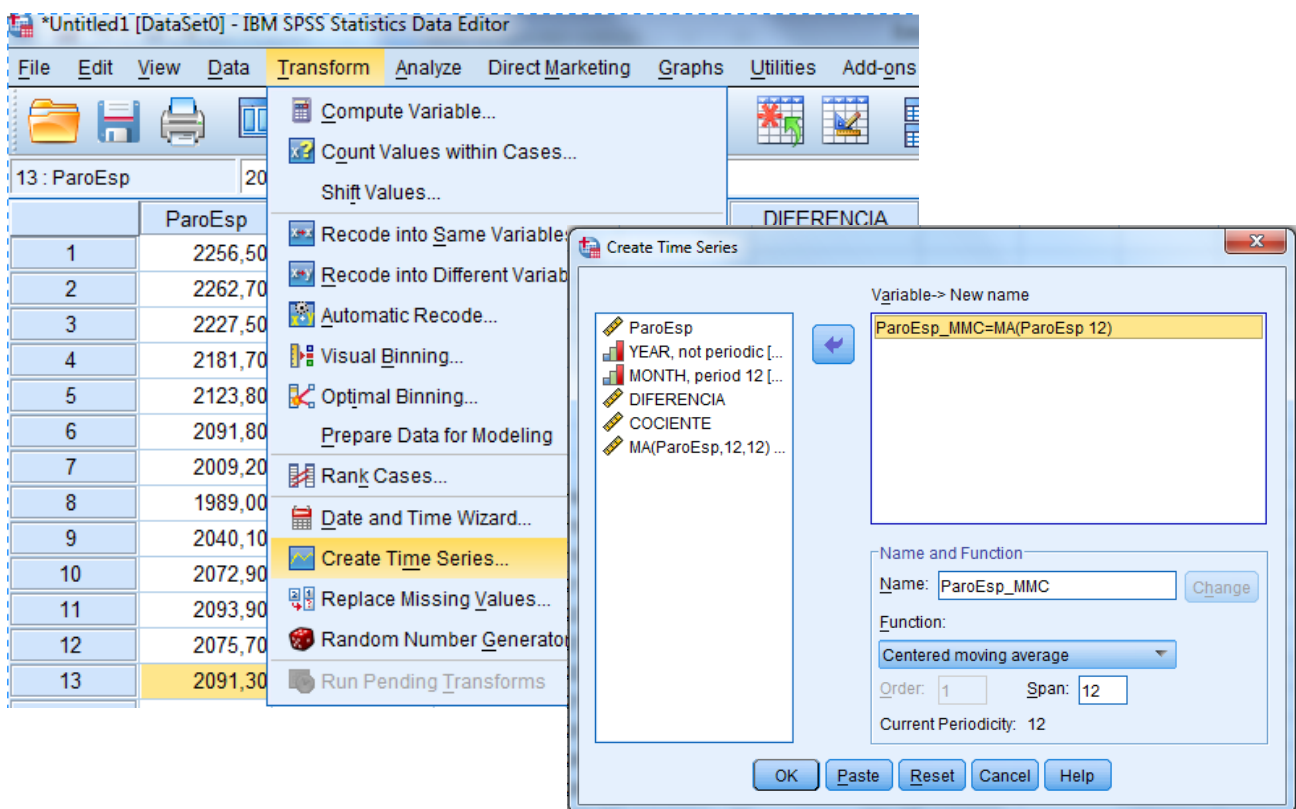


Figura 81.- SPSS Image 10

To obtain correctly the centered moving average, it is necessary select function *centered moving average* over selected variable (in this example ParoEsp) as Figure 81 shows, adjusting number of observed station by period (12 months in this example).

The result is the following, figure 82:

	ParoEsp	YEAR_	MONTH_	DATE_	DIFERENCIA	COCIENTE	ParoEsp_MMC
1	2256,50	1997	1	JAN 1997	-	-	-
2	2262,70	1997	2	FEB 1997	-	-	-
3	2227,50	1997	3	MAR 1997	-	-	-
4	2181,70	1997	4	APR 1997	-	-	-
5	2123,80	1997	5	MAY 1997	-	-	-
6	2091,80	1997	6	JUN 1997	-	-	-
7	2009,20	1997	7	JUL 1997	-	-	2111,85
8	1989,00	1997	8	AUG 1997	-	-	2096,85
9	2040,10	1997	9	SEP 1997	-	-	2080,88
10	2072,90	1997	10	OCT 1997	-	-	2064,12

64	1636,30	2002	4	APR 2002	101,20	1,07	1599,05
65	1589,00	2002	5	MAY 2002	110,90	1,08	1607,67
66	1567,40	2002	6	JUN 2002	106,80	1,07	1616,77
67	1548,40	2002	7	JUL 2002	96,90	1,07	-
68	1552,00	2002	8	AUG 2002	93,00	1,06	-
69	1590,30	2002	9	SEP 2002	101,70	1,07	-
70	1641,70	2002	10	OCT 2002	101,70	1,07	-
71	1678,00	2002	11	NOV 2002	105,20	1,07	-
72	1688,10	2002	12	DEC 2002	113,30	1,07	-

Figura 82.- SPSS Image 11

## 6. - Factorization.

Figure 83 shows dialogue frame, it is explained below.

Model type was obtained in step 4.

Time series ponderation will be selected according data quantity:

- Every points with as equal when number of periods is odd.
- Final points pondered as 0.5 when number of period is couple.

This process will add four new columns:

STC\_1: Trend-cycle (T x C) show trend an cycle of the series. Disjoin both them is really difficult so it is considered as just one component.

SAF\_1: Seasonal correction factors (E). Its interpretation depend if this seasonal decomposition is considered as an additive or multiplicative scheme:

Additive.- Zero value (0) indicates there is not seasonality in the concrete period. When this is positive, variable has values higher than average in this period, else, variable has values fewer than average in this period.

Multiplicative.- One value (1) indicates there is not seasonality in the concrete period. When this is positive, variable has values higher than average in this period, else, variable has values fewer than average in this period.

SAS\_1: Values of factorized series. This is obtained with ratio among series and seasonal factors (multiplicative scheme) or difference among series and seasonal factors (additive).

ERR\_1: Residual component or noise (A). These are values which remain after remove previous components.

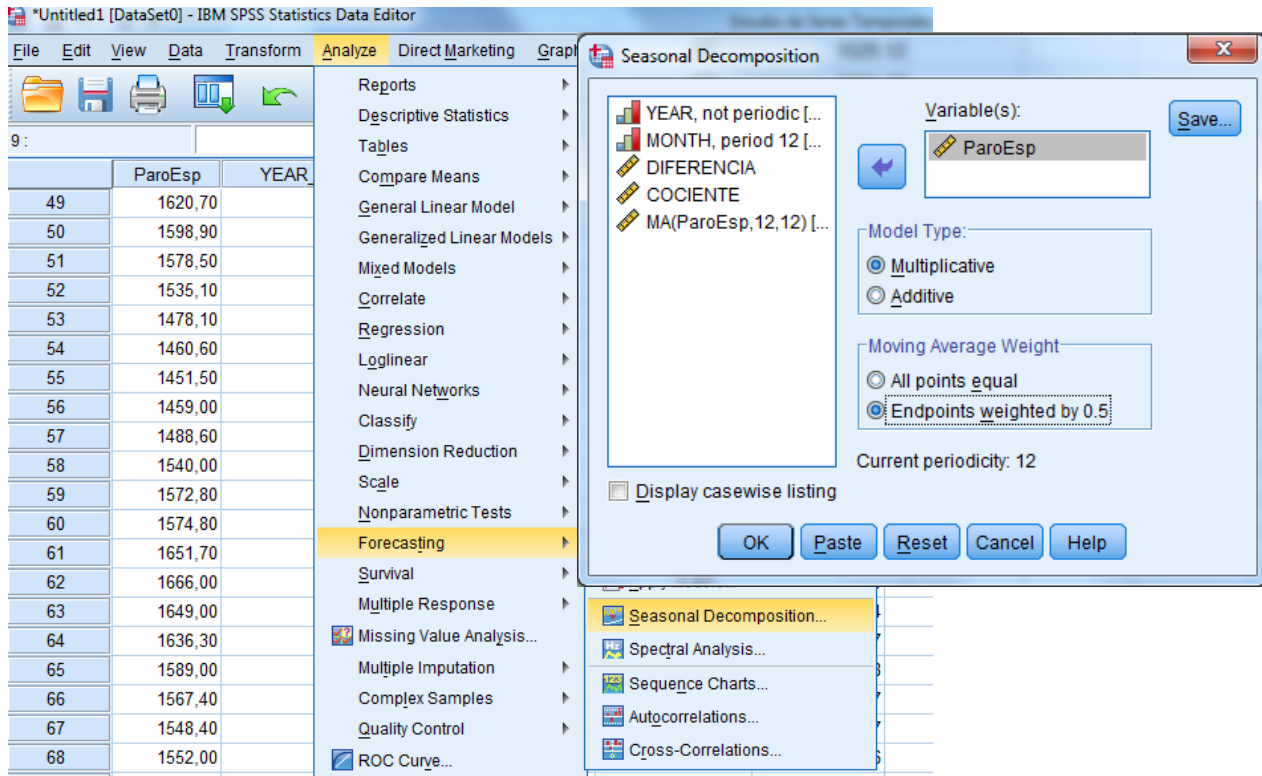


Figura 83.- SPSS image 12

The result is the following, figure 84:

	ParoEsp	YEAR_	MONTH_	DATE_	ERR_1	SAS_1	SAF_1	STC_1
1	2256,50	1997	1	JAN 1997	,99840	2150,31386	1,04938	2153,76114
2	2262,70	1997	2	FEB 1997	1,00355	2162,05957	1,04655	2154,40284
3	2227,50	1997	3	MAR 1997	,99775	2150,83508	1,03564	2155,68622
4	2181,70	1997	4	APR 1997	1,00127	2159,49805	1,01028	2156,75655
5	2123,80	1997	5	MAY 1997	1,00177	2155,24166	,98541	2151,43019
6	2091,80	1997	6	JUN 1997	1,00899	2158,10179	,96928	2138,87747
7	2009,20	1997	7	JUL 1997	,99642	2111,11200	,95173	2118,68873
8	1989,00	1997	8	AUG 1997	,99188	2083,38652	,95470	2100,44678
9	2040,10	1997	9	SEP 1997	1,00398	2096,64426	,97303	2088,32608
10	2072,90	1997	10	OCT 1997	1,00135	2080,24714	,99647	2077,44733
11	2093,90	1997	11	NOV 1997	1,00307	2066,62261	1,01320	2060,29142
12	2075,70	1997	12	DEC 1997	1,00563	2046,36427	1,01434	2034,90991
13	2091,30	1998	1	JAN 1998	,99321	1992,88782	1,04938	2006,51007
14	2067,80	1998	2	FEB 1998	,99647	1975,82834	1,04655	1982,82626

Figura 84.- SPSS image 13



## 7.- Prediction.

Once factorized series is obtained, is necessary establishing the regression curve in order to determine which type best suits, Figure 85:

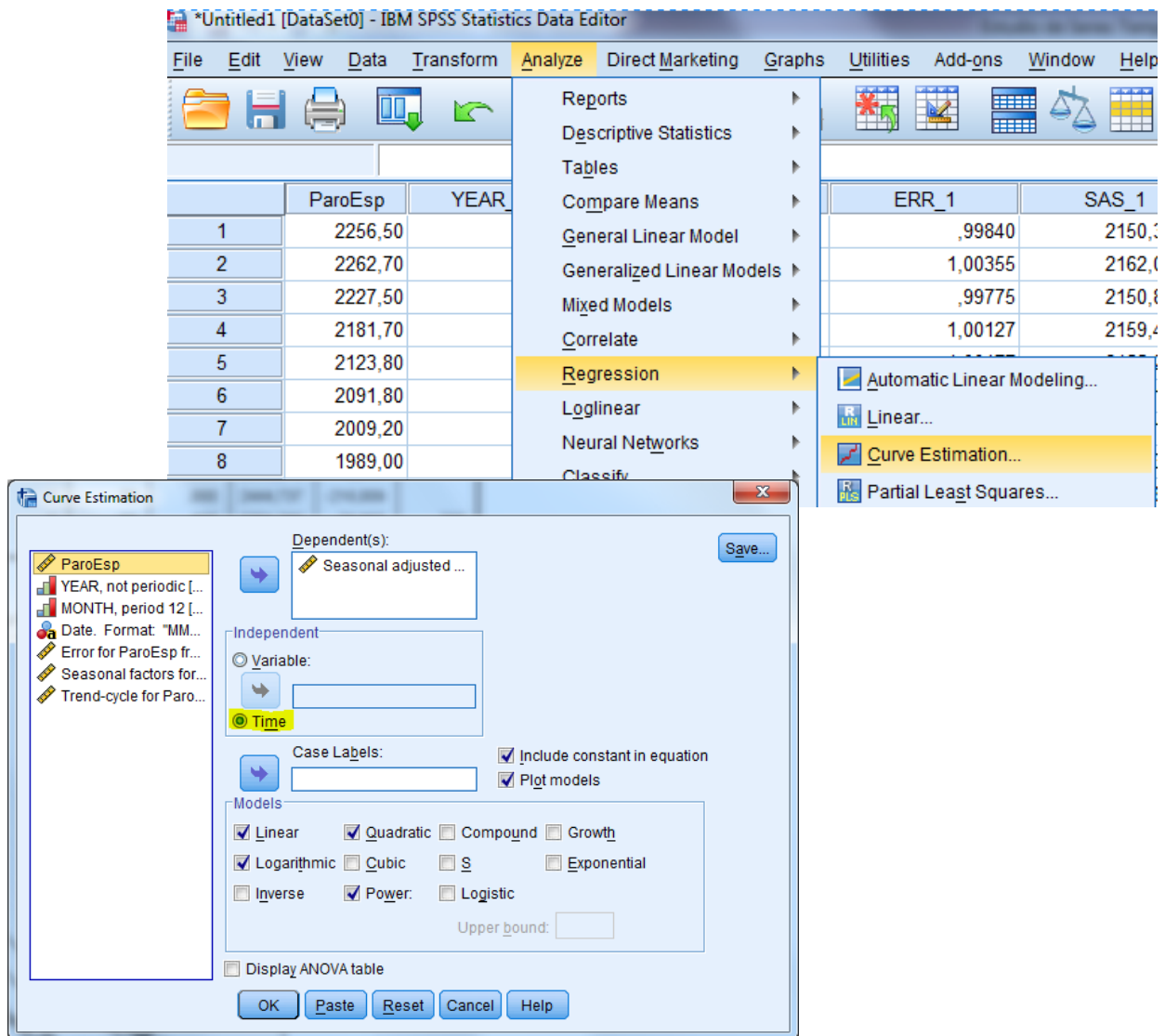


Figura 85.- SPSS Image 14

To establish chart, as dependent variable will be selected factorized series (SAS) and as independent series will be selected time (highlighted at figure 85). Then, many schemes will be selected to determine which best suits.

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	,706	167,727	1	70	,000	2043,272	-8,644	
Logarithmic	,836	355,777	1	70	,000	2444,737	-216,009	
Quadratic	<b>,980</b>	1657,364	2	69	,000	2304,280	-29,807	,290
Power	,820	318,353	1	70	,000	2540,335	-,118	



As observed in next chart, quadratic method shows a R square with a value of 0.98, which is a well adjust.

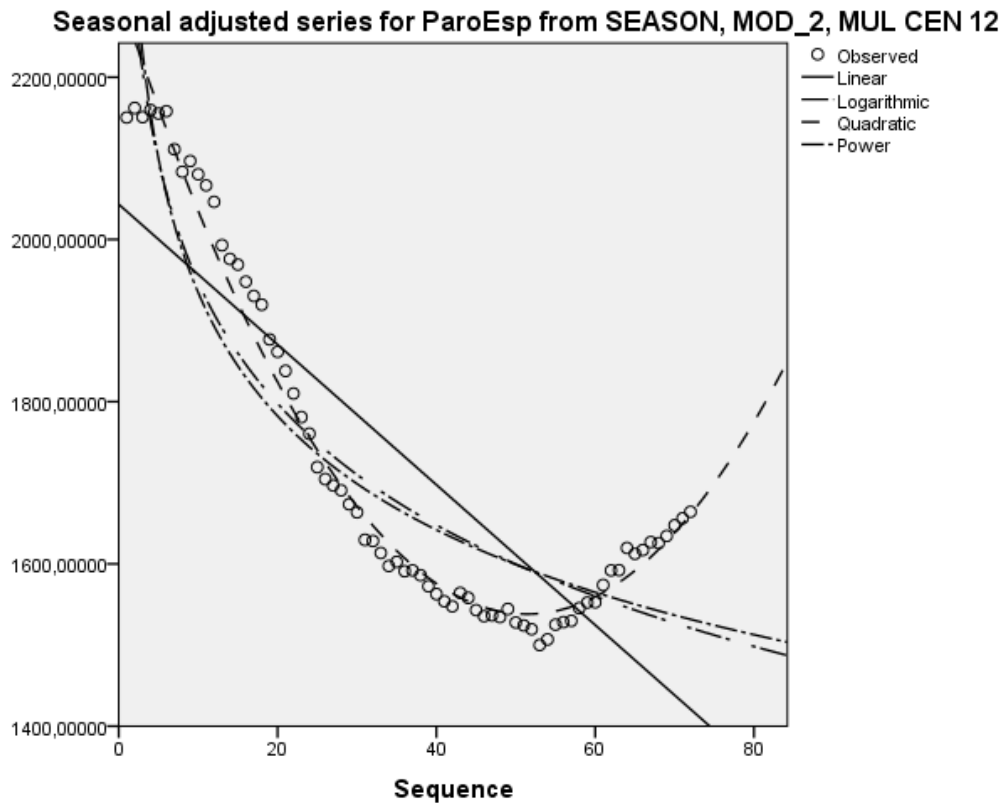


Figura 86.- SPSS Image 15

As mentioned in the previous paragraph, series suits to a quadratic scheme.

Then, data for the first 2003 quarter will be estimated. To this end, is necessary establish a relationship depending on the scheme best fit.

SPSS allows realize this estimation using same option previously used to get best scheme.

In dialog frame below (Figure 87) will be selected predicted values (highlighted) and predict until time limit to estimate. Now just is necessary select scheme previously obtained (quadratic).

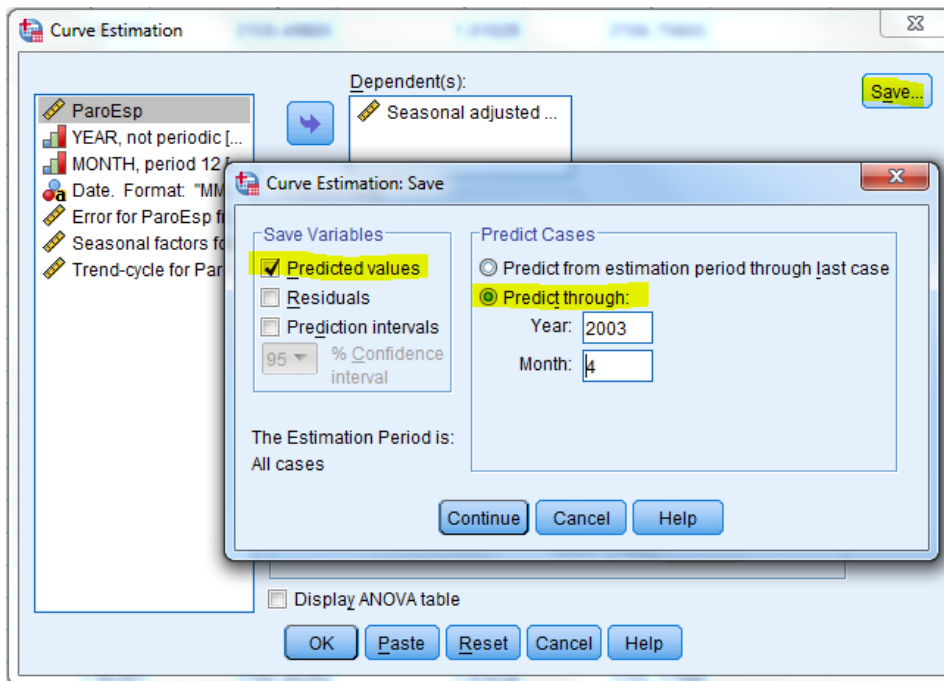


Figura 87.- SPSS Image 16

Thus it was obtained trend data will first four months of next year, but remember that this is a series seasonalized therefore, is necessary to relate the seasonal component (IVE) with trend data will follow during the estimated period. The power of SPSS can perform this step without calculating IVE, unlike other platforms.

Thus, process estimation consist in multiply (multiplicative scheme) SAF series with estimated series.

73	.	.	JAN 2003	.	1,04938	.	1670,45673
74	.	.	FEB 2003	.	1,04655	.	1683,45971
75	.	.	MAR 2003	.	1,03564	.	1697,05391
76	.	.	APR 2003	.	1,01028	.	1711,23931

Figura 88.- SPSS Image 17

To estimate the data, is necessary defining a new variable resulting product for the series estimated SAF:

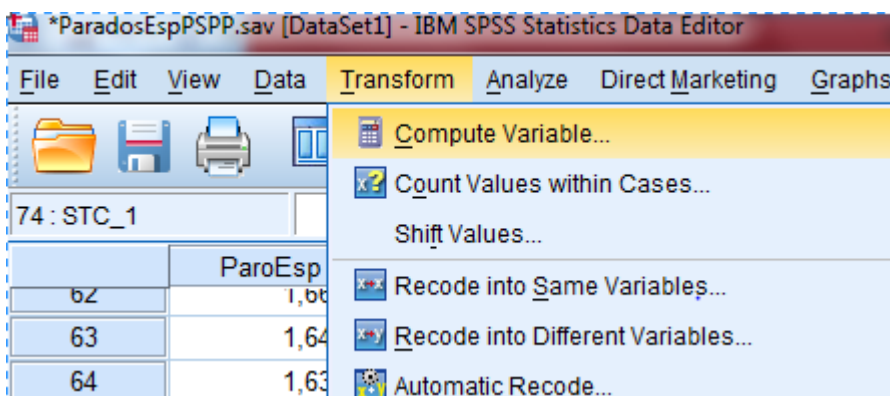


Figura 89.- SPSS Image 18

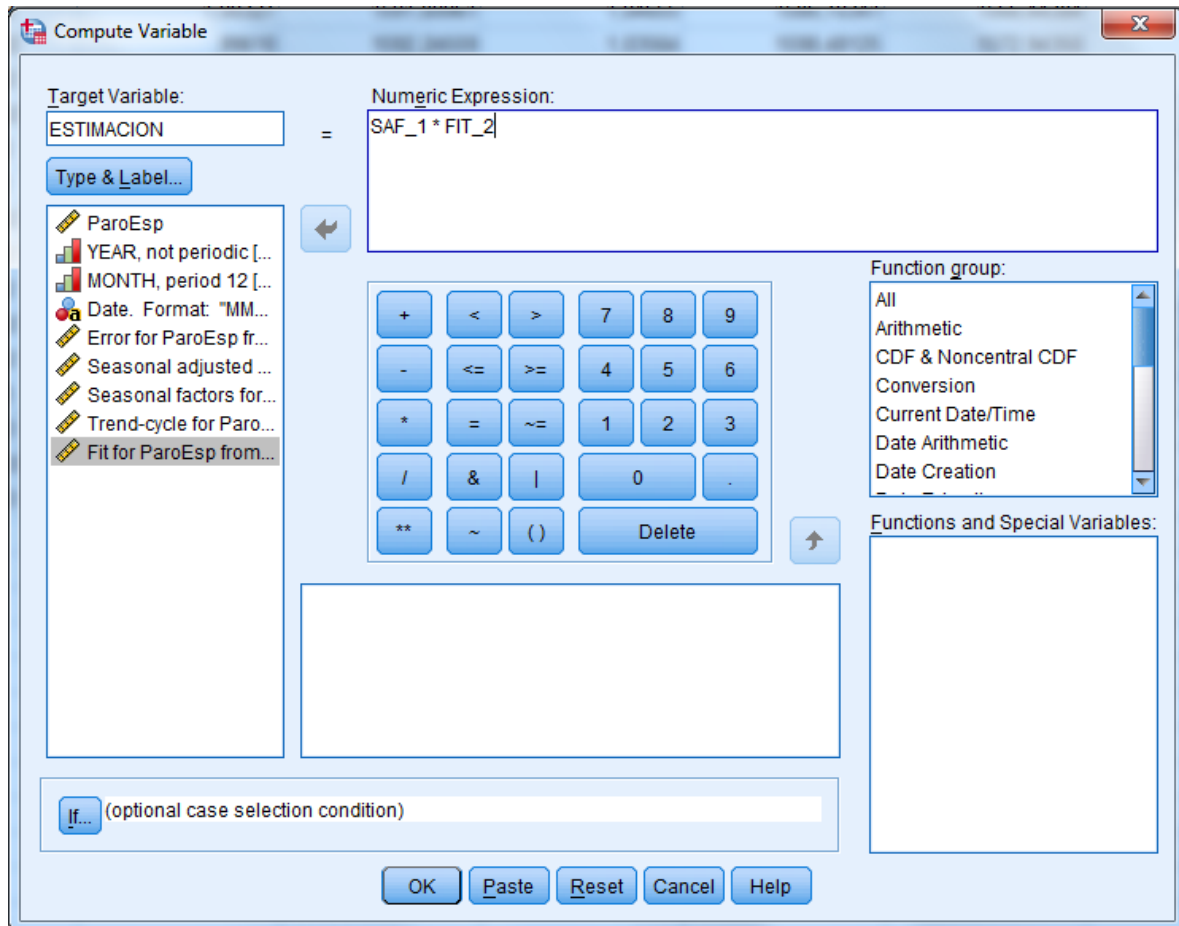


Figura 90.- SPSS Image 19

70	1,641.70	2002	10 OCT 2002	1,00096	1647,51880	,99647	1645,94262	1638,33237	1632,55
71	1,678.00	2002	11 NOV 2002	1,00010	1656,14057	1,01320	1655,96719	1649,40182	1671,17
72	1,688.10	2002	12 DEC 2002	1,00196	1664,24220	1,01434	1660,97948	1661,05108	1684,86
73	.	.	JAN 2003	.	.	1,04938	.	1673,28013	1755,91
74	.	.	FEB 2003	.	.	1,04655	.	1686,08899	1764,57
75	.	.	MAR 2003	.	.	1,03564	.	1699,47765	1760,05
76	.	.	APR 2003	.	.	1,01028	.	1713,44611	1731,06

Figura 91.- SPSS Image 20

## AI12.- Time series study using Excel.

AS first step, is necessary remember that Excel is not a statistics software, thus an important part of the process have to be made manually. However, there are statistics tools which is necessary activate.

