



# **Implantació d'una solució BI Open Source**

**Nadal Salamanca Nadal**  
Grau de Enginyeria Informàtica

**Consultor: Humberto Andrés Sanz**

Data Lliurament

## **Copyright**

© (l'autor/a)

Reservats tots els drets. Està prohibit la reproducció total o parcial d'aquesta obra per qualsevol mitjà o procediment, compresos la impressió, la reprografia, el microfilm, el tractament informàtic o qualsevol altre sistema, així com la distribució d'exemplars mitjançant lloguer i préstec, sense l'autorització escrita de l'autor o dels límits que autoritzi la Llei de Propietat Intel•lectual.

## FITXA DEL TREBALL FINAL

<b>Títol del treball:</b>	<i>IMPLANTACION PLATAFORMA BI</i>
<b>Nom de l'autor:</b>	<i>Nadal Salamanca Nadal</i>
<b>Nom del consultor:</b>	<i>Humberto Andrés Sanz</i>
<b>Data de lliurament (mm/aaaa):</b>	<i>01/2017</i>
<b>Àrea del Treball Final:</b>	<i>Business Intelligence</i>
<b>Titulació:</b>	<i>Grau d'enginyeria Informàtica</i>
<b>Resum del Treball (màxim 250 paraules):</b>	
<p>El presente Trabajo muestra el proceso de una implantación de una plataforma Open Source de BI (Business Intelligence). El estudio pasa por todas las fases de un proyecto de BI, desde la selección de la plataforma, su instalación, personalización, selección de fuentes de datos, transformación, modelado y presentación de resultados.</p> <p>El objetivo es demostrar la utilidad de la utilización de un sistema de BI para la gestión de una empresa o una comunidad, en este caso nos basamos en datos reales obtenidos de la observación de Wifi pública en la ciudad de Palma de Mallorca.</p> <p>El trabajo va a servir de base para una próxima implantación en el Ayuntamiento de Palma de una plataforma de BI, con datos procedentes de las múltiples fuentes disponibles.</p> <p>No se van a desarrollar todas las posibilidades de la plataforma, puesto que el tiempo disponible para este cometido no es suficiente, pero si va a dar una idea muy aproximada de los problemas que se pueden encontrar y como resolverlos.</p>	

**Abstract (in English, 250 words or less):**

This work shows the process of an implementation of an Open Source BI platform (Business Intelligence). The study goes through all phases of a BI project, from the selection of the platform, its installation, and customization, selection of data sources, transformation, modeling and presentation of results.

The objective is to demonstrate the utility of the use of a BI system for the management of a company or a community, in this case we rely on real data obtained from the observation of public Wifi in the city of Palma de Mallorca. The work will serve as the basis for a future implementation in the City of Palma of a BI platform, with data from the multiple sources available. Not all the possibilities of the platform will be developed, since the time available for this task is not enough, but it will give a very approximate idea of the problems that can be found and how to solve them.

**Paraules clau (entre 4 i 8):**

BI, SpagoBI, implantación, resultados, utilidad, funcionamiento.

# Índice

1. Introducción.....	1
1.1 Contexto y justificación del Trabajo.....	1
1.2 Objetivos del Trabajo. ....	1
1.3 Enfoque y método seguido.....	2
1.3 Planificación del Trabajo.....	3
1.5 Breve resumen de productos obtenidos. ....	3
1.6 Breve descripción de los otros capítulos de la memoria.....	3
2. Desarrollo del proyecto.....	5
2.1 Revisión de conceptos BI y minería de datos.....	5
2.2 Qué significa Business Intelligence .....	6
2.3 Etapas de BI.....	7
2.4 Definición de la tarea de minería de datos. ....	10
2.5 Selección de datos. ....	14
2.6 Software elegido.....	15
2.7 Instalación de SpagoBI Server. ....	17
2.8 Creación y prueba del conector de datos. ....	21
2.9 Preparación de datos: .....	23
2.10 Modelado de datos:.....	27
2.11 Elaboración de gráficos y cuadro de mandos.....	39
2.12 Uso del Cubo OLAP para creación de informes de forma dinámica. ....	50
2.13 Interpretación de resultados. ....	54
2.14 Conclusión.....	55
3. Conclusiones.....	57
3.1 Lecciones aprendidas.....	57
3.2 Logro de objetivos: .....	57
3.3 Seguimiento de la planificación. ....	58
3.4 Líneas de trabajo futuras.....	58
4. Glosario .....	59
5. Bibliografía .....	60
6. Anexos .....	61

## Lista de figuras

1 Diagrama de Gantt	3
2 Jerarquía datos-conocimiento	5
3 Mapa de zonas WIFI	10
4 Estructura de SpagoBI	15
5 Pantalla inicio SpagoBI	20
6 Conexión de datos	21
7 Configuración conector MySQL	21
8 Esquema MySQL 1	23
9 Esquema MySQL 2	24
10 Estructura WifiClients	25
11 Cardinalidad de las tablas	25
12 Esquema lógico	26
13 Creación modelo datos1	27
14 Conexión SpagoBi Meta SpagoBi Server	28
15 Creación modelo de datos1	28
16 Creación modelo de datos 2	29
17 Selección de la conexión de datos	30
18 Selección de tablas físicas	31
20 Asociación tablas físicas y lógicas	32
21 Creación dimensiones y cubo 1	32
22 Creación dimensiones y cubo 2	32
23 Creación modelo de datos jerarquías 1	33
24 Creación modelo de datos jerarquías 2	33
25 Creación modelo de datos cubo	34
26 Definición de atributos y medidas	35
27 Subir modelo a SpagoBI Server 1	36
28 Subir modelo a SpagoBI Server 2	36
29 Creación esquema mondrian	37
30 Subir al servidor SpagoBI el modelo OLAP	38
31 Verificación modelo datos	39
32 Creación de gráficos 1	39
33 Creación de gráficos 2	40
34 Creación de gráficos 3	40
35 Creación de gráficos 4	41
36 Creación de gráficos 5	42
37 Creación de gráficos 6	42
38 Creación de gráficos 7	42
39 Gráfico uso por SO todas las áreas	43
40 Gráfico uso por SO Palma Ciudad	43
41 Gráfico uso por SO Playa de Palma	44
42 Gráfico usuarios por día todas las áreas	44
43 Gráfico usuarios por día de la semana Palma Ciudad	45
44 Gráfico usuarios por día Playa de Palma	45
45 Gráfico usuarios por día todas las áreas	46
46 Gráfico usuarios por hora todas las áreas	46
47 Informe puntos acceso con más usuarios	47

48 Informe usuarios por fabricante	47
49 Informe usuarios por día AP-Zona	48
50 Mapa puntos acceso zona 0 Playa de Palma	49
51 Cuadro de mandos	50
52 Comprobación esquema mondrian	50
53 Selección fuente de datos	51
54 Ejemplo OLAP 1	51
55 Ejemplo OLAP 2	52
56 Ejemplo OLAP 3	52
57 Ejemplo OLAP 4	53
58 Ejemplo OLAP 5	53
59 Ejemplo OLAP 6	54
60 Grafico comparativo 3 puntos de acceso por hora del día	55

# 1. Introducción

## 1.1 Contexto y justificación del Trabajo.

En la actualidad, debido a al bajo coste del espacio de almacenamiento de datos, se tiende a guardar información prácticamente sin límite aun sin saber si se va a utilizar en algún momento (big data), esta información para que sea útil debe convertirse en conocimiento, para ello se han desarrollado herramientas de BI, las cuales necesitan ser convenientemente utilizadas para que la extracción de conocimiento sea útil. En el Ayuntamiento de Palma se dispone de una red Wifi, con puntos de acceso repartidos por toda la ciudad, especialmente en las zonas turísticas, donde el visitante puede conectarse mediante su dispositivo móvil (portátil, Smartphone, Tablet, etc.) para acceder a la Web. Las redes Wifi están implementadas con una solución Cisco Meraki [1], el cual guarda un log para cada acceso a la Wifi con distintos datos del usuario. Cisco Meraki proporciona herramientas de BI, como mapas de calor, gráficos de utilización, tiempos de conexión, etc. Debido a que esta plataforma es de pago y es propiedad de la empresa concesionaria, el Ayuntamiento, el cual es el propietario de los datos, se plantea implementar una plataforma de BI, la cual, utilizando los datos antes mencionados, permita ejecutar diferentes estudios de comportamiento de la población en las zonas Wifi municipales, así mismo esta experiencia piloto servirá para incluir otros fuentes de datos municipales en la plataforma y elaborar un DataWarehouse municipal, mediante el que se obtendrán conocimientos de distintas fuentes de datos que ayuden a la gestión municipal.

Las premisas para la implementación de esta plataforma son las siguientes:

- Debe ser una plataforma Open Source
- Debe permitir una implantación rápida de la solución
- Fácil manejo de la misma con el fin de minimizar la curva de aprendizaje de los usuarios finales.
- Posibilidad de manejar diferentes fuentes de datos (en principio se usara para los datos de Wifi, pero existen otras fuentes de datos municipales que son susceptibles de incluirse en la plataforma)
- Funcionamiento en navegador Web, sin instalación.

## 1.2 Objetivos del Trabajo.

El principal objetivo del trabajo es crear un prototipo funcional con el fin de demostrar la utilidad de la explotación de la información almacenada y ver como esta puede ser usada para la toma de decisiones. Para ello se documentaran todos los pasos de un proyecto de BI.



Los datos de los que se parte son reales, por lo tanto los resultados serán evaluables con los que proporciona la plataforma Cisco Meraki. No se usarán datos de carácter personal con el fin de proteger la confidencialidad.

Los objetivos principales son:

- Selección de una plataforma Open Source.
- Instalar y aprender el funcionamiento de la plataforma.
- Conectar esta plataforma con un repositorio de datos.
- Cargar datos en el repositorio
- Elaborar gráficos e informes.
- Interpretar resultados.
- Ayudar a tomar decisiones.

### 1.3 Enfoque y método seguido.

El trabajo que se va a realizar va a servir como guion para ser implantado de forma real en el Ayuntamiento de Palma, aun cuando la plataforma seleccionada puede ser distinta a la utilizada en este proyecto. Debido a la gran cantidad de plataformas Open Source existentes en el mercado y dado que el objeto de este trabajo no es escoger una plataforma de BI, se utilizará **SpagoBI** la cual cumple con los requisitos exigidos al igual que otras.

Otras plataformas Open Source que se han estudiado antes de elegir la mencionada han sido:

Rapid Miner.  
Jaspersoft para Business Intelligence.  
Pentaho  
BIRT  
SpagoBI

El motivo de elegir SpagoBI es principalmente, ser una plataforma modular, a la cual se le pueden incorporar complementos a medida que se necesiten, su posibilidad de integrar en una web, y sobre todo que sus requerimientos de hardware permiten que se elabore el prototipo sobre un ordenador personal medio, sin grandes prestaciones.

Los motivos para no elegir las otras cuatro estudiadas han sido los siguientes:

- Pentaho. No soportada por el equipo disponible, si bien cumple todas los requisitos.
- Jaspersoft. La versión community no ofrece las mismas prestaciones que las demás, no tiene motor de GIS, no funciona en web sino en una plataforma eclipse, etc.
- Rapid Miner. No funciona en web, debe instalarse en cada equipo. La versión gratuita tiene limitada la capacidad de datos.
- BIRT. Es más bien un software java para definición de informes que pueden ser incrustados o llamados desde otras plataformas.

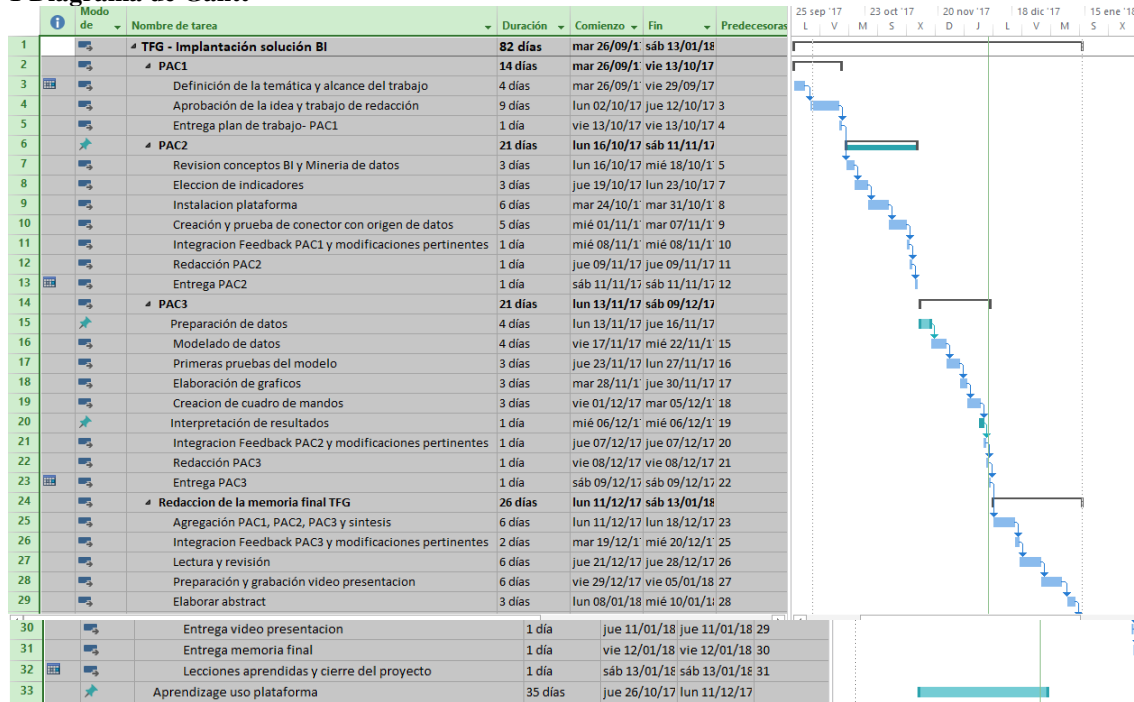
El método de trabajo va a ser, definir los objetivos a obtener, hacer la instalación completa de esta en un ordenador personal (Pentium IV Quad Core

con 4 GB de RAM y 1 Tb de disco), sabiendo que las prestaciones no son equiparables a las que se pueden obtener en la implantación real. Una vez hecha la instalación se aprenderá su funcionamiento, se personalizará, se estructurarán, modificarán y modelarán los datos para conseguir los resultados requeridos. Se seguirá la metodología aprendida en la asignatura de **Minería de datos** para la ejecución del proyecto

### 1.3 Planificación del Trabajo.

Las entregas principales consistirán en tres entregas parciales coincidiendo con las fechas establecidas en el plan docente con las Pruebas de Evaluación Continuada (PEC) numeradas 1, 2 y 3 y finalmente la entrega de esta memoria final del Trabajo de Fin de Grado (TFG), debidamente acompañado de una breve presentación de defensa del proyecto en formato audiovisual. Las fechas e hitos a conseguir se muestran en el diagrama de Gantt a continuación.

#### 1 Diagrama de Gantt



### 1.5 Breve resumen de productos obtenidos.

Los resultados obtenidos son:

- Instalación y personalización de la plataforma.
- Creación del repositorio de datos.
- Obtención de gráficos e informes.
- Interpretación de resultados.

### 1.6 Breve descripción de los otros capítulos de la memoria

En el capítulo 2 se exponen las diferentes fases de la ejecución del proyecto.

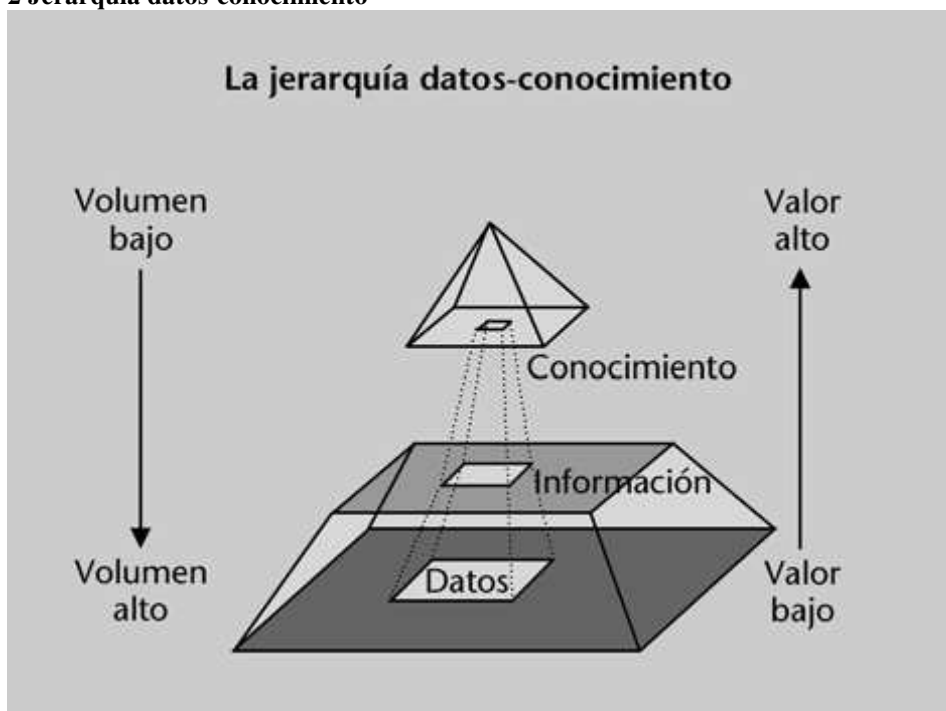
- Apartados 2.1 hasta 2.3: se hace una revisión de conceptos de BI y minería de datos, con objeto de sentar las bases teóricas de la ejecución del proyecto.
- Apartado 2.4: explica la situación de partida y describe los datos y recursos disponibles para la implementación de la solución.
- Apartado 2.5: se describe la segunda etapa del proceso de BI, la selección de los datos a incluir en el repositorio DataWarehouse, elección de fuentes y atributos.
- Apartado 2.6: describe las prestaciones del software elegido así como sus funcionalidades.
- Apartado 2.7: Describe el proceso de instalación del producto y su personalización.
- Apartado 2.8: Creación del conector de datos con el repositorio MySQL:
- Apartado 2.9: Tercera fase del proceso de BI, preparación de datos.
- Apartado 2.10: Creación de los modelos de datos. Fase de minería de datos, cuarta fase del proceso de BI. (creación de un modelo QBE y un esquema OLAP).
- Apartado 2.11: Descripción del proceso de creación de gráficos e informes.
- Apartado 2.12: Uso del Cubo OLAP para la creación de informes dinámicos.
- Apartado 2.13: Interpretación de resultados. Obtención de conocimiento a partir de los gráficos e informes obtenidos en los puntos anteriores. Quinta fase del proceso de BI.
- Apartado 2.14: Conclusiones del uso de la herramienta de BI SpagoBI.

## 2. Desarrollo del proyecto.

### 2.1 Revisión de conceptos BI y minería de datos.

Las empresas mediante sus sistemas de información producen un gran número de datos, estos datos explotados de la forma adecuada, generan conocimiento y permiten obtener ventajas competitivas a la hora de tomar decisiones y planificar estrategias. Para ello se necesitan herramientas que organicen y agrupen estos datos de forma que puedan entenderse y utilizarse. Esta información suele estar inconexa, en diferentes fuentes de datos y no accesible en línea. Para poder planificar y obtener conocimiento de esta información nació la ciencia que vino en llamarse “inteligencia de negocio” o “Business Intelligence” (BI).

#### 2 Jerarquía datos-conocimiento



## 2.2 Qué significa Business Intelligence

Existen numerosas definiciones de lo que es BI, como ejemplo unas cuantas:

*Business Intelligence (BI) o inteligencia de negocios se define como la habilidad corporativa para tomar decisiones. Esto se logra mediante el uso de metodologías, aplicaciones y tecnologías que permiten reunir, depurar, transformar datos, y aplicar en ellos técnicas analíticas de extracción de conocimiento (Parr, 2000) [2].*

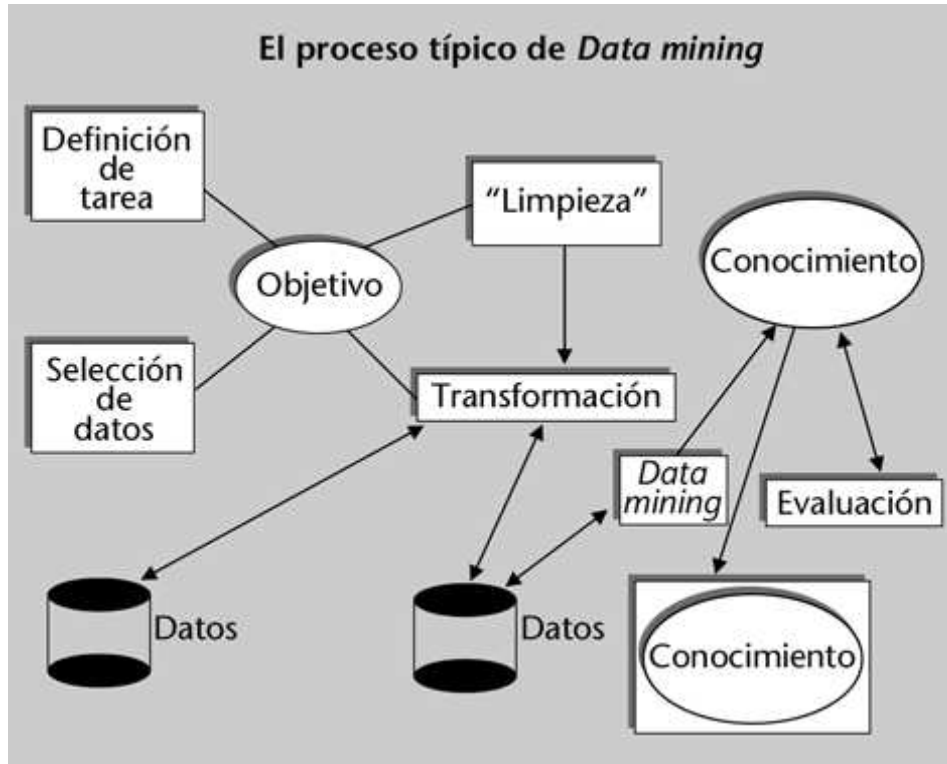
*BI se puede definir como el uso de los datos recopilados con el fin de generar mejores decisiones de negocio, esto implica accesibilidad, análisis y revelar nuevas oportunidades (Almeida, Ishikawa, Reinschmidt, & Roeber, 1999). [2]*

*BI se puede ver como el proceso en el que intervienen personas y sistemas con la meta de obtener, recopilar, analizar y presentar la información que soporte de mejor forma la toma de decisiones de negocio. El proceso se puede dividir en 4 etapas: extracción, consolidación, explotación y visualización (Dávila, 2006). [2]*

Como vemos todas las definiciones van en la misma línea, establecer un método mediante el cual se ayuda a la toma de decisiones. En la última definición (Dávila, 2006), se introduce el concepto de las etapas del proceso de BI. La implementación de un sistema de BI en una organización se debe iniciar mediante la selección de la información relevante para la toma de decisiones, esto requiere contar con la participación de personal en los niveles operativo, táctico y gerencial.

