



# Aplicaciones de inteligencia de negocio



I. Introducción

II. Objetivos

III. Aplicaciones de inteligencia de negocio

IV. Herramientas de inteligencia de negocio

V. Herramientas de inteligencia de negocio: Pentaho Business Analytics

VI. Resumen

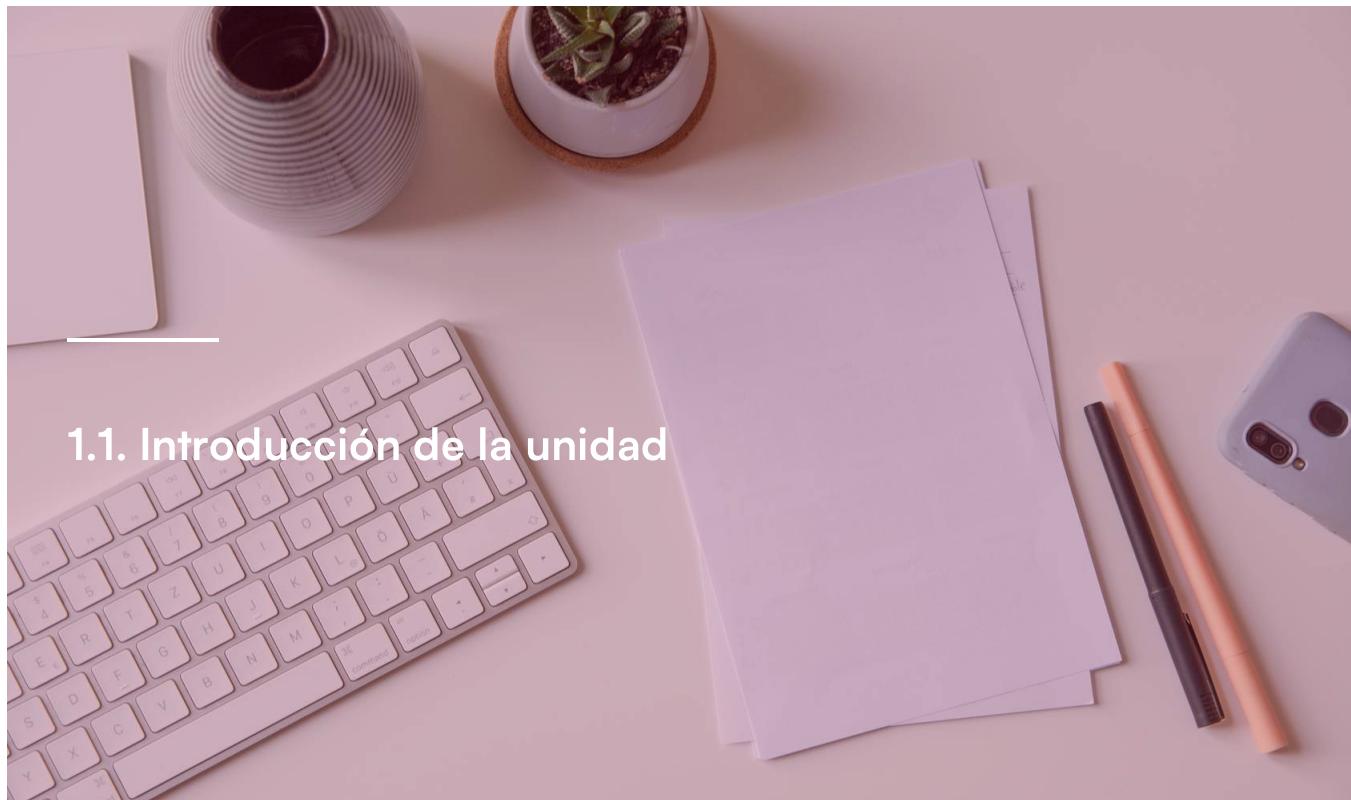
VII. Caso práctico con solución

VIII. Glosario

IX. Anexos

# I. Introducción

---



## 1.1. Introducción de la unidad

La mejor forma de entender con claridad la capacidad de la inteligencia de negocio es analizar casos de uso y aplicaciones reales y tratar de concebir e implementar un modelo inicial sencillo para ir asimilando los conceptos adquiridos. Este módulo tiene un doble objetivo: introducir al alumno en algunos ejemplos reales de aplicación de una solución de inteligencia de negocio y realizar un caso práctico con Pentaho Business Analytics.