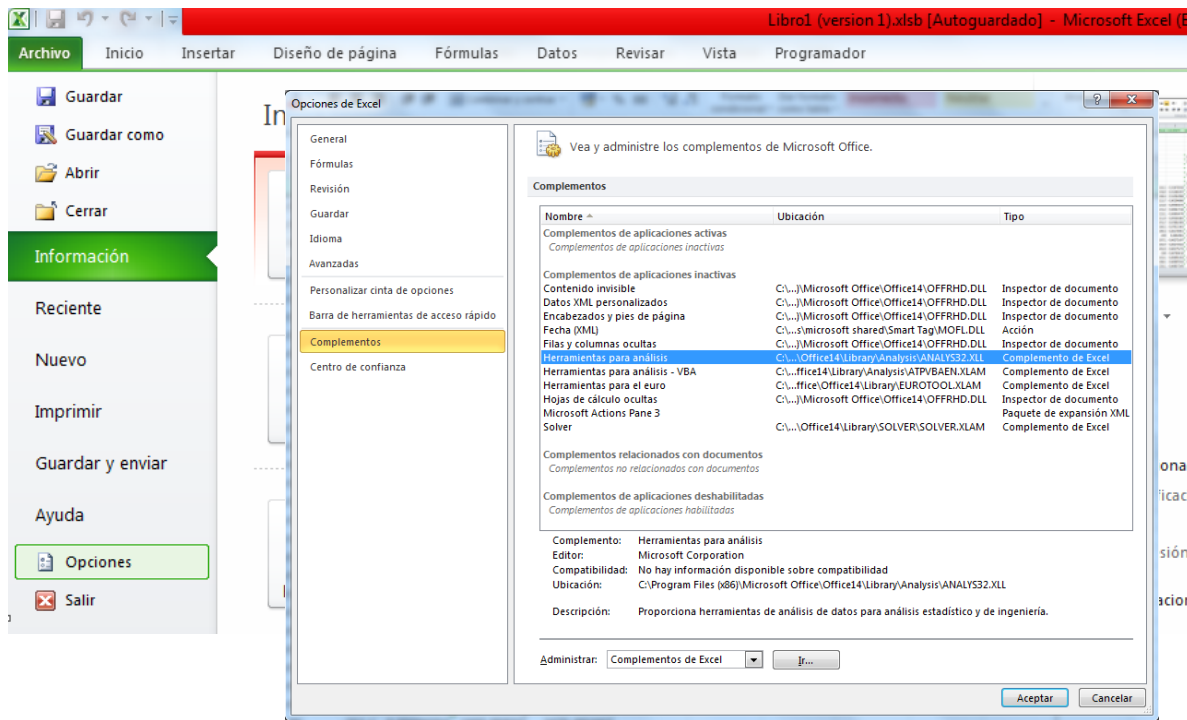


Figura 92.- Excel Image 1

### 1.-Insert data

Excel allow insert data manually, or using external files.

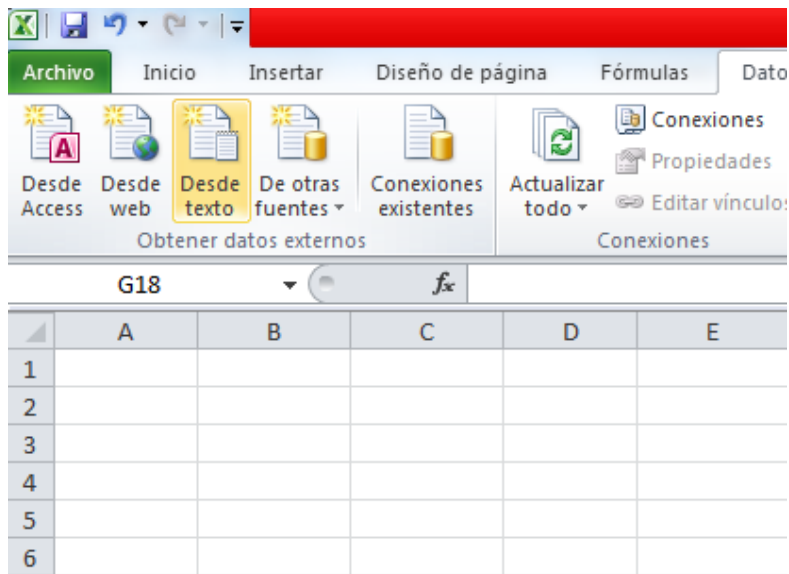


Figura 93.- Excel Image 2

## 2.- Defining dates of samples.

It is necessary define dates where data were observed. Excel has no function to define it automatically; however is simply define it manually. Is possible define first period (in the example, January 1997 -> Jan 1997 (MMM AAA)) and drag it until end. Is interesting set other two columns with months and years.

## 3.- Generating Sequence chart.

Es importante generar el gráfico de secuencia de la serie. A través de él es posible apreciar si existe estacionalidad en la serie así como estimar el posible modelo que sigue.

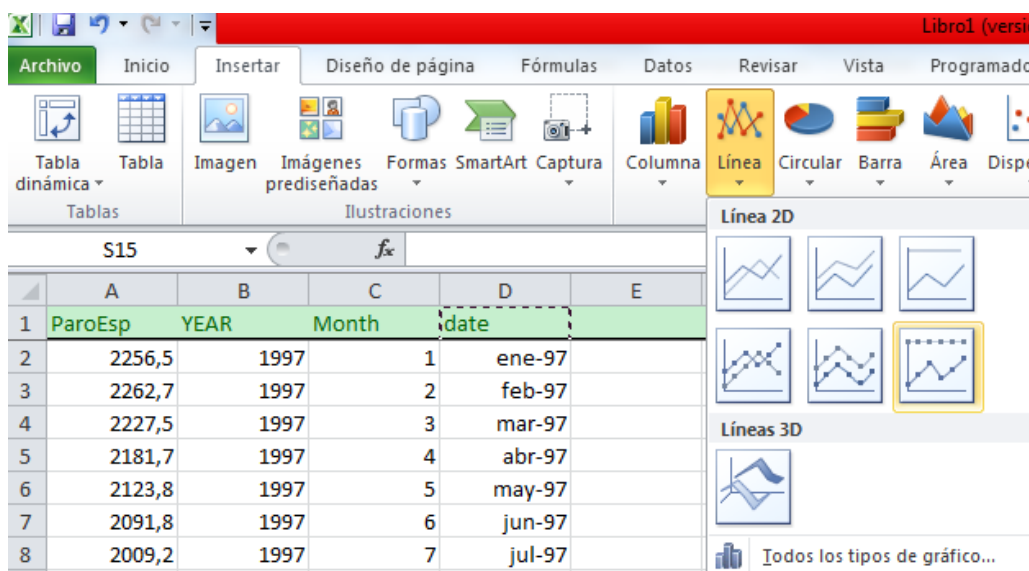


Figura 94.- Excel Image 3

Once type of chart is selected data will be selected in popup.

As first step, is necessary select the time series where data is (ParoEsp in examples).

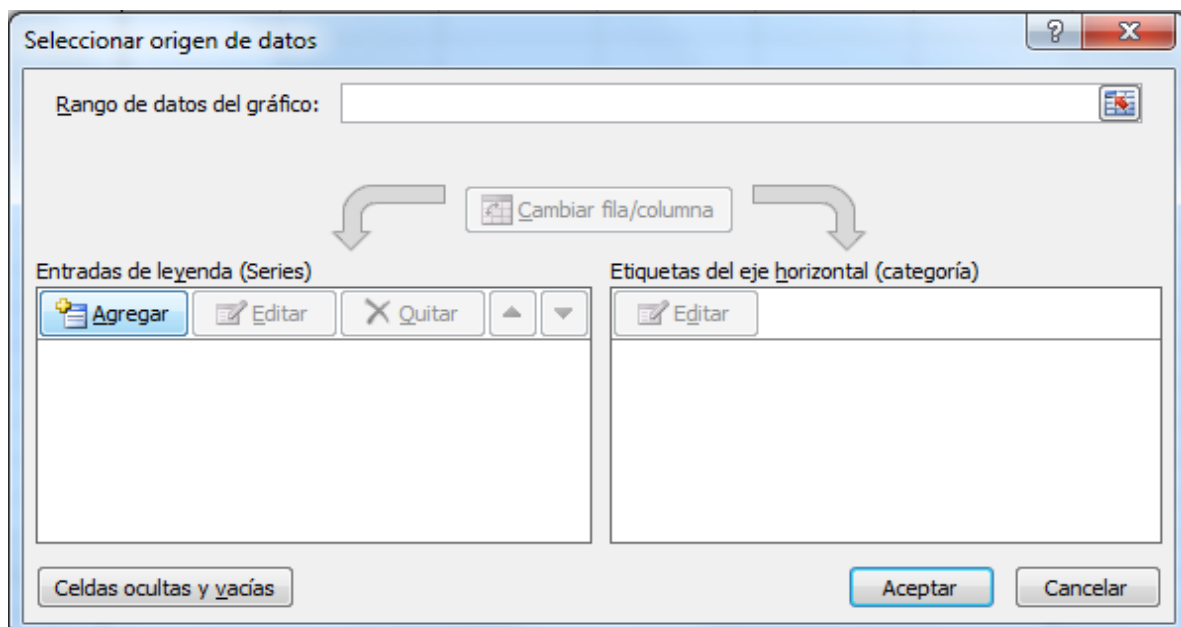


Figura 95.- Excel Image 4

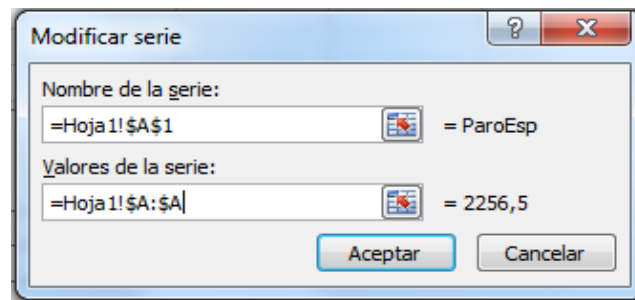


Figura 96.- Excel Image 5

Subsequently labels are selected for the X-axis:

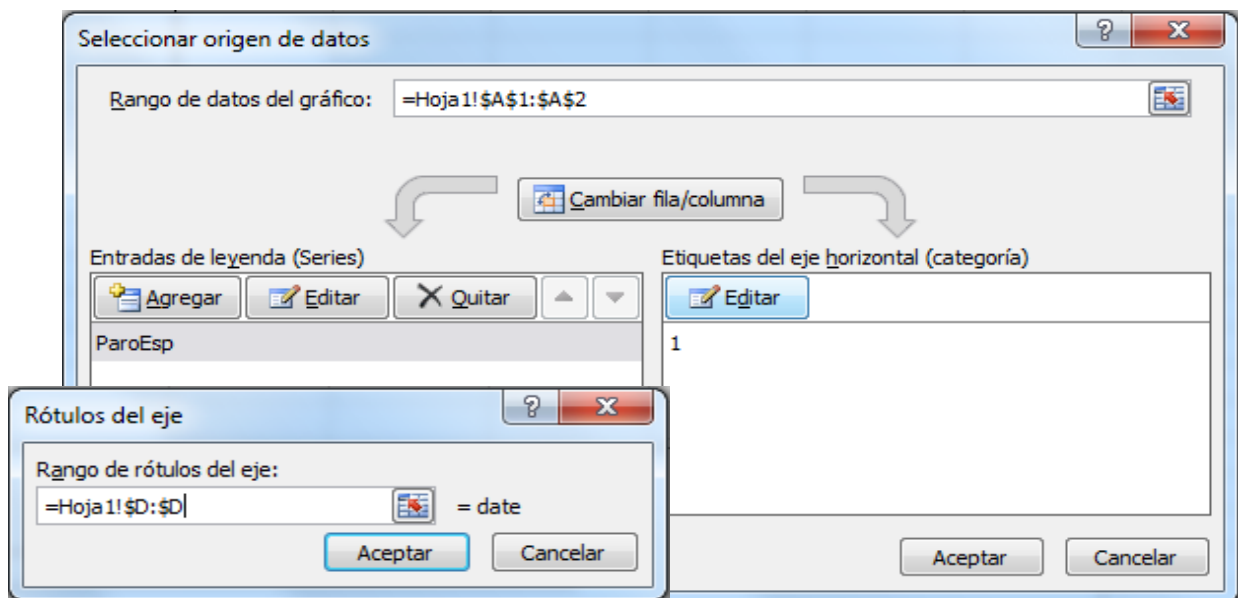


Figura 97.- Excel Image 6

Result is the following, Figure 98:

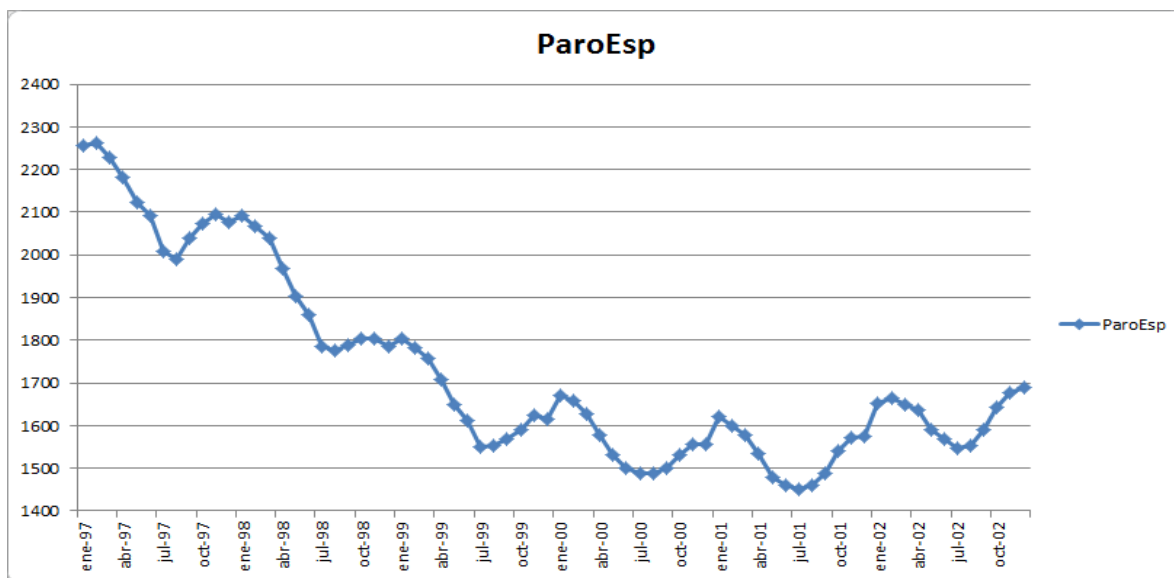


Figura 98.- Excel Image 7

In chart above (Figure 98) is possible observe that there is seasonality and probably has a multiplicative method. However, to strengthen the latter conclusion will be made a study in order to determine the scheme of time series.

#### 4. - Determining method of time series.

To determining if time series has a multiplicative or additive scheme, will be used method variation coefficient differences and ratios.

Por this, two new series will be defined. Thus, in the example, among 1997 January and 1998 January will be establish difference and ratio (always higher against fewer)

One of the strengths of Excel is the great maneuverability available, it makes it possible to directly define a formula in a cell and dragging down generate a complete series. The result is as follows in figure 8:

A	B	C	D	E	F
ParoEsp	YEAR	Month	date	DIFERENCIA	COCIENTE
2256,5	1997	1	ene-97		
2262,7	1997	2	feb-97		
2227,5	1997	3	mar-97		
2181,7	1997	4	abr-97		
2123,8	1997	5	may-97		
2091,8	1997	6	jun-97		
2009,2	1997	7	jul-97		
1989	1997	8	ago-97		
2040,1	1997	9	sep-97		
2072,9	1997	10	oct-97		
2093,9	1997	11	nov-97		
2075,7	1997	12	dic-97		
2091,3	1998	1	ene-98	-165,2	0,92678928
2067,8	1998	2	feb-98	-194,9	0,91386397
2039,1	1998	3	mar-98	-188,4	0,91542088
1968	1998	4	abr-98	-213,7	0,90204886

Figura 99.- Excel Image 8

Once defined the series, is necessary calculate the mean and standard deviation of both series. This will be carried with the average (mean) and DEVEST.M functions.

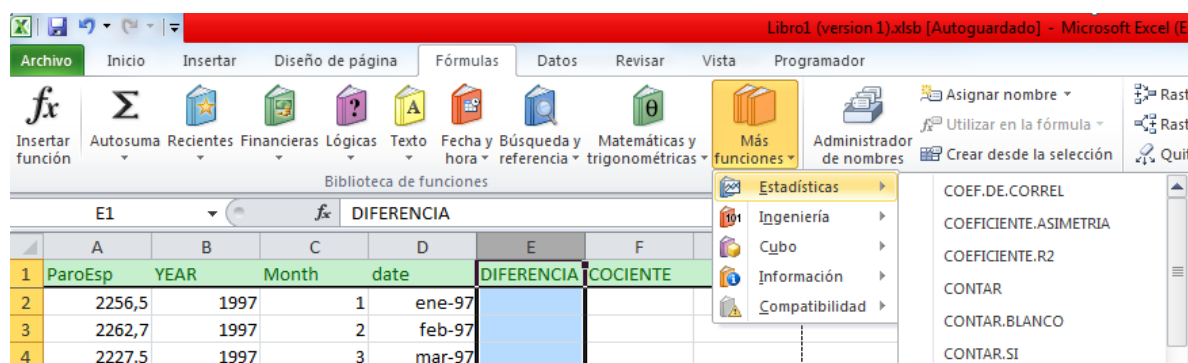


Figura 100.- Excel Image 9

Result is the following:

DIFERENCIA	media	-99,4483333
	DesEst	129,670028
COCIENTE	media	0,95149723
	DesEst	0,07054069

Thus, coefficients are:

$$CV_{diferencia} = 129.67 / -99.44 = -1.30$$

$$CV_{cociente} = 0.06957 / 0.9505 = 0.073$$

To compare ratios, the absolute value will be applied, regardless of signs.

In this case  $CV_{diferencia}$  exceeds  $CV_{cociente}$  is therefore multiplicative scheme, as previous graph shows.

## 5. - Factorizing.

Generally applications statistical field have automatic methods for seasonal adjustment of a time series.

However, as was initially warned, Excel is not a purely statistical tool, so the seasonal adjustment process in a series to be done manually.

For this, the first step is to calculate moving averages not centered in the series, that is, for an amount of data equal to a full term (12 months -samples-) in this case calculate the arithmetic mean. Due to the nature of the process, a certain amount of data, which depend on the number of samples is lost. It is important to note this fact because the means should be applied only to those records that should not be missed. Excel has tools for calculating the moving average.

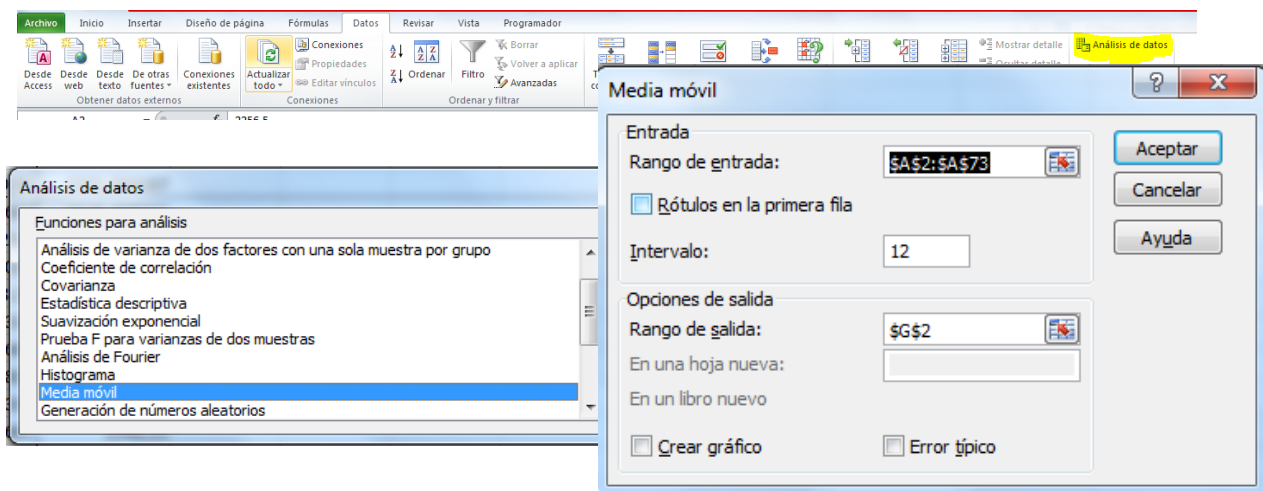


Figura 101.- Excel Image 10



The interval is the number of station by period (12 in this case)

Result is the following :

	A	B	C	D	E	F	G
1	ParoEsp	YEAR	Month	date	DIFERENCIA	COCIENTE	Med Móviles
2	2256,5	1997	1	ene-97			#N/A
3	2262,7	1997	2	feb-97			#N/A
4	2227,5	1997	3	mar-97			#N/A
5	2181,7	1997	4	abr-97			#N/A
6	2123,8	1997	5	may-97			#N/A
7	2091,8	1997	6	jun-97			2118,73333
8	2009,2	1997	7	jul-97			2104,96667
9	1989	1997	8	ago-97			2088,725
10	2040,1	1997	9	sep-97			2073,025

Figura 102.- Excel Image 11

Subsequently, because it has an even number of data should be performed the average between two moving averages to obtain the moving average centered:

	A	B	C	D	E	F	G	H
1	ParoEsp	YEAR	Month	date	DIFERENCIA	COCIENTE	Med Móviles	Med Mov Cent
2	2256,5	1997	1	ene-97			#N/A	#N/A
3	2262,7	1997	2	feb-97			#N/A	#N/A
4	2227,5	1997	3	mar-97			#N/A	#N/A
5	2181,7	1997	4	abr-97			#N/A	#N/A
6	2123,8	1997	5	may-97			#N/A	#N/A
7	2091,8	1997	6	jun-97			2118,73333	#N/A
8	2009,2	1997	7	jul-97			2104,96667	2111,85
9	1989	1997	8	ago-97			2088,725	2096,845833
10	2040,1	1997	9	sep-97			2073,025	2080,875
11	2072,9	1997	10	oct-97			2055,21667	2064,120833
12	2093,9	1997	11	nov-97			2036,75	2045,983333

Figura 103.- Excel Image 12

Once get this point. Is necessary calculating series composed by seasonality and error, this could be realized using ratio (multiplicative) among sample and its correspondence MMC.