En este trabajo se van a seguir las etapas según Fayyad, como se estudió en la asignatura de **Minería de datos**.

## Proceso de minería de datos



Estas etapas [2], [3] como se ve en el gráfico son: definición de la tarea de minería de datos, selección de los datos, preparación de datos, minería de datos, evaluación e interpretación del modelo, e integración. Estas etapas no presentan un recorrido lineal, sino que pueden repetirse a medida que nuevos cambios en la situación pueden provocar que nuestro conocimiento sea erróneo y debamos replantearnos el extraer nuevos datos.

### 2.3 Etapas de BI

#### 2.3.1 Definición de la tarea.

Es la fase más importante, donde nos planteamos a donde queremos llegar, para ello existen diferentes metodologías para llegar al modelo que queremos representar (encontrar similitudes, clasificar objetos, predecir resultados, describir y encontrar asociaciones significativas o casuales entre diversas variables, explicar las razones por las que se da determinado comportamiento).

#### 2.3.2 Selección de datos.

En esta fase la tarea consiste en encontrar los datos que necesitamos de las fuentes disponibles y hacer que tengan el mismo significado.

*Desde una perspectiva ideal, la tecnología de data warehousing –literalmente, almacenes de datos– (Inmon, 1996) está especialmente orientada a facilitar la localización de los datos dentro de una empresa en relación con varios tipos de utilidades. Un DataWarehouse integra datos procedentes de los distintos datos de cada departamento de una empresa. [2]*

### 2.3.3 Preparación de datos.

Una vez localizadas las fuentes de datos, debemos proceder a prepararlas para que se les puedan aplicar los métodos o herramientas que construirán al modelo deseado. Esta fase, aunque parezca sencilla, junto con la de selección de datos, consume el 70% del esfuerzo en los proyectos de minería de datos de nueva implantación.

### 2.3.4 Minería de datos.

En la fase de minería de datos tenemos los datos con la calidad adecuada, en el formato adecuado y hemos seleccionado los atributos y los registros aparentemente necesarios y relevantes. Tenemos decidido qué tipo de modelo queremos obtener. Por lo tanto, ahora hay que elegir un método de construcción de modelos entre la multitud de métodos que permiten obtener el tipo de modelo que nos interesa.

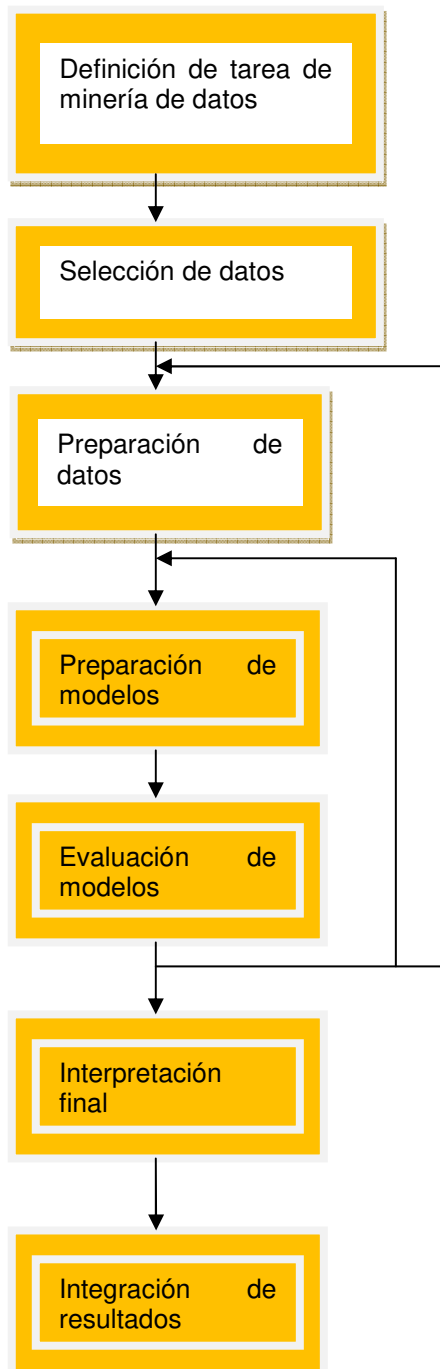
### 2.3.5 Evaluación e interpretación del modelo.

El final de la fase de minería de datos es un modelo que representa un tipo determinado de conocimiento sobre el dominio que estábamos estudiando.

### 2.3.6 Integración de los resultados en el proceso.

El último paso consiste en integrar los resultados del data mining en el proceso típico del sistema de información en el que se esté aplicando.

De forma gráfica se puede representar así:



**Etapas proceso BI**



## 2.4 Definición de la tarea de minería de datos.

### Descripción del escenario.

En el municipio de Palma se dispone de 197 puntos de acceso Wifi (AP), públicos sobre una plataforma Cisco Meraki, la cual administra estos puntos de acceso. Estos puntos de acceso están agrupados por zonas, 8 zonas en la ciudad numeradas del 0 al 7 (en verde en el mapa) y 4 zonas en la playa de palma donde se concentran la mayoría de hoteles del municipio numeradas del 0 al 3 (en rojo en el mapa). [1]

### 3 Mapa de zonas WIFI



Los teléfonos inteligentes (Smartphone) dotados de Wifi se pueden usar como un indicador de la presencia del usuario gracias a un mecanismo que es común en todas las solicitudes de sondeo de dispositivos. Las tramas de negociación de la norma 802.11 se transmiten a intervalos regulares desde dispositivos Wifi. Estas tramas contienen información que puede ser usada para identificar la presencia, el tiempo dedicado y las visitas repetidas dentro del alcance de un punto de acceso Wifi. Estos dispositivos pueden detectarse mediante puntos de acceso Wifi independientemente de su estado de asociación, lo que significa que incluso si un usuario no conecta su dispositivo a la red inalámbrica, la presencia del dispositivo aún se puede detectar mientras el dispositivo está dentro del alcance de la red y la antena Wifi del mismo está encendida.

Esta característica es aprovechada por el sistema Cisco Meraki el cual monitoriza las conexiones Wifi públicas en la ciudad de Palma y a través una API nos proporciona un fichero en formato JSON con la siguiente información:

Nombre	Tipo	Descripción
AP observado	String	Nombre del Ap Observado
Data	date	Fecha de la observación
Observations	[string]	Array con todas las observaciones del AP
apMac	string	Dirección MAC del AP observado
apTags	[string]	Array JSON de todas las etiquetas aplicadas en el AP en el tablero
apFloors	[string]	Array JSON de todos los nombres de plano en los que aparece este AP
clientMac	string	MAC del dispositivo
ipv4	string	Dirección IPv4 del cliente y hostname, en formato "hostname/address" ; solo "/address" si no hay hostname, nulo si no está disponible
ipv6	string	Dirección IPv6 del cliente y hostname, en formato "hostname/address" ; solo "/address" si no hay hostname, nulo si no está disponible
seenTime	ISO 8601 date string	Fecha y hora de la observación en formato UTC; ejemplo: "1970-01-01T00:00:00Z"
seenEpoch	integer	Tiempo de observación en milisegundos
ssid	string	Nombre SSID del cliente nulo si el dispositivo no está conectado
rssi	integer	RSSI de dispositivo que ha según lo visto por el AP
manufacturer	Stringh	Fabricante del dispositivo; nulo si no puede ser determinado
os	string	Sistema operativo de dispositivo; nulo si no puede ser determinado
Location	location	Geolocalización del dispositivo; nulo si no puede ser determinada
lat	decimal	Latitud del dispositivo en grados Norte desde el ecuador
lng	decimal	Longitud en grados desde el Este del primer meridiano
unc	decimal	Desviación posible en metros
x	[decimal]	Array JSON de desplazamiento x (en metros) desde la esquina inferior izquierda de cada plano
y	[decimal]	Array JSON de desplazamiento y (en metros) desde la esquina inferior izquierda de cada h plano

La implantación de la API se efectuará sobre el modelo real una vez que se adopte la plataforma definitiva.

Además del fichero JSON, podemos obtener archivos CSV, desde la plataforma Meraki, con las observaciones del último mes, última semana o último día.

Estos CSV son los que se van a utilizar en este trabajo, creando para ello una estructura de datos apropiada para ser interrogada.

Dado que los teléfonos inteligentes tienen más de un 50% de penetración en la población, las solicitudes de sondeo pueden usarse para generar y detectar un conjunto de datos estadísticamente significativo con respecto a la presencia de dispositivos habilitados para Wifi dentro del alcance de un punto de acceso dado.

Los dispositivos Wifi normalmente emiten una solicitud de sondeo a intervalos regulares en función del estado del dispositivo (consultar la Tabla 1). Los teléfonos inteligentes envían solicitudes de sondeo para descubrir las redes inalámbricas circundantes, de modo que puedan poner las redes a disposición del usuario.

Estado del dispositivo	Intervalo de sondeo (Smartphone)
Dormido (pantalla apagada)	~ algunas por minuto
Espera (pantalla encendida)	10 - 15 veces por minuto
Asociado a la red	Variable, puede requerir que el usuario busque manualmente las redes.

**Tabla1**

Las tramas de datos recibidas de todos los dispositivos Wifi conectados y las solicitudes de sondeo detectadas desde todos los dispositivos que se ven dentro del alcance (típicamente hasta 30 metros o más) generan eventos de "dispositivos vistos" en los AP de Cisco Meraki. Los AP de triple radio tienen una radio de escaneo dedicado que escucha las solicitudes de sondeo 24 horas al día, 7 días a la semana, en todos los canales.

Una vez recibida esta trama por la nube Meraki, se agregan las firmas de presencia de todos los puntos de acceso en una red. Después de la agregación, los datos de cada dispositivo observado se someten a una serie de cálculos para categorizarlos para su presentación posterior. Por ejemplo, los comercios deben comprender la tasa de captura, que es la proporción de personas que pasan por la tienda o que entran realmente. Para determinar la tasa de captura, la nube de Meraki analiza la intensidad de la señal de cada dispositivo, junto con el tiempo que pasa dentro de esa ubicación (una señal de alta intensidad por sí sola puede no indicar un visitante si simplemente está pasando por la tienda rápidamente).

El comprender patrones como el tráfico diario según el momento del día y cómo varía la tasa de captura en diferentes sitios es muy importante para que los departamentos de TI pueden entender mejor el uso y las tendencias de la red, y los departamentos que no son de TI, como los equipos de marketing e inteligencia comercial, pueden sugerir ideas y preguntas como, por ejemplo, mi nueva campaña de marketing en el sitio A está basada en los números de tráfico a pie o necesito asignar más personal al sitio B durante las horas punta. Algunos de los diferentes casos de uso para los que pueden usarse estos datos se relacionan a continuación:

- Detectar visitas totales de clientes

- Optimizar el personal por hora del día
- Analizar el tiempo de permanencia del visitante y la frecuencia de repetición.
- Ver si en una parada de autobús hay demasiada gente esperando y optimizar la frecuencia de estos según la ocupación horaria.
- Comparar los sitios o tomar promedios para conjuntos de sitios para comprender el tráfico a pie por comercio por debajo o por encima del promedio, el tiempo de permanencia y la frecuencia de repetición.
- Optimizar y ejecutar pruebas A / B para ver si los cambios en una variable afectan el resultado de los parámetros medibles (por ejemplo, la tasa de captura).
- Preparar la red para las fluctuaciones semanales o estacionales mediante la optimización de las políticas de acceso.

Vistos los datos disponibles establecemos cual será nuestra tarea de minería de datos.

- Ver la utilización que se hace de los puntos de acceso, tanto en tiempo de permanencia como en número de usuarios.
- Comparar utilización de Wifi según áreas, zonas y Ap.
- Descubrir patrones de comportamiento observando la utilización por días de la semana y hora del día.
- Asimismo se desarrollaran gráficos comparativos de sistemas operativos usados por los Smartphone con el fin de ver cuál es la penetración de cada uno de ellos en el mercado.
- Elaborar un cuadro de mandos que permita ver en una sola pantalla los principales parámetros de utilización.

## 2.5 Selección de datos.

Desde la plataforma Meraki se han exportado los datos siguientes:

Observaciones de cada AP de la última semana (del 27-10 al 2-11) en formato CSV.

<b>Nombre</b>	<b>Tipo</b>	<b>Descripción</b>
<b>Last seen</b>	Fecha	Fecha y hora ultima observación
<b>Usage</b>	Texto corto	Tiempo de uso en milisegundos
<b>OS</b>	Texto corto	Sistema operativo del dispositivo
<b>IPv4 address</b>	Texto corto	Dirección IP
<b>Policy</b>	Texto corto	Política aplicada
<b>MAC address</b>	Texto corto	Mac address del dispositivo
<b>Manufacturer</b>	Texto corto	Fabricante del dispositivo
<b>Connected to</b>	Texto corto	Punto de acceso conectado
<b>First seen</b>	Fecha	Primera fecha y hora de acceso

Puntos de acceso de cada zona en formato CSV con la siguiente información:

<b>Nombre</b>	<b>Tipo</b>	<b>Descripción</b>
<b>MAC</b>	Texto corto	MAC address del AP
<b>Name</b>	Texto corto	Nombre del AP
<b>Model</b>	Texto corto	Modelo
<b>Serial</b>	Texto corto	Número de serie
<b>Latitude</b>	Simple	Latitud en grados
<b>Product description</b>	Texto corto	Descripción del AP
<b>Internet address</b>	Texto corto	IP AP
<b>Longitude</b>	Simple	Longitud en grados
<b>LAN IP</b>	Texto corto	IP de la LAN

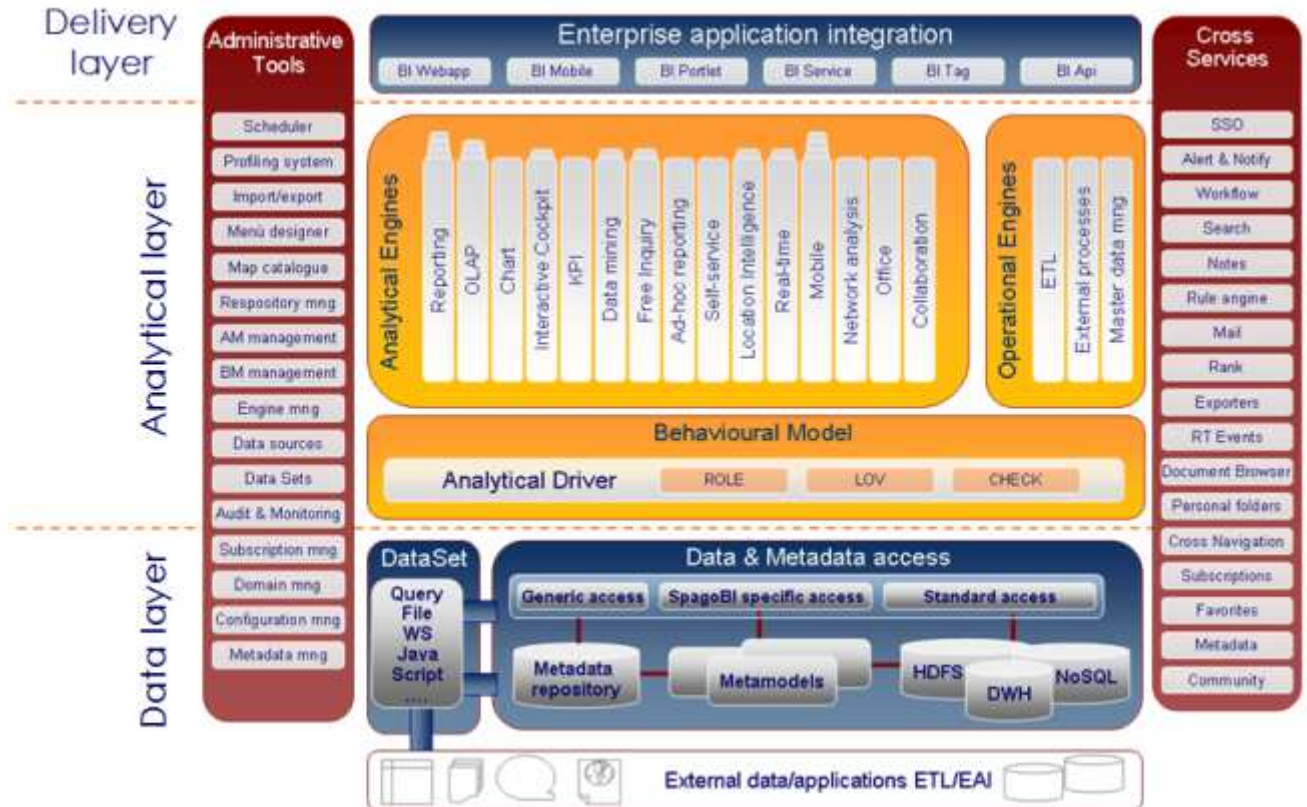
## 2.6 Software elegido.

Como ya se indicó en punto 1.3 el software elegido es la suite de BI SpagoBI

### 2.6.1 Características de SpagoBI.

Arquitectura de SpagoBI [7]

#### 4 Estructura de SpagoBI



- Es una aplicación web desarrollada para un servidor de aplicaciones J2EE (Tomcat, JBoss, WebSphere, etc.)
- Se puede ejecutar en cualquier sistema operativo que soporte JVM 1.5 (1.6 recomendado)
- Funciona con un repositorio privado alojado en DBMS (MySQL, Postgree, Oracle, Ingres, HSQL, etc.)
- La aplicación se puede acceder a través de casi todos los principales navegadores (IE, Firefox, Opera, Safari)
- Se ejecuta de forma nativa en varios portal server (liferay, exo portal, WebSphere, etc.)
- Es un entorno que ofrece todas las funcionalidades típicas del BI:
  - Reporting
  - OLAP
  - Charting
  - KPI (modeller and navigator)
  - Interactive cockpits

- Ad hoc reporting (Worksheet)
- Self-service BI
- Location Intelligence
- Free Inquiry (QbE)
- Data Filtering (Smart Filter)
- Data Mining
- Real-time dashboards and consoles
- Accessible reporting
- Collaboration (analytical dossier and workflow)
- Office integration
- ETL
- Mobile BI
- External processes
- Master Data Management
- Network Analysis.

- Permite gestionar el producto a través de una interfaz web
- Permite gestionar el ciclo de vida del objeto dentro del repositorio de BI.
- Expone herramientas y servicios transversales, tales como la programación, la colaboración, etc.

## **El software SPAGOBI consta de diferentes módulos:**

**SpagoBi Server:** Es el núcleo central de SpagoBi el cual integra la funcionalidad de los diferentes motores de base de datos, la gestión de usuarios y roles, integración de módulos open Source y propietarias, funcionalidades de análisis, gestión para el control de versiones y la aprobación del flujos de trabajo.

**SpagoBi Studio:** Es un entorno de desarrollo integrado que permite a los desarrolladores diseñar y editar las plantillas de los distintos motores de análisis, tales como informes, gráficos, análisis geográfico y los informes del tablero de instrumentos. Este módulo es compatible con la fase de despliegue, donde se ponen a prueba los materiales de análisis publicados en el SpagoBI Server, y con el que SpagoBI Estudio interactúa con SpagoBI SDK.

**SpagoBi Meta:** Este módulo se ha sido diseñado específicamente para la gestión y consulta de metadatos. La plataforma permite la gestión de metadatos técnicos y los metadatos de negocio, permitiendo a los administradores y usuarios de interpretar adecuadamente sus fuentes de datos. Los modelos generados con SpagoBI Meta, se exportan a SpagoBI Server para que puedan ser interrogados y usados para la generación de todas las funciones de análisis.

**SpagoBI SDK:** se utiliza para la integración de los servicios prestados por SpagoBI Server. Permite la integración de los documentos a través de una gama de servicios web y publicar documentos en portales SpagoBI y aplicaciones externas.

### 2.6.2 Módulos SpagoBI a instalar.

Para el desarrollo de este TFG se han considerado dos módulos de la suite, SpagoBI Server el cual es el núcleo de la suite y SpagoBI Meta, para la creación de los modelos de datos.

### 2.7 Instalación de SpagoBI Server.

La instalación de SpagoBI Server [3] es bastante sencilla, solo hay que descargar el paquete de la versión más reciente de SpagoBI desde OW2 Consortium Project Forge (<http://forge.ow2.org>).

Este paquete contiene una instalación estándar de SpagoBI que corre sobre Tomcat, todos los motores ya configurados y una demo de las funcionalidades de SpagoBI en un repositorio HSQLDB.

Por defecto la suite de SpagoBI utiliza el motor de base de datos HSQLDB, para este Trabajo crearemos un conector para MySQL y eliminaremos todos los datos de ejemplo pre configurados de la suite, a fin de obtener una versión limpia solo con los informes y gráficos necesarios para este proyecto.