## Business Intelligence applied to tourism | 2014 -2015

	A	B	C	D	E	F	G	H	I
1	ParoEsp	YEAR	Month	date	DIFERENCIA	COCIENTE	Med Móviles	Med Mov Cent	Eij X Aij
2	2256,5	1997	1	ene-97					#DIV/0!
3	2262,7	1997	2	feb-97					#DIV/0!
4	2227,5	1997	3	mar-97					#DIV/0!
5	2181,7	1997	4	abr-97					#DIV/0!
6	2123,8	1997	5	may-97					#DIV/0!
7	2091,8	1997	6	jun-97			2118,73333		#DIV/0!
8	2009,2	1997	7	jul-97			2104,96667	2111,85	95,13933281
9	1989	1997	8	ago-97			2088,725	2096,845833	94,85675906
10	2040,1	1997	9	sep-97			2073,025	2080,875	98,04048778
11	2072,9	1997	10	oct-97			2055,21667	2064,120833	100,4253223
12	2093,9	1997	11	nov-97			2036,75	2045,983333	102,3419871
13	2075,7	1997	12	dic-97			2017,48333	2027,116667	102,3966718
14	2091,3	1998	1	ene-98	-165,2	0,92678928	1998,89167	2008,1875	104,1386823
15	2067,8	1998	2	feb-98	-194,9	0,91386397	1981,23333	1990,0625	103,9062844
16	2039,1	1998	3	mar-98	-188,4	0,91542088	1960,25833	1970,745833	103,4684415
17	1968	1998	4	abr-98	-213,7	0,90204886	1937,825	1949,041667	100,9727003

Figura 104.- Excel Image 13

Then the IBVE factor that is identified as the average of the same MMC different periods (every May, June all ...) shall be calculated. The result is as follows (Figure 105):

	A	B	C	D	E	F	G	H	I	J
1	ParoEsp	YEAR	Month	date	DIFERENCIA	COCIENTE	Med Móviles	Med Mov Cent	Eij X Aij	IBVE
2	2256,5	1997	1	ene-97						104,7074081
3	2262,7	1997	2	feb-97						104,4177912
4	2227,5	1997	3	mar-97						103,3531336
5	2181,7	1997	4	abr-97						101,0941204
6	2123,8	1997	5	may-97						98,14924331
7	2091,8	1997	6	jun-97			2118,73333			96,73652564
8	2009,2	1997	7	jul-97			2104,96667	2111,85	95,13933281	95,00071295
9	1989	1997	8	ago-97			2088,725	2096,845833	94,85675906	95,34655232
10	2040,1	1997	9	sep-97			2073,025	2080,875	98,04048778	97,21707291
11	2072,9	1997	10	oct-97			2055,21667	2064,120833	100,4253223	99,50105907
12	2093,9	1997	11	nov-97			2036,75	2045,983333	102,3419871	101,3667585
13	2075,7	1997	12	dic-97			2017,48333	2027,116667	102,3966718	101,3543005
14	2091,3	1998	1	ene-98	-165,2	0,92678928	1998,89167	2008,1875	104,1386823	

Figura 105.- Excel Image 14

Following, is necessary calculating anual average of IBVE

IBVE	media	99,8537232
------	-------	------------

And with that value the IVE Seasonal Variation Index is represented as each party IBVE value by the value of the annual average is calculated by 100.

	A	B	C	D	E	F	G	H	I	J	K
1	ParoEsp	YEAR	Month	date	DIFERENCIA	COCIENTE	Med Móviles	Med Mov Cent	Eij X Aij	IBVE	IVE
2	2256,5	1997	1	ene-97						104,7074081	104,860795
3	2262,7	1997	2	feb-97						104,4177912	104,570754
4	2227,5	1997	3	mar-97						103,3531336	103,504537
5	2181,7	1997	4	abr-97						101,0941204	101,242214
6	2123,8	1997	5	may-97						98,14924331	98,2930232
7	2091,8	1997	6	jun-97			2118,73333			96,73652564	96,878236
8	2009,2	1997	7	jul-97			2104,96667	2111,85	95,13933281	95,00071295	95,1398805
9	1989	1997	8	ago-97			2088,725	2096,845833	94,85675906	95,34655232	95,4862265
10	2040,1	1997	9	sep-97			2073,025	2080,875	98,04048778	97,21707291	97,3594872
11	2072,9	1997	10	oct-97			2055,21667	2064,120833	100,4253223	99,50105907	99,6468192
12	2093,9	1997	11	nov-97			2036,75	2045,983333	102,3419871	101,3667585	101,515252
13	2075,7	1997	12	dic-97			2017,48333	2027,116667	102,3966718	101,3543005	101,502776
14	2091,3	1998	1	ene-98	-165,2	0,92678928	1998,89167	2008,1875	104,1386823		

Figura 106.- Excel Image 15

At point, is necessary adjust seasonally series, it is necessary for the ratio of the original series and the corresponding value IVE station multiplied by 100.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ParoEsp	YEAR	Month	date	DIFERENCIA	COCIENTE	Med Móviles	Med Mov Cent	Eij X Aij	IBVE	IVE	IVE (REP)	SAS
2	2256,5	1997	1	ene-97						104,7074081	104,860795	104,860795	2151,90052
3	2262,7	1997	2	feb-97						104,4177912	104,570754	104,570754	2163,79811
4	2227,5	1997	3	mar-97						103,3531336	103,504537	103,504537	2152,07958
5	2181,7	1997	4	abr-97						101,0941204	101,242214	101,242214	2154,93114
6	2123,8	1997	5	may-97						98,14924331	98,2930232	98,2930232	2160,68235
7	2091,8	1997	6	jun-97			2118,73333			96,73652564	96,878236	96,878236	2159,20529
8	2009,2	1997	7	jul-97			2104,96667	2111,85	95,13933281	95,00071295	95,1398805	95,1398805	2111,83784
9	1989	1997	8	ago-97			2088,725	2096,845833	94,85675906	95,34655232	95,4862265	95,4862265	2083,02294
10	2040,1	1997	9	sep-97			2073,025	2080,875	98,04048778	97,21707291	97,3594872	97,3594872	2095,4301
11	2072,9	1997	10	oct-97			2055,21667	2064,120833	100,4253223	99,50105907	99,6468192	99,6468192	2080,24703
12	2093,9	1997	11	nov-97			2036,75	2045,983333	102,3419871	101,3667585	101,515252	101,515252	2062,64572
13	2075,7	1997	12	dic-97			2017,48333	2027,116667	102,3966718	101,3543005	101,502776	101,502776	2044,96871
14	2091,3	1998	1	ene-98	-165,2	0,92678928	1998,89167	2008,1875	104,1386823			104,860795	1994,35833
15	2067,8	1998	2	feb-98	-194,9	0,91386397	1981,23333	1990,0625	103,9062844			104,570754	1977,41713

Figura 107.- Excel Image 16

## 6. - Calculating linear regression

After calculating the seasonally adjusted series, it is possible to calculate the regression line that best fits the representation of the points. This regression line is very important, because is the function which points of series will continue. This means that using this feature to move forward in time, it will be possible to predict future behavior of the series.

	A	B
1	Tiempo	ParoEsp
2	1	2256,5
3	2	2262,7
4	3	2227,5
5	4	2181,7
6	5	2123,8
7	6	2091,8
8	7	2009,2
9	8	1989
10	9	2040,1
11	10	2072,9

Figura 108.- Excel Image 17

For this, first is necessary generating this line. Excel has one tool, Regression, which is used for this:

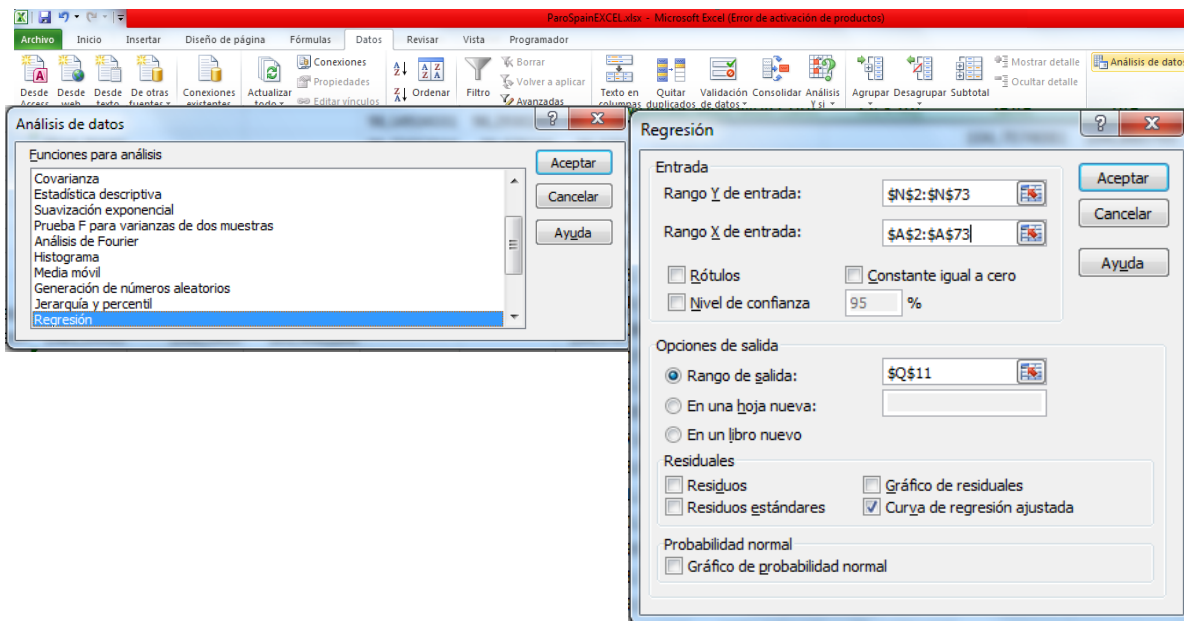


Figura 109.- Excel Image 18

The variable Y is the dependent variable (SAS) and the X variable is the independent variable (time). As a confidence level of 95% is pretty good and it is interesting to leaving the marked fitted regression curve.

The result is the following (Figure 110):

Resumen							
<b>Estadísticas de la regresión</b>							
Coeficiente	0,84027948						
Coeficiente	0,7060696						
R <sup>2</sup> ajustado	0,7018706						
Error típico	117,662705						
Observacion	72						
<b>ANÁLISIS DE VARIANZA</b>							
	<i>Grados de libertad</i>	<i>de cuadrado de los cua</i>	<i>F</i>	<i>valor crítico de F</i>			
Regresión	1	2327977,14	2327977,14	168,151621	2,751E-20		
Residuos	70	969115,843	13844,512				
Total	71	3297092,98					
	<i>Coeficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>nferior 95,0%</i>
Intercepción	2043,59012	28,0247923	72,9207943	9,1336E-68	1987,69643	2099,4838	1987,69643
Variable X 1	-8,65213179	0,66722544	-12,967329	2,751E-20	-9,98287097	-7,32139261	-9,98287097

Figura 110.- Excel Image 19

Above is the mathematical result, after that is showed graphical result.

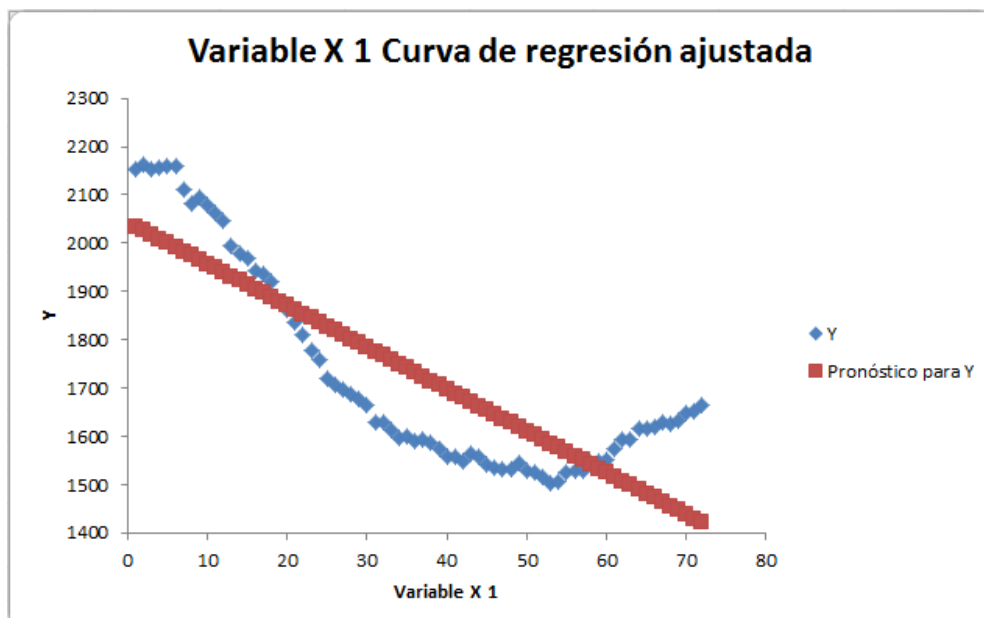
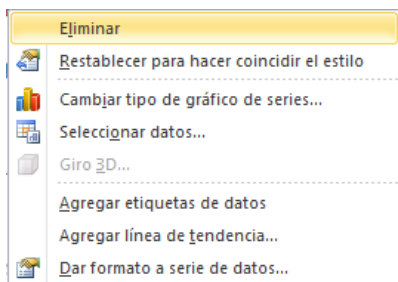
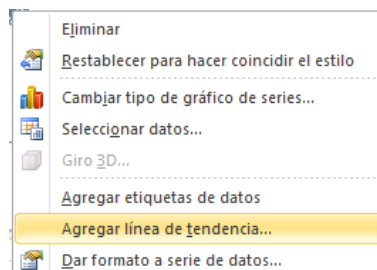


Figura 111.- Excel Image 20

As shown, the automatic line is not particularly suited to the function of the series, so it must be replaced. To do this, first remove the line, click anywhere on the line and delete (a) to add a new line below trend, for it click on any point of the series and select "Add Trendline".



A



b

In the dialog box that appears is necessary compare different lines in order to choose the one that best fits the series.

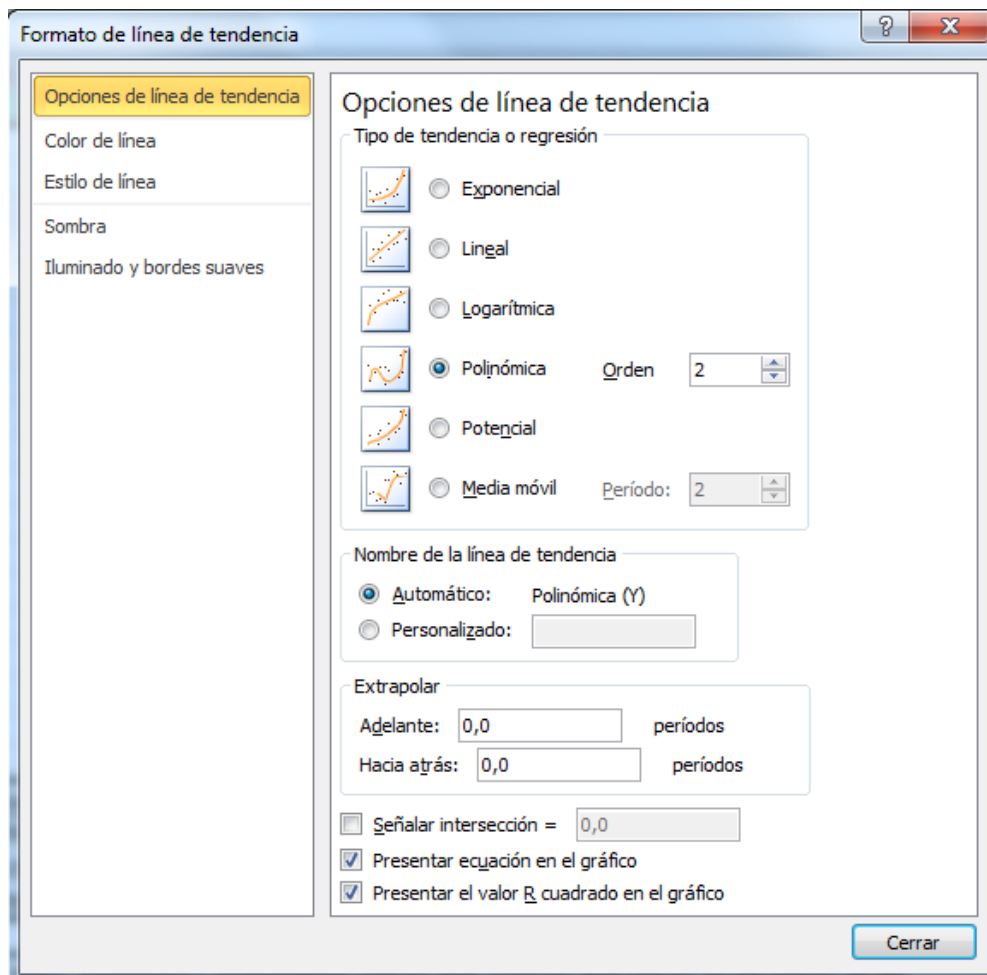


Figura 112.- Excel Image 21

In this case, the best value is a straight-order polynomial 2. For the equation is necessary check "making equation on the graph" box. It is also advisable to check "Display R-squared value on the graph" to check the fit of the line box (Figure 113).

Final graph is the following:

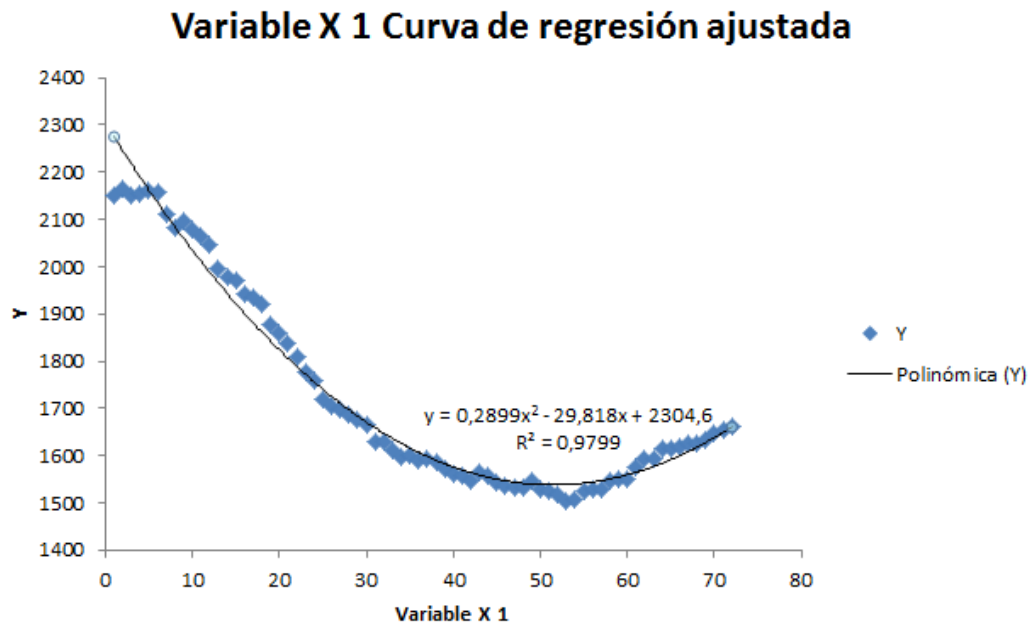


Figura 113.- Excel Image 22

Thus, function is the following:

$$y = 0,2899x^2 - 29,818x + 2304,6$$

This means that for estimating future values, simply replace the value of x. The original series came to x = 72, to estimate the first 3 months of the year after that it will be used is x = 73, x = 74 and x = 75. X=73

$$0,2899 (73)^2 - 29,818 (73) + 2304,6 = 1672,7631$$

$$X=74$$

$$0,2899 (74)^2 - 29,818 (74) + 2304,6 = 1685,5604$$

$$X=75$$

$$0,2899 (75)^2 - 29,818 (75) + 2304,6 = 1698,9375$$

Is added to the column of time the new values (73, 74, 75). ESTIMATE new column is created with the formula by replacing the X values by time and that will be the predicted straight.

The result is following in Figue 114:

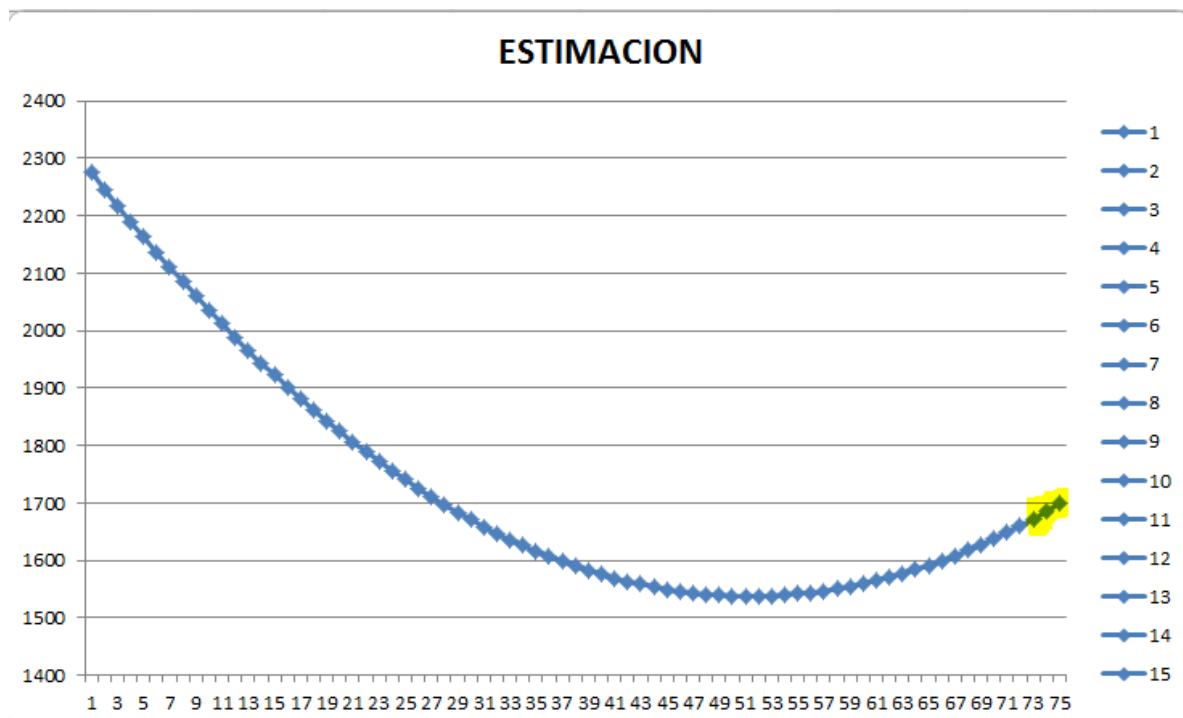


Figura 114.- Excel Image 23



### AI13. - Time Series Study using R

R is a programming language with scripting format mainly oriented statistical work. His feature script language allows to work at a very low level from the IT point of view, while the possibility of using already defined in R packages both by others, such as yourself, work at a very high level computer, without being necessary to have in-depth knowledge of programming. Also, this feature allows users to work with R for deep knowledge of statistics or for occasional users with basic knowledge.

As previous step is import load package called "forecast" which make predictions

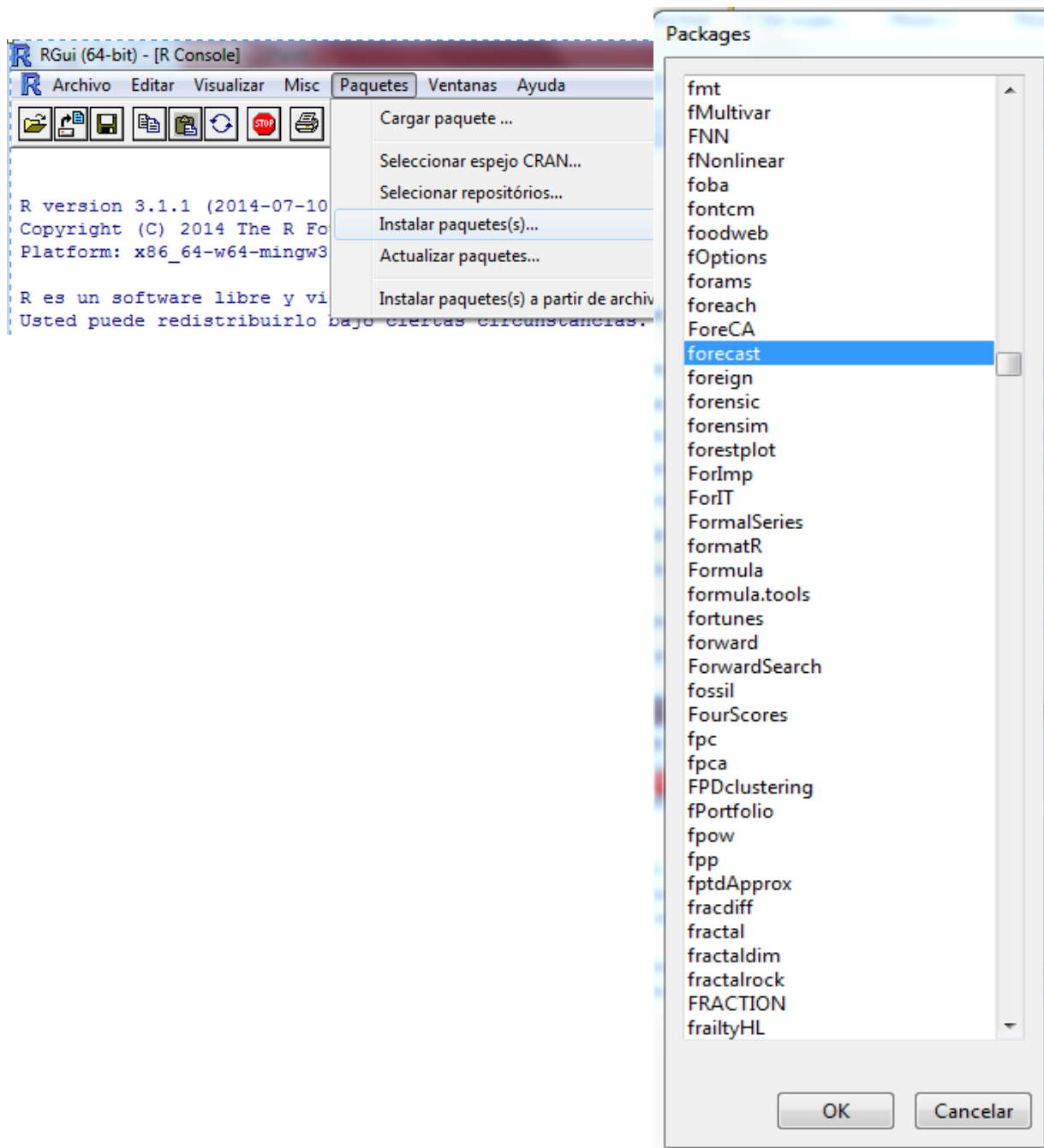
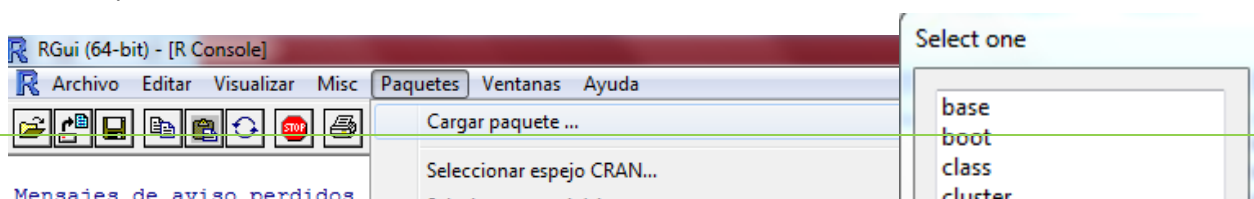


Figura 115.- R Image 1

After selecting the package and click on 'OK' packages are installed and configured automatically. It will be necessary to activate them.



Again, once package is selected and after check 'OK', selection will be load and will be ready to use.

Once package are loaded, is necessary load working directory where files are.

```
> setwd("C:/Users/ven02604/Universidad")
> dir()
[1] "~$ummex.docx"
[3] "2012-04-18_mujer-emprendedora.jpg"
[5] "Master (APROBADO)"
[7] "ParadosSpain9702.csv"
[9] "QODAT_20120717"
[11] "QOUSER_20120717"
"-WRL0782.tmp"
"ITIS (APROBADA) - Ingenieria Técnica en Informática de Sistemas"
"Master TIC"
"ParadosSpain9702.txt"
"QOSYS_20120717"
```

Figura 117.- R Image 3

With command `setwd(path)`, where 'path' refers the path where working directory is. Moreover, instruction `dir()`, shows all files located in this working directory.

Once working directory is established, is possible obtain data from time series. For this, R has commands with different kind of files, .csv in this case.

```
> ParadosESP<- read.csv(file="ParadosSpain9702.csv",head=TRUE,sep=",")
> ParadosESP
  paroEsp YEAR MONTH   DATE
1   2256.5 1997    1 jan-1997
2   2262.7 1997    2 feb-1997
3   2227.5 1997    3 mar-1997
4   2181.7 1997    4 apr-1997
5   2123.8 1997    5 may-1997
6   2091.8 1997    6 jun-1997
7   2009.2 1997    7 jul-1997
8   1989.0 1997    8 aug-1997
9   2040.1 1997    9 sep-1997
10  2072.9 1997   10 oct-1997
11  2093.9 1997   11 nov-1997
12  2075.7 1997   12 dec-1997
13  2091.3 1998    1 jan-1998
14  2067.8 1998    2 feb-1998
15  2039.1 1998    3 mar-1998
16  1968.0 1998    4 apr-1998
17  1902.2 1998    5 may-1998
18  1860.6 1998    6 jun-1998
19  1786.1 1998    7 jul-1998
20  1777.1 1998    8 aug-1998
21  1788.4 1998    9 sep-1998
```

Figura 118.- R Image 4

As seen in Figure 118, the instruction to use is `read.csv`. This statement uses as parameters file, which is the file to read, head, if variable is TRUE indicates that the file contains header, and Sep is the separator file, which in this case is a comma. As shown, it is important to overturn the result of reading a variable to use, in this case the variable is called `ParadosESP`.

Now is possible start time series study.

### 1.- Turning into time series.

First, you need to convert data in a time series. R has functions for this purpose. The file that is loaded contains several columns for the study of time series, the first column (where the data is) will therefore only necessary with which to work.

```
> tsParoESP <- ts(ParadosESP[1], frequency=12, start=c(1997,1))
> tsParoESP
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1997	2256.5	2262.7	2227.5	2181.7	2123.8	2091.8	2009.2	1989.0	2040.1	2072.9	2093.9	2075.7
1998	2091.3	2067.8	2039.1	1968.0	1902.2	1860.6	1786.1	1777.1	1788.4	1803.7	1804.5	1785.7
1999	1804.2	1783.9	1757.2	1708.0	1649.1	1612.5	1551.0	1554.5	1570.0	1591.7	1623.7	1613.8
2000	1670.6	1659.8	1628.5	1578.9	1531.2	1500.1	1488.8	1487.6	1501.4	1530.1	1556.9	1556.4
2001	1620.7	1598.9	1578.5	1535.1	1478.1	1460.6	1451.5	1459.0	1488.6	1540.0	1572.8	1574.8
2002	1651.7	1666.0	1649.0	1636.3	1589.0	1567.4	1548.4	1552.0	1590.3	1641.7	1678.0	1688.1

Figura 119.- R Image 5

The function to be used is 'ts'. This function uses the following parameters: first the time series, in this case ParadosESP [1] (how it looks, has been used only the first column), frequency, which is the periodicity of the series, 12 in this case, start, variable indicating the beginning of the series, January 1997 in this case. As before the result of the transformation is stored in the variable tsParoESP.

To can observe how is that series, in important show the chart.

For this, exists funtion "plot".

```
> plot(tsParoESP)
```

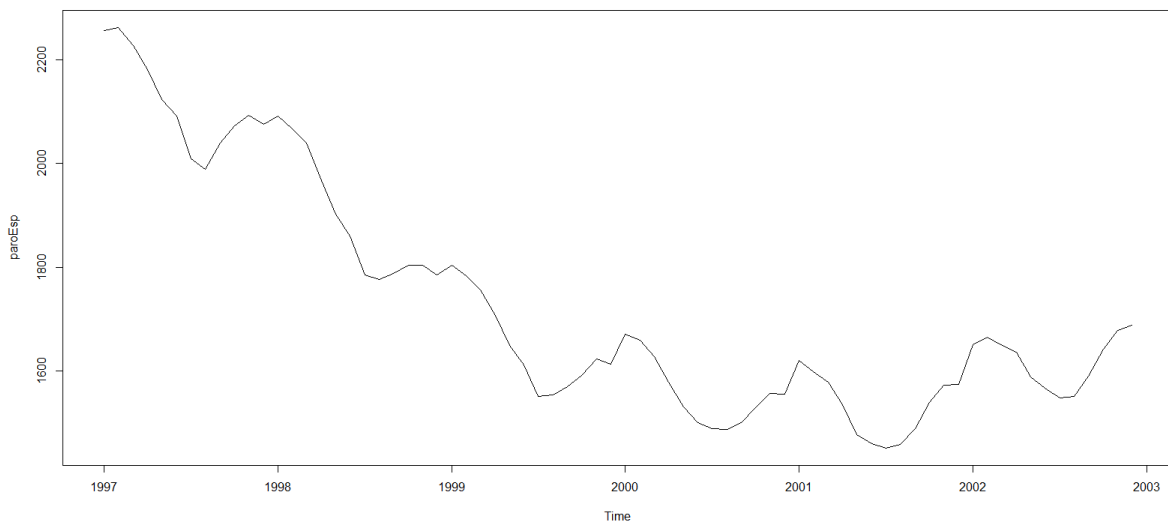


Figura 120.- R Image 6

## 2.- Decomposition.

The time series with which it is working is a seasonal series; this means that should decompose to see what their components. R has a function to perform this decomposition easily.

## Business Intelligence applied to tourism | 2014 -2015

```
> ParoESPComp <- decompose(tsParoESP)
> ParoESPComp
$X
      Jan  Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
1997 2256.5 2262.7 2227.5 2181.7 2123.8 2091.8 2009.2 1989.0 2040.1 2072.9 2093.9 2075.7
1998 2091.3 2067.8 2039.1 1968.0 1902.2 1860.6 1786.1 1777.1 1788.4 1803.7 1804.5 1785.7
1999 1804.2 1783.9 1757.2 1708.0 1649.1 1612.5 1551.0 1554.5 1570.0 1591.7 1623.7 1613.8
2000 1670.6 1659.8 1628.5 1578.9 1531.2 1500.1 1488.8 1487.6 1501.4 1530.1 1556.9 1556.4
2001 1620.7 1598.9 1578.5 1535.1 1478.1 1460.6 1451.5 1459.0 1488.6 1540.0 1572.8 1574.8
2002 1651.7 1666.0 1649.0 1636.3 1589.0 1567.4 1548.4 1552.0 1590.3 1641.7 1678.0 1688.1

$seasonal
      Jan  Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
1997 81.087431 76.149097 58.719097 20.860764 -27.420069 -50.404236 -84.502569 -78.370069 -44.316736 -4.970903 26.710764 26.457431
1998 81.087431 76.149097 58.719097 20.860764 -27.420069 -50.404236 -84.502569 -78.370069 -44.316736 -4.970903 26.710764 26.457431
1999 81.087431 76.149097 58.719097 20.860764 -27.420069 -50.404236 -84.502569 -78.370069 -44.316736 -4.970903 26.710764 26.457431
2000 81.087431 76.149097 58.719097 20.860764 -27.420069 -50.404236 -84.502569 -78.370069 -44.316736 -4.970903 26.710764 26.457431
2001 81.087431 76.149097 58.719097 20.860764 -27.420069 -50.404236 -84.502569 -78.370069 -44.316736 -4.970903 26.710764 26.457431
2002 81.087431 76.149097 58.719097 20.860764 -27.420069 -50.404236 -84.502569 -78.370069 -44.316736 -4.970903 26.710764 26.457431

$trend
      Jan  Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
1997 NA NA NA NA NA NA NA 2111.850 2096.846 2080.875 2064.121 2045.983 2027.117
1998 2008.188 1990.062 1970.746 1949.042 1925.767 1901.625 1877.579 1853.787 1830.213 1807.633 1786.254 1765.371
1999 1745.237 1726.167 1707.792 1689.858 1673.492 1658.796 1646.067 1635.329 1624.796 1614.054 1603.763 1594.167
2000 1586.892 1581.512 1575.867 1570.442 1565.092 1559.917 1555.446 1550.829 1546.208 1542.300 1538.263 1534.404
2001 1531.204 1528.458 1526.733 1526.612 1527.688 1529.117 1531.175 1535.263 1540.996 1548.150 1556.987 1566.058
2002 1574.546 1582.458 1590.571 1599.046 1607.667 1616.771 NA NA NA NA NA NA

$random
      Jan  Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec
1997 NA NA NA NA NA NA NA -18.1474306 -29.4757639 3.5417361 13.7500694 21.2059028 22.1259028
1998 2.0250694 1.5884028 9.6350694 -1.9024306 3.8534028 9.3792361 -6.9765972 1.6825694 2.5042361 1.0375694 -8.4649306 -6.1282639
1999 -22.1249306 -18.4157639 -9.3107639 -2.7190972 3.0284028 4.1084028 -10.5640972 -2.4590972 -10.4790972 -17.3832639 -6.7732639 -6.8240972
2000 2.6209028 2.1384028 -6.0857639 -12.4024306 -6.4715972 -9.4124306 17.8567361 15.1409028 -0.4915972 -7.2290972 -8.0732639 -4.4615972
2001 8.4084028 -5.7074306 -6.9524306 -12.3732639 -22.1674306 -18.1124306 4.8275694 2.1075694 -8.0790972 -3.1790972 -10.8982639 -17.7157639
2002 -3.9332639 7.3925694 -0.2899306 16.3934028 8.7534028 1.0334028 NA NA NA NA NA NA

$figure
[1] 81.087431 76.149097 58.719097 20.860764 -27.420069 -50.404236 -84.502569 -78.370069 -44.316736 -4.970903 26.710764 26.457431

$type
[1] "additive"

attr(,"class")
[1] "decomposed.ts"
```

Figura 121.- R Image 7

The function is 'decompose', this function only as an input parameter is the time series of work. As you can see, the observed series, seasonality, trend and the random component is displayed (Figure 121).

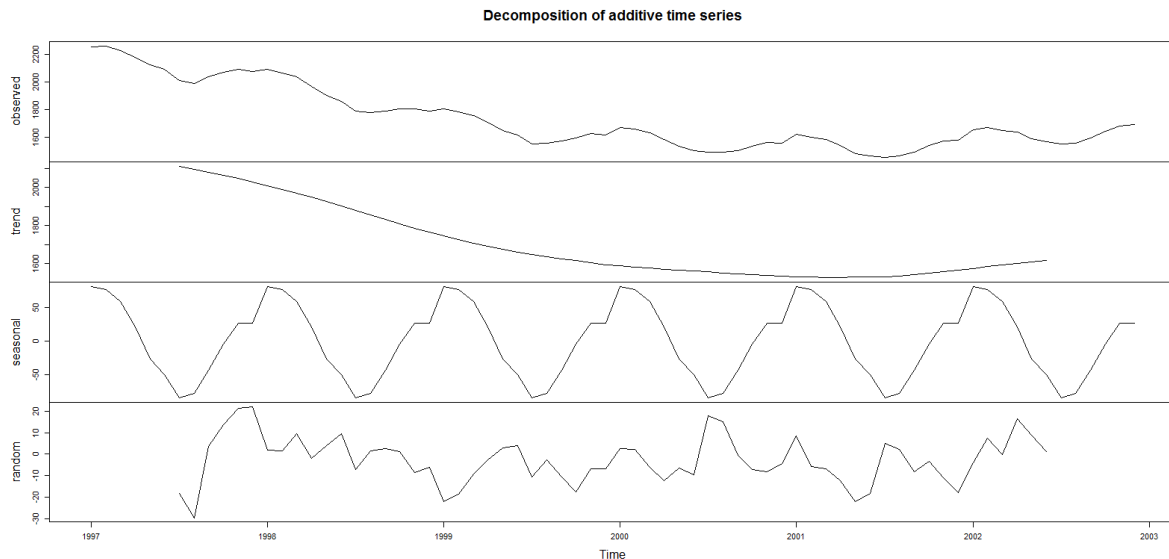


Figura 122.- R Image 8

### 3.- Forecast.

This is done using the preloaded package 'forecast'.

```
> ForeParoESP <- forecast(tsParoESP,4)
> ForeParoESP
```

	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2003		1746.246	1729.081	1763.411	1719.995	1772.498
Feb 2003		1752.148	1725.745	1778.552	1711.768	1792.529
Mar 2003		1741.998	1707.239	1776.757	1688.839	1795.157
Apr 2003		1711.889	1669.069	1754.710	1646.401	1777.378

Figura 123.- R Image 9

The forecast function uses as a parameter the time series with the number of values to be estimated. As shown, the points that were expected to estimate, with certain margins of error are obtained. To see graphically how the estimate is possible use plot function over the variable that stores the prediction.

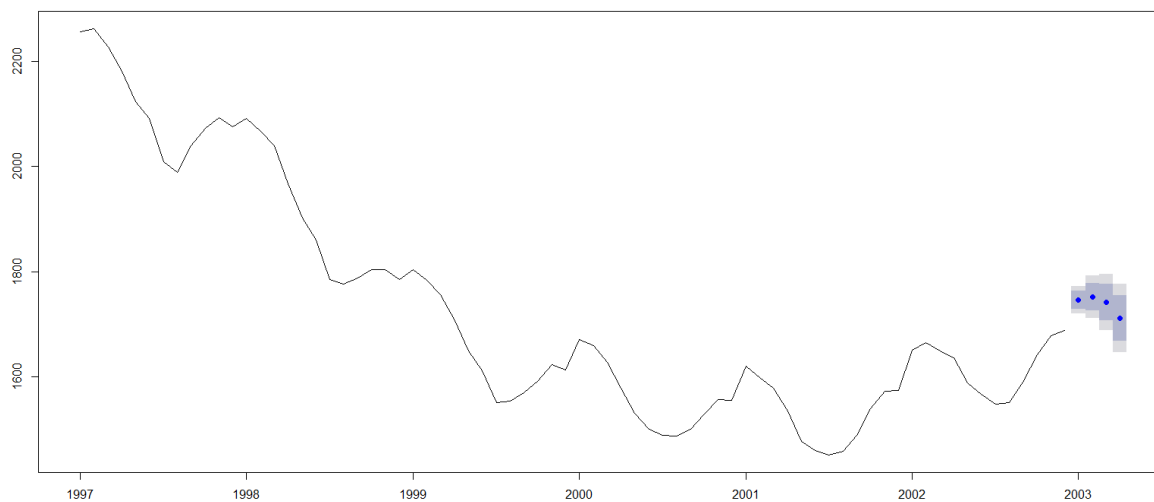


Figura 124.- R Image 10.

### III4.- Time Series study using Tableau

Tableau is a tool which allow obtain data from many different sources, in the case of this study, this source will be a separated by commas file o .csv. For the proper functioning of this tool, is necessary to include a reference to date when data was recorded in this case MMM-yyyy (ene-1997), as following image shows (Figure 125)