## Pasos a seguir para configurar SpagoBI con repositorio MySQL.

1 – Descargar el conector MySQL desde la página <http://dev.mysql.com/get/Downloads/Connector-J/mysql-connector-java-5.1.34.tar.gz> y copiarlo en la carpeta \All-In-One-SpagoBI-5.2.0\lib

2 – Descargar los siguientes scripts sql desde [http://forge.ow2.org/project/download.php?group\\_id=204&file\\_id=20982](http://forge.ow2.org/project/download.php?group_id=204&file_id=20982)  
MySQL\_create.sql  
MySQL\_create\_quartz\_schema.sql  
MySQL\_create\_social.sql

3 – Desde MySQL crear un esquema llamado “spagobi”

Ejecutar los siguientes scripts descargados en el paso 2 con el fin de crear las tablas necesarias para el repositorio:

MySQL\_create.sql  
MySQL\_create\_quartz\_schema.sql  
MySQL\_create\_social.sql

4 – Modificar el fichero “/conf/server.xml” con el siguiente contenido:

```
<Resource name="jdbc/spagobi" auth="Container"  
type="javax.sql.DataSource" driverClassName="com.mysql.jdbc.Driver"  
url="jdbc:mysql://localhost:3306/spagobi"  
username="root" password="lucifer" maxActive="20" maxIdle="10"  
maxWait="-1"/>
```

```
<Resource name="jdbc/mySql" auth="Container"  
type="javax.sql.DataSource"  
driverClassName="com.mysql.jdbc.Driver"  
url="jdbc:mysql://localhost:3306/nadal"  
username="root"  
password="lucifer"  
maxActive="20" maxIdle="10"  
maxWait="-1"/>
```

5 – Modificar el fichero “/webapps/SpagoBI/META-INF/context.xml” con el contenido siguiente:

```
<Context docBase="SpagoBIProject" path="/SpagoBI" privileged="true"  
reloadable="true" source="org.eclipse.jst.j2ee.server:SpagoBIProject">  
<ResourceLink global="jdbc/spagobi" name="jdbc/spagobi"  
type="javax.sql.DataSource"/>  
<ResourceLink global="spagobi_resource_path" name="spagobi_resource_path"  
type="java.lang.String"/>  
<ResourceLink global="spagobi_sso_class" name="spagobi_sso_class"  
type="java.lang.String"/>
```

```

<ResourceLink global="spagobi_host_url" name="spagobi_host_url"
type="java.lang.String"/>
</Context>

```

6 – Modificar el fichero “ /opt/spagobi/webapps/SpagoBI/WEB-INF/classes/hibernate.cfg.xml” configurando el conector de base de datos hacia MySQL

Actualizar solo las líneas siguientes:

```

<hibernate-configuration>
<session-factory name="HibernateSessionFactoryMySQL">
<property
name="hibernate.connection.datasource">java:/comp/env/jdbc/spagobi</property
>
<property
name="hibernate.dialect">org.hibernate.dialect.MySQLDialect</property>
<property name="hibernate.cache.use_second_level_cache">>false</property>
<property name="hibernate.cache.use_query_cache">>false</property>
<!--
Dejar del resto de la configuración tal como está.
-->
</session-factory>
</hibernate-configuration>

```

7 – Modificar el fichero “/opt/spagobi/webapps/SpagoBI/WEB-INF/classes/jbpm.hibernate.cfg.xml”  
Dejar el resto de tags como están.

```

<hibernate-configuration>
<session-factory>
<property name="hibernate.dialect">org.hibernate.dialect.MySQLDialect</property>
<property
name="hibernate.connection.datasource">java:comp/env/jdbc/spagobi</property>
<property name="hibernate.cache.use_second_level_cache">>false</property>
<property name="hibernate.cache.use_query_cache">>false</property>
<!--
Leave rest of the Configuration as it is.
-->
</session-factory>
</hibernate-configuration>

```

8 – Modificar el fichero “/opt/spagobi/webapps/SpagoBI/WEB-INF/classes/quartz.properties”  
Comentar la línea siguiente:

```

org.quartz.jobStore.driverDelegateClass=org.quartz.impl.jdbcjobstore.HSQLDBDe
legate

```

Quitar comentario a la línea siguiente:

```

Org.quartz.jobStore.driverDelegateClass=org.quartz.impl.jdbcjobstore.StdJDBCDe
ele

```

9 - Crear la variable de entorno JAVA\_HOME y CATALINA\_HOME.

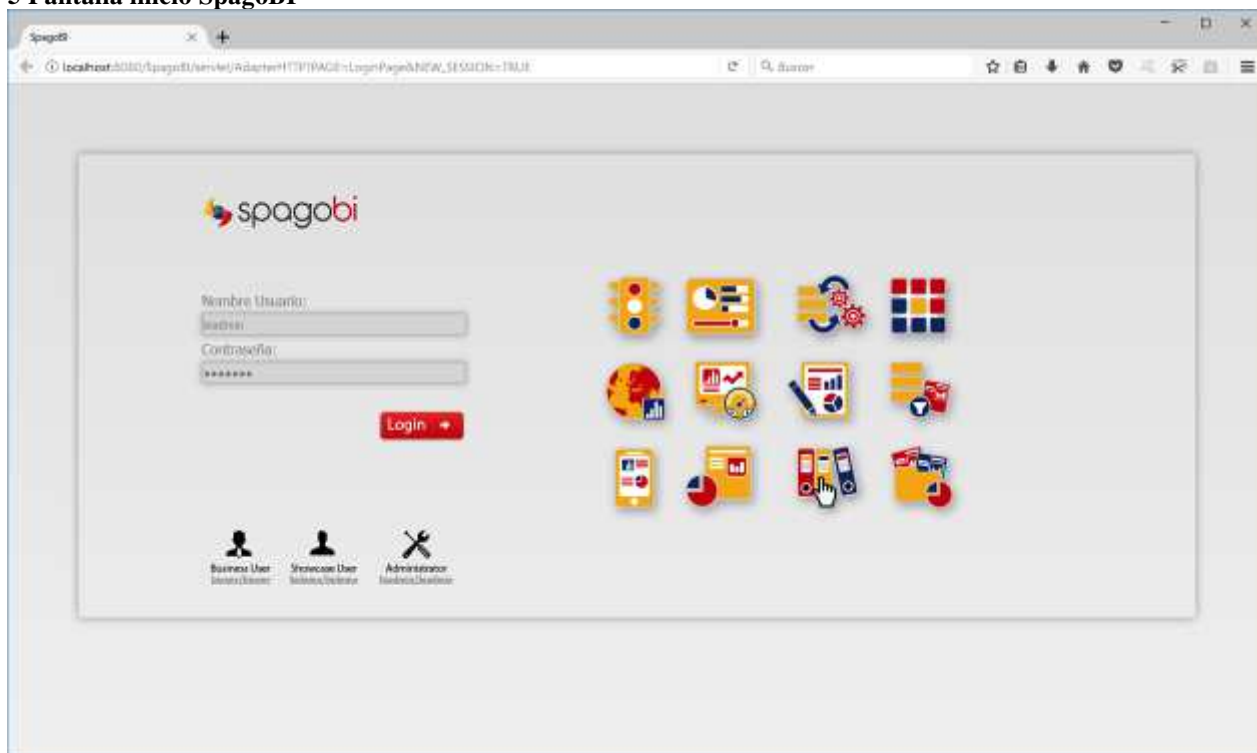
10- Por último modificar el arranque de SpagoBi “SpagoBIStartup.bat”, para que no ejecute la base de datos por defecto.

```
rem set CD=%cd%
rem cd %CATALINA_HOME%\database
rem start "HSQLDB" /max cmd.exe /c start.bat
```

### Arranque de SpagoBI Server

Ejecutar el script “SpagoBIStartup.bat” que se encuentra en la carpeta “All-In-One-SpagoBI-5.2.0\apache-tomcat-6.0.18\bin”, y una vez arrancado el servidor Tomcat abrir el navegador y escribir la url “http://localhost:8080/SpagoBI”. A continuación iniciar sesión usando la siguiente cuenta de administrador: biadmin (contraseña: biadmin).

### 5 Pantalla inicio SpagoBI



### Instalación del módulo SpagoBI Meta.

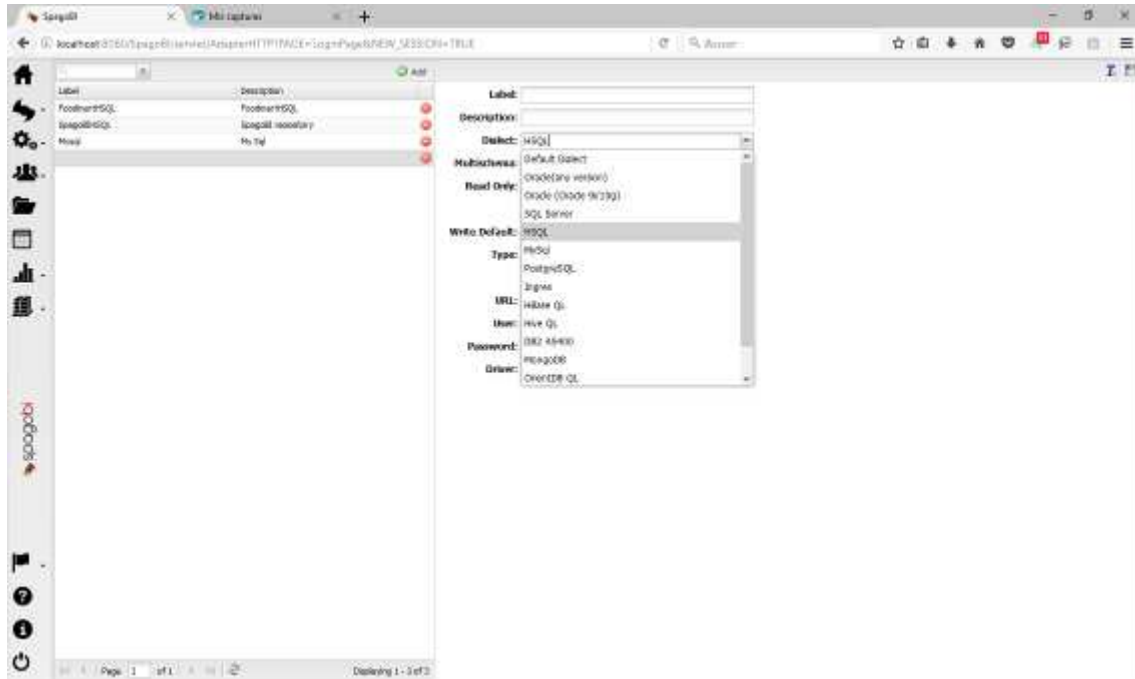
Descargar el fichero zip SpagoBIMeta\_5.2.0\_win64\_12042016.zip, desde la web <http://forge.ow2.org>

Descomprimir en una carpeta y arrancar con el ejecutable “SpagoBI.exe”

## 2.8 Creación y prueba del conector de datos.

El primer paso es definir la conexión de datos que vamos a utilizar (MySQL):

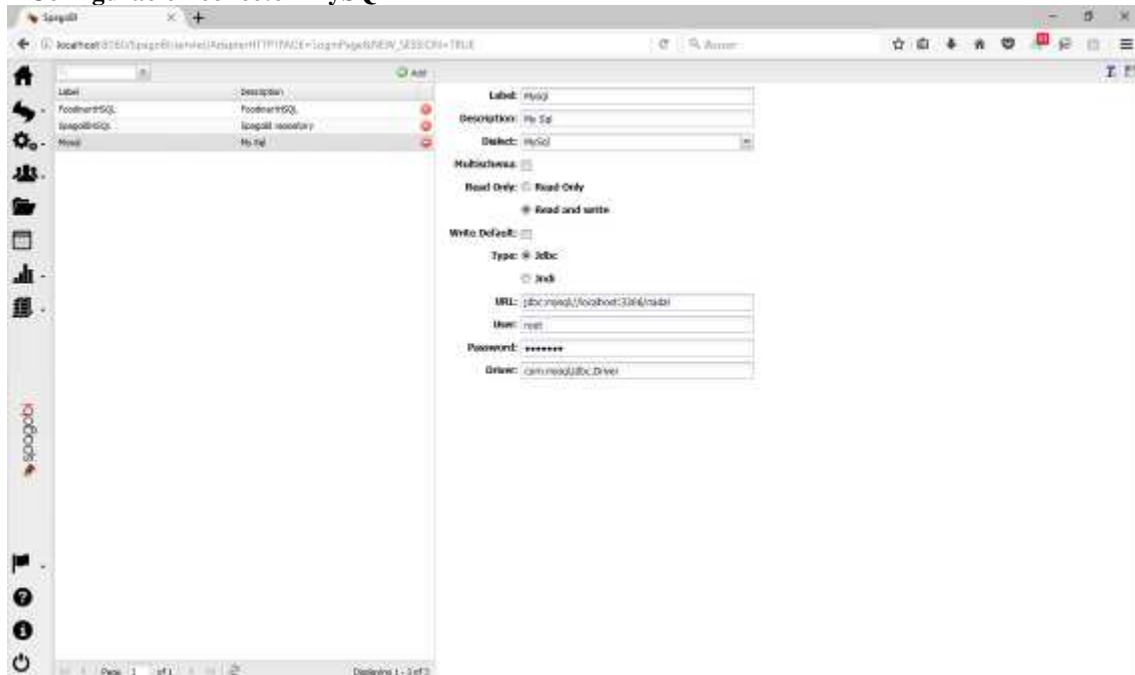
### 6 Conexión de datos



Se ha añadido a la plataforma la fuente de datos MySQL puesto que no existe ningún conector para Microsoft Access, desde donde se han depurado los datos de entrada.

En SpagoBI se ha configurado el conector con MySQL de esta forma (configuración data Source):

### 7 Configuración conector MySQL



A partir de aquí podemos ejecutar las siguientes fases del proyecto:

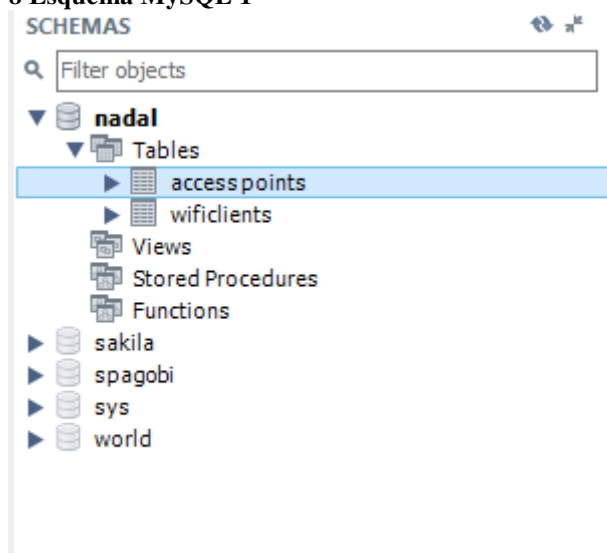
- Preparación de datos
- Creación del modelo (Tarea de minería de datos propiamente dicha).
- Evaluación e interpretación del modelo
- Integración de los resultados en el proceso

## 2.9 Preparación de datos:

Gracias a que la plataforma de BI SpagoBi acepta datos de diferentes sistemas de bases de datos, se ha instalado la base de datos MySQL versión community (la decisión de instalar MySQL viene dada por el conocimiento de esta herramienta y por ser Open Source) . Los ficheros CSV obtenidos de la plataforma Meraki se han importado a MySQL y se ha trabajado en la depuración y modelado de los datos, los cuales mediante el conector definido en el punto 2.8 se utilizarán para la creación del modelo de BI.

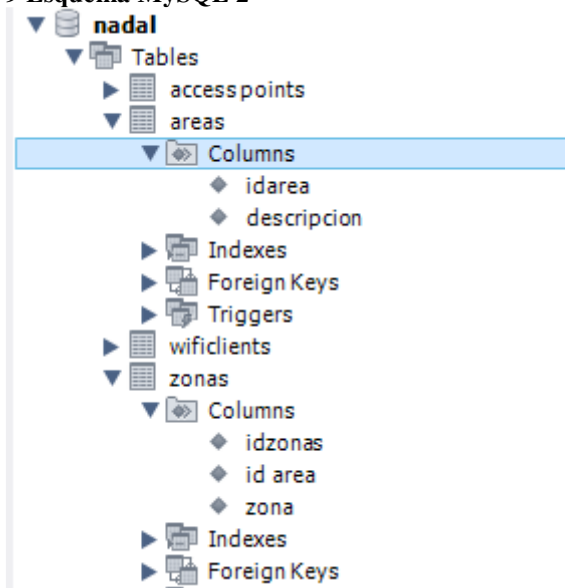
Los datos obtenidos de la plataforma Meraki en formato CSV se han importado en un esquema MySQL “nadal” creando las siguientes tablas:

### 8 Esquema MySQL 1



Una vez importados estos datos se ha decidido crear dos tablas más con el objetivo de crear un modelo relacional más adecuado para explotar los mismos. Para ello se han extraído los diferentes valores de área y zona y se han asignado claves foráneas a la tabla de ‘Access points’, quedando el esquema con las tablas que se muestran a continuación:

## 9 Esquema MySQL 2



Se han eliminado datos no significativos e información personal (mac address del usuario observado).

Mediante la utilización de herramientas de este software (consultas) se ha procedido a su depuración y unificación. Para ello se han realizado las siguientes tareas:

- Eliminación de registros duplicados o con información nula.
- Conversión de formatos de campos (fechas, numéricos, etc.)

En esta fase y como trabajo de preparación de datos se ha visto que era imprescindible añadir atributos para poder hacer estudios de uso por franjas horarias y días.

Se han añadido tres nuevos atributos los cuales se calculan automáticamente dependiendo de los datos de la fila:

**Dia\_sem\_Is:** Contiene el día de la semana en formato texto del atributo **Last seen**.

**Hora\_dia\_Is:** Contiene la hora obtenida del atributo **Last seen**.

**Días\_transc:** Es un campo calculado con la diferencia en días transcurridos entre **Last seen** y **First seen**.

Las tablas principales han quedado con los atributos siguientes:

### Table: wificlients

#### Columns:

Last seen	datetime
Usage	int(11)
OS	varchar(25)
IPv4 address	varchar(20)
MAC address	varchar(20)
Manufacturer	varchar(40)
Connected to	varchar(60)
First seen	datetime
Dia_sem_ls	varchar(10)
hora_dia_ls	varchar(2)
dias_transc	double

### Table: access points

#### Columns:

zona	int(11)
Area	int(11)
<u>Name</u>	varchar(60) PK
MAC	varchar(20)
Model	varchar(20)
Product description	varchar(20)
Latitude	double
Longitude	double
Internet address	varchar(20)
LAN IP	varchar(20)

Aun cuando SpagoBi nos proporciona herramientas ETL (engine Talend), se ha decidido hacer la transformación de datos en MySQL, puesto que para este cometido es necesario usar otro módulo “Talend Open Studio” y no es objeto de este trabajo integrar esta herramienta. Estos atributos de han creado en MySQL y la tabla de observaciones han quedado con la siguiente estructura:

## 10 Estructura WifiClients

Column	Type	Default Value	Nullable	Character Set	Collation	Privileges	Extra
Last seen	datetime		YES			select,insert,update,references	
Usage	int(11)		YES			select,insert,update,references	
OS	varchar(25)		YES	utf8	utf8_general_ci	select,insert,update,references	
IPv4 address	varchar(20)		YES	utf8	utf8_general_ci	select,insert,update,references	
MAC address	varchar(20)		YES	utf8	utf8_general_ci	select,insert,update,references	
Manufacturer	varchar(40)		YES	utf8	utf8_general_ci	select,insert,update,references	
Connected to	varchar(60)		NO	utf8	utf8_general_ci	select,insert,update,references	
First seen	datetime		YES			select,insert,update,references	
Dia_sem_ls	varchar(10)		YES	utf8	utf8_general_ci	select,insert,update,references	
hora_dia_ls	varchar(2)		YES	utf8	utf8_general_ci	select,insert,update,references	
dias_transc	double		YES			select,insert,update,references	

Este será el origen de datos o repositorio desde donde se creará el modelo de datos que será utilizado por SpagoBi para desarrollar el trabajo de BI.

Las dimensiones de las tablas son:

- Access points: 197 filas, una para cada punto de acceso.
- WifiClients: 28746 filas, una para cada observación correspondiente a la última semana del mes de Octubre de 2017.
- Areas: 2 filas una para cada area.
- Zonas:

## 11 Cardinalidad de las tablas

#	Time	Action	Message
1	19:34:52	SELECT * FROM nadal.'access points' LIMIT 0, 50000	197 row(s) returned
2	19:36:32	SELECT * FROM nadal.wificlients LIMIT 0, 50000	28746 row(s) returned

Se puede ver a continuación una muestra de los datos contenidos en cada tabla.

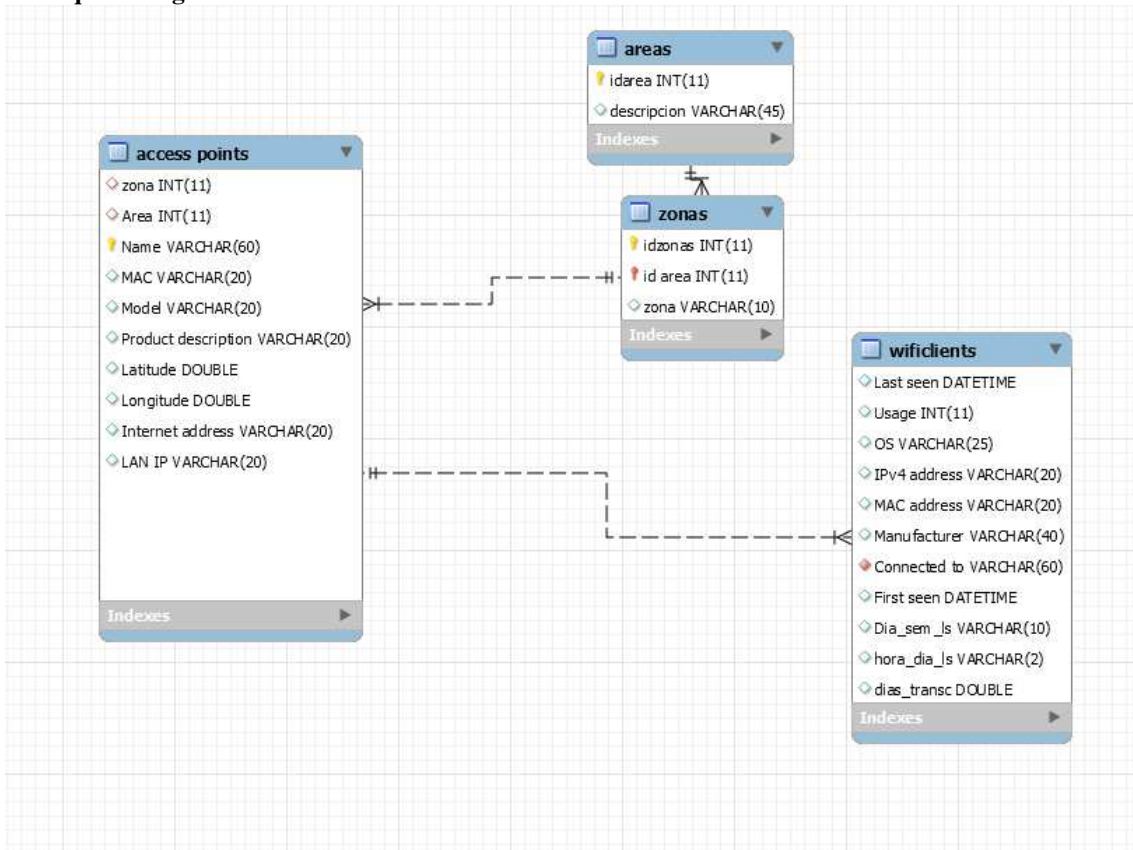


zona	Area	Name	MAC	Model	Product description
7-Palma	Palma	Anima Beach - Gat. Anima	00:18:0a:03:0d:06	MR62	Meraki MR62 Cloud Ma
7-Palma	Palma	Anima Beach - Gat. Plava	00:18:0a:03:0c:6c	MR62	Meraki MR62 Cloud Ma
7-Palma	Palma	Anima Beach - Gat. Plava de Gesa	88:15:44:2c:4c:d0	MR72	Meraki MR72 Cloud Ma
0-Palma	Palma	Evento Skoda Mar de Nudos - Nudos Club	e0:55:3d:ee:42:7b	MR30H	Meraki MR30H Cloud M
0-Palma	Palma	Evento Skoda Mar de Nudos - Sal3n	e0:55:3d:ee:4a:91	MR30H	Meraki MR30H Cloud M
2-Palma	Palma	Gat. Antoni Maura - Enco	88:15:44:2c:19:90	MR72	Meraki MR72 Cloud Ma
1-Palma	Palma	Gat. Antoni Maura - Huoo Boss	88:15:44:2c:1d:c0	MR72	Meraki MR72 Cloud Ma
1-Palma	Palma	Gat. Antoni Maura - Lennox Pub	00:18:0a:5b:60:e0	MR72	Meraki MR72 Cloud Ma
1-Palma	Palma	Gat. Antoni Maura - Palma on Bike	00:18:0a:5b:86:a0	MR72	Meraki MR72 Cloud Ma

Last seen	Usage	OS	IPv4 address	MAC address	Manufacturer	Connected to
2017-10-24 11:41:05	5930	Android	10.181.194.184	c8:14:51:c3:e8:b6	Huawei Technoloaies	Gat. The Guinness House - Almudaina
2017-10-24 11:43:31	3816	Android	10.251.114.12	84:11:9e:ef:fb:b1	Samsuna	Gat. Pesuero
2017-10-24 11:44:30	377	Apple iPhone	10.232.141.153	d4:f4:6f:ef:4d:24	Apple	Gat. Antoni Maura - Lennox Pub
2017-10-24 11:44:41	0	Apple iPhone	10.57.132.212	1c:91:48:61:35:7d	Apple	Gat. Plaza Juan Carlos I
2017-10-24 11:44:54	1194	Android	10.169.187.7	70:4d:7b:ab:94:29	ASUS	Gat. EM4 Facturacion - Central
2017-10-24 11:45:12	4632	Android	10.28.248.162	30:19:66:d8:85:cf	Samsuna	Gat. EM1 Embaroue 9-10
2017-10-24 11:46:16	34503	Apple iPhone	10.68.31.110	9c:35:eb:81:8a:ac	Apple	Gat. EM1 Embaroue 5
2017-10-24 11:47:46	1763	iOS	10.186.136.178	70:70:0d:0e:3c:8d	Apple	Gat. EM1 Embaroue 11
2017-10-24 11:47:46	5669	iOS	10.173.130.200	68:db:ca:91:b9:4a	Apple	Gat. EM1 Embaroue 11
2017-10-24 11:47:47	412	Apple iPhone	10.108.231.70	40:4d:7f:a5:c4:d3	Apple	Gat. Plaza Juan Carlos I

Por lo tanto el esquema l3gico que tenemos en MySQL es el siguiente:

## 12 Esquema l3gico



## 2.10 Modelado de datos:

El modelado de datos es esencial para las posteriores fases del BI, consiste en crear un modelo lógico de datos a partir de los datos físicos, el cual va a servir para estudiar el conocimiento de los datos y generar información.