```

paroEsp, YEAR, MONTH, DATE
2256.50, 1997, 1, ene-1997
2262.70, 1997, 2, feb-1997
2227.50, 1997, 3, mar-1997
2181.70, 1997, 4, abr-1997
2123.80, 1997, 5, may-1997
2091.80, 1997, 6, jun-1997
2009.20, 1997, 7, jul-1997

```

Figura 125.- Tableau Image 1

#### 1.- Data acquisition

On the home screen of Tableau, the option “Text File” will be selected, as this image shows (Figure 126).

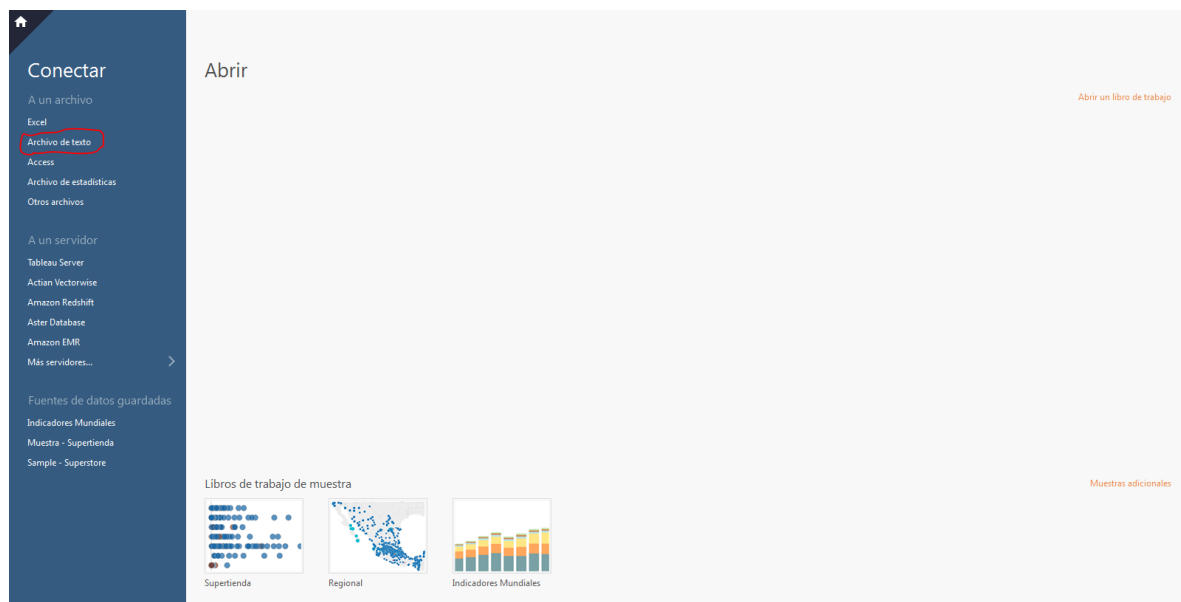


Figura 126.- Tableau Image 2

From this window, option “Personalized division” will be selected on the menu.

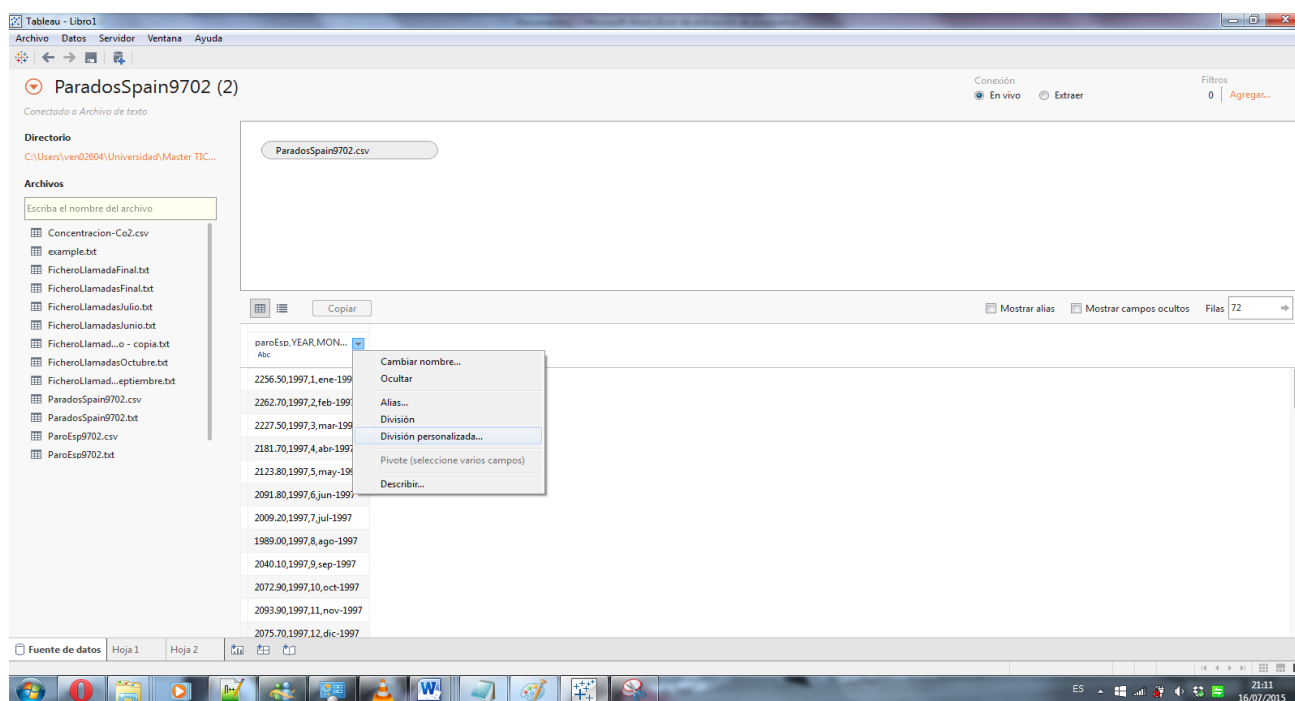


Figura 127.- Tableau Image 3

This option will open the following menu (Figure 128), where is possible select separator, same as column range to stablish.

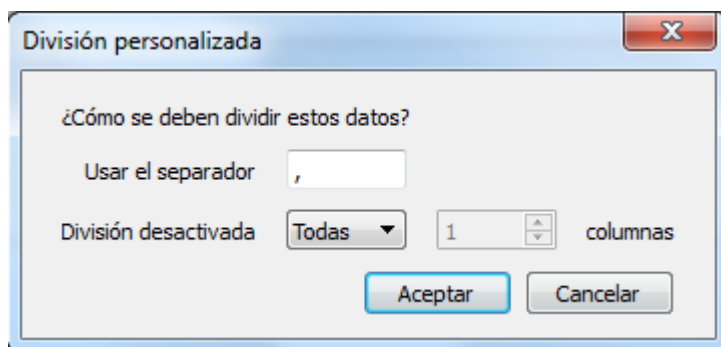


Figura 128.- Tableau Image 4

Then, column name change for make work easy.

paroEsp, YEAR, MON...	ParoESP	Year	Month	Date
Abc ParadosSpain9702....	=Abc	=Abc	=Abc	=Abc
2256.50,1997,1,ene-1997	2256.50	1997	1	ene-1997
2262.70,1997,2,feb-1997	2262.70	1997	2	feb-1997
2227.50,1997,3,mar-1997	2227.50	1997	3	mar-1997
2181.70,1997,4,abr-1997	2181.70	1997	4	abr-1997
2123.80,1997,5,may-1997	2123.80	1997	5	may-1997

Figura 129.- Tableau Image 5



## 2.- Variable Formatting

To do this, right-click on the left side of the screen in the section dimensions on the variable date. From the dropdown menu select "Create" and the sub-menu "Calculated Field" (Figure 130).

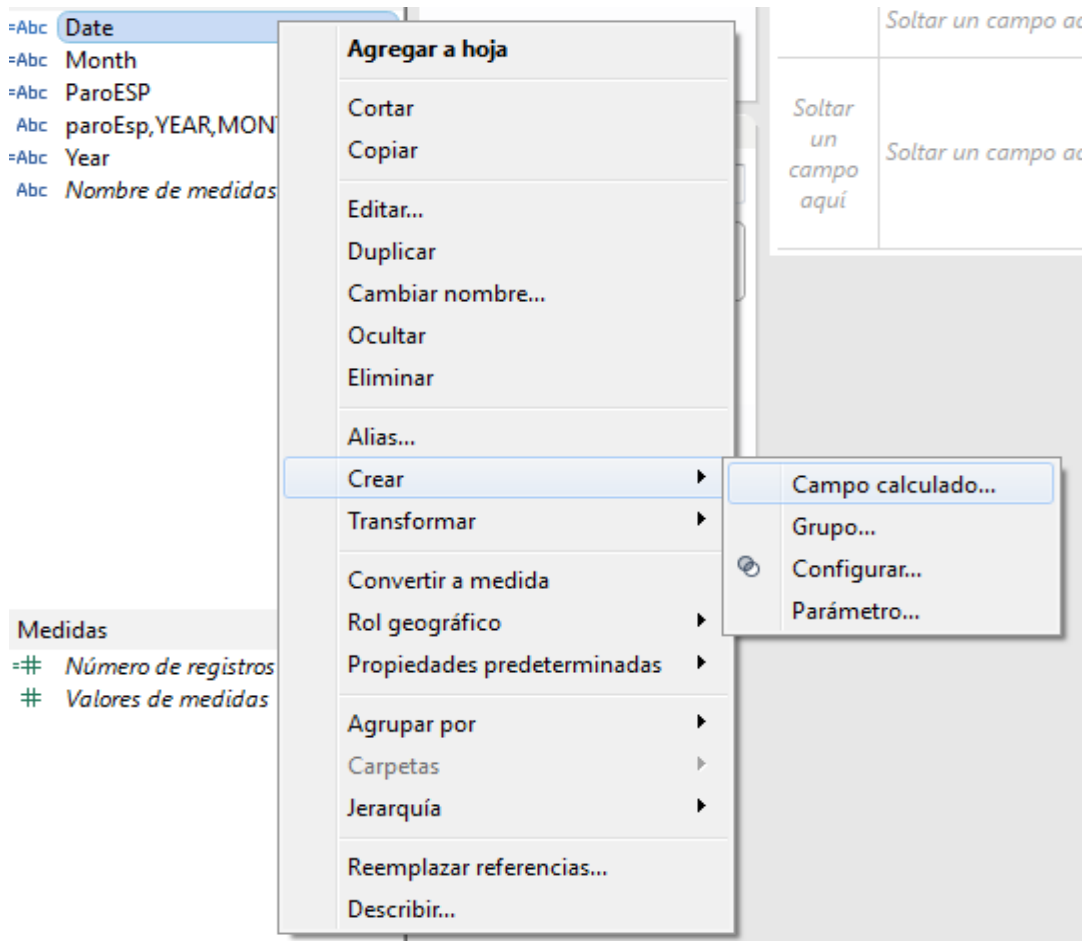


Figura 130.- Tableau Image 6

Then, on the following window (Figure 131) must type this line:.

`DATEPARSE("MMM-yyyy",[Date])`

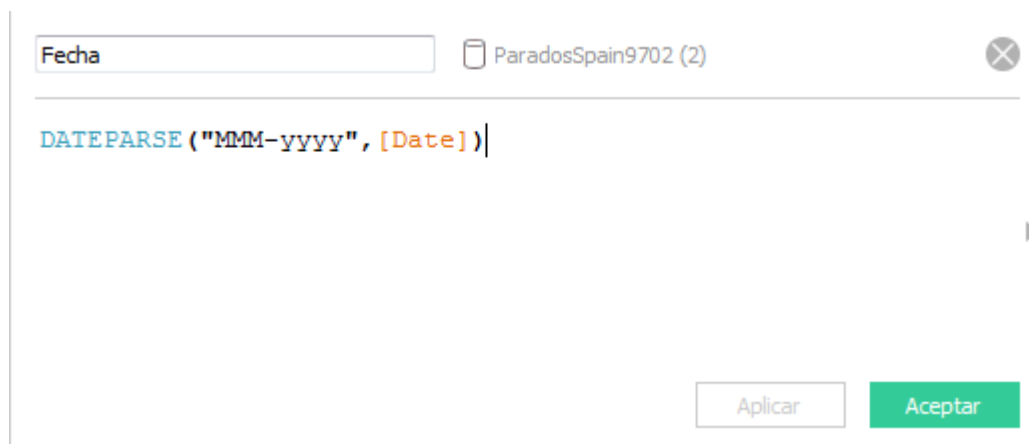


Figura 131.- Tableau Image 7

Del mismo modo se cambiará el tipo del dato de “ParoESP”.

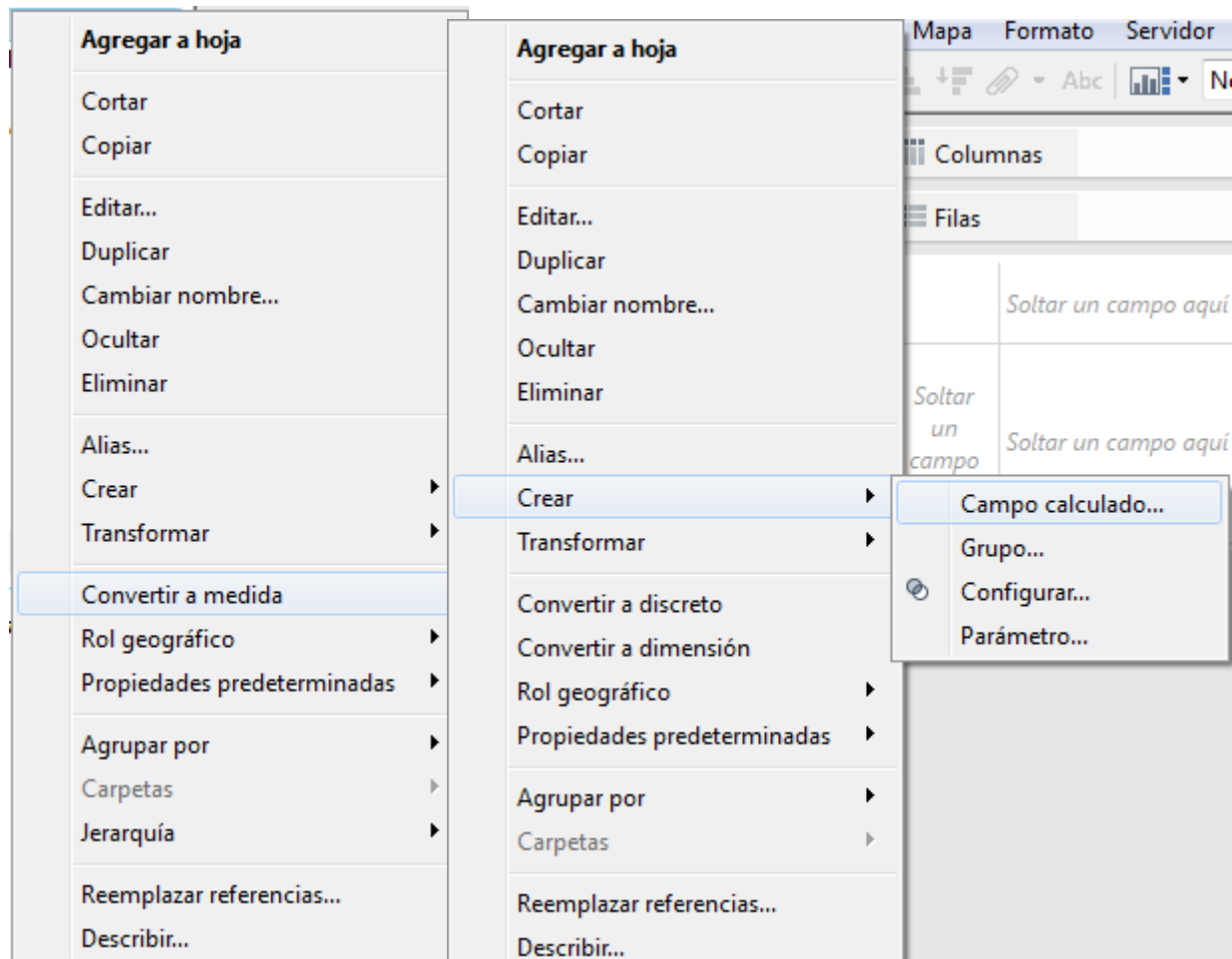


Figura 132.- Tableau Image 8

In the following window (Figure 133), type this line.

FLOAT([ParoESP])

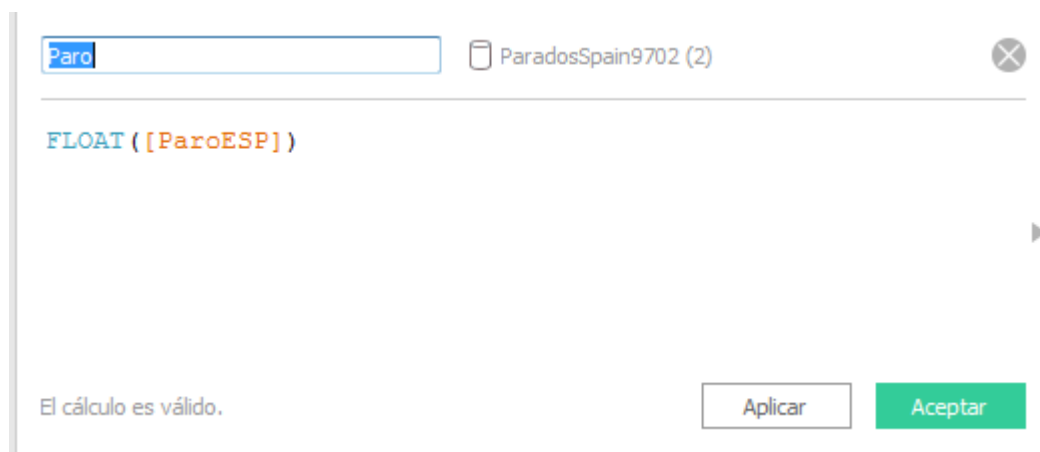


Figura 133.- Tableau Image 9

### 3.- Generating graph.

Once formatted fields will be placed on the graph. In the X axis variable created from the date it will be disaggregated quarter and once more a month, while the axis is selected and the value of the variable "Stop" will apply. Leaving the graph follows (Figure 134).

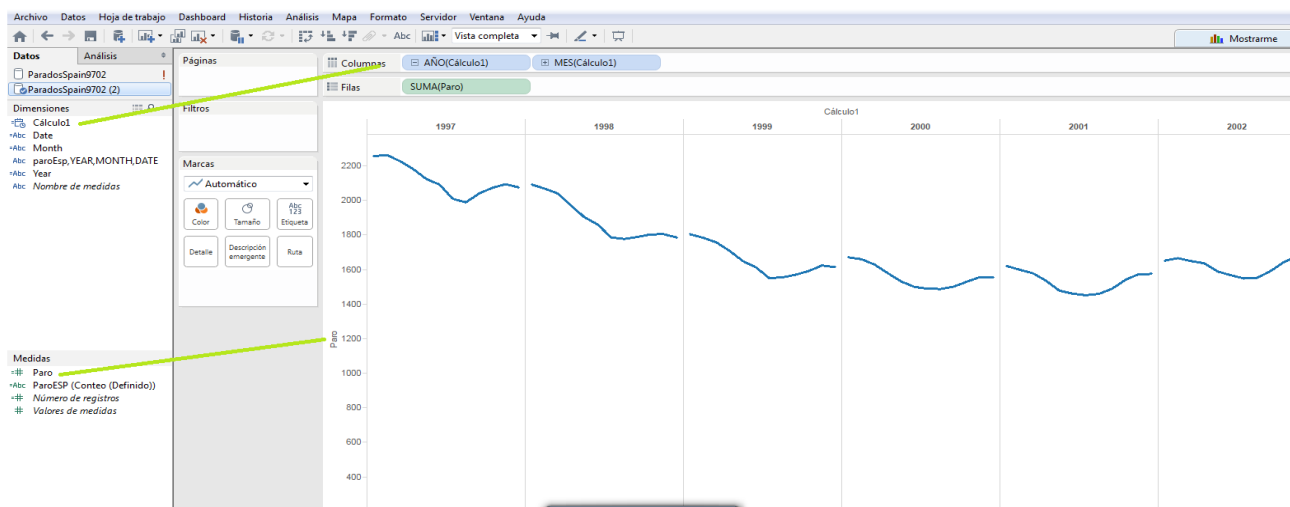


Figura 134.- Tableau Image 10

### 4.- Forecasting.

In the tool bar, select menu "Analysis", then option "Forecast" and in the submenu, select option "Show forecast"

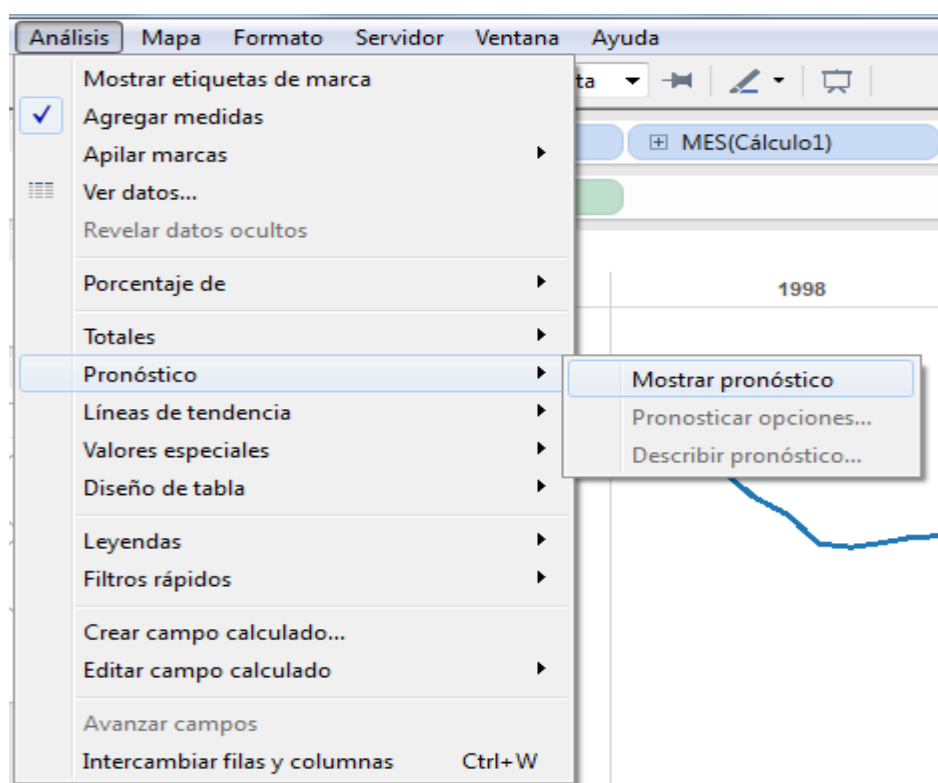


Figura 135.- Tableau Image 11

Result graph is as follow (Figure 136):

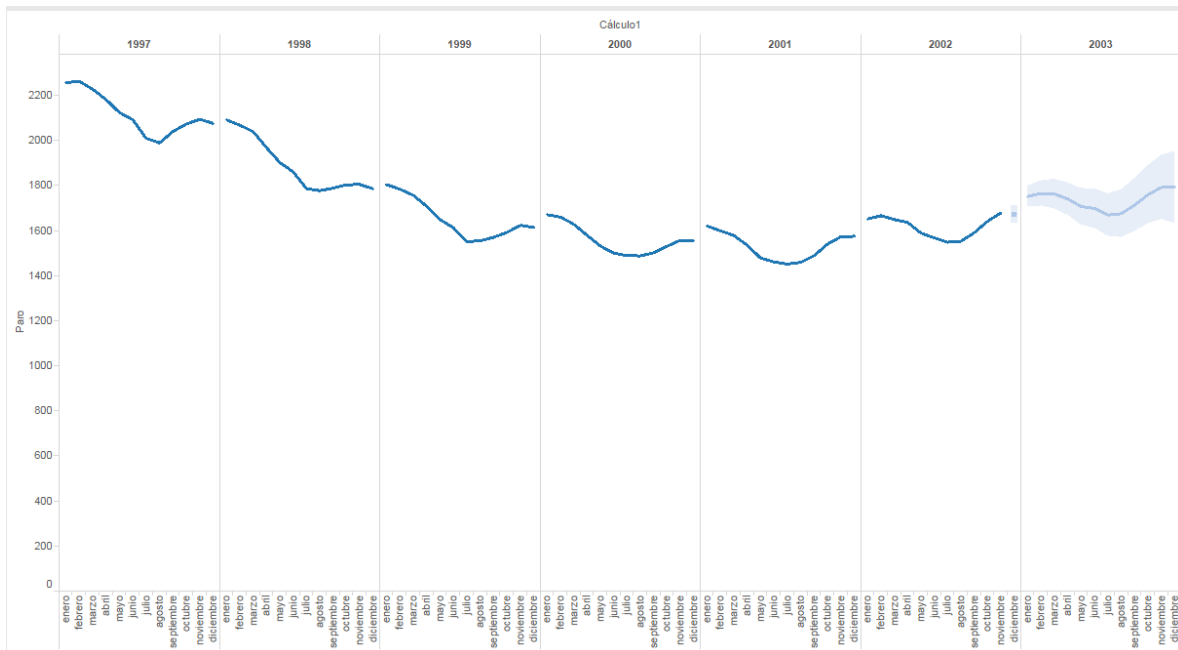


Figura 136.- Tableau Image 12

For adjust necessities, select at the same previous menu “Forecast options”

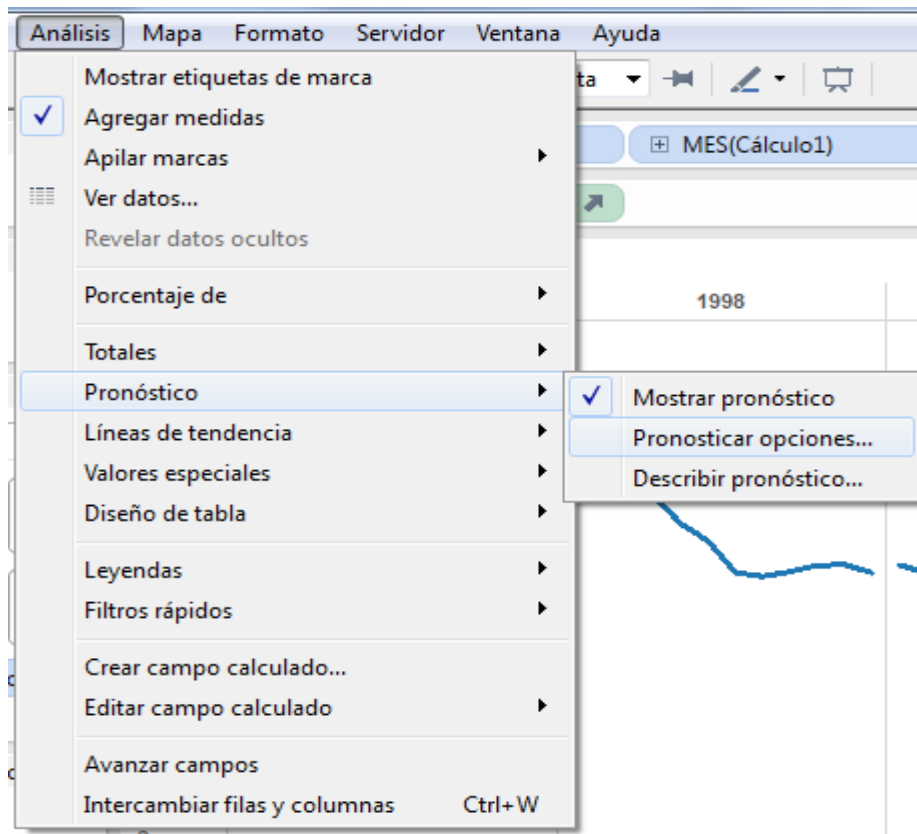


Figura 137.- Tableau Image 13

Then, opens the following menu (Figure 138):

**Opciones de pronóstico**

**Longitud del pronóstico**

☒ Auto Sigüientes 12 meses

☐ Exactamente 1 Años

☐ Hasta 1 Años

**Fuente de datos**

Agregar por: Automático (Meses)

Ignorar último: 0 Meses

☒ Completar los valores que faltan con ceros

**Modelo de pronóstico**

Automático

Selecciona automáticamente un modelo de suavizado exponencial para datos que pueden tener una tendencia y que pueden tener un patrón de temporada.

☒ Mostrar 95% intervalos de predicción

Actualmente se están usando datos de fuente de ene 1997 a dic 2002 para crear un pronóstico a través de dic 2003. Buscar patrones de temporada potenciales cada 12 Meses.

[Obtenga más información sobre las opciones de pronóstico](#)

Aceptar

Figura 138.- Tableau Image 14

Ignore cases = 0

Select exactly value needed.

Final result is the following (Figure 139).

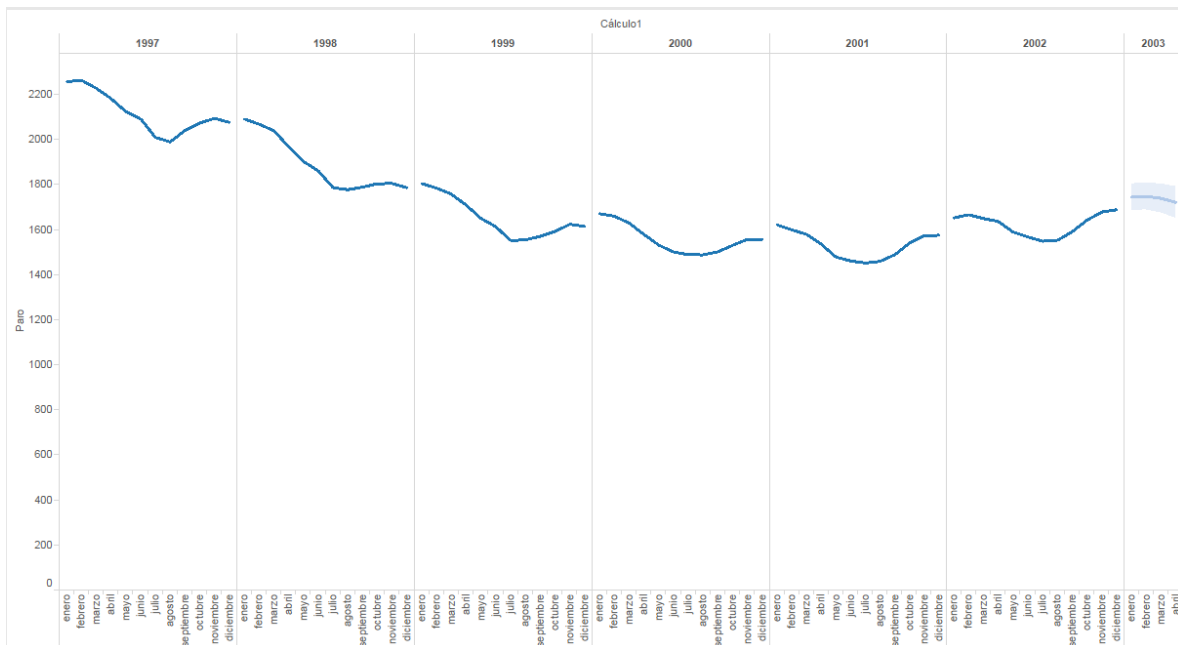


Figura 139.- Tableau Image 15

## 5. - Getting data.

Selecting data forecasted in graph above (Figure 140).

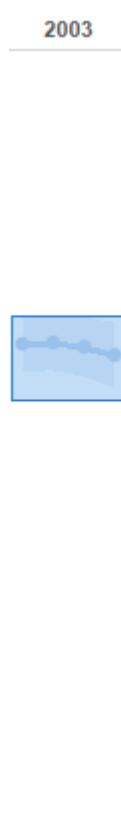


Figura 140.- Tableau Image 16

Right button, select data and following menu appears (Figure 144), and then export data:

Año de Cálculo1	Indicador de pronóstico	Mes de Cálculo1	Paro
2003	Cálculo	enero	1.744,4562
2003	Cálculo	febrero	1.747,5826
2003	Cálculo	marzo	1.738,3848
2003	Cálculo	abril	1.720,3579

Figura 141.- Tableau Image 17

### AI15.- Time series study using WEKA.

Weka is a tool oriented to Data Mining distributed under openGL license (free software).

As first step, after install it is necessary activate plugin for forecast. It is made directly in editor (although is possible make it through commands)

After open this program, select in the tool bar, menu tolls, and after select “Package Manager”.

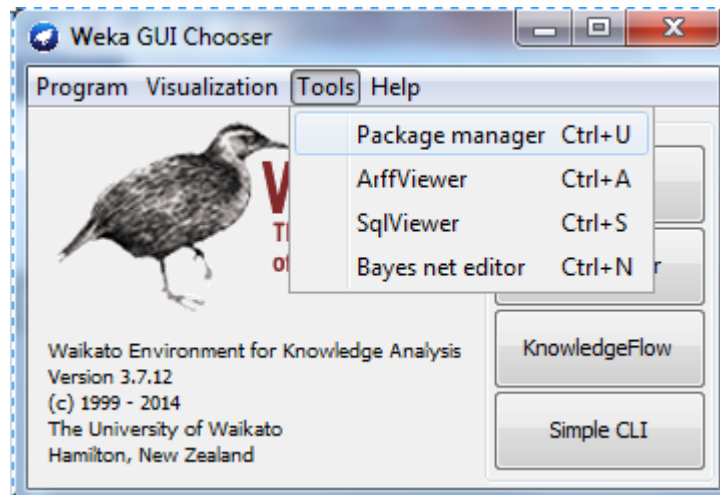


Figura 142.- Weka Image 1

After that, select package “TimeSeriesForecasting”, then click on “install”.

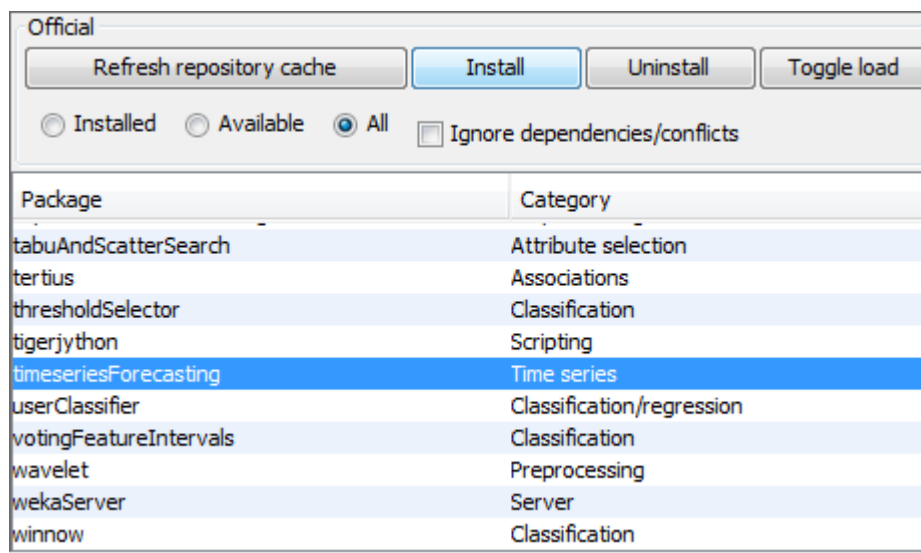


Figura 143.- Weka Image 2

At this point, is possible start the time series study.



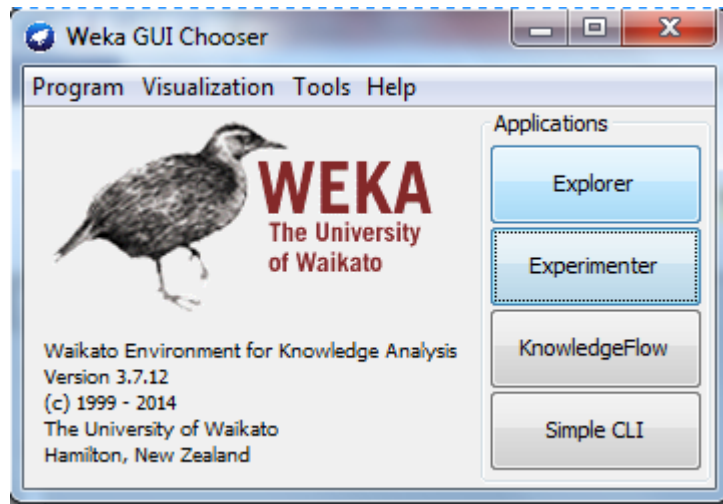


Figura 144.- Weka Image 3

### 1.- Loading Data

In this example, data are located into one file, although is possible get it from data base o even an URL. Is important notice that Weka doesn't admit every kind of files, it is recommended use .csv.

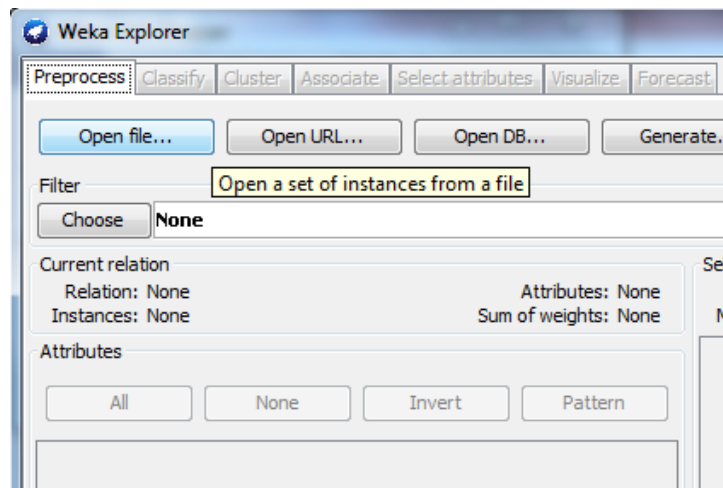


Figura 145.- Weka Image 4

To carry out this study, select in the tool bar menu Forecast, then select "Variable to treat", as second step, select periodicity, if don't know dates when data were recorded, select on the menu "use an artificial time index" and periodicity "monthly". As last step select number of unit to predict and click button START to start study.

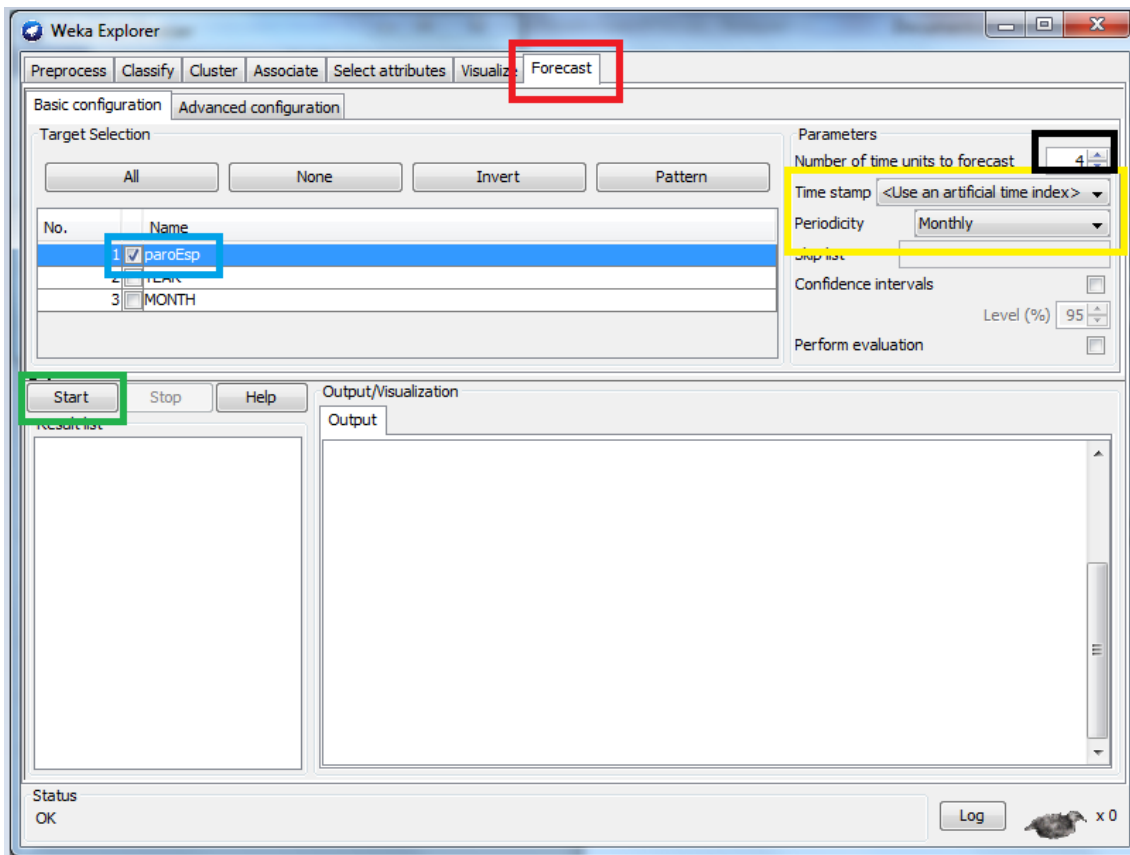


Figura 146.- Weka Image 5

Result is the following (Figure 147):

Output	Train future pred.
48	1598.4
49	1620.7
50	1598.9
51	1578.5
52	1535.1
53	1478.1
54	1460.6
55	1451.5
56	1459
57	1488.6
58	1540
59	1572.8
60	1574.8
61	1651.7
62	1666
63	1649
64	1636.3
65	1589
66	1567.4
67	1548.4
68	1552
69	1590.3
70	1641.7
71	1678
72	1688.1
73*	1719.0274
74*	1731.5777
75*	1718.6447
76*	1699.6007

Figura 147.- Weka Image 6

Graphically:

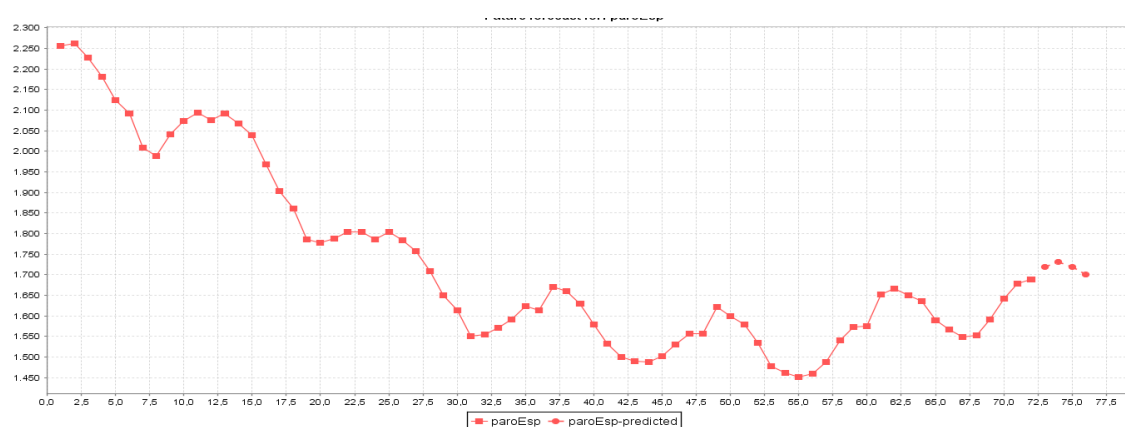


Figura 148.- Weka Image 7



### AIII.- Datos de estudio. Series temporales de pernoctaciones en hoteles rurales por provincias.

#### Álava

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	2000	1743	3287	4265	3674	3771	4728	9656	4246	6965	3075	5529
2011	1578	2257	3064	5713	3502	3650	5612	8632	3078	4069	2756	3452
2012	2309	2297	3358	6164	4234	3720	5454	7240	3863	4991	3218	3346
2013	851	1625	4139	2610	3450	3696	4799	7299	3444	4249	3287	3581
2014	1369	1248	2378	5122	4218	3517	4484	5839	2775	2961	2615	3136
2015	1439	1402	2887	5162	4332							

#### Albacete

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	5580	5021	9388	14402	7667	7602	12160	29578	7439	12107	7298	20371
2011	6036	5220	9449	14838	5978	4876	9466	27979	6708	10537	6561	9300
2012	2474	3511	9377	14146	4817	4602	6783	27791	9272	11258	11935	17226
2013	1847	2519	8974	3903	4341	4097	8674	19896	5753	7402	8212	9325
2014	3975	3960	5397	10731	6346	7259	12777	28424	6344	9068	8837	15260
2015	6305	5869	7043	12941	7049							

#### Alicante

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	5137	1525	3754	5275	3264	5087	8072	11411	2524	5974	2223	9923
2011	3253	3092	2465	5938	3272	5596	7764	12454	2868	3473	2197	5140
2012	2517	1316	2163	6261	3323	3332	6479	8376	2686	4120	3405	7932
2013	1714	1297	4768	2649	2736	4659	8509	11714	3292	2961	3070	6780
2014	2427	2926	3534	5425	4156	5350	7996	8718	5138	5965	4139	6041
2015	3809	2838	4807	5782	4858							

#### Almería

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	912	980	1235	1726	1478	2088	3536	7527	1885	2153	1117	2577
2011	1268	821	1247	1523	1941	2445	3329	5877	1665	1542	1118	2245
2012	1174	1648	942	2077	997	1565	1556	7014	1896	2343	1334	2476
2013	616	885	2022	1144	1182	1706	3400	8979	2083	1185	1719	2069
2014	874	1486	815	2222	1353	1635	3828	6749	2149	1696	1243	2016
2015	1279	997	1775	2010	1416							

**Asturias**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	11572	12873	23672	44365	19560	28575	115258	228578	52230	44114	8948	35539
2011	9928	9848	12670	50602	24177	42461	138293	213817	56001	32559	14897	21424
2012	9736	14299	13736	50177	22905	28122	92317	196212	47962	36916	23591	23284
2013	9872	8044	31466	11639	18170	26626	88650	186434	36899	18920	13470	19313
2014	7325	8396	13142	44844	32963	39719	115866	204668	57983	25465	17703	26149
2015	7502	10259	19033	42247	26868							

**Ávila**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	13373	14463	21763	23399	19537	16930	24458	49365	16956	33476	14976	31874
2011	10731	16881	15987	30199	15198	22742	38151	50719	15993	24141	16769	25209
2012	9604	10729	16705	28329	15488	16816	31963	43919	12028	23933	18701	27017
2013	10238	10228	28820	13806	15989	19631	29999	43064	15469	17676	20013	24500
2014	10318	11700	12526	26592	19817	19354	35749	43016	17715	20814	18210	28130
2015	14860	17483	17537	27666	22566							

**Badajoz**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	1007	1576	1593	2863	1788	1895	1492	3119	1172	2795	778	1766
2011	713	1131	1466	2414	1254	1576	2028	3598	1676	1736	1146	2111
2012	1392	1318	1902	4195	1571	1366	2637	3784	2486	3249	2562	3586
2013	1690	2108	3342	2226	2376	2328	3691	4101	2810	2510	2970	3692
2014	1320	2460	2209	4987	3240	2937	3844	6875	3368	3643	2407	4336
2015	1792	2467	3353	4520	3780							

**Baleares**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	1805	2875	5769	9150	15251	18523	22788	35668	19686	12261	2988	3934
2011	2889	3680	7434	14862	16703	22978	29985	38826	24220	14297	2729	2483
2012	2354	3834	7821	15994	21454	24621	34420	45513	29616	16739	4090	2765
2013	1630	3659	8932	13756	20923	25151	30946	38561	27759	17663	3121	2473
2014	1886	3686	7583	16436	21396	28317	32266	45344	31021	19855	4221	3123
2015	2837	4465	8627	18244	25819							

## Barcelona

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	8666	8843	13078	18177	15623	20150	17602	35148	13685	19863	10810	26098
2011	5232	11096	14814	21748	17979	23486	29396	39775	15414	19995	15975	26951
2012	5536	8555	8810	18831	11207	16214	18479	29491	11134	14806	14370	18967
2013	3865	6832	15369	8029	12121	17578	18228	30184	13075	14622	14713	19735
2014	8344	7063	10708	16732	13908	20381	22602	35124	17884	13928	15604	23706
2015	7725	9306	10930	19801	19785							

## Bizkaia

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	2622	3327	5376	6730	5621	5498	8623	15727	6199	6855	3777	6322
2011	2276	2528	5181	8606	4756	6552	11283	19157	5649	4710	5109	5331
2012	2272	2305	3518	7555	4340	5546	11229	19764	6045	5453	3766	4816
2013	2512	2093	5761	3600	4702	5632	12035	19986	6462	5225	4228	5604
2014	2607	2250	3242	6986	6825	6882	13120	15369	5360	4196	3029	4842
2015	1731	1986	4170	7063	6056							

## Burgos

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	3785	6517	11938	15740	13171	14248	21290	30745	9617	18565	9534	13574
2011	4046	5681	7439	16773	9913	12382	21550	30394	10892	18366	9980	10750
2012	6014	5186	9015	19780	10846	9376	16211	25649	12031	18731	12646	14576
2013	2981	4224	16533	10392	9409	9051	13886	22689	8429	10224	10220	12126
2014	3254	3821	7982	17613	13409	10460	14123	27794	12502	14118	12542	14537
2015	2332	3440	5498	18240	14380							

## Cáceres

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	8320	11303	22983	33818	12901	12980	19720	30682	13648	25709	11384	13765
2011	7627	11884	19101	36534	13329	14750	23608	38779	16668	18942	14124	16505
2012	9009	11789	17538	37751	11325	13656	19082	38095	11920	17482	14877	16785
2013	5285	8274	24548	19080	15993	14934	19122	37305	16312	14605	16920	19921
2014	7890	10996	16522	30542	19605	17131	23370	50541	17031	17764	14285	22444
2015	9152	13717	22525	30271	20914							

## Cádiz

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	2616	2349	2546	3150	2078	2463	5655	10405	3803	4083	2411	4419
2011	1722	3067	1829	5167	2908	3011	6066	10719	4300	3177	2944	3359
2012	1252	2181	2255	4872	3219	4215	5944	9841	4777	4359	2400	3909
2013	961	1672	3765	2935	3442	3454	6206	10482	3452	3757	3632	4193
2014	1652	3276	3473	3578	3519	3463	6712	10897	4002	3322	2683	3280
2015	2559	2481	2665	4457	4130							

## Cantabria

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	9873	13812	28122	34233	24971	32129	70491	134664	41929	32491	11804	23593
2011	8012	13454	18092	31943	24087	33954	88010	116683	49616	26158	13545	18140
2012	9009	13296	13802	38107	24432	28981	68930	110134	39864	22487	20858	15312
2013	4709	7919	26534	12684	18973	24963	60464	114295	34876	19535	15373	14156
2014	8426	8908	11304	31777	24399	26539	65313	119340	39720	21562	14885	16123
2015	9394	12674	19839	31252	27042							

## Castellón

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	8750	5604	7713	11552	6408	8428	14790	24451	6931	14793	5989	16890
2011	6256	4654	6813	16290	6304	7074	12594	28857	6105	7376	5153	13784
2012	3582	3820	5616	11185	5070	4361	13976	22005	3162	7328	4178	9953
2013	2355	3851	9714	4377	4824	7254	10160	21395	5421	5751	7270	10758
2014	2757	4298	5866	12913	5909	5616	12185	22016	6047	7820	5829	11026
2015	6378	4392	7253	13830	7453							

## Ciudad Real

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	2275	4154	6461	7510	5872	4892	6957	8345	6020	6162	4262	8089
2011	2103	3656	2386	5596	3280	4444	4963	9050	2487	5019	2578	4633
2012	1848	2217	2106	3977	2489	2660	5161	5942	3120	3857	3801	3830
2013	1484	1759	5051	3249	3533	4879	4650	8166	4410	4273	5185	6696
2014	2039	2508	2965	5361	3951	4910	6975	9532	3405	5214	5404	6592
2015	3651	4099	3453	5311	5118							



**Córdoba**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	4263	2889	2835	4055	2557	4070	6127	13938	4687	6591	3550	9020
2011	1978	3821	1932	6036	3134	2191	7810	10399	2871	5930	2694	7662
2012	2870	5291	5350	5019	2667	3246	3942	8321	4114	4613	2854	10140
2013	2155	2646	6447	3757	4506	4545	5474	9656	3897	3589	5065	8455
2014	1358	2926	3486	7669	5539	3542	7596	10891	4340	4572	5193	8777
2015	1493	2690	2322	4598	4473							

**Coruña**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	2628	2848	5214	8113	8577	10150	19557	31110	13009	12555	4969	6130
2011	1404	1619	2132	6491	4740	7170	10695	26448	5828	5260	2103	3642
2012	1837	1510	2038	7268	6135	5751	12111	23676	8238	5125	2531	4072
2013	841	2002	4349	2616	4428	6880	12831	23884	9031	4629	3657	4054
2014	1431	1707	3788	6696	7922	8158	16972	29196	11891	5935	2826	3791
2015	1481	2402	5217	8491	8445							

**Cuenca**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	3792	4492	6589	10412	6883	4856	6823	5757	4747	10727	5316	11125
2011	3831	3436	5740	10770	6018	5600	6550	10648	4294	8530	6507	8909
2012	2639	5826	6397	16835	4767	5790	6552	12056	4682	7659	7448	13466