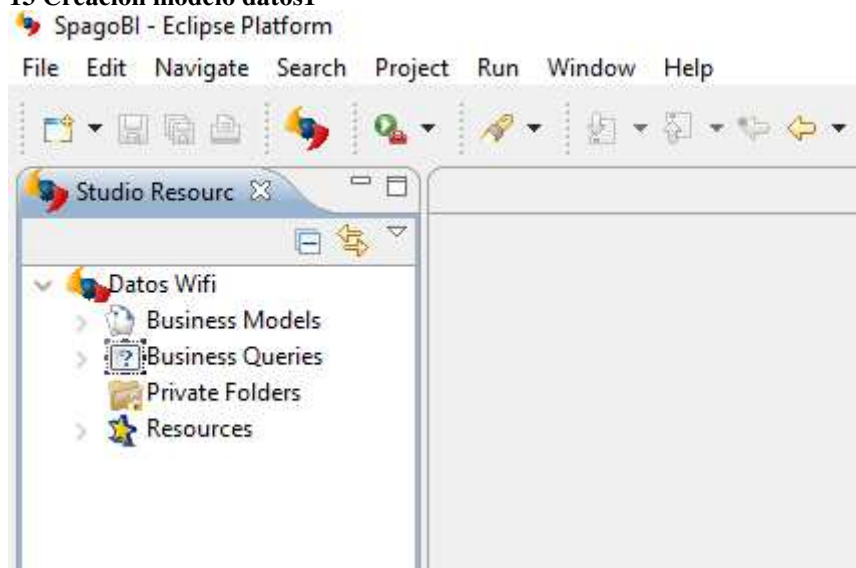
Uno de los modelos de datos más utilizados es el cubo OLAP [9] (*OnLine Analytical Processing*): El cubo OLAP tiene origen en la definición que hizo [Edgar Frank Codd](#), un científico inglés considerado el padre de las bases de datos relacionales. La propuesta de **Codd** consistía en realizar una disposición de los datos en vectores para permitir un análisis rápido de los mismos. Estos vectores son los llamados cubos, donde a cada una de sus caras se le llama dimensión. Esta forma de ver los datos supera las limitaciones de las bases de datos relacionales, ya que podemos crear cubos de más de tres dimensiones llamados hipercubos.

Cada una de las dimensiones de un cubo puede resumirse en una jerarquía, lo cual nos permite explorar los datos de forma que descendiendo en la jerarquía podemos llegar al detalle que nos permite identificar una singularidad. Un ejemplo de jerarquía sería la ubicación física de un hecho, el cual podemos jerarquizar por continente, país, región y, ciudad de forma que podemos ver este agrupado por cada uno de estos niveles.

En SpagoBi disponemos de una herramienta SpagoBIMeta para crear cubos OLAP, Querys y otros modelos de datos.

La forma que tenemos en SpagoBIMeta de crear modelos de datos es la siguiente:

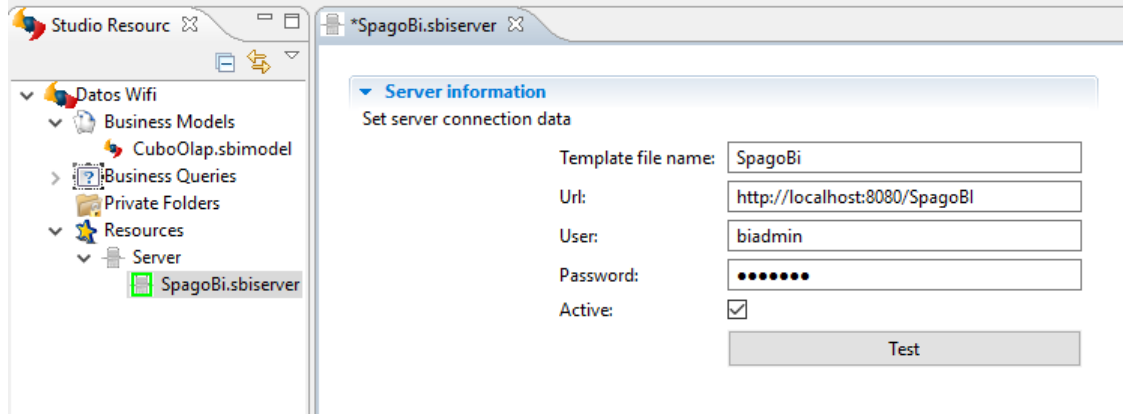
### 13 Creación modelo datos1



Desde la opción File creamos un nuevo proyecto al que llamamos Datos Wifi, este proyecto puede tener diferentes modelos, querys almacenadas y recursos.

En primer lugar hay que definir la conexión con el servidor SpagoBI, de forma que podemos subir los modelos creados a este.  
La conexión al servidor se define en “Resources”.

#### 14 Conexión SpagoBi Meta SpagoBi Server



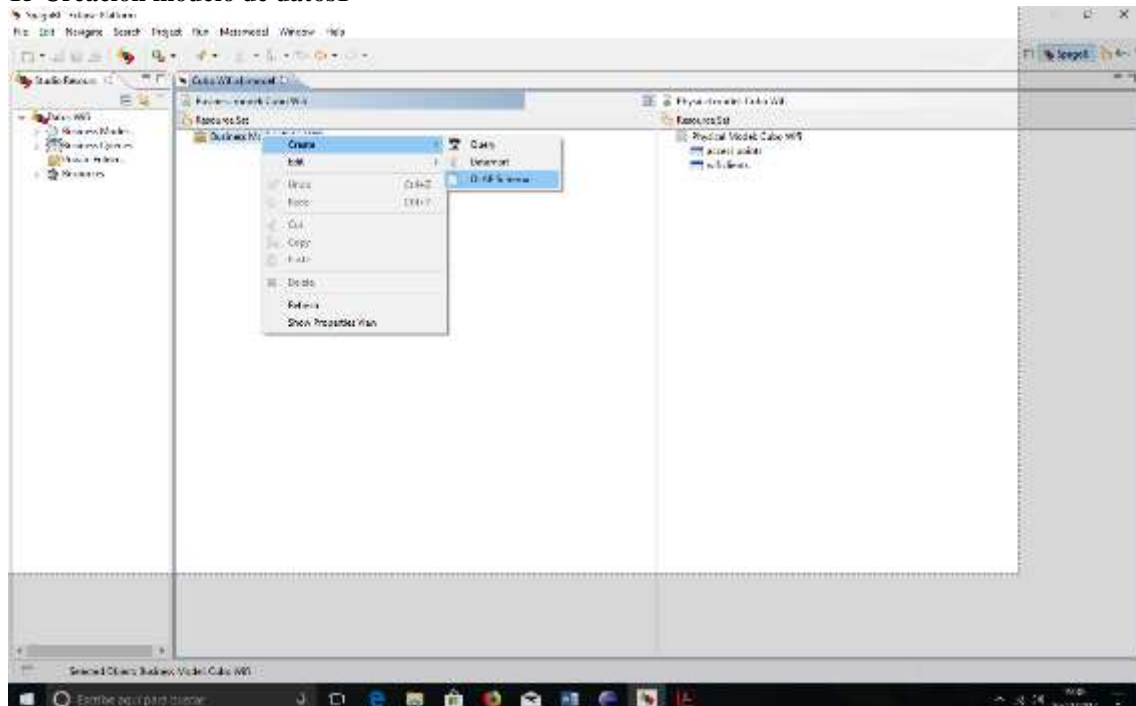
En esta opción le informamos a SpagoBi Meta donde se encuentra el servidor SpagoBi Server y probamos la conexión.  
Con esto ya estamos en disposición de comunicarnos con el servidor SpagoBi y actualizar los modelos que vayamos creando.

Seguidamente creamos el primer modelo.

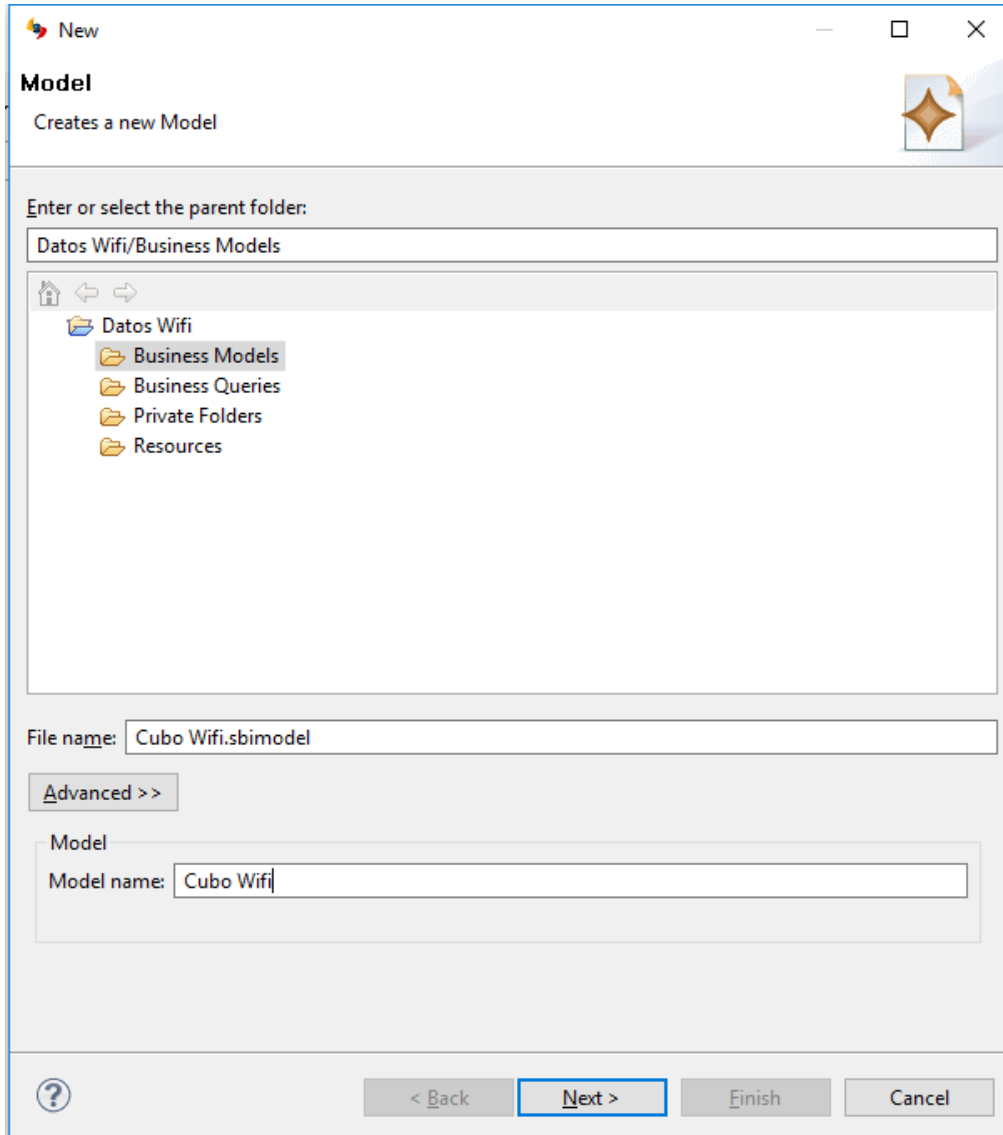
En Business Models crearemos el cubo OLAP, pulsando el botón derecho en Business Models y eligiendo “new model”.

A partir de aquí seleccionamos con el botón derecho sobre Business Models, la opción OLAP Schema.

#### 15 Creación modelo de datos1

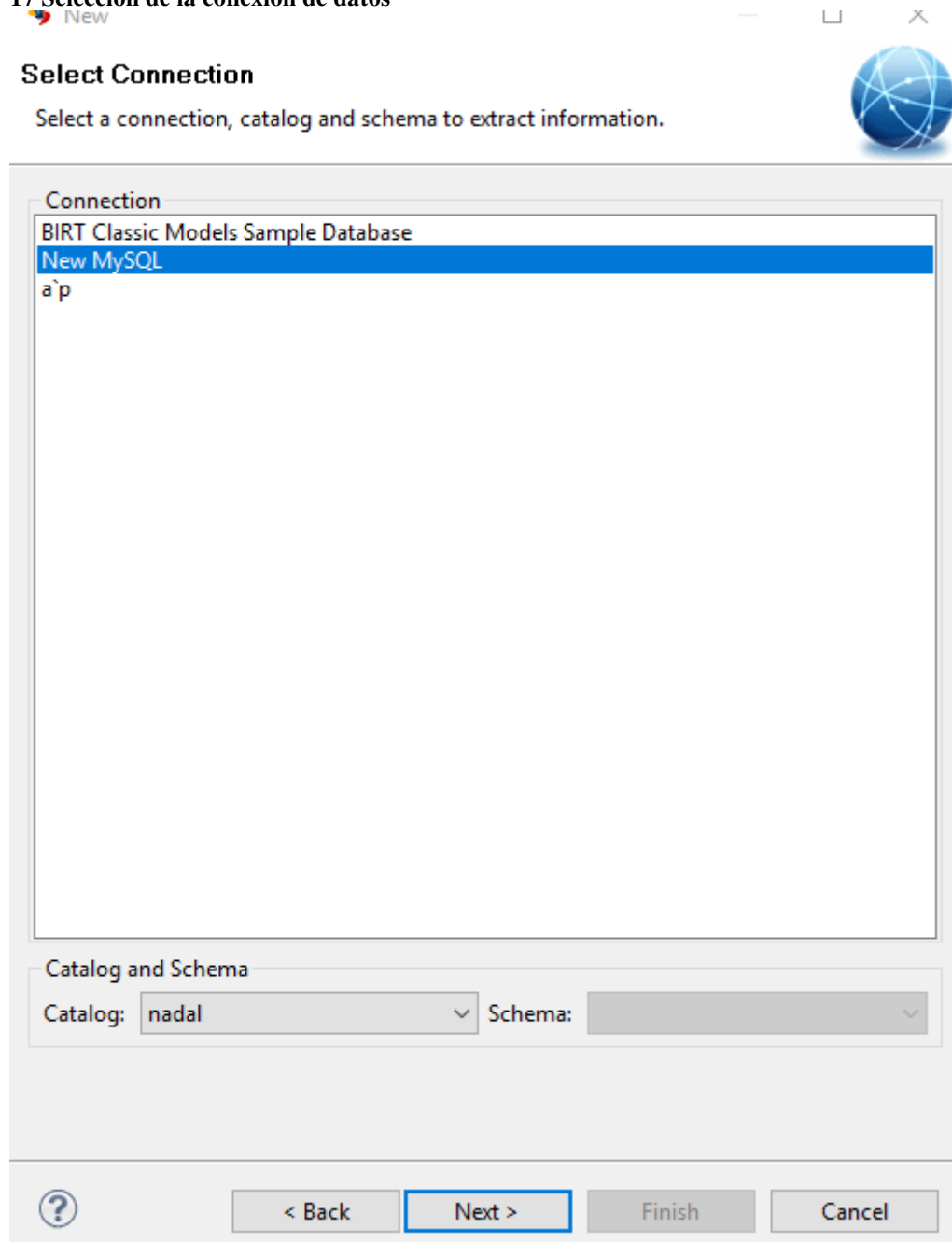


## 16 Creación modelo de datos 2



Llamamos a nuestro modelo Cubo Wifi y al pulsar Next, nos pedirá la conexión de datos que vamos a usar, de la cual se obtendrán los datos físicos para crear el modelo. Dado que nuestros datos se encuentran en MySQL, seleccionamos esta opción:

## 17 Selección de la conexión de datos



Una vez seleccionada la conexión nos mostrara las tablas contenidas en el esquema:

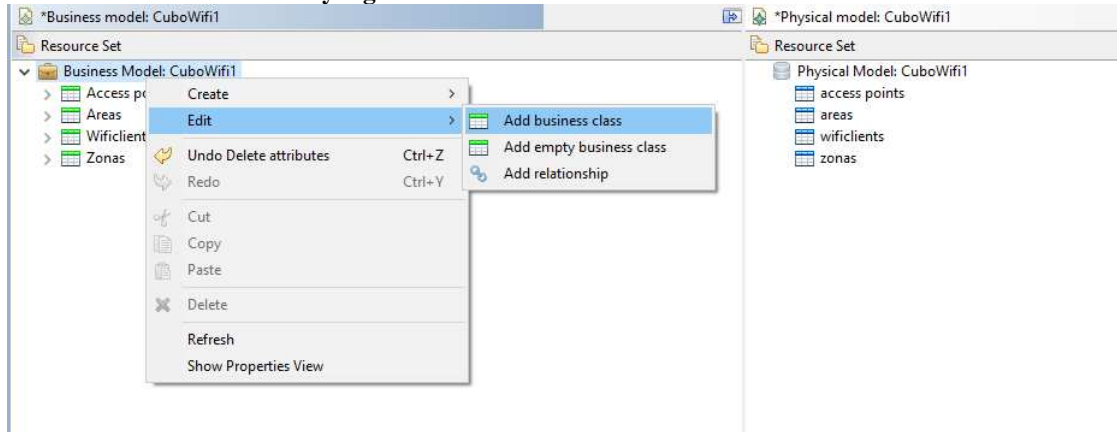
## 18 Selección de tablas físicas



Podemos ver que nos muestra las tablas que hay en el esquema seleccionado y nos permite seleccionarlas, en nuestro caso vamos a seleccionar todas las tablas.

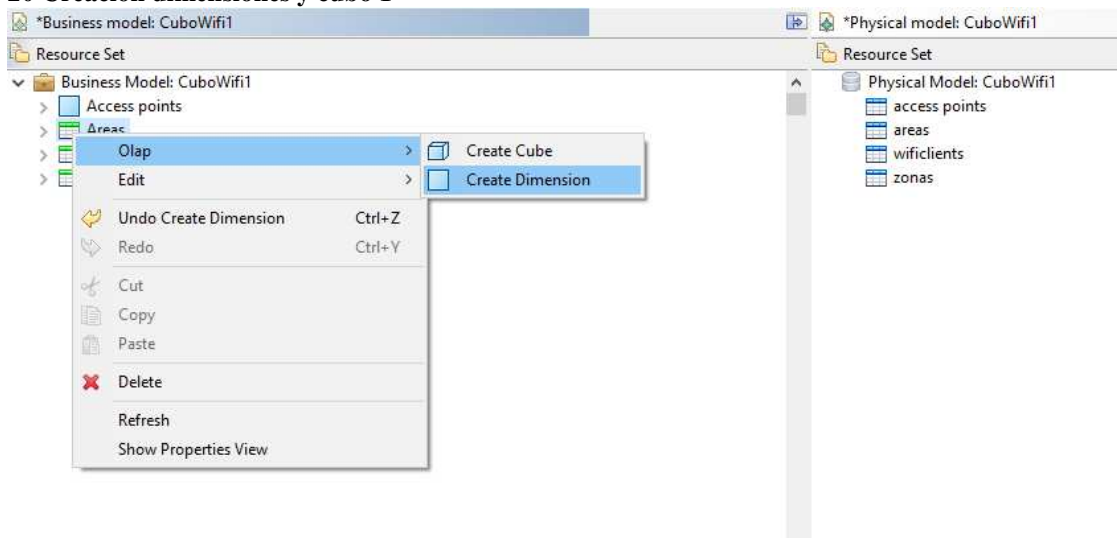
Una vez hecho esto vamos a asociar las tablas físicas con las lógicas pulsando edit en el icono Business Model y seleccionando "add business class": hasta tener las tablas lógicas:

## 19 Asociación tablas físicas y lógicas



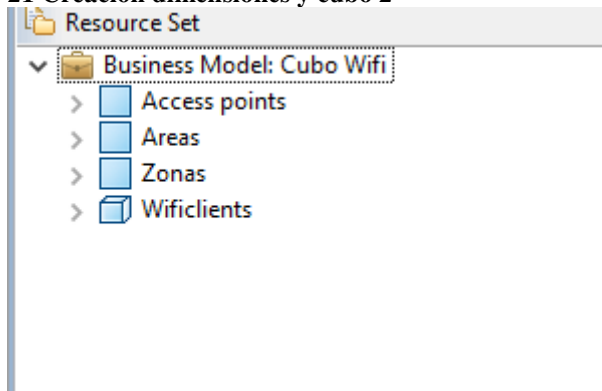
Vamos a proceder a crear las dimensiones del cubo y el cubo.

## 20 Creación dimensiones y cubo 1



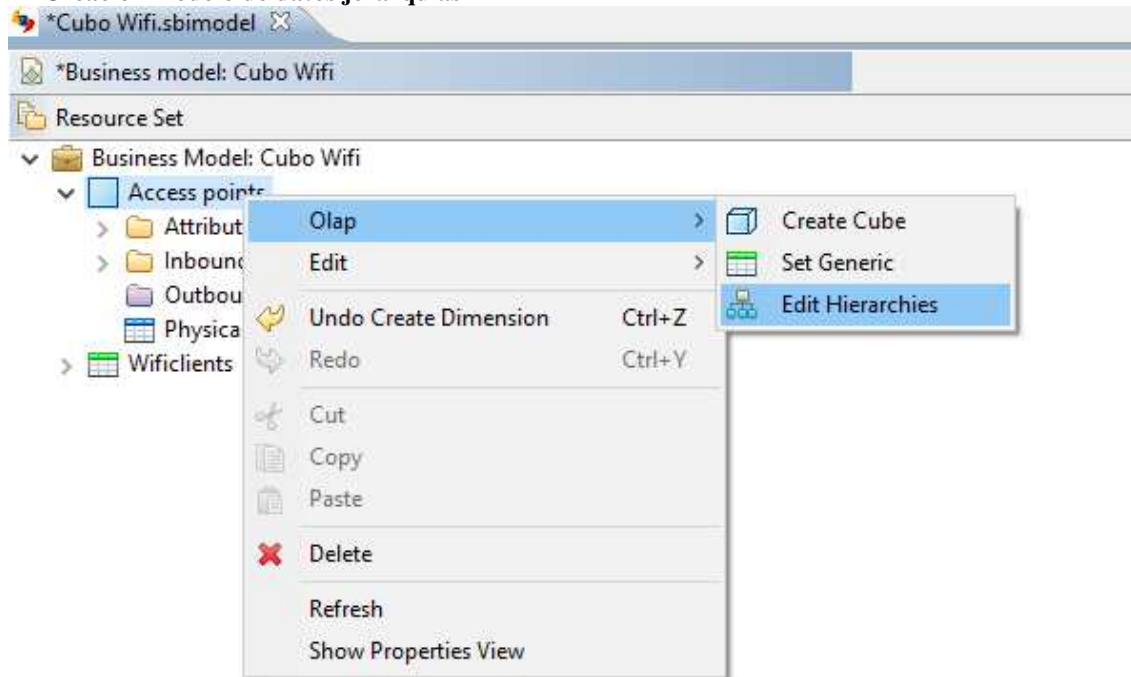
Continuamos creando dimensiones y el cubo hasta que tenemos esta estructura:

## 21 Creación dimensiones y cubo 2

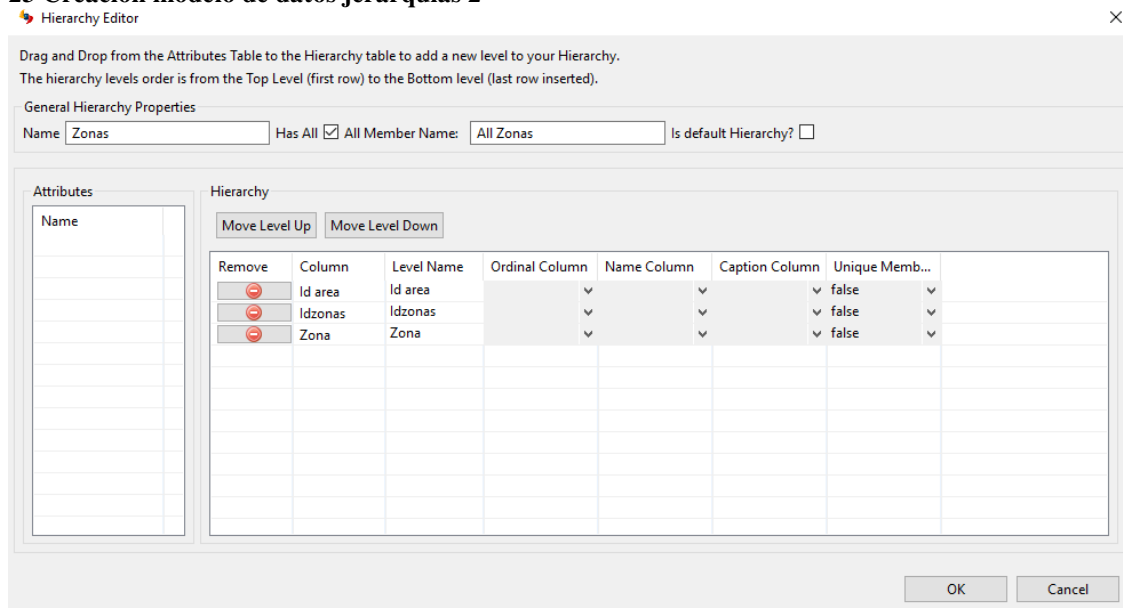


En las dimensiones creamos jerarquías

## 22 Creación modelo de datos jerarquías 1



## 23 Creación modelo de datos jerarquías 2



Vemos que hemos creado una jerarquía donde id Area es el primer nivel y idzonas es el segundo nivel.

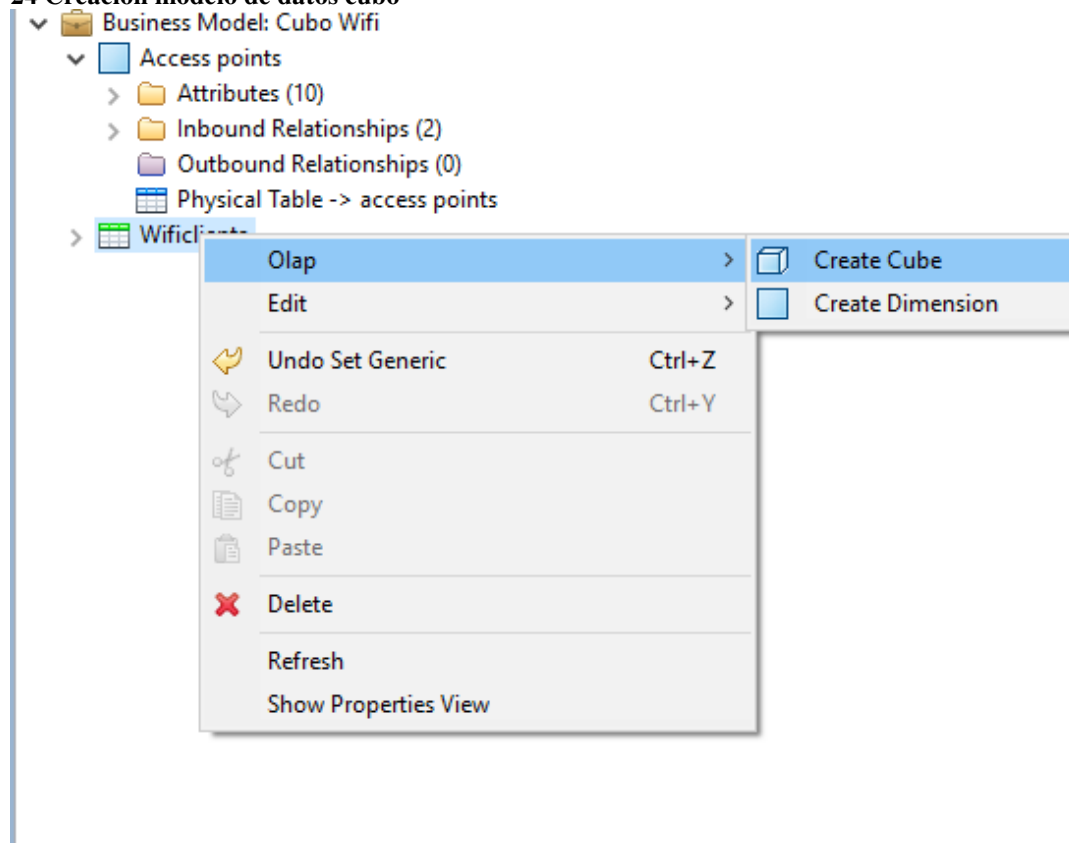
Este paso lo repetimos para cada dimensión

El siguiente paso es definir la clase WifiClients como cubo.

También aprovechamos para eliminar atributos que no vamos a usar en el análisis y a renombrar algunos para que sean más entendibles.

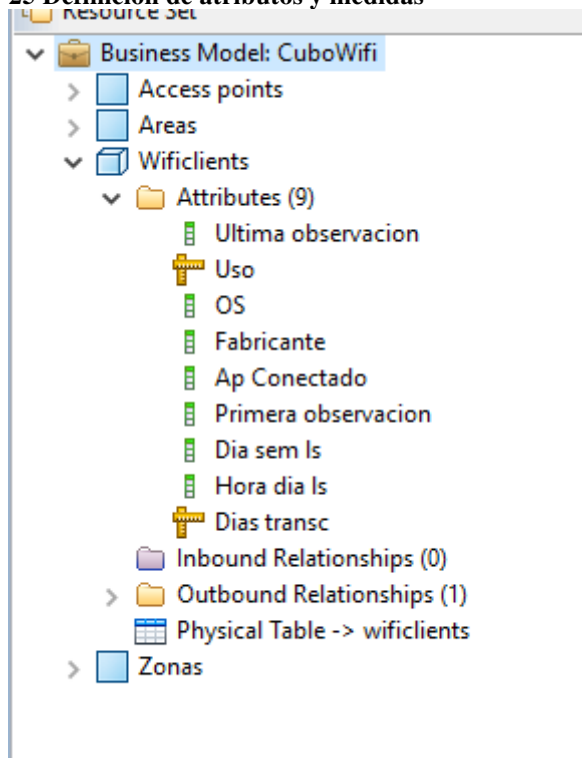


## 24 Creación modelo de datos cubo



En el cubo a cada campo le indicamos si lo vamos a usar como medida o como atributo.

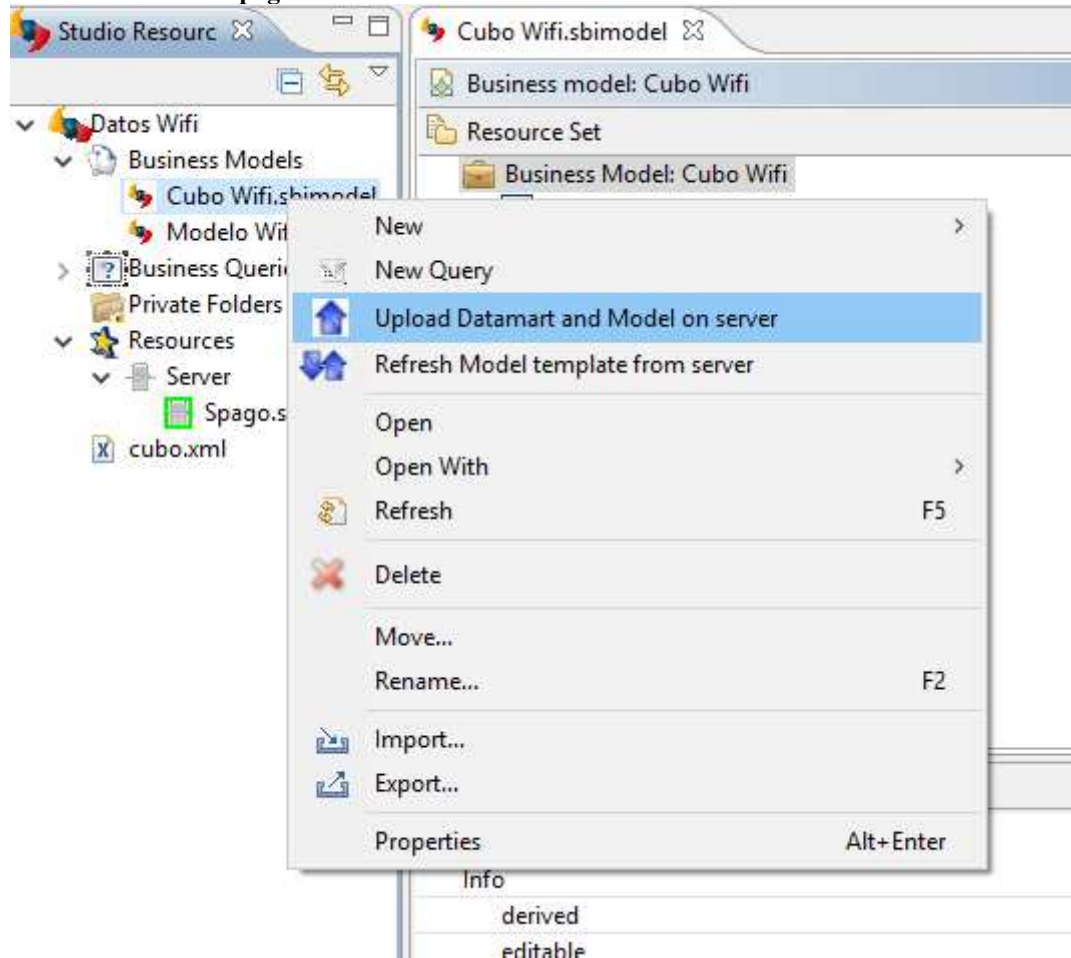
## 25 Definición de atributos y medidas



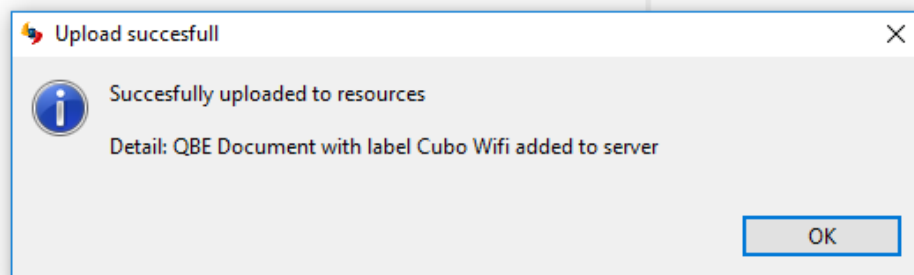
En este caso hemos definido **Uso** y **Dias transc** como medida y el resto de campos como atributos.

Una vez hemos terminado las definiciones del modelo hay que subirlo al servidor SpagoBI Server para poder ser utilizado, para ello utilizamos la opción "Update Data Mart and Model on Server":

## 26 Subir modelo a SpagoBI Server 1



## 27 Subir modelo a SpagoBI Server 2

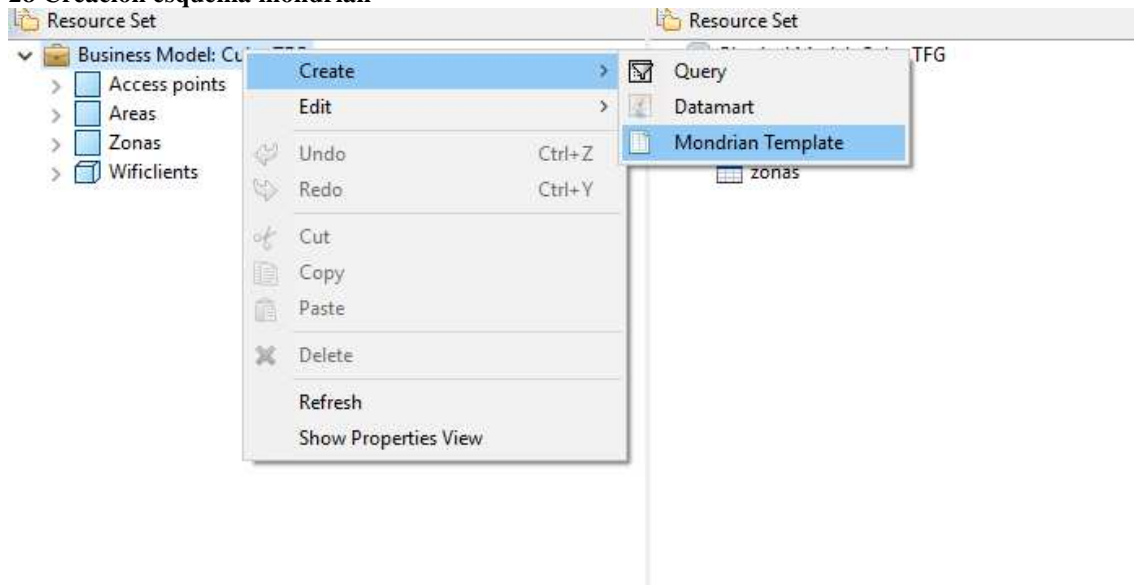


Con este mensaje nos indica que el modelo está disponible en SpagoBI Server. Este modelo lo usaremos como QBE (Query by example), y nos servirá para diseñar Query's de un modo muy sencillo.

Una vez tenemos este modelo de datos que nos servirá para crear los diferentes gráficos e informes requeridos vamos a crear realmente el cubo OLAP, como un esquema "mondrian" XML.

Para crear el esquema modrian o cubo, seleccionamos la opción "Create Mondrian Template".

## 28 Creación esquema mondrian



Este paso nos genera un documento XML el cual contiene la definición del cubo OLAP.

```
<?xml version="1.0" ?>
<SCHEMA name="Cubo TFG">
<Dimension name="Access points">
<Hierarchy hasAll="true"
  allMemberName="Todos los puntos de Acceso"
  primaryKey="Name"
  name="Puntod de Acceso"
  >
  <Table name="access points"/>
  <Level name="Area" column="Area"
    nameColumn="Area"
    ordinalColumn="Area"
    uniqueMembers="false"
  />
  <Level name="Zona" column="zona"
    uniqueMembers="false"
  />
  <Level name="Nombre" column="Name"
    uniqueMembers="false"
  />
</Hierarchy>
</Dimension>
<Dimension name="Areas">
<Hierarchy hasAll="true"
  allMemberName="Todas las Areas"
  primaryKey="idarea"
  name="Areas"
  >
  >
</Hierarchy>
</Dimension>
```

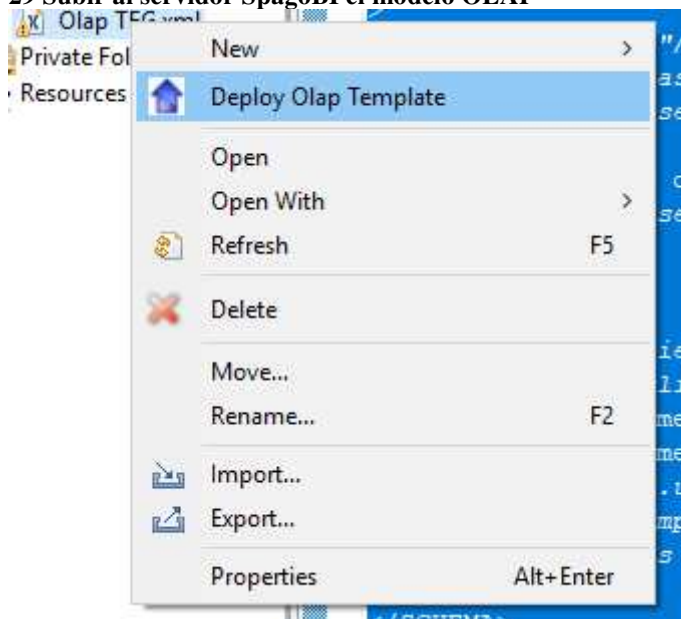
```

<Table name="areas"/>
  <Level name="Area" column="descripcion"
    uniqueMembers="false"
  />
</Hierarchy>
</Dimension>
<Dimension name="Zonas">
<Hierarchy hasAll="true"
  allMemberName="Todas las Zonas"
  primaryKey="id area"
  name="Zonas"
  >
<Table name="zonas"/>
  <Level name="Idzonas" column="idzonas"
    uniqueMembers="false"
  />
  <Level name="Zona" column="zona"
    uniqueMembers="false"
  />
</Hierarchy>
</Dimension>
<Cube name="Wificlients">
<Table name="wificlients"/>
  <DimensionUsage name="Access points" source="Access points"
    foreignKey="Connected to"/>
  <Measure name="Num.usuarios" column="Last seen"
    aggregator="count" formatString="#,###"/>
  <Measure name="Tiempo Uso" column="Usage" aggregator="sum"
    formatString="#,###"/>
  <Measure name="Dias transc" column="dias_transc"
    aggregator="count" formatString="#,###"/>
</Cube>
</SCHEMA>

```

Este XML también lo subimos al servidor SpagoBI utilizando la opción siguiente:

#### 29 Subir al servidor SpagoBI el modelo OLAP

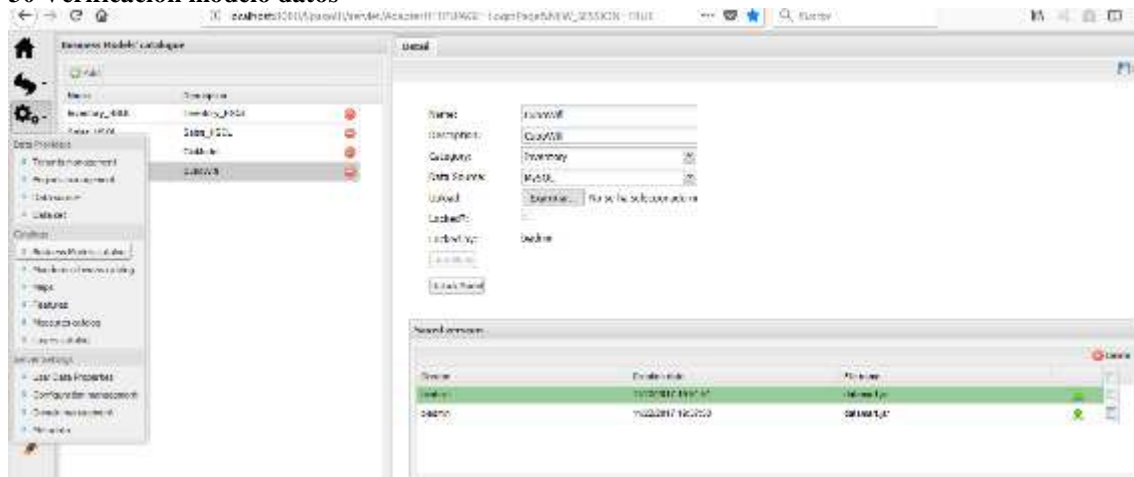


A partir de aquí pasamos a SpagoBi Server para crear los gráficos.