2013	4249	4568	14786	6094	7300	7156	9529	16426	5776	7321	10469	11436
2014	3827	5749	6358	11095	6525	6549	5487	13242	5796	9388	9427	13428
2015	4415	4015	7171	11984	8974							

**Guipúzcoa**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	3643	5460	8826	11391	8373	9652	20865	35834	13943	12299	5435	10482
2011	3128	5428	7824	15320	10463	11775	26006	38004	15229	9912	5426	8967
2012	3809	4334	7268	15155	7506	9899	25430	36619	13368	9558	7012	5781
2013	3436	4199	13246	7669	9069	9215	22909	35653	13193	7336	7756	6862
2014	3888	4519	6602	12888	10194	11486	27205	35889	13544	8903	6878	7271
2015	4433	5339	8866	13292	12617							

**Girona**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	16939	11504	16079	24155	19444	28903	36502	63644	23583	25174	17961	27586
2011	11386	12387	15720	28150	18838	28462	34768	69869	19680	22079	17455	31622
2012	13910	11335	17027	25951	19520	21103	38029	70958	24070	21584	16349	32864
2013	7626	8407	22026	11029	18466	22352	26341	56455	20168	18340	18129	24469
2014	8307	11333	13567	23862	21560	30608	34347	64042	25960	20909	20292	33696
2015	11561	15707	19066	30050	29968							

**Granada**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	1716	2200	2440	4303	1772	1442	3940	7439	2864	2559	1418	4869
2011	2570	3962	2603	4805	2103	1804	4595	10067	3607	3428	1899	4853
2012	3274	2754	2242	4063	2631	2976	5158	7281	3515	3940	2457	4911
2013	1582	2190	4248	2146	2687	2552	4222	6107	3167	2743	2469	4459
2014	1543	2976	1838	4333	3248	3427	3954	8003	3419	3791	3686	6462
2015	1896	3483	3061	4705	3470							

**Guadalajara**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	4230	4078	7829	7873	6308	6681	6742	8594	3165	7987	5543	10418
2011	3129	4704	5127	9109	7508	8928	8366	5451	3763	8184	8707	6564
2012	2336	3044	6782	9504	4762	6277	9766	6730	3869	6623	5844	8004
2013	1908	2959	5551	3130	3609	5647	11251	10551	3717	7284	7891	6408
2014	2135	2785	4565	7304	5522	5528	8899	8750	3798	7623	7005	6323
2015	2243	2755	5654	7950	7804							

**Huelva**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	2244	2878	1833	2225	1247	628	2137	5256	2872	3702	2329	5635
2011	2643	2333	1795	3144	1654	1279	2338	6412	1398	3900	2682	4957
2012	1370	2411	2148	2396	814	2038	4379	8399	2100	4206	3678	5040
2013	1935	2084	3006	2279	2291	2535	5637	7099	2645	3507	5490	6378
2014	2801	3032	2548	4125	2131	2210	3613	6891	4131	5757	5583	8006
2015	3493	4928	4390	4877	4116							

**Huesca**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	12341	7328	13084	17940	8303	15395	29237	64627	13565	19034	5261	25264
2011	11999	10200	10148	19960	8361	15593	30065	59781	10179	14297	6901	19393
2012	11100	6927	10096	20183	6699	10031	25001	48053	8093	11164	6499	17735
2013	6152	6342	19621	6272	8552	12149	24638	45819	10101	7623	7579	17278
2014	6481	6348	6994	18151	11292	12730	24062	50402	11973	9882	6251	20948
2015	10664	8357	11946	20271	13118							

**Jaén**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	3148	1311	1903	4158	1881	2703	2764	8800	2785	3020	1473	6389
2011	2643	2305	1475	4568	739	1427	3871	6590	2379	2697	2043	4325
2012	2731	1428	2683	3701	1817	1626	3577	9105	2149	4023	3275	4648
2013	1155	1433	3791	2905	5110	3849	9150	7835	2510	2735	2788	3723
2014	1391	1676	1492	4253	2496	2405	4409	10008	3581	4206	3356	5489
2015	1820	2839	2597	5110	3297							

**La Rioja**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	4060	1691	5079	5965	4391	3497	4101	9572	4329	7586	4519	7231
2011	1556	3049	3319	7071	4262	5230	6190	12784	4997	6420	5150	5825
2012	3720	2719	4794	8250	4537	4132	5684	9526	4947	6331	4546	7193
2013	1621	2186	6216	3360	4344	5137	6098	10242	4758	5290	4533	6030
2014	1459	1493	2573	5563	4506	3295	4560	9799	3816	5346	5120	6240
2015	2763	2212	3958	7321	4889							

**Las Palmas**

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	3684	2562	3002	2617	1476	1368	2773	4628	2793	2726	1944	3954
2011	2732	3298	2446	2444	2239	1790	2201	4998	2529	3286	2574	2463
2012	2163	2326	2020	2396	1963	2014	2141	4997	2330	3653	2452	2943
2013	2626	2477	3116	2054	1683	1360	2054	4230	2387	2383	3067	3830
2014	2935	2794	2691	2699	2014	1712	3270	5057	2900	3256	2907	2891
2015	3088	3420	2889	2640	2245							

## León

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	8954	9022	13429	15845	12345	12865	18504	40752	10842	14762	7208	14002
2011	5756	7902	6263	17498	12589	13492	20787	38736	13570	15409	6834	10590
2012	5599	5787	6721	20190	11250	11733	19089	34932	13013	12263	6378	12395
2013	5045	5475	13833	8467	13085	12600	17961	32658	13755	11015	8648	10257
2014	3443	4950	5799	15509	14896	11885	20565	43659	18818	13449	9122	11006
2015	3584	4560	9274	15547	12117							

## Lleida

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	10649	10675	11641	15137	9240	12381	23512	57465	11194	19704	10277	27079
2011	8710	7001	10488	14050	5433	12977	19941	49443	11884	11378	7546	21596
2012	6706	8739	6077	15812	8044	10106	19210	42357	11845	10596	10289	22127
2013	5796	6364	13315	5631	7592	9723	14005	37173	8132	8189	9045	16251
2014	6132	8028	8121	14190	7610	12442	18653	45508	10766	10015	8208	18747
2015	7773	8664	7775	14385	11739							

## Lugo

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	1249	1449	4607	6189	8246	7131	12441	22841	9242	8415	3808	5201
2011	1100	577	2827	4886	4766	5209	10529	19638	6886	6723	3466	3278
2012	1382	1514	2133	5904	4180	4620	8541	16333	8265	7199	3301	1719
2013	610	1062	3575	2238	4498	4515	8426	20021	6824	4131	2730	2394
2014	795	977	1797	5777	6022	6734	11013	22243	8591	4589	2537	2737
2015	1209	2137	3230	5195	6674							

## Madrid

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	10169	14642	14183	13845	12903	15498	17905	25331	14384	19187	14106	23717
2011	12063	15181	14312	17942	16387	16985	20171	29286	19086	19684	15762	18887
2012	13561	16112	18176	25263	15985	19323	23927	26795	17713	17642	15717	19970
2013	10895	12062	20697	12059	14319	17575	19455	23366	15038	15195	14992	17100
2014	12267	13215	13572	18169	18016	17146	20524	25947	15478	12264	12078	17141
2015	11013	12603	13324	18013	15596							

## Málaga

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	7962	9902	5750	5938	6204	9217	14280	29889	7927	9883	5315	14126
2011	5131	6467	3786	5298	3309	6209	12582	30159	8786	6728	5329	7596
2012	4946	7250	5672	8097	6029	9069	11673	18894	4261	4569	4250	9859
2013	3707	3216	6558	5499	5692	7301	11992	16325	8116	6040	6388	11582
2014	4482	6036	6743	9499	7943	9962	14285	20963	10033	8843	7575	13482
2015	6564	6648	6939	10223	9218							

## Murcia

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	3973	3437	5567	5263	3185	4144	11054	18776	4382	7755	4792	10950
2011	2092	5338	2066	9548	2962	3191	11448	16231	2850	5476	3419	6998
2012	7473	4220	5760	8812	3220	5197	10225	24947	5574	6008	7004	13702
2013	2730	2683	7056	2618	2548	3766	5159	13463	3358	3049	4740	7204
2014	2254	3339	2948	5189	2967	3602	7212	14604	4068	4264	4010	7336
2015	4791	3035	4050	6555	4425							

## Navarra

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	12635	13627	24794	33383	16363	21267	43320	76086	21751	32466	13746	36220
2011	8775	10699	18438	41917	21789	21183	42687	81064	23976	27776	16797	32213
2012	8071	10340	16063	37496	16331	16680	38594	70195	16044	23963	17320	24214
2013	7060	6033	26309	12231	14038	16692	30282	60608	17409	19536	15126	23092
2014	5638	5575	9001	30801	22252	15675	29871	62022	20166	20665	16974	25993
2015	9331	10024	16969	33039	21403							

## Ourense

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	1902	2274	3212	4761	4364	4099	5585	10264	4273	5549	3612	4179
2011	1604	1563	2707	4629	2762	3490	5956	8710	4705	4189	2973	2527
2012	1332	1727	2074	3858	2137	3139	3843	7747	3033	3365	2064	3033
2013	1468	2229	3737	1672	2881	2032	3877	7824	4011	3083	3604	2896
2014	1830	2213	3119	3698	3303	2816	5031	9350	3486	2928	2571	2905
2015	1785	2057	2676	3262	3038							

### Palencia

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	3784	5088	5358	8010	6899	5766	7366	23861	7373	9689	4476	8212
2011	2210	1828	3125	8916	4281	6163	7859	18811	5391	6803	4636	4920
2012	1921	3182	3276	11150	5411	4078	9007	15866	4117	6218	4301	5779
2013	1052	2143	6486	3486	5129	4948	7537	15557	4068	4105	3387	5012
2014	1006	1809	1988	6325	4568	4502	7232	14582	4887	6061	4277	5654
2015	2746	2780	3438	6225	4784							

### Pontevedra

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	2160	1990	4355	7185	5682	7169	15778	27967	8262	7989	3414	5326
2011	1179	1875	2337	5494	4661	6064	12156	28571	8180	5153	2576	4971
2012	649	1791	2862	5080	4237	5445	9208	19168	5417	5292	2810	4248
2013	861	1301	3626	1547	3083	4277	9882	21735	6609	3086	2030	2668
2014	784	1338	2105	4130	3714	4886	10955	21773	8101	4127	2186	3402
2015	1029	1569	2509	5025	5217							

### Salamanca

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	8292	8335	15796	16312	10436	8289	13446	21259	7995	13450	5367	18550
2011	6114	7741	11031	17769	8486	9207	14755	31467	8260	11296	4967	11515
2012	4983	10755	9122	21404	6977	8414	15189	33233	6247	8648	7834	10047
2013	5015	7601	17778	9634	12617	9762	14595	23753	9554	9706	9927	16171
2014	6800	8739	11628	17963	12007	9337	18206	37493	10807	12919	10911	19410
2015	6337	11289	12298	18273	13636							

### Santa Cruz

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	5324	4646	5334	5950	4334	3722	12891	15868	4666	5899	4688	7495
2011	6208	5262	6191	6078	5219	4725	7547	14001	6767	6290	6139	6750
2012	5272	6671	7003	8647	7063	5247	7559	13675	6561	6485	7061	9052
2013	6401	6358	7472	4873	4719	4320	6930	9096	5717	6039	6439	6725
2014	6840	6637	7949	6302	4707	4268	6144	11773	7865	6964	7171	7963
2015	8365	9412	9627	9498	6982							

## Segovia

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	8403	11320	16516	17466	13039	12058	15029	19408	11711	20030	11200	23686
2011	5567	11825	15101	21046	16796	15043	17336	22431	10813	18519	13474	22560
2012	7502	9909	13822	25151	13831	11668	17797	20584	12038	21205	16944	20183
2013	5457	10012	18512	10388	11995	13175	11953	18778	11980	15767	16576	22633
2014	5974	9437	11361	16583	16354	13777	14144	20823	13283	18951	19466	23624
2015	7972	11256	14012	21836	17673							

## Sevilla

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	840	1099	1854	1846	1228	1981	3452	4100	1498	1609	1116	2970
2011	1179	1703	1310	1492	1032	894	3390	3036	1200	1745	1430	3888
2012	1791	1088	2151	3132	1231	1950	3240	5977	1983	4079	3812	7055
2013	1578	2719	4375	2849	3171	3008	4918	7784	3746	3309	3516	5499
2014	2241	2235	3425	3474	3810	3042	6089	8904	4938	4127	3050	5226
2015	2477	2634	2537	3350	3159							

## Soria

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	3312	4893	11801	15224	9796	10753	12942	31656	9378	20788	9804	14265
2011	2580	4341	5735	18320	8810	11744	15298	35128	9104	16349	7838	13160
2012	3800	3256	7446	23740	7452	8782	15464	28718	9538	15109	10731	13726
2013	2402	3620	15619	6018	9069	9956	13354	26325	8581	13453	11396	9902
2014	2527	4215	6019	14971	11532	8490	13798	25750	8447	15932	10931	11198
2015	5190	4859	8194	17641	11665							

## Tarragona

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	2697	4880	5606	8979	4211	5266	7097	12840	6152	6791	3529	7017
2011	2215	3710	4542	10007	5347	8498	9522	15924	3994	5492	5108	6776
2012	1836	2887	4613	9596	6004	5266	5940	14055	5190	5945	2604	4458
2013	1339	1015	6538	2734	3655	6265	5562	10755	3738	2683	2932	3681
2014	1348	2094	3727	7967	6111	6377	7957	12498	5567	3931	4094	7501
2015	2257	1951	2968	6730	5360							

### Teruel

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	5085	2397	4337	10004	3815	2290	8302	18044	6121	8042	2081	9726
2011	2555	3178	2753	7502	4677	5606	8035	24567	5278	7809	4178	9637
2012	3771	5034	5238	11094	6176	3539	9506	22702	5682	10136	6155	12151
2013	3970	4318	12447	4598	4076	4277	9221	22674	6052	5922	5085	10185
2014	3268	4687	3331	10271	5414	3712	8426	22290	5778	6130	3889	9933
2015	5590	4177	5788	11460	6519							

### Toledo

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	1915	2487	4084	5759	4856	5541	4487	5575	4679	4892	3941	7006
2011	1751	1842	4348	4392	5569	7050	7066	7843	4465	5588	5066	7869
2012	2860	3528	3942	7711	5680	5363	5975	8528	3111	7861	5968	7995
2013	2031	2367	5855	4320	3992	5383	5865	9343	4281	4846	4532	6841
2014	1558	1602	3070	5948	5581	8565	16058	7430	5140	4968	4837	8594
2015	2549	2747	5823	7161	7746							

### Valencia

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	3584	2956	7446	15298	10411	13960	21467	23430	6046	6862	4372	8920
2011	4776	3838	6591	12855	13749	14200	23377	27232	6404	8091	9744	9148
2012	3169	3292	6091	13419	11717	16855	16233	20682	6291	5032	4764	6967
2013	2498	3025	10038	8770	14328	18166	19553	22768	6335	7050	6995	9390
2014	2000	1928	4290	18593	17215	19415	22214	19894	4715	5261	4927	9078
2015	3831	2745	5313	15692	14102							

### Valladolid

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	3375	2374	4571	6764	4915	5273	7264	9537	6522	6997	4538	7579
2011	2980	3104	5368	8196	5101	5771	5684	9721	5260	7494	5432	7351
2012	4016	3289	6225	7534	4261	4476	6140	6325	4943	6092	4106	6171
2013	1686	1750	6173	3525	4203	3520	4804	6563	4670	5296	5478	6262
2014	1805	2315	3567	6134	4695	4073	5550	8423	5081	5994	5186	6179
2015	3074	2866	4454	7719	5383							



### Zamora

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	2871	3016	7269	11539	7522	7507	10801	21340	9750	10477	6006	6632
2011	3031	5182	7777	10497	9818	10042	15529	24919	10958	8961	7046	7697
2012	3930	3679	7000	14960	7065	6103	13313	24203	7659	9951	5786	6251
2013	3329	3723	10938	5919	8368	8808	12148	22526	8547	7764	7170	7503
2014	4077	3474	5999	10989	8637	8493	11921	21491	6323	7149	6002	6988
2015	2935	3797	7012	9530	6956							

### Zaragoza

	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
2010	3343	2799	4749	6608	3796	3639	4918	8646	4903	7977	3090	7678
2011	1925	3318	3711	7472	4465	4995	5090	9164	4373	6725	4104	7008
2012	2462	2238	3683	8547	4859	5249	4457	7185	4568	6361	4823	6866
2013	1606	2046	6986	3605	4249	5018	4921	8932	4495	4582	5335	6587
2014	1199	1503	3256	7094	5830	4235	4437	8378	4006	4200	3646	7384
2015	3119	2466	4887	7690	7381							



## Bibliografía

- Ajenstat, F. (2014). *www.betterbuys.com*. Recuperado el 06 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Ajuntament de Barcelona. (Noviembre de 2012). *bcn.cat*. Recuperado el 09 de 05 de 2015, de [http://w27.bcn.cat/porta22/images/es/Barcelona\\_treball\\_capsula\\_sectorial\\_telecomunicacions\\_y\\_TIC\\_noviembre2012\\_es\\_tcm24-22920.pdf](http://w27.bcn.cat/porta22/images/es/Barcelona_treball_capsula_sectorial_telecomunicacions_y_TIC_noviembre2012_es_tcm24-22920.pdf)
- annalect. (23 de 05 de 2013). *www.annalect.com*. Recuperado el 26 de 05 de 2015, de <http://www.annalect.com/data-is-the-fuel-that-powers-business-intelligence/>