## 2.11 Elaboración de gráficos y cuadro de mandos.

El primer paso es conectarnos a SpagoBI Server y gracias a la conexión de datos que hemos definido en el apartado 2.8, podemos comprobar que el modelo que hemos creado de SpagoBIMeta ha quedado actualizado en el servidor.

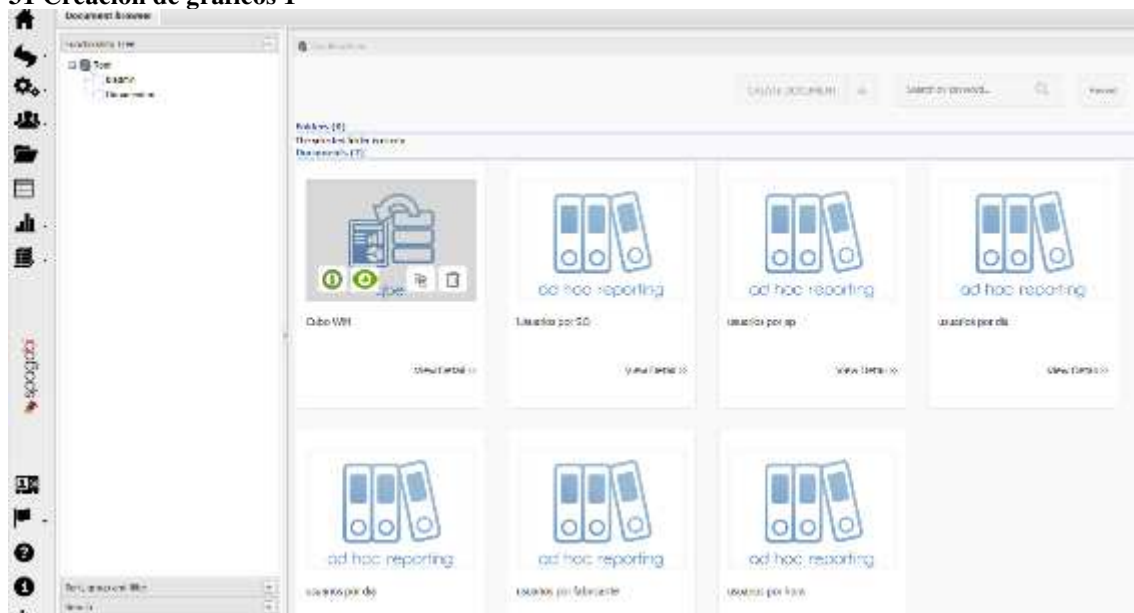
### 30 Verificación modelo datos



Vemos que el modelo de datos está correctamente subido al servidor y la conexión utilizada es la definida anteriormente.

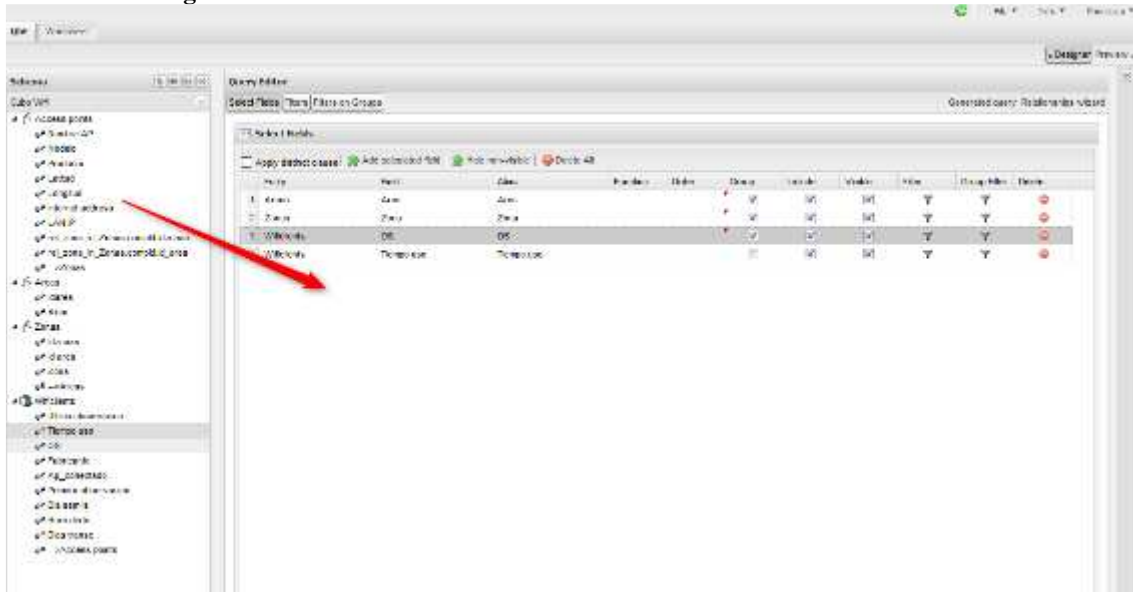
La metodología para realizar un gráfico es muy sencilla:

### 31 Creación de gráficos 1



Seleccionamos el modelo de datos sobre el que vamos a trabajar en este caso el Cubo Wifi.

### 32 Creación de gráficos 2



SpagoBI nos muestra los datos que existen en nuestro modelo y sus relaciones, simplemente vamos arrastrando los datos a la ventana de campos seleccionados, definimos las agrupaciones que necesitamos y la función de cada campo.

### 33 Creación de gráficos 3



Esta es la Query que nos genera.

### 34 Creación de gráficos 4

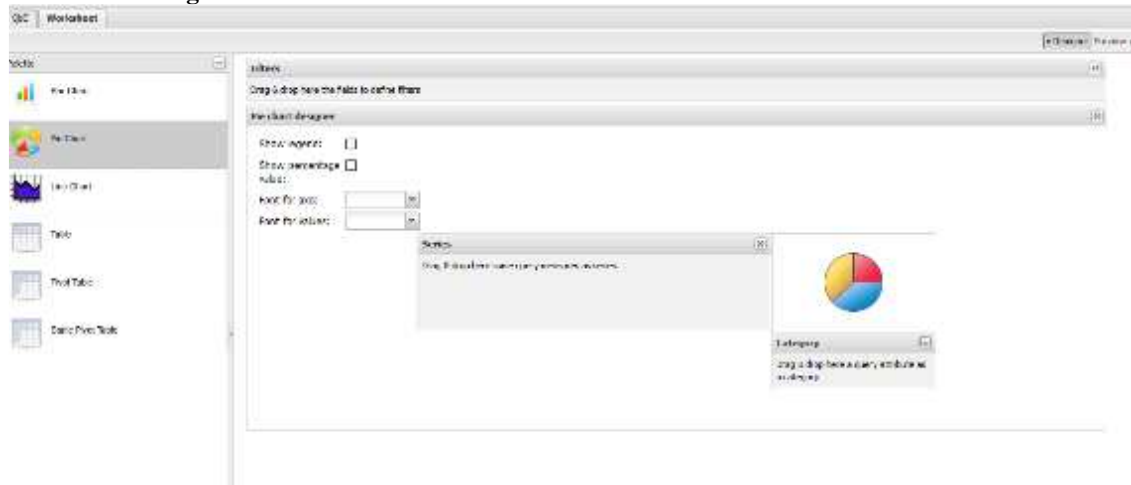
	Area	Zona	OS	Tiempo uso
1	Palma Ciudad	0-PALMA	Android	507
2	Palma Ciudad	0-PALMA	Apple iPad	22,718
3	Palma Ciudad	0-PALMA	Apple iPhone	102,832
4	Palma Ciudad	0-PALMA	BlackBerry	16,652
5	Palma Ciudad	0-PALMA	Chrome OS	0
6	Palma Ciudad	0-PALMA	Ellipsis 8	68,099
7	Palma Ciudad	0-PALMA	iOS	194
8	Palma Ciudad	0-PALMA	Mac OS X	449,248
9	Palma Ciudad	0-PALMA	Mac OS X 10.10	0
10	Palma Ciudad	0-PALMA	Mac OS X 10.11	274,799
11	Palma Ciudad	0-PALMA	Mac OS X 10.12	427,211
12	Palma Ciudad	0-PALMA	Mac OS X 10.9	2,649
13	Palma Ciudad	0-PALMA	Nexus	79,194
14	Palma Ciudad	0-PALMA	Other	917
15	Palma Ciudad	0-PALMA	RIM BlackBerry	571
16	Palma Ciudad	0-PALMA	Samsung TV	9,520
17	Palma Ciudad	0-PALMA	Windows	5,371
18	Palma Ciudad	0-PALMA	Windows 10	73,109
19	Palma Ciudad	0-PALMA	Windows 8	34,762
20	Palma Ciudad	0-PALMA	Xbox	0
21	Palma Ciudad	1-PALMA	Android	407,209
22	Palma Ciudad	1-PALMA	Apple iPad	53,286
23	Palma Ciudad	1-PALMA	Apple iPhone	0
24	Palma Ciudad	1-PALMA	BlackBerry	7,272
25	Palma Ciudad	1-PALMA	Ellipsis 8	244

Y estos los resultados.

Una vez que comprobamos que los resultados son los deseados pasamos a la pestaña "Worksheet"

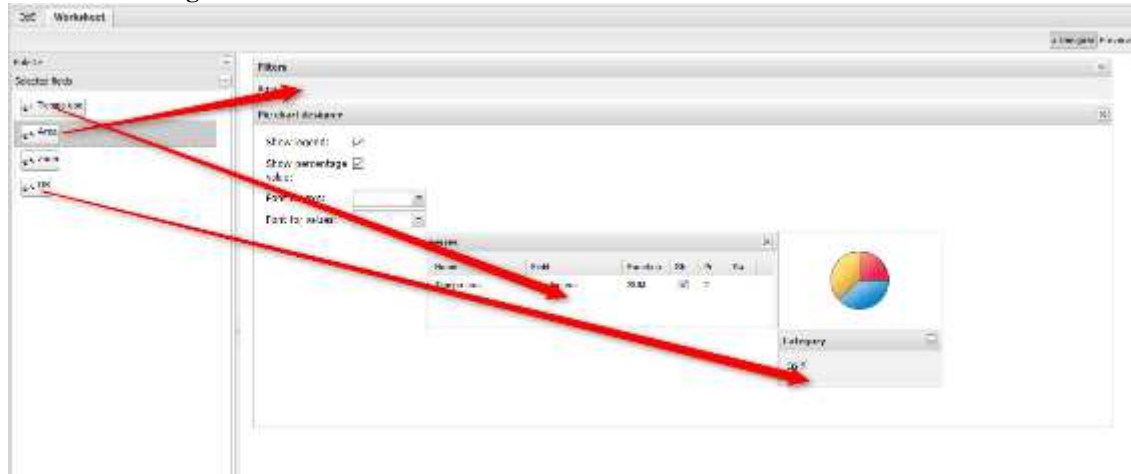


### 35 Creación de gráficos 5



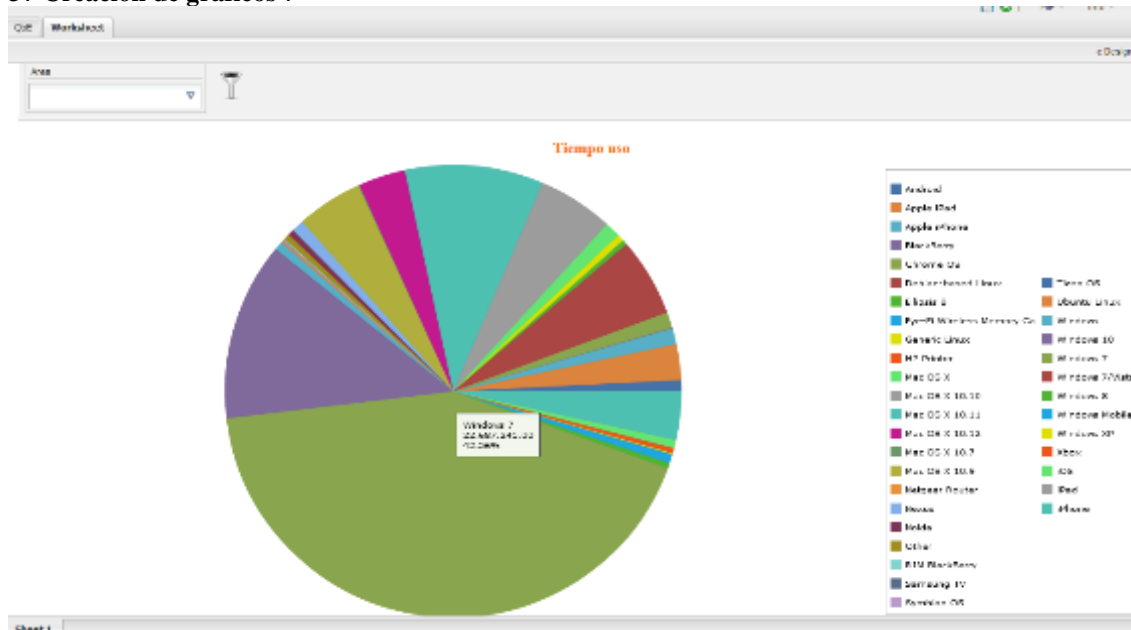
Seleccionamos el tipo de gráfico que deseamos y vamos arrastrando campos a sus respectivos lugares.

### 36 Creación de gráficos 6



Y pulsando la opción "Preview" vemos los resultados.

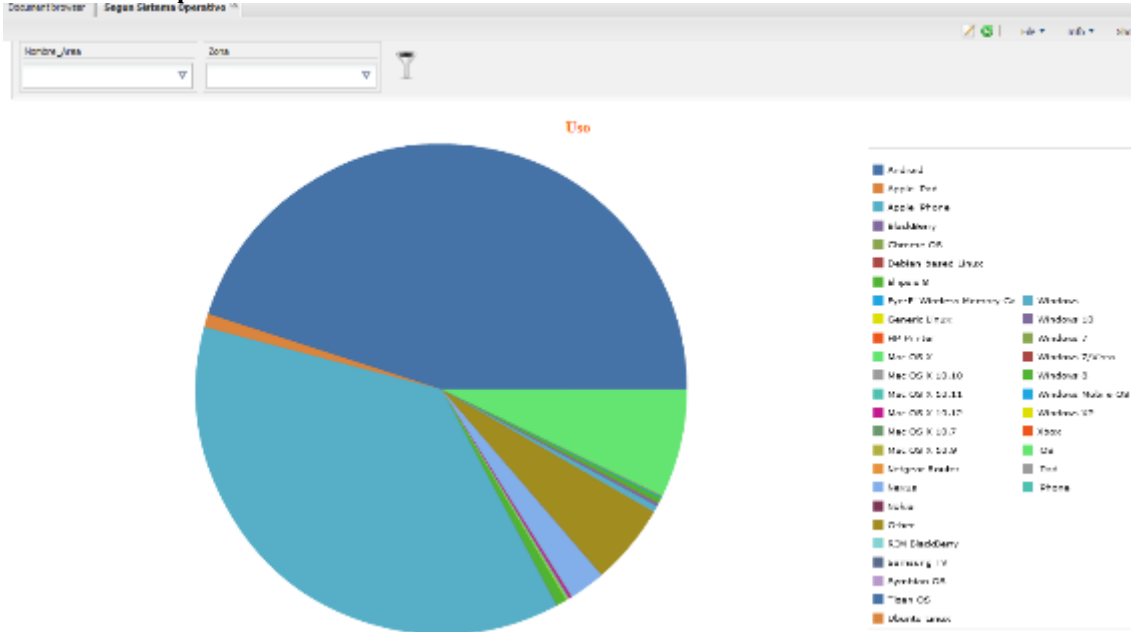
### 37 Creación de gráficos 7



En este gráfico hemos representado el tiempo de uso por SO, y podemos ver que el mayor tiempo de uso corresponde a dispositivos con Windows7.

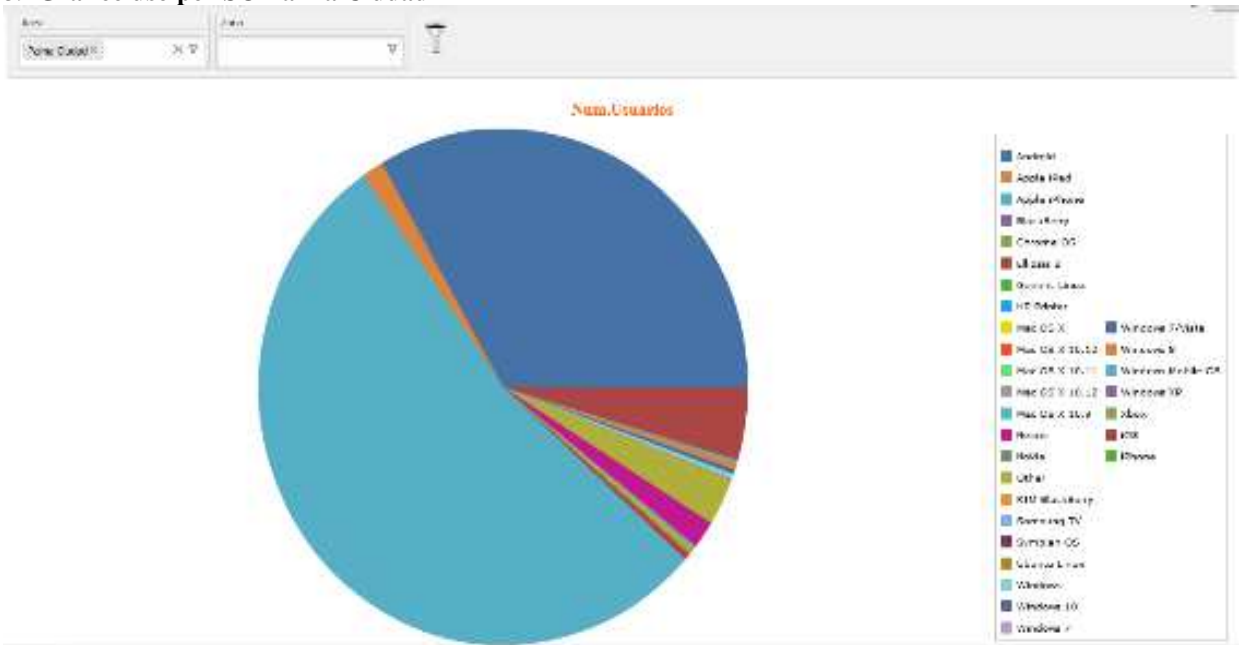
A partir de aquí definimos los gráficos siguientes:  
 Clientes por sistema operativo:  
 De todas las áreas

38 Gráfico uso por SO todas las áreas

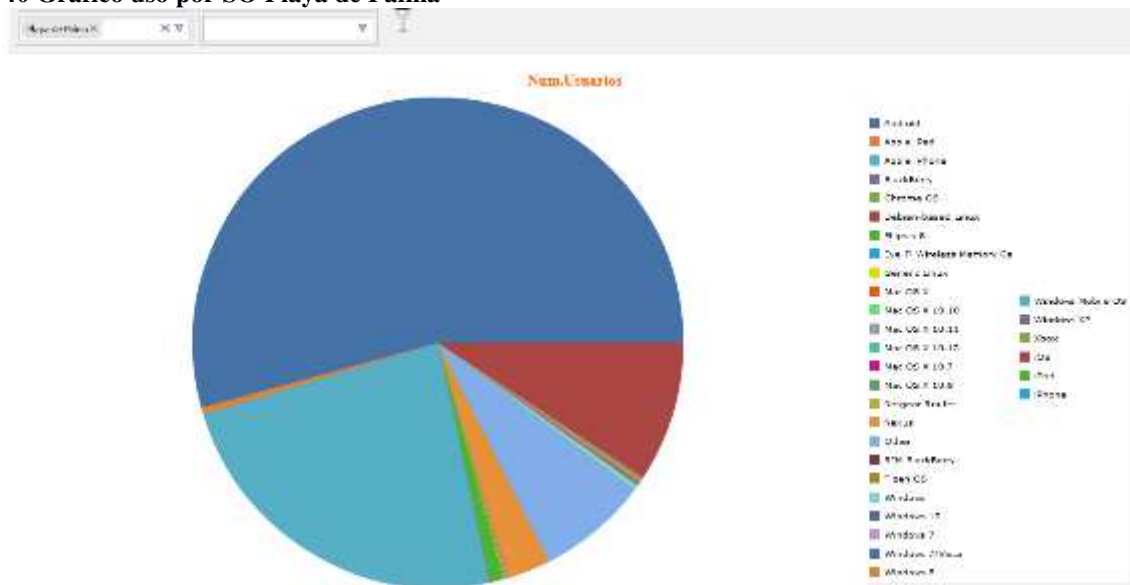


En este mismo grafico hemos seleccionado el área Palma Ciudad

39 Gráfico uso por SO Palma Ciudad

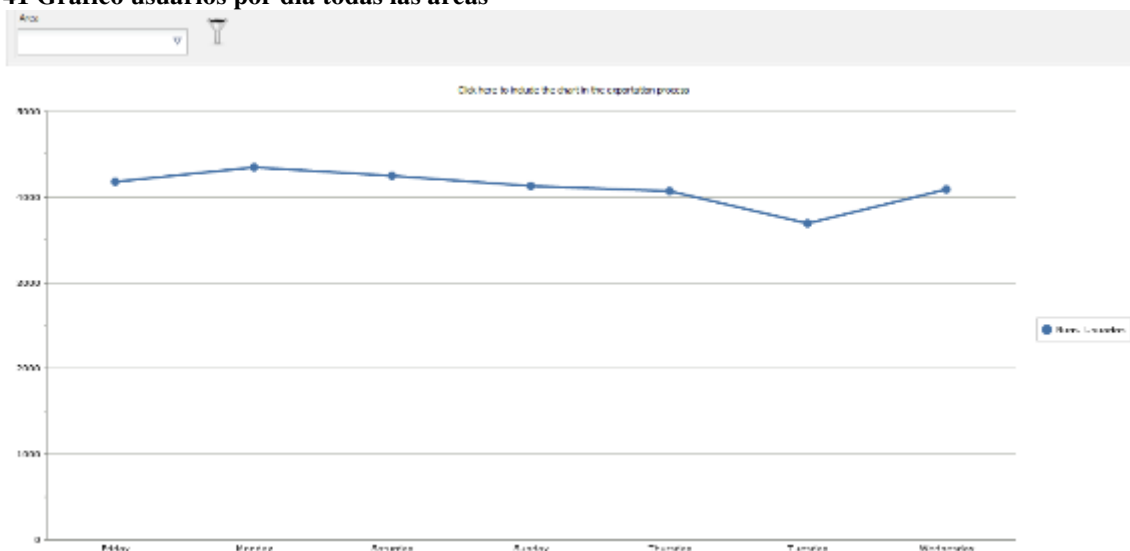


Y en este otro el Área de Playa de Palma.  
**40 Gráfico uso por SO Playa de Palma**



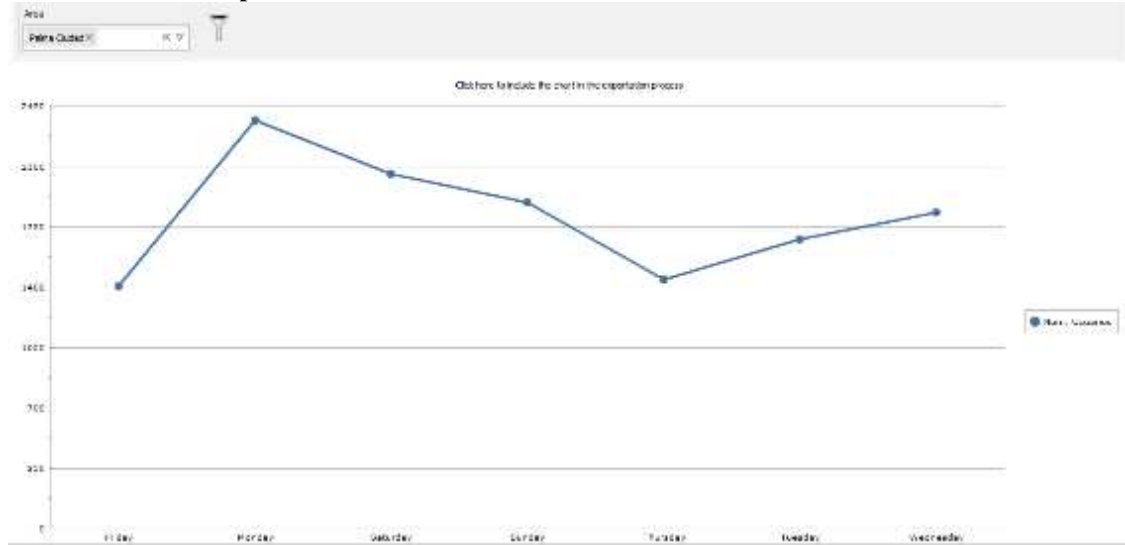
Aquí se ha elaborado un gráfico con el número de usuarios por día de la semana en todas las áreas.