- Bhudia, D. (2014). *www.betterbuys.com*. Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Biskup, T. (2014). *www.betterbuys.com*. Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- BMN. (23 de 02 de 2015). *BMN.es*. Recuperado el 07 de 05 de 2015, de <https://www.dropbox.com/s/v eb7hv4dx dzrc2o/Leading%20Spanish%20Bank%20Banco%20Mare%20Nostrum%20Implements%20MicroStrategy%20to%20Drive%20Greater%20Analysis%20and%20Insight.docx?dl=0>
- Business Analytics. (15 de 03 de 2013). <http://www.businessanalytics.com>. Recuperado el 19 de 05 de 2015, de <http://www.businessanalytics.com/difference-between-business-analytics-and-business-intelligence/>
- Business and Technology. (08 de 09 de 2013). *lmax.com*. Recuperado el 06 de 05 de 2015, de <http://www.lmax.com/blog/business-and-technology/2013/04/08/latency-worth-it/>
- Business Dictionary. (2014). *www.businessdictionary.com*. Recuperado el 13 de 06 de 2015, de <http://www.businessdictionary.com/definition/decision-making.html>
- Chichilla, R. (2011). Mercados de datos. *Tecnología en Marcha*, 55-56.
- Cursor Software. (29 de 06 de 2015). <http://www.cursor-distribution.de>. Recuperado el 23 de 07 de 2015, de <http://www.cursor-distribution.de/en/portfolio-and-sales/sales-cognos/price-lists-cognos>
- Datamology. (2010). *www.datamology.com*. Recuperado el 03 de 06 de 2015, de <http://www.datamology.com/exploration.shtml>
- Denvens, R. M. (1865). *Cyclopædia of Commercial and Business Anecdotes*.
- Elliot, T. (09 de 03 de 2011). <http://timoelliott.com>. Recuperado el 19 de 05 de 2015, de <http://timoelliott.com/blog/2011/03/business-analytics-vs-business-intelligence.html>
- Elliot, T. (2014). *www.betterbuys.com*. Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>

- EscapadaRural. (2015). <http://www.escapadarural.com/>. Recuperado el 19 de 07 de 2015, de <http://www.escapadarural.com/>
- Esparza Catalán, C. (2001). *humanidades.cchs.csic.es*. Recuperado el 20 de 06 de 2015, de [http://humanidades.cchs.csic.es/cchs/web\\_UAE/tutoriales/PDF/SeriesTemporales.pdf](http://humanidades.cchs.csic.es/cchs/web_UAE/tutoriales/PDF/SeriesTemporales.pdf)
- Evelson, N. (2008). <https://www.forrester.com>. Recuperado el 13 de 06 de 2015, de <https://www.forrester.com/Topic+Overview+Business+Intelligence/fulltext/-/E-RES39218>
- Exelia. (18 de 11 de 2014). *www.exeliatech.com*. Recuperado el 26 de 05 de 2015, de <http://www.exeliatech.com/assets1/data-cleaning-in-the-etl-process/>
- García González, D. (25 de 04 de 2014). <http://necsia.es/>. Recuperado el 21 de 07 de 2015, de <http://necsia.es/big-data-vs-business-intelligence-complemento-o-sustituto/#>
- García, I. M. (2009). *Series Temporales*. Cáceres: Materiales UEX.
- Gartner. (2009). *Gartner Says Worldwide Business Intelligence and Analytics Software Market Grew 8 Percent in 2013*. STAMFORD: NewsRoom.
- Gartner. (2014). <http://www.gartner.com>. Recuperado el 21 de 07 de 2015, de <http://www.gartner.com/it-glossary/big-data>
- Hagerty, S. R. (2012). *www.gartner.com*. Recuperado el 13 de 06 de 2015, de <http://www.gartner.com/technology/reprints.do?id=1-1982NPD&ct=120208&st=sb>
- Hartzen, D. (2010). <http://infogoal.com>. Recuperado el 26 de 05 de 2015, de [http://infogoal.com/datawarehousing/data\\_sources.htm](http://infogoal.com/datawarehousing/data_sources.htm)
- IBM. (03 de 2012). <http://www-01.ibm.com>. Recuperado el 11 de 07 de 2015, de <http://www-01.ibm.com/software/es/analytics/spss/>
- IBM. (06 de 2014). <http://www-01.ibm.com>. Recuperado el 13 de 07 de 2015, de <http://www-01.ibm.com/software/es/analytics/cognos/>
- INE. (2002). *Estadística mensual Paro 1997 2002*.
- INE. (23 de 01 de 2014). *www.ine.es*. Recuperado el 09 de 07 de 2015
- INE. (05 de 2015). Encuesta de ocupación en alojamientos turísticos. 2015, 07, 18: INE.
- ine. (06 de 2015). <http://www.ine.es/>. Recuperado el 18 de 07 de 2015, de <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft11%2Fe162eotr&file=inebase&L=0>
- Jansen, C. (2014). *www.techopedia.com*. Recuperado el 02 de 06 de 2015, de <http://www.techopedia.com/definition/28789/data-exploration>
- Kotorov, R. (2014). *www.betterbuys.com*. Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>

- Lixto. (2005). *www.dbai.tuwien.ac.at*. Recuperado el 02 de 06 de 2015, de <http://www.dbai.tuwien.ac.at/proj/lixtto/WebBI.pdf>
- Major, R. (10 de 2014). *halobi.com*. Recuperado el 08 de 06 de 2015, de <http://halobi.com/2014/10/descriptive-predictive-and-prescriptive-analytics-explained/>
- MarketingDirecto. (25 de 08 de 2007). *Marketingdirecto.com*. Recuperado el 09 de 05 de 2015, de <http://www.marketingdirecto.com/actualidad/marketing/nestle-elige-cognos-8-bi-para-complementar-sus-solucion-de-reporting-sap-bw/>
- Martínez, S. (22 de 12 de 2014). *Mundo ERP*. Recuperado el 23 de 04 de 2015, de <http://mundoerp.com/blog/funcionamiento-entorno-business-intelligence-bi/>
- Mauricio, J. A. (11 de 11 de 2013). *www.ucm.es*. Recuperado el 14 de 06 de 2015, de <https://www.ucm.es/data/cont/docs/518-2013-11-11-JAM-IAST-Libro.pdf>
- Mearian. (2011). By 2020, there will be 5,200 GB of data for every person on Earth. *ComputerWorld*.
- Microsoft. (06 de 2014). <https://products.office.com/>. Recuperado el 11 de 07 de 2015, de <https://products.office.com/es-es/excel>
- Milton. (2011). *www.forbes.com*. Recuperado el 13 de 06 de 2015, de <http://www.forbes.com/sites/>
- minetur. (02 de 2014). <http://www.minetur.gob.es>. Recuperado el 09 de 07 de 2015, de [http://www.minetur.gob.es/turismo/es-ES/Novedades/Documents/005b\\_Plan%20Integral%20de%20Turismo%20Rural%202014.pdf](http://www.minetur.gob.es/turismo/es-ES/Novedades/Documents/005b_Plan%20Integral%20de%20Turismo%20Rural%202014.pdf)
- MySoftware. (12 de 04 de 2011). *MySoftware.wordpress*. Recuperado el 23 de 04 de 2015, de <https://mysoftwar.wordpress.com/2011/04/12/¿cual-es-la-principal-diferencia-entre-el-erp-y-crm/>
- Nelson, G. (2007). SAS. Recuperado el 23 de 04 de 2015, de <http://www2.sas.com/proceedings/forum2007/207-2007.pdf>
- Oracle. (2007). <http://docs.oracle.com>. Recuperado el 30 de 05 de 2015, de [http://docs.oracle.com/html/E10312\\_01/dm\\_concepts.htm](http://docs.oracle.com/html/E10312_01/dm_concepts.htm)
- rainmakerworks. (27 de 04 de 2012). *www.rainmakerworks.com*. Recuperado el 28 de 05 de 2015, de <http://datawarehouse4u.info/OLTP-vs-OLAP.html>
- Rijmenam, M. v. (2014). *www.betterbuys.com*. Recuperado el 06 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Roche, P. (2014). *www.betterbuys.com*. Recuperado el 06 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Rouse, M. (10 de 2014). <http://searchdatamanagement.techtarget.com>. Recuperado el 19 de 05 de 2015, de <http://searchdatamanagement.techtarget.com/definition/business-intelligence>
- Rouse, M. (04 de 2014). <http://searchsap.techtarget.com/>. Recuperado el 13 de 07 de 2015, de <http://searchsap.techtarget.com/definition/mySAP>

- SAP. (2014). *help.sap.com*. Recuperado el 11 de 06 de 2015, de [http://help.sap.com/saphelp\\_nw70/helpdata/EN/8d/2b4e3cb7f4d83ee10000000a114084/content.htm](http://help.sap.com/saphelp_nw70/helpdata/EN/8d/2b4e3cb7f4d83ee10000000a114084/content.htm)
- SAP. (13 de 02 de 2014). <http://go.sap.com/>. Recuperado el 12 de 07 de 2015, de <http://go.sap.com/hk/product/analytics/lumira.html>
- Sinnexus. (2012). <http://www.sinnexus.com>. Recuperado el 30 de 05 de 2015, de [http://www.sinnexus.com/business\\_intelligence/olap\\_avanzado.aspx](http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx)
- Sinnexus. (2012). *www.sinnexus.com*. Recuperado el 13 de 05 de 2015, de [http://www.sinnexus.com/business\\_intelligence/](http://www.sinnexus.com/business_intelligence/)
- Skirus, V. y. (2012). Business Intelligence in the Process of Decision Making: Changes and Trends. Lituania.
- Skyrius, V. y. (2012). Business Intelligence in the process of decision making. Lituania.
- Standen, J. (02 de 12 de 2008). <http://www.datamartist.com>. Recuperado el 19 de 05 de 2015, de <http://www.datamartist.com/data-warehouse-vs-data-mart>
- Steinbach, K. (08 de 05 de 2005). <http://www-users.cs.umn.edu>. Recuperado el 02 de 06 de 2015, de [http://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap3\\_data\\_exploration.pdf](http://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap3_data_exploration.pdf)
- stratebi. (2013). 2013. Recuperado el 07 de 06 de 2015, de [www.stratebi.es](http://www.stratebi.es): [http://www.stratebi.es/todobi/May13/Nuevas\\_Tendencias\\_BI.pdf](http://www.stratebi.es/todobi/May13/Nuevas_Tendencias_BI.pdf)
- stratebi. (05 de 2013). *www.stratebi.com*. Recuperado el 09 de 05 de 2015, de [http://www.stratebi.com/c/document\\_library/get\\_file?uuid=2e79e125-b78d-406d-9235-80cd5cbd87c2&groupId=10157](http://www.stratebi.com/c/document_library/get_file?uuid=2e79e125-b78d-406d-9235-80cd5cbd87c2&groupId=10157)
- tableau. (03 de 2014). <http://get.tableau.com/es-es/>. Recuperado el 12 de 07 de 2015, de <http://get.tableau.com/es-es/trial/tableau-software.html?cid=70160000000YEzj&ls=Paid%20Search&lsd=Google%20AdWords%20-%20Tableau%20-%20EMEA%20-%20ES%20-%20Free%20Trial&adgroup=Tableau%20-%20Exact&kw=tableau&adused=63871546582&distribution=search&gclid=Cj0>
- Tech-faq. (24 de 02 de 2011). <http://www.tech-faq.com>. Recuperado el 28 de 05 de 2015, de <http://www.tech-faq.com/data-warehouse.html>
- thearling. (2011). *www.thearling.com*. Recuperado el 03 de 06 de 2015, de <http://www.thearling.com/text/dmtechniques/dmtechniques.htm>
- Third Nature. (2010). <http://www.slideshare.net>. Recuperado el 21 de 07 de 2015, de <http://www.slideshare.net/oktopuslu/bi-comparison-of-open-source-and-traditional-vendor-4327259>
- tinko. (25 de 10 de 2012). <https://startdatawarehousingwithtinko.wordpress.com>. Recuperado el 19 de 05 de 2015, de <https://startdatawarehousingwithtinko.wordpress.com/author/vimalantinko/>

- Triad. (05 de 20105). <http://triadtechpartners.com>. Recuperado el 23 de 07 de 2015, de <http://triadtechpartners.com/wp-content/uploads/Tableau-GSA-Pricelist-May-2015.pdf>
- Universidad Pablo de Olavide. (27 de 03 de 2009). <http://www.entreculturas.uma.es>. Recuperado el 09 de 07 de 2015, de <http://www.entreculturas.uma.es/n1pdf/articulo24.pdf>
- UOC. (2012). <http://informatica.blogs.uoc.edu>. Recuperado el 13 de 06 de 2015, de <http://informatica.blogs.uoc.edu/2012/04/23/reducir-el-fracaso-en-los-proyectos-de-business-intelligence-ii/>
- UOP. (13 de 02 de 2014). Inteligencia de Negocio y Big Data. *Inteligencia de Negocio y Big Data: Negocio, Tecnología y Personas*. Madrid, Madrid, España: EIMT.OUC.EDU.
- Waikato. (05 de 2013). <http://www.cs.waikato.ac.nz>. Recuperado el 11 de 07 de 2015, de <http://www.cs.waikato.ac.nz/ml/weka/>
- WebMining Consultores. (13 de 03 de 2012). <http://www.webmining.cl>. Recuperado el 19 de 05 de 2015, de <http://www.webmining.cl/2012/03/business-analytics-versus-business-intelligence/>
- Xerox. (10 de 04 de 2015). [Xerox.com](http://www.xerox.com). Recuperado el 07 de 05 de 2015, de <http://www.xerox.es/servicios/transport-solutions/eses.html>
- Ajenstat, F. (2014). [www.betterbuys.com](http://www.betterbuys.com). Recuperado el 06 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- annalect. (23 de 05 de 2013). [www.annalect.com](http://www.annalect.com). Recuperado el 26 de 05 de 2015, de <http://www.annalect.com/data-is-the-fuel-that-powers-business-intelligence/>
- Bhudia, D. (2014). [www.betterbuys.com](http://www.betterbuys.com). Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Biskup, T. (2014). [www.betterbuys.com](http://www.betterbuys.com). Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Business Analytics. (15 de 03 de 2013). <http://www.businessanalytics.com>. Recuperado el 19 de 05 de 2015, de <http://www.businessanalytics.com/difference-between-business-analytics-and-business-intelligence/>
- Business Dictionary. (2014). [www.businessdictionary.com](http://www.businessdictionary.com). Recuperado el 13 de 06 de 2015, de <http://www.businessdictionary.com/definition/decision-making.html>
- Datamology. (2010). [www.datamology.com](http://www.datamology.com). Recuperado el 03 de 06 de 2015, de <http://www.datamology.com/exploration.shtml>
- Elliot, T. (09 de 03 de 2011). <http://timoelliott.com>. Recuperado el 19 de 05 de 2015, de <http://timoelliott.com/blog/2011/03/business-analytics-vs-business-intelligence.html>
- Elliot, T. (2014). [www.betterbuys.com](http://www.betterbuys.com). Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>

- Evelson, N. (2008). <https://www.forrester.com>. Recuperado el 13 de 06 de 2015, de <https://www.forrester.com/Topic+Overview+Business+Intelligence/fulltext/-/E-RES39218>
- Exelia. (18 de 11 de 2014). [www.exeliatech.com](http://www.exeliatech.com). Recuperado el 26 de 05 de 2015, de <http://www.exeliatech.com/assets1/data-cleaning-in-the-etl-process/>
- García González, D. (25 de 04 de 2014). <http://necsia.es/>. Recuperado el 21 de 07 de 2015, de <http://necsia.es/big-data-vs-business-intelligence-complemento-o-sustituto/#>
- Gartner. (2014). <http://www.gartner.com>. Recuperado el 21 de 07 de 2015, de <http://www.gartner.com/it-glossary/big-data>
- Hartzen, D. (2010). <http://infogoal.com>. Recuperado el 26 de 05 de 2015, de [http://infogoal.com/datawarehousing/data\\_sources.htm](http://infogoal.com/datawarehousing/data_sources.htm)
- Jansen, C. (2014). [www.techopedia.com](http://www.techopedia.com). Recuperado el 02 de 06 de 2015, de <http://www.techopedia.com/definition/28789/data-exploration>
- Kotorov, R. (2014). [www.betterbuys.com](http://www.betterbuys.com). Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Lixto. (2005). [www.dbai.tuwien.ac.at](http://www.dbai.tuwien.ac.at). Recuperado el 02 de 06 de 2015, de <http://www.dbai.tuwien.ac.at/proj/lixta/WebBI.pdf>
- Major, R. (10 de 2014). [halobi.com](http://halobi.com). Recuperado el 08 de 06 de 2015, de <http://halobi.com/2014/10/descriptive-predictive-and-prescriptive-analytics-explained/>
- Oracle. (2007). <http://docs.oracle.com>. Recuperado el 30 de 05 de 2015, de [http://docs.oracle.com/html/E10312\\_01/dm\\_concepts.htm](http://docs.oracle.com/html/E10312_01/dm_concepts.htm)
- rainmakerworks. (27 de 04 de 2012). [www.rainmakerworks.com](http://www.rainmakerworks.com). Recuperado el 28 de 05 de 2015, de <http://datawarehouse4u.info/OLTP-vs-OLAP.html>