**41 Gráfico usuarios por día todas las áreas**



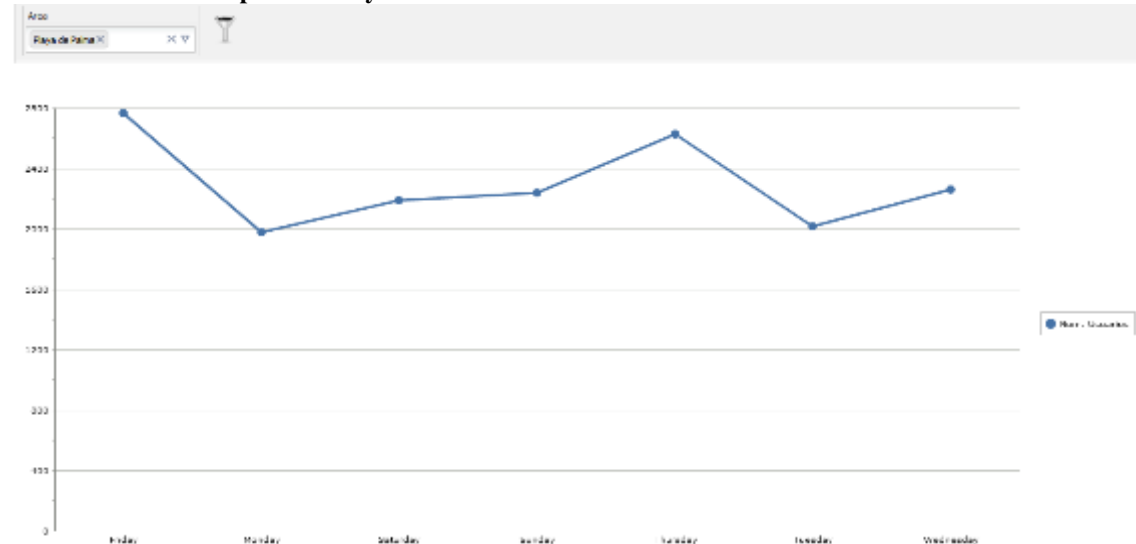
El mismo gráfico con los datos del Área de Palma Ciudad.

42 Gráfico usuarios por día de la semana Palma Ciudad



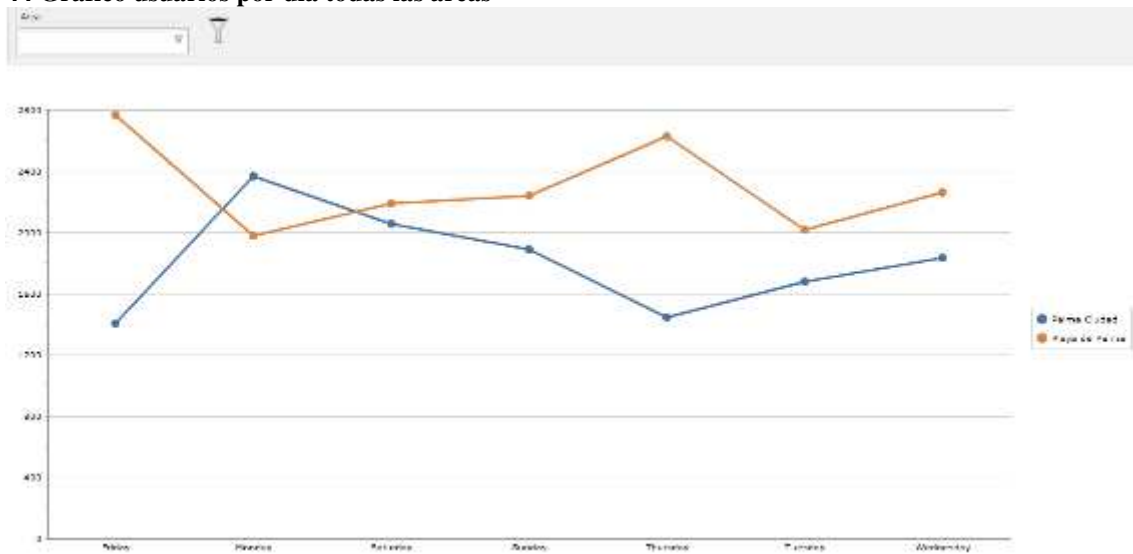
Y seguidamente con los datos del Área de Playa de Palma.

43 Gráfico usuarios por día Playa de Palma



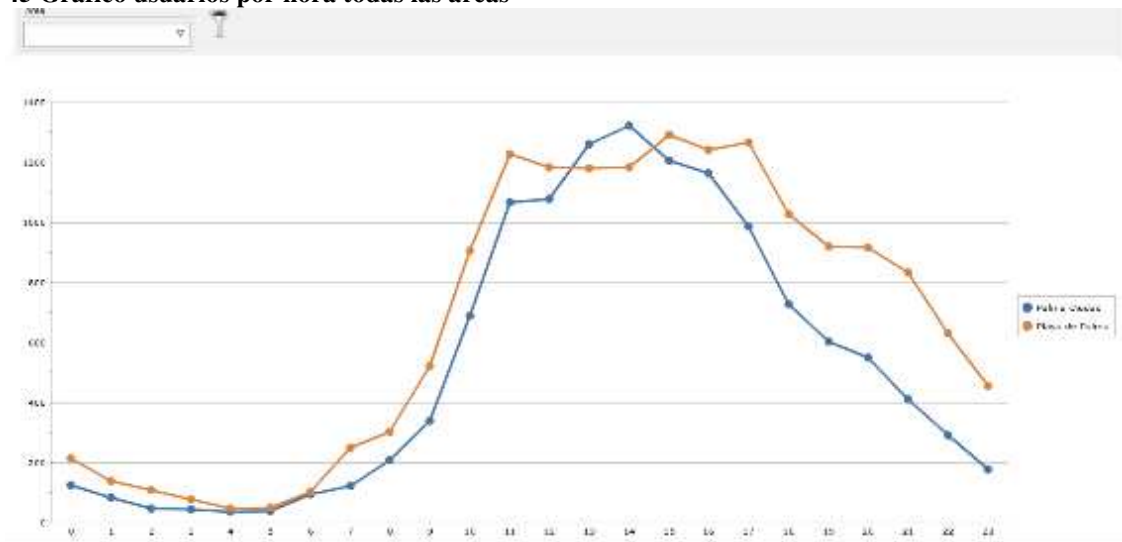
Los mismos datos de las dos áreas como series diferentes nos permiten comparar el número de usuarios por día de la semana en cada una de ellas.

44 Gráfico usuarios por día todas las áreas



En este caso vemos el número de usuarios por hora del día en las dos áreas.

45 Gráfico usuarios por hora todas las áreas



En el gráfico inferior hemos hecho un informe con los puntos de acceso ordenador por mayor número de usuarios.

#### 46 Informe puntos acceso con más usuarios

Area	Zona	Ap_conectado	Num. Usuarios
Playa de Palma	0-PLAYA	Gat. Café del Sol	1,353
Playa de Palma	1-PLAYA	Gat. Balneario 15 - Playa	924
Palma Ciudad	0-PALMA	Gat. Plaza Juan Carlos I	862
Playa de Palma	3-PLAYA	Gat. Grill Meister / Bierstrasse 1	778
Playa de Palma	0-PLAYA	Gat. MG Cafe Lounge	764
Playa de Palma	1-PLAYA	Gat. Balneario 15 - Hotel Rodes	653
Playa de Palma	0-PLAYA	Gat. MG Litte Italy	615
Playa de Palma	3-PLAYA	Gat. Farola 129 - Hotel Fontanellas	584
Palma Ciudad	7-PALMA	Anima Beach - Gat. Playa de Gesa	562
Playa de Palma	3-PLAYA	Gat. CNA Azotea Varadero Playa, Tiendas	546
Playa de Palma	0-PLAYA	Gat. Bierkonig Grill Konig	545
Palma Ciudad	6-PALMA	Gat. Plaza España - Banco Popular	524
Playa de Palma	3-PLAYA	Gat. Grill Meister / Calle: Sant Ramon Nonat	506
Palma Ciudad	1-PALMA	Gat. Paseo del Born 22 - Zara	495
Playa de Palma	1-PLAYA	Gat. Balneario 08 - Tiendas MCDonalds	493
Playa de Palma	0-PLAYA	Gat. MG Cafe - Terraza Chillout	482
Palma Ciudad	6-PALMA	Gat. Plaza España	450
Playa de Palma	3-PLAYA	Gat. CNA Azotea Varadero Entrada Parkinn	450

Y seguidamente por fabricante.

#### 47 Informe usuarios por fabricante

Fabricante	Num. Usuarios
Apple	13,190
Samsung	4,684
Samsung(THAILAND)	3,275
Huawei Technologies	2,096
Sony Mobile...	830
Murata Manufacturing	755
LG Electronics (Mobile...	473
Motorola Mobility LLC, a...	394
BQ	365
HTC	262
Xiaomi Communications	253
TCT mobile	224
zte	218
Shenzhen TINNO Mobile...	162
undefined	113
OnePlus Tech (Shenzhen)	111
Microsoft	107
WISOL	100

Una de las funcionalidades más potentes es la tabla de referencias cruzadas, donde aplicando filtros podemos analizar por ejemplo el número de usuarios por día en cada AP de una zona seleccionada.

#### 48 Informe usuarios por día AP-Zona

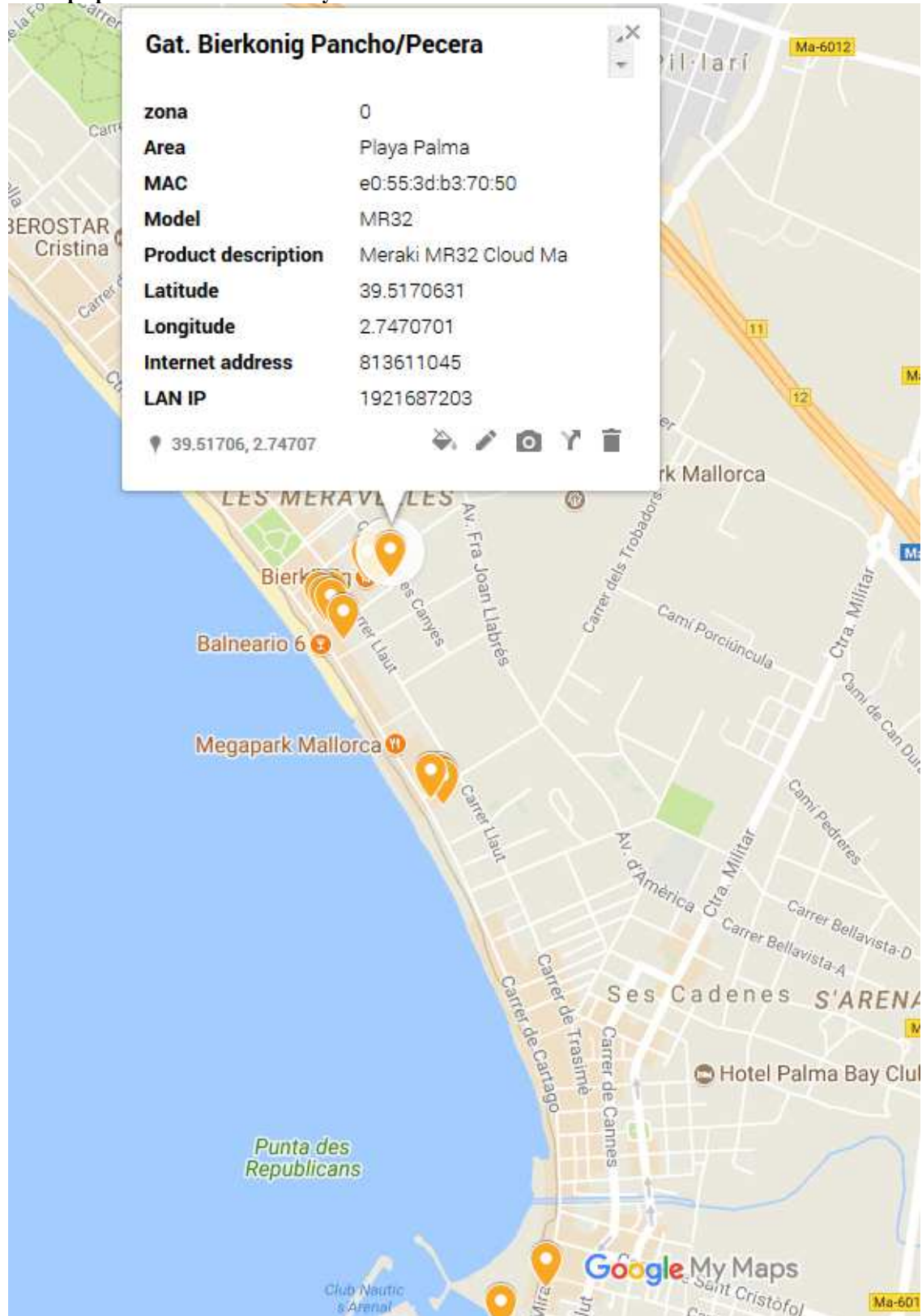
The screenshot shows a web interface with a filter for 'Zona' set to '0-PLAYA'. Below the filter is a table with columns for 'Nombre AP', 'Zona', and 'DÍA' (Friday, Monday, Saturday, Sunday, Thursday, Tuesday, Wednesday). The table lists various APs and their corresponding user counts for each day of the week.

Nombre AP	Zona	DÍA						
		Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
Gat. Bierkonig - Entrada	0-PLAYA	16.00	7.00	23.00	12.00	11.00	22.00	11.00
Gat. Bierkonig Carrusel 1	0-PLAYA	14.00	39.00	48.00	38.00	24.00	29.00	29.00
Gat. Bierkonig Carrusel 3	0-PLAYA	27.00	42.00	34.00	31.00	31.00	59.00	40.00
Gat. Bierkonig Entrada	0-PLAYA	47.00	60.00	65.00	59.00	44.00	62.00	37.00
Gat. Bierkonig Grill Konig	0-PLAYA	71.00	93.00	85.00	83.00	65.00	85.00	63.00
Gat. Bierkonig	0-PLAYA	12.00	24.00	16.00	17.00	12.00	24.00	16.00
Gat. Bierkonig Pecera	0-PLAYA	2.00	3.00	4.00	1.00	3.00	5.00	3.00
Gat. Bierkonig Restaurants	0-PLAYA	20.00	46.00	27.00	27.00	29.00	53.00	24.00
Gat. Café del Sol	0-PLAYA	203.00	235.00	189.00	172.00	198.00	210.00	146.00
Gat. MG Cafe - HPG Piscina	0-PLAYA	2.00	10.00	3.00	4.00	NA	5.00	4.00
Gat. MG Cafe - Terraza	0-PLAYA	67.00	100.00	82.00	67.00	57.00	52.00	57.00
Gat. MG Cafe Lounge	0-PLAYA	102.00	167.00	79.00	157.00	83.00	103.00	73.00
Gat. MG Cafe Piscina Lateral	0-PLAYA	29.00	49.00	29.00	27.00	17.00	34.00	15.00
Gat. MG Litte Italy	0-PLAYA	101.00	144.00	109.00	121.00	36.00	51.00	53.00
Gat. Nova Beach Park -	0-PLAYA	3.00	4.00	1.00	3.00	6.00	3.00	5.00
Gat. Nova Beach Park Azotea	0-PLAYA	19.00	30.00	31.00	34.00	7.00	25.00	20.00
Gat. Nova Beach Park	0-PLAYA	60.00	67.00	72.00	64.00	45.00	56.00	54.00
Gat. Varadero Beach	0-PLAYA	4.00	NA	2.00	3.00	NA	NA	NA

Aquí hemos seleccionado la zona 0 de la playa de palma, y vemos todos sus puntos de acceso con las conexiones por día de la semana.

Gracias a que disponemos de las coordenadas geográficas de cada punto de acceso podemos representarlo en forma de mapa con la aplicación "maps" de Google exportando en formato "csv" los registros que nos interesa representar.

#### 49 Mapa puntos acceso zona 0 Playa de Palma

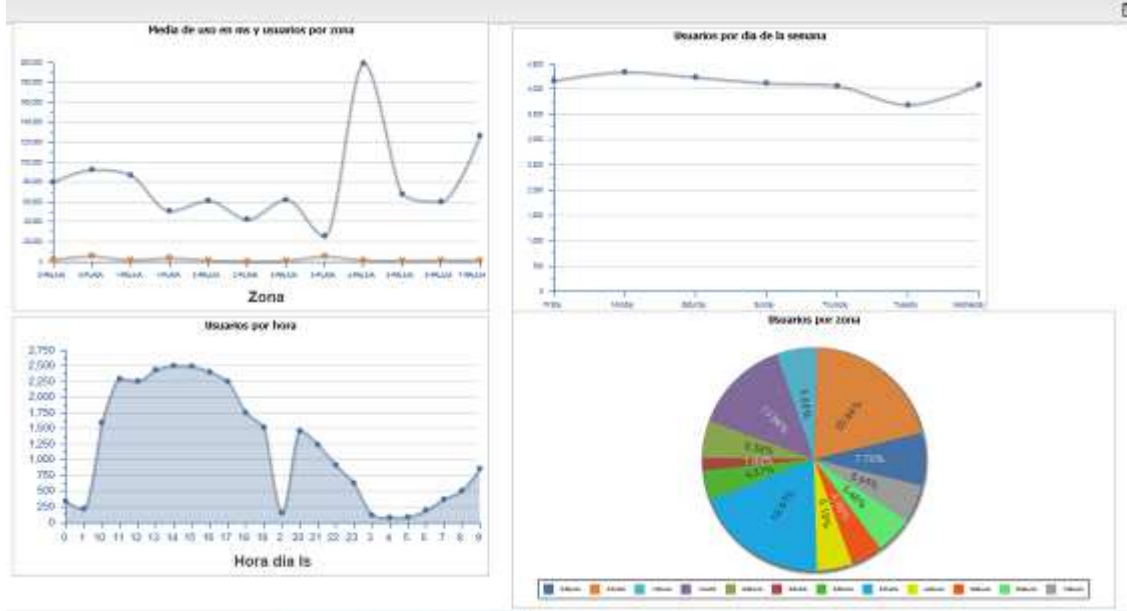




## Cuadro de mandos.

Podemos elaborar un cuadro de mandos donde se ven una serie de gráficos al mismo tiempo.

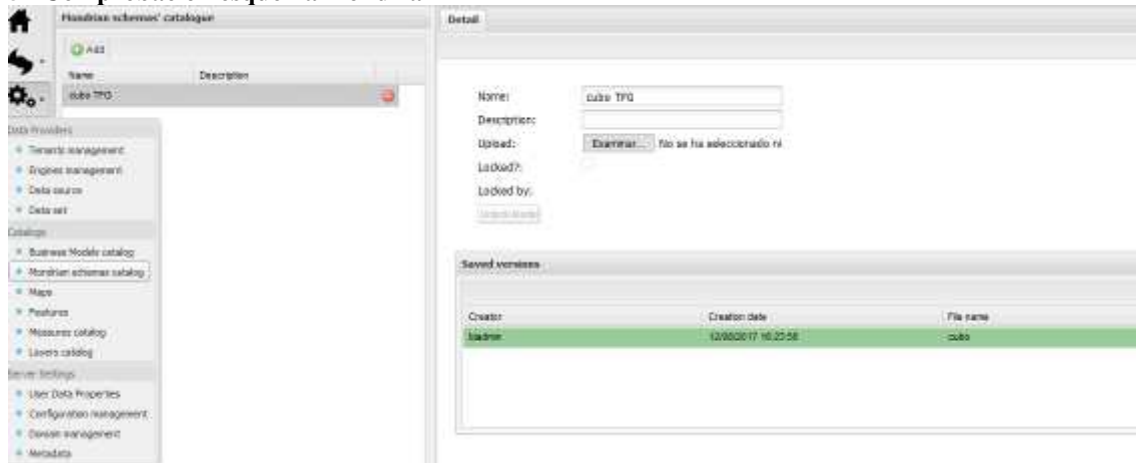
### 50 Cuadro de mandos



## 2.12 Uso del Cubo OLAP para creación de informes de forma dinámica.

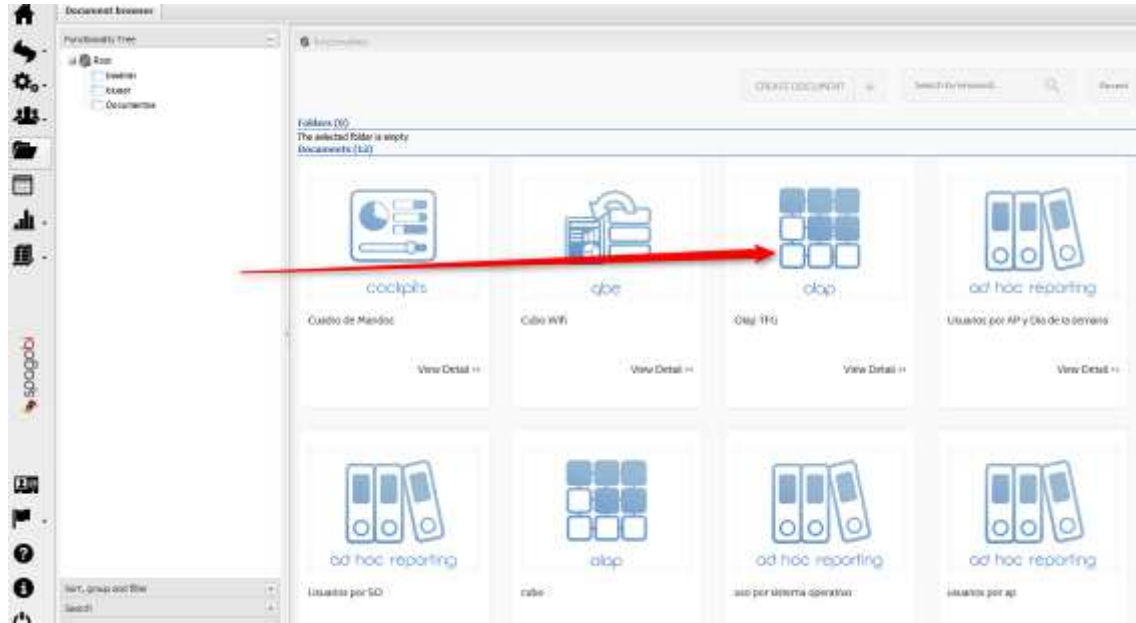
Para usar el esquema mondrian creado en el punto 2.11 primero comprobamos que es este subido al servidor seleccionando SpagoBi la opción “Mondrian schema catalog”

### 51 Comprobación esquema mondrian



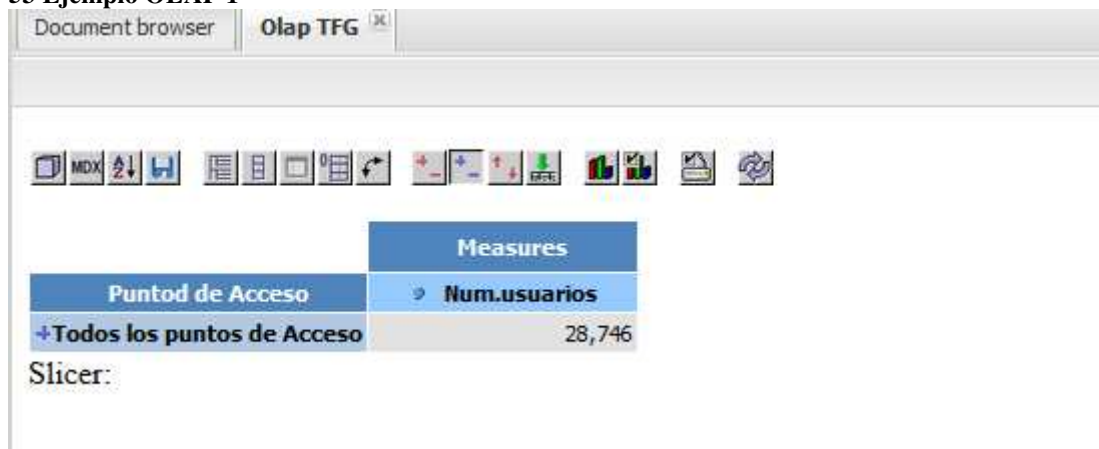
Una vez comprobado este punto veremos que en Document development ya figura como cubo Olap y podemos trabajar sobre el.

## 52 Selección fuente de datos



Pulsando sobre Olap TFG nos aparece esta ventana, la cual permite ver información del cubo según las dimensiones especificadas y las medidas incluidas en su definición

## 53 Ejemplo OLAP 1



Podemos expandir la selección pulsando sobre el símbolo +.

### 54 Ejemplo OLAP 2



Measures	
Access points	Usuarios
-All Access points	28,746
-1	12,672
+0	2,214
+1	1,724
+2	1,542
+3	1,239
+4	1,493
+5	1,321
+6	1,576
+7	1,563
-2	16,074
+0	5,934
+1	4,013
+2	529
+3	5,598

Slicer:

Añadir medidas.

### 55 Ejemplo OLAP 3



**Measures** ✖

Usuarios

Uso

None    Group    OK    Cancel

Measures		
Access points	Usuarios	Uso
-All Access points	28,746	2,101,658,147
-1	12,672	1,179,728,566
+0	2,214	177,485,298
+1	1,724	149,778,991
+2	1,542	94,766,167
+3	1,239	77,301,673
+4	1,493	297,546,984
+5	1,321	89,767,526
+6	1,576	95,272,380
+7	1,563	197,809,547
-2	16,074	921,929,581
+0	5,934	548,889,498
+1	4,013	204,877,080
+2	529	22,561,558
+3	5,598	145,601,445

Slicer:

Comparar datos de puntos de acceso seleccionados.  
56 Ejemplo OLAP 4



**Access points** [X]

- All Access points

- 1

- 0

- Evento Skoda Mar de Nudos - Nudos Club ...

• Evento Skoda Mar de Nudos - Nudos Club

• Evento Skoda Mar de Nudos - Salón

• Gat. Azotea Pesquero - Paseo Sagrera / Lonja

• Gat. Azotea Terraza Pesquero

• Gat. Moll Vell - Amarres

• Gat. Moll Vell - Restaurante Port Blanc

• Gat. Paseo Marítimo - Darsena

• Gat. Pesquero

• Gat. Plaza Juan Carlos I

• Gat. RCNP

• Gat. The Guinness House - Almudaina

• Gat. The Guinness House - Catedral

- Mar de Nudos - Gat. Mar de Nudos ...

• Mar de Nudos - Gat. Mar de Nudos

• Mar de Nudos - Gat. Terraza

+ 1

+ 2

+ 3

+ 4

+ 5

57 Ejemplo OLAP 5

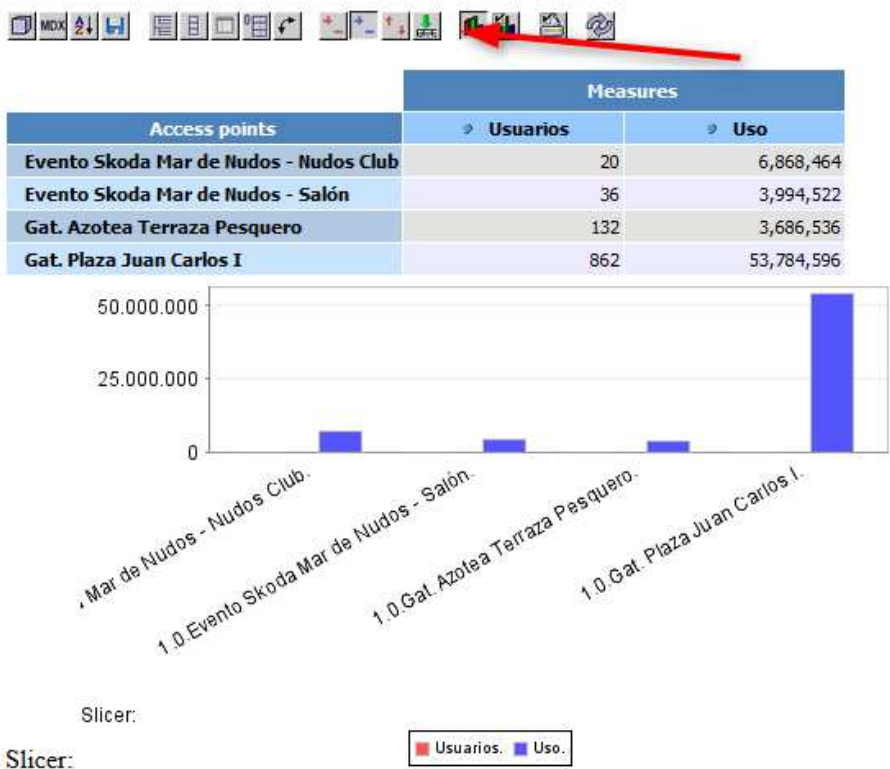


Access points	Measures	
	Usuarios	Uso
Evento Skoda Mar de Nudos - Nudos Club	20	6,868,464
Evento Skoda Mar de Nudos - Salón	36	3,994,522
Gat. Azotea Terraza Pesquero	132	3,686,536
Gat. Plaza Juan Carlos I	862	53,784,596

Slicer:

Y elaborar un grafico de barras pulsando un solo botón:

### 58 Ejemplo OLAP 6



Como se demuestra en estos ejemplos el uso de un cubo OLAP es muy potente e intuitivo.

### 2.13 Interpretación de resultados.

#### De los resultados obtenidos podemos deducir el siguiente conocimiento:

Los usuarios que más tiempo usan la Wifi son los de Windows 7, seguido de dispositivos Windows 10 y en tercer lugar MAC OS, por lo tanto el mayor uso es de dispositivos tipo PC (sobremesa o portátiles).

Las horas del día con mayor número de usuarios están en la franja entre las 10 y las 18.

Los sistemas operativos más usados por los clientes son Android y IOS.

El viernes es el día con más usuarios en el Área de playa de Palma, mientras que es el día con menos usuarios en el Área de Palma Ciudad. Podemos deducir que los usuarios se trasladan a la playa de palma los viernes.

Los puntos de acceso con mayor número de usuarios se concentran en la playa de palma, de los 10 primeros 8 son de esta área. Y si nos fijamos los dos de Palma Ciudad que figuran en los lugares 3 y 9 son:

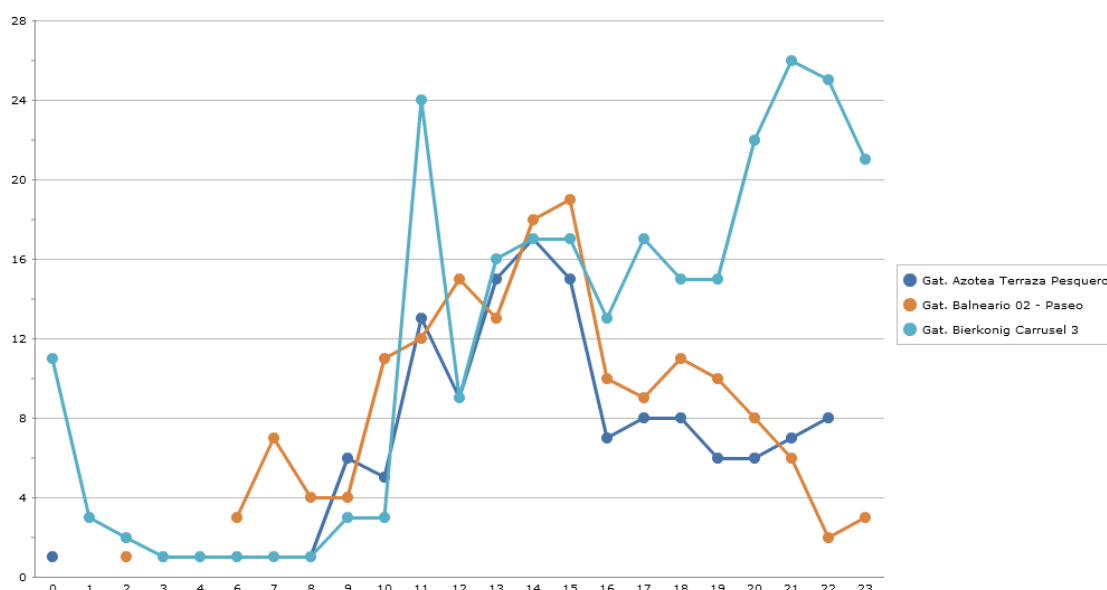
Plaza Juan Carlos I: Zona de terrazas en el centro de la ciudad.

Playa de Gesa: Zona de playa dentro de la ciudad.

Los fabricantes más utilizados son Apple, Samsung y Huawei.

Otro estudio que se puede hacer con este grafico puede ser comparar el número de usuarios por horas del día en diferentes puntos de acceso.

## 59 Grafico comparativo 3 puntos de acceso por hora del día



Se han seleccionado tres puntos de acceso con características muy diferentes:

- Azotea terraza pesquero. Bar del paseo marítimo muy popular entre la gente local.
- Balenario 2. Zona playa de palma en plena playa con afluencia sobre todo de alemanes.
- Bierkonig Carrusel. Conocido como BierStrasse (calle de la cerveza, zona más bien nocturna).

Vemos que en la zona de BierStrasse la mayor afluencia de usuarios se produce a partir de las 21 horas y hay una punta a las 11.

En la zona de playa la afluencia va subiendo desde las 10 a las 15 (horas de playa más habituales) y desciende por la tarde coincidiendo con el ascenso de BierStrasse. Los turistas se desplazan de la playa a la zona de ocio.

En cambio en la zona del paseo marítimo hay un incremento en el horario de comida habitual para los españoles y vuelve a subir a partir de las 22, después de la cena, pero muy por debajo de la zona turística.

Del conocimiento obtenido se pueden tomar decisiones en los siguientes aspectos:

- Los comercios pueden adaptar su horario de apertura al movimiento de los clientes.
- Se pueden reforzar las frecuencias de paso de los autobuses a determinadas horas y determinados días.
- En los puntos de acceso con mayor número de usuarios se puede plantear potenciar los equipos wifi.

### 2.14 Conclusión.

SpagoBI, una vez creados los modelos de datos es muy fácil de usar, además existe un motor de datos geográficos (GEOEngine), el cual permite posicionar resultados sobre un mapa y obtener resultados sobre este. No se han

elaborado ejemplos sobre este motor por falta de tiempo, ya que preparar los modelos necesarios para hacer las representaciones sobre mapas es bastante laborioso.

La fase más costosa es la de buscar las fuentes de datos, procesarlos y crear el modelo. Este proceso suele ser recursivo, ya que a medida que vamos obteniendo resultados, se nos ocurren más ideas que requieren replantearse el modelo.

El producto analizado permite crear todas las conexiones con bases de datos necesarias y los modelos que necesitemos sobre los mismos datos o sobre otros, lo cual nos permite cruzar información de diferentes fuentes en un mismo cuadro de mandos.

Gracias a la interface Web del que dispone, solo hay que hacer la instalación sobre un servidor y el usuario no requiere ninguna instalación adicional.

La gestión de roles de usuario es muy flexible y permite configurar carpetas para cada usuario o grupo de estos, de modo que la confidencialidad está garantizada.

El conocimiento que se obtiene analizando los gráficos e informes es muy elevado, permite establecer estrategias y predecir comportamientos. Una herramienta de BI bien utilizada puede proporcionar grandes beneficios y es una gran ayuda para la gestión de la ciudad.

El producto SpagoBI se puede implementar en el Ayuntamiento de Palma, para ello habría que destinar un servidor para SpagoBI Server, otro servidor MySQL o Postgree como repositorio para el DataWarehouse, utilizar la API de Meraki para descargar datos de forma programada, y crear un proceso ETL para su transformación.

Para otras fuentes de datos, se debería crear un proceso ETL para cada una de ellas.

## 3. Conclusiones

### 3.1 Lecciones aprendidas.

El desarrollo del presente trabajo me ha permitido enfrentarme a diferentes instalaciones de productos, las cuales han sido relativamente sencillas gracias a lo aprendido en diferentes asignaturas de la carrera. Como ejemplos:

- La instalación de la base de datos MySQL y la creación de esquemas, tablas, claves foráneas, etc. Lo aprendido en las diferentes asignaturas de Base de datos ha sido imprescindible.
- Configuración de SpagoBi Server: el entender la estructura de los ficheros XML estudiados en Base de datos ha facilitado la labor de configuración.
- Creación de modelos de datos: Lo aprendido en la asignatura de Minería de datos se ha podido aplicar a hechos reales y ha facilitado la comprensión de las diferentes fases del modelado.

La informática es una ciencia con muchas ramas y especialidades, no se puede pretender conocerlo todo, ya que es inabarcable, la especialización hoy en día es imprescindible. Cuando empecé en este mundo en el año 1978, era posible controlar casi todos los aspectos de un ordenador, desde cómo funcionaba, donde estaba físicamente la memoria (podías ver los núcleos de ferrita organizados en placas de circuitos impresos), como se organizaba esta cuando cargábamos un programa, los diferentes lenguajes de programación que existían (Ensamblador, Cobol, RPG, y poco más).

Gracias al desarrollo del hardware y las comunicaciones es posible disponer de volúmenes de datos inimaginables hace solo 20 años, y el desarrollo de las bases de datos relacionales ha facilitado el acceso a estos aun no sabiendo donde se encuentran físicamente, todo esto ha permitido crear software tan interesante como el de BI.

En este campo, donde todavía hay mucho que explorar, he aprendido que existen dos labores muy diferenciadas, por una parte el trabajo del informático, que permite saber dónde están los datos, como transformarlos y como crear un repositorio, y, por otra parte el trabajo del experto en minería de datos, el cual tiene que saber que datos necesita y como organizarlos para extraer conocimiento.

Son dos labores complementarias, el uno necesita al otro para obtener resultados de los procesos de BI.

### 3.2 Logro de objetivos:

En el transcurso del TFG se han logrado los objetivos propuestos al inicio: se ha implementado la plataforma open Source, se han creado los modelos de datos y se han creado gráficos e informes que permiten ver la utilidad del proyecto.

Quizás donde se ha quedado pobre, es en la parte de interpretación de resultados, como se menciona en el punto anterior, hace falta la figura de un experto en minería de datos, el cual hubiera



ayudado a crear gráficos o informes de acuerdo con las necesidades reales para lograr una interpretación de resultados más interesante, y sobre todo obtener predicciones de comportamiento.

### 3.3 Seguimiento de la planificación.

En cuanto al seguimiento de la planificación inicial no ha habido demasiadas desviaciones. Simplemente las comentadas en los informes que se han presentado juntamente con las PAC2 y PAC3.

La metodología prevista ha sido adecuada y ha permitido seguir los pasos de forma metódica.

Los cambios introducidos han sido los normales de un proceso de BI, cuando se ven los primeros resultados se plantea cambiar los procesos de extracción de datos y transformación de cara a crear modelos más adecuados para los resultados a obtener.

### 3.4 Líneas de trabajo futuras.

Vistos los resultados del trabajo me hubiera gustado tener más tiempo para incorporar estudios geográficos con representación sobre mapas. Ya que se disponía de datos de situación geográfica de los puntos de acceso e incluso de los metros de distancia de las observaciones realizadas, se hubiera podido hacer un estudio de paso de los usuarios. También se disponía de su MAC address, con lo cual podríamos ver el comportamiento a través del tiempo de un usuario o grupo de ellos, cuales son los puntos donde la gente permanece más tiempo y en qué horas sucede, y compararlo con históricos de meteorología de los mismos días y horas.

El campo de BI combinado con sistemas de Información Geográfica (GIS) lo veo muy interesante y la plataforma usada en este TFG lo permite.

Otra línea interesante de futuro es estudiar cómo funcionan los motores ETL de cara a automatizar el proceso de extracción y preparación de datos.

## 4. Glosario

**OLAP:** (On-Line Analytical Processing). Procesamiento analítico en línea. Es una solución utilizada en el campo de la llamada Inteligencia de negocios (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos [9]

**BI.** Business Intelligence. Inteligencia de negocio. Habilidad corporativa para tomar decisiones [2]

**ETL:** (Extract, Transform and load). El proceso de Extraer, transformar y cargar datos en el DataWarehouse.

**GIS:** (Geographical Information System). Sistema de información geográfica. Cualquier sistema de información capaz de integrar, almacenar, editar, analizar, compartir y mostrar la información geográficamente referenciada [11].

**XML:** (Extended markup Language). Lenguaje extendido de marcas.

**OS u SO:** (Operating System). Sistema operativo.

**API:** (Aplication program interface). Interfase de programación de aplicaciones. Conjunto de rutinas que permiten comunicar aplicaciones.

**CSV:** (Comma separated values). Valores separados por comas. Estructura de fichero con los campos separados por comas.

**Esquema Mondrian:** Contiene la definición de un cubo OLAP en formato XML.

## 5. Bibliografía

- [1] Cisco Meraki. White papers Location Analytics. (PDF)
- [2] El procés de descobriment de coneixement a partir de dades Objectius, fases i problemática Ramón Sangüesa i Solé PID\_00165718
- [3] La Inteligencia de Negocios: Etapas del proceso (PDF) Vásquez Castrillón John Bayron, Sucerquia Osorio Andrés. Maestría en ingeniería de sistemas, Universidad Tecnológica de Pereira , Colombia 2011
- [4] Introducción al business intelligence (epub) Jordi Conesa Caralt (coord.) Josep Curto Díaz Editorial UOC
- [5] Inicio Rápido - Un proyecto con SpagoBI 5. (PDF) Copyright © 2014 Engineering Group.
- [6] Learn SpagoBI in two hours. (PDF) Stephen Ogutu
- [7] SpagoBi User Manual Release 5.X. 2014 Engineering Group, SpagoBI Labs.
- [8] How to install SpagoBI 5.x Engineering Group, SpagoBI Labs - spagobi.org-
- [9] [https://es.wikipedia.org/wiki/Cubo\\_OLAP](https://es.wikipedia.org/wiki/Cubo_OLAP) Fecha visita: 20/10/2017
- [10] <http://culturacrm.com/business-intelligence/5-herramientas-business-intelligence/> Fecha visita: 23/9/2017
- [11] [https://es.wikipedia.org/wiki/Sistema\\_de\\_informaci%C3%B3n\\_geogr%C3%A1fica](https://es.wikipedia.org/wiki/Sistema_de_informaci%C3%B3n_geogr%C3%A1fica) Fecha visita: 15/12/2017
- [12] SpagoBI JPivot Engine Davide Zerbetto (PDF). 2014/07/09 17:461
- [13] BIRT [https://es.wikipedia.org/wiki/Business\\_Intelligence\\_and\\_Reporting\\_Tools](https://es.wikipedia.org/wiki/Business_Intelligence_and_Reporting_Tools) Fecha visita: 24/9/2017
- [14] JasperReports <https://community.jaspersoft.com/documentation/tibco-jaspersoft-studio-user-guide/v640/concepts-jasperreports> Fecha visita: 24/9/2017
- [15] Rapid Miner <https://rapidminer.com/> Fecha visita: 24/9/2017
- [16] Pentaho <http://www.pentaho.com/> Fecha visita:25/9/2017

## 6. Anexos

Llistat d'apartats que són massa extensos per incloure dins la memòria i tenen un caràcter autocontingut (per exemple, manuals d'usuari, manuals d'instal·lació, etc.)

Depenent del tipus de treball, és possible que no calgui afegir cap annex.