- Rijmenam, M. v. (2014). [www.betterbuys.com](http://www.betterbuys.com). Recuperado el 06 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Roche, P. (2014). [www.betterbuys.com](http://www.betterbuys.com). Recuperado el 06 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Rouse, M. (10 de 2014). <http://searchdatamanagement.techtarget.com>. Recuperado el 19 de 05 de 2015, de <http://searchdatamanagement.techtarget.com/definition/business-intelligence>
- SAP. (2014). [help.sap.com](http://help.sap.com). Recuperado el 11 de 06 de 2015, de [http://help.sap.com/saphelp\\_nw70/helpdata/EN/8d/2b4e3cb7f4d83ee10000000a114084/content.htm](http://help.sap.com/saphelp_nw70/helpdata/EN/8d/2b4e3cb7f4d83ee10000000a114084/content.htm)
- Sinnexus. (2012). <http://www.sinnexus.com>. Recuperado el 30 de 05 de 2015, de [http://www.sinnexus.com/business\\_intelligence/olap\\_avanzado.aspx](http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx)



- Sinnexus. (2012). *www.sinnexus.com*. Recuperado el 13 de 05 de 2015, de [http://www.sinnexus.com/business\\_intelligence/](http://www.sinnexus.com/business_intelligence/)
- Skirus, V. y. (2012). *Business Intelligence in the Process of Decision Making: Changes and Trends*. Lituania.
- Standen, J. (02 de 12 de 2008). <http://www.datamartist.com>. Recuperado el 19 de 05 de 2015, de <http://www.datamartist.com/data-warehouse-vs-data-mart>
- Steinbach, K. (08 de 05 de 2005). <http://www-users.cs.umn.edu>. Recuperado el 02 de 06 de 2015, de [http://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap3\\_data\\_exploration.pdf](http://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap3_data_exploration.pdf)
- stratebi. (2013). *2013*. Recuperado el 07 de 06 de 2015, de [www.stratebi.es](http://www.stratebi.es): [http://www.stratebi.es/todobi/May13/Nuevas\\_Tendencias\\_BI.pdf](http://www.stratebi.es/todobi/May13/Nuevas_Tendencias_BI.pdf)
- Tech-faq. (24 de 02 de 2011). <http://www.tech-faq.com>. Recuperado el 28 de 05 de 2015, de <http://www.tech-faq.com/data-warehouse.html>
- thearling. (2011). *www.thearling.com*. Recuperado el 03 de 06 de 2015, de <http://www.thearling.com/text/dmtechniques/dmtechniques.htm>
- tinko. (25 de 10 de 2012). <https://startdatawarehousingwithtinko.wordpress.com>. Recuperado el 19 de 05 de 2015, de <https://startdatawarehousingwithtinko.wordpress.com/author/vimalantinko/>
- UOC. (2012). <http://informatica.blogs.uoc.edu>. Recuperado el 13 de 06 de 2015, de <http://informatica.blogs.uoc.edu/2012/04/23/reducir-el-fracaso-en-los-proyectos-de-business-intelligence-ii/>
- WebMining Consultores. (13 de 03 de 2012). <http://www.webmining.cl>. Recuperado el 19 de 05 de 2015, de <http://www.webmining.cl/2012/03/business-analytics-versus-business-intelligence/>
- Ajenstat, F. (2014). *www.betterbuys.com*. Recuperado el 06 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Ajuntament de Barcelona. (Noviembre de 2012). *bcn.cat*. Recuperado el 09 de 05 de 2015, de [http://w27.bcn.cat/porta22/images/es/Barcelona\\_treball\\_capsula\\_sectorial\\_telecomunicacions\\_y\\_TIC\\_noviembre2012\\_es\\_tcm24-22920.pdf](http://w27.bcn.cat/porta22/images/es/Barcelona_treball_capsula_sectorial_telecomunicacions_y_TIC_noviembre2012_es_tcm24-22920.pdf)
- annalect. (23 de 05 de 2013). *www.annalect.com*. Recuperado el 26 de 05 de 2015, de <http://www.annalect.com/data-is-the-fuel-that-powers-business-intelligence/>
- Bhudia, D. (2014). *www.betterbuys.com*. Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Biskup, T. (2014). *www.betterbuys.com*. Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- BMN. (23 de 02 de 2015). *BMN.es*. Recuperado el 07 de 05 de 2015, de <https://www.dropbox.com/s/veb7hv4dxdzrc2o/Leading%20Spanish%20Bank%20Banco%20Mare%20Nostrum%20Implements%20MicroStrategy%20to%20Drive%20Greater%20Analysis%20and%20Insight.docx?dl=0>

- Business Analytics. (15 de 03 de 2013). <http://www.businessanalytics.com>. Recuperado el 19 de 05 de 2015, de <http://www.businessanalytics.com/difference-between-business-analytics-and-business-intelligence/>
- Business and Technology. (08 de 09 de 2013). *Imax.com*. Recuperado el 06 de 05 de 2015, de <http://www.lmax.com/blog/business-and-technology/2013/04/08/latency-worth-it/>
- Business Dictionary. (2014). *www.businessdictionary.com*. Recuperado el 13 de 06 de 2015, de <http://www.businessdictionary.com/definition/decision-making.html>
- Chichilla, R. (2011). Mercados de datos. *Tecnología en Marcha*, 55-56.
- Cursor Software. (29 de 06 de 2015). <http://www.cursor-distribution.de>. Recuperado el 23 de 07 de 2015, de <http://www.cursor-distribution.de/en/portfolio-and-sales/sales-cognos/price-lists-cognos>
- Datamology. (2010). *www.datamology.com*. Recuperado el 03 de 06 de 2015, de <http://www.datamology.com/exploration.shtml>
- Denvens, R. M. (1865). *Cyclopædia of Commercial and Business Anecdotes*.
- Elliot, T. (09 de 03 de 2011). <http://timoelliott.com>. Recuperado el 19 de 05 de 2015, de <http://timoelliott.com/blog/2011/03/business-analytics-vs-business-intelligence.html>
- Elliot, T. (2014). *www.betterbuys.com*. Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- EscapadaRural. (2015). <http://www.escapadarural.com/>. Recuperado el 19 de 07 de 2015, de <http://www.escapadarural.com/>
- Esparza Catalán, C. (2001). *humanidades.cchs.csic.es*. Recuperado el 20 de 06 de 2015, de [http://humanidades.cchs.csic.es/cchs/web\\_UAE/tutoriales/PDF/SeriesTemporales.pdf](http://humanidades.cchs.csic.es/cchs/web_UAE/tutoriales/PDF/SeriesTemporales.pdf)
- Evelson, N. (2008). <https://www.forrester.com>. Recuperado el 13 de 06 de 2015, de <https://www.forrester.com/Topic+Overview+Business+Intelligence/fulltext/-/E-RES39218>
- Exelia. (18 de 11 de 2014). *www.exeliatech.com*. Recuperado el 26 de 05 de 2015, de <http://www.exeliatech.com/assets1/data-cleaning-in-the-etl-process/>
- García González, D. (25 de 04 de 2014). <http://necsia.es/>. Recuperado el 21 de 07 de 2015, de <http://necsia.es/big-data-vs-business-intelligence-complemento-o-sustituto/#>
- García, I. M. (2009). *Series Temporales*. Cáceres: Materiales UEX.
- Gartner. (2009). *Gartner Says Worldwide Business Intelligence and Analytics Software Market Grew 8 Percent in 2013*. STAMFORD: NewsRoom.
- Gartner. (2014). <http://www.gartner.com>. Recuperado el 21 de 07 de 2015, de <http://www.gartner.com/it-glossary/big-data>
- Hagerty, S. R. (2012). *www.gartner.com*. Recuperado el 13 de 06 de 2015, de <http://www.gartner.com/technology/reprints.do?id=1-1982NPD&ct=120208&st=sb>

- Hartzen, D. (2010). <http://infogoal.com>. Recuperado el 26 de 05 de 2015, de [http://infogoal.com/datawarehousing/data\\_sources.htm](http://infogoal.com/datawarehousing/data_sources.htm)
- IBM. (03 de 2012). <http://www-01.ibm.com>. Recuperado el 11 de 07 de 2015, de <http://www-01.ibm.com/software/es/analytics/spss/>
- IBM. (06 de 2014). <http://www-01.ibm.com>. Recuperado el 13 de 07 de 2015, de <http://www-01.ibm.com/software/es/analytics/cognos/>
- INE. (2002). *Estadística mensual Paro 1997 2002*.
- INE. (23 de 01 de 2014). [www.ine.es](http://www.ine.es). Recuperado el 09 de 07 de 2015
- INE. (05 de 2015). Encuesta de ocupación en alojamientos turísticos. 2015, 07, 18: INE.
- ine. (06 de 2015). <http://www.ine.es/>. Recuperado el 18 de 07 de 2015, de <http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft11%2Fe162eotr&file=inebase&L=0>
- Jansen, C. (2014). [www.techopedia.com](http://www.techopedia.com). Recuperado el 02 de 06 de 2015, de <http://www.techopedia.com/definition/28789/data-exploration>
- Kotorov, R. (2014). [www.betterbuys.com](http://www.betterbuys.com). Recuperado el 07 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Lixto. (2005). [www.dbai.tuwien.ac.at](http://www.dbai.tuwien.ac.at). Recuperado el 02 de 06 de 2015, de <http://www.dbai.tuwien.ac.at/proj/lixtto/WebBI.pdf>
- Major, R. (10 de 2014). [halobi.com](http://halobi.com). Recuperado el 08 de 06 de 2015, de <http://halobi.com/2014/10/descriptive-predictive-and-prescriptive-analytics-explained/>
- MarketingDirecto. (25 de 08 de 2007). [Marketingdirecto.com](http://www.marketingdirecto.com). Recuperado el 09 de 05 de 2015, de <http://www.marketingdirecto.com/actualidad/marketing/nestle-elige-cognos-8-bi-para-complementar-sus-solucion-de-reporting-sap-bw/>
- Martínez, S. (22 de 12 de 2014). *Mundo ERP*. Recuperado el 23 de 04 de 2015, de <http://mundoerp.com/blog/funcionamiento-entorno-business-intelligence-bi/>
- Mauricio, J. A. (11 de 11 de 2013). [www.ucm.es](http://www.ucm.es). Recuperado el 14 de 06 de 2015, de <https://www.ucm.es/data/cont/docs/518-2013-11-11-JAM-IAST-Libro.pdf>
- Mearian. (2011). By 2020, there will be 5,200 GB of data for every person on Earth. *ComputerWorld*.
- Microsoft. (06 de 2014). <https://products.office.com/>. Recuperado el 11 de 07 de 2015, de <https://products.office.com/es-es/excel>
- Milton. (2011). [www.forbes.com](http://www.forbes.com). Recuperado el 13 de 06 de 2015, de <http://www.forbes.com/sites/>
- minetur. (02 de 2014). <http://www.minetur.gob.es>. Recuperado el 09 de 07 de 2015, de [http://www.minetur.gob.es/turismo/es-ES/Novedades/Documents/005b\\_Plan%20Integral%20de%20Turismo%20Rural%202014.pdf](http://www.minetur.gob.es/turismo/es-ES/Novedades/Documents/005b_Plan%20Integral%20de%20Turismo%20Rural%202014.pdf)

- MySoftware. (12 de 04 de 2011). *MySoftware.wordpress*. Recuperado el 23 de 04 de 2015, de <https://mysoftwar.wordpress.com/2011/04/12/¿cual-es-la-principal-diferencia-entre-el-erp-y-crm/>
- Nelson, G. (2007). SAS. Recuperado el 23 de 04 de 2015, de <http://www2.sas.com/proceedings/forum2007/207-2007.pdf>
- Oracle. (2007). <http://docs.oracle.com>. Recuperado el 30 de 05 de 2015, de [http://docs.oracle.com/html/E10312\\_01/dm\\_concepts.htm](http://docs.oracle.com/html/E10312_01/dm_concepts.htm)
- rainmakerworks. (27 de 04 de 2012). *www.rainmakerworks.com*. Recuperado el 28 de 05 de 2015, de <http://datawarehouse4u.info/OLTP-vs-OLAP.html>
- Rijmenam, M. v. (2014). *www.betterbuys.com*. Recuperado el 06 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Roche, P. (2014). *www.betterbuys.com*. Recuperado el 06 de 06 de 2015, de <https://www.betterbuys.com/bi/business-intelligence-vs-business-analytics/>
- Rouse, M. (10 de 2014). <http://searchdatamanagement.techtarget.com>. Recuperado el 19 de 05 de 2015, de <http://searchdatamanagement.techtarget.com/definition/business-intelligence>
- Rouse, M. (04 de 2014). <http://searchsap.techtarget.com/>. Recuperado el 13 de 07 de 2015, de <http://searchsap.techtarget.com/definition/mySAP>
- SAP. (2014). *help.sap.com*. Recuperado el 11 de 06 de 2015, de [http://help.sap.com/saphelp\\_nw70/helpdata/EN/8d/2b4e3cb7f4d83ee10000000a114084/content.htm](http://help.sap.com/saphelp_nw70/helpdata/EN/8d/2b4e3cb7f4d83ee10000000a114084/content.htm)
- SAP. (13 de 02 de 2014). <http://go.sap.com/>. Recuperado el 12 de 07 de 2015, de <http://go.sap.com/hk/product/analytics/lumira.html>
- Sinnexus. (2012). <http://www.sinnexus.com>. Recuperado el 30 de 05 de 2015, de [http://www.sinnexus.com/business\\_intelligence/olap\\_avanzado.aspx](http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx)
- Sinnexus. (2012). *www.sinnexus.com*. Recuperado el 13 de 05 de 2015, de [http://www.sinnexus.com/business\\_intelligence/](http://www.sinnexus.com/business_intelligence/)
- Skirus, V. y. (2012). Business Intelligence in the Process of Decision Making: Changes and Trends. Lituania.
- Skyrius, V. y. (2012). Business Intelligence in the process of decision making. Lituania.
- Standen, J. (02 de 12 de 2008). <http://www.datamartist.com>. Recuperado el 19 de 05 de 2015, de <http://www.datamartist.com/data-warehouse-vs-data-mart>
- Steinbach, K. (08 de 05 de 2005). <http://www-users.cs.umn.edu>. Recuperado el 02 de 06 de 2015, de [http://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap3\\_data\\_exploration.pdf](http://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap3_data_exploration.pdf)
- stratebi. (2013). 2013. Recuperado el 07 de 06 de 2015, de [www.stratebi.es](http://www.stratebi.es): [http://www.stratebi.es/todobi/May13/Nuevas\\_Tendencias\\_BI.pdf](http://www.stratebi.es/todobi/May13/Nuevas_Tendencias_BI.pdf)

- stratebi. (05 de 2013). *www.stratebi.com*. Recuperado el 09 de 05 de 2015, de [http://www.stratebi.com/c/document\\_library/get\\_file?uuid=2e79e125-b78d-406d-9235-80cd5cbd87c2&groupId=10157](http://www.stratebi.com/c/document_library/get_file?uuid=2e79e125-b78d-406d-9235-80cd5cbd87c2&groupId=10157)
- tableau. (03 de 2014). <http://get.tableau.com/es-es/>. Recuperado el 12 de 07 de 2015, de <http://get.tableau.com/es-es/trial/tableau-software.html?cid=70160000000YEzj&ls=Paid%20Search&lsd=Google%20AdWords%20-%20Tableau%20-%20EMEA%20-%20ES%20-%20Free%20Trial&adgroup=Tableau%20-%20Exact&kw=tableau&adused=63871546582&distribution=search&gclid=Cj0>
- Tech-faq. (24 de 02 de 2011). <http://www.tech-faq.com>. Recuperado el 28 de 05 de 2015, de <http://www.tech-faq.com/data-warehouse.html>
- thearling. (2011). *www.thearling.com*. Recuperado el 03 de 06 de 2015, de <http://www.thearling.com/text/dmtechniques/dmtechniques.htm>
- Third Nature. (2010). <http://www.slideshare.net>. Recuperado el 21 de 07 de 2015, de <http://www.slideshare.net/oktopuslu/bi-comparison-of-open-source-and-traditional-vendor-4327259>
- tinko. (25 de 10 de 2012). <https://startdatawarehousingwithtinko.wordpress.com>. Recuperado el 19 de 05 de 2015, de <https://startdatawarehousingwithtinko.wordpress.com/author/vimalantinko/>
- Triad. (05 de 2010). <http://triadtechpartners.com>. Recuperado el 23 de 07 de 2015, de <http://triadtechpartners.com/wp-content/uploads/Tableau-GSA-Pricelist-May-2015.pdf>
- Universidad Pablo de Olavide. (27 de 03 de 2009). <http://www.entreculturas.uma.es>. Recuperado el 09 de 07 de 2015, de <http://www.entreculturas.uma.es/n1pdf/articulo24.pdf>
- UOC. (2012). <http://informatica.blogs.uoc.edu>. Recuperado el 13 de 06 de 2015, de <http://informatica.blogs.uoc.edu/2012/04/23/reducir-el-fracaso-en-los-proyectos-de-business-intelligence-ii/>
- UOP. (13 de 02 de 2014). Inteligencia de Negocio y Big Data. *Inteligencia de Negocio y Big Data: Negocio, Tecnología y Personas*. Madrid, Madrid, España: EIMT.OUC.EDU.
- Waikato. (05 de 2013). <http://www.cs.waikato.ac.nz>. Recuperado el 11 de 07 de 2015, de <http://www.cs.waikato.ac.nz/ml/weka/>
- WebMining Consultores. (13 de 03 de 2012). <http://www.webmining.cl>. Recuperado el 19 de 05 de 2015, de <http://www.webmining.cl/2012/03/business-analytics-versus-business-intelligence/>
- Xerox. (10 de 04 de 2015). *Xerox.com*. Recuperado el 07 de 05 de 2015, de <http://www.xerox.es/servicios/transport-solutions/eses.html>