**Pentaho es una herramienta de *business intelligence* que ofrece una edición empresarial y otra *open source* del software. Pentaho proporciona integración y extracción de datos, así como todas las funcionalidades necesarias para realizar un proceso ETL, servicios OLAP, informes y cuadros de mando.**

De forma general, los requisitos de las empresas se identifican y analizan mediante modelos (mapas de procesos, diagramas de actividad, modelos de datos comerciales, modelos de dimensiones, diagramas de estado, mapas ETL, etc.), así como utilizando técnicas de análisis de negocios y sistemas. Una vez que los requisitos de negocio de inteligencia de negocio se identifican, analizan y validan a través de los modelos/diagramas, los requisitos de negocio de inteligencia de negocio pueden destilarse y organizarse en el formato de caso de uso de BI. En esta unidad del módulo se citarán y plantearán, con mayor o menor profundidad, casos de uso reales, en los que se ha participado de forma activa, donde empresas de diferentes sectores han implementado soluciones de inteligencia de negocio para poder afrontar con éxito distintos escenarios o problemas.

Así, en cada uno de los casos se plantearán la situación de partida de la empresa, los medios y tecnología de los que dispone, cuáles son los retos y necesidades que hay que resolver y cuál es la solución propuesta para cumplir con los requisitos demandados. Se divide la unidad en diferentes apartados en función del sector o área a la que pertenezca la empresa.

Por último, es importante destacar que el alumno ha debido entender y comprender con claridad los conceptos impartidos en las tres unidades anteriores antes de estudiar esta. Una buena base de conocimiento teórico es fundamental para poder entender la aplicación en la práctica real de la inteligencia de negocio.

Adicionalmente, como complemento al contenido de esta unidad, IMF ofrece a sus estudiantes<sup>1</sup> la posibilidad de conocer y manejar una herramienta de visualización muy interesante: Qlik Sense. Durante el tiempo que estén matriculados con IMF, los estudiantes podrán acceder de forma gratuita a los materiales

audiovisuales y escritos de la herramienta, así como a laboratorios prácticos, en un entorno de la nube ofrecido por Qlik. Para registrarse, se deberán seguir los pasos que se indican en [este anexo](#).

---

<sup>1</sup>Para poder registrarse en el programa, los estudiantes deberán acreditar que están vinculados con IMF en la actualidad. Para ello, pueden adjuntar en el proceso de registro un correo o imagen de la inscripción o matrícula en IMF.



**Anexo\_QlikSense.pdf**

320.5 KB



## II. Objetivos

---



### 1.2. Objetivos de la unidad

- 1 Entender la aplicación de una solución de inteligencia de negocio en un contexto real.
- 2 Identificar futuros escenarios susceptibles de la aplicación de una solución de inteligencia de negocio.
- 3 Conocer la herramienta Pentaho Business Analytics con el objetivo de realizar un ejemplo práctico.

## III. Aplicaciones de inteligencia de negocio

---

A continuación, se presentarán ejemplos y casos reales de aplicación de sistemas de inteligencia de negocio, con mayor o menor medida de detalle.

En todos los casos, se implementó una solución de inteligencia de negocio, en ocasiones, integrada en un ecosistema *big data* o *analytics*.

### 3.1. Sector automoción

#### 3.1.1. Contexto

El sector de automoción ha sido uno de los grandes sectores que siempre han estado interesados en los sistemas de soporte de toma de decisiones, inteligencia de negocio, *analytics* y, más recientemente, sistemas inteligentes.

Siempre preocupados por la gestión de flotas, servicios de alto valor para el usuario, optimización del vehículo y un largo etcétera, el análisis y seguimiento de información, tanto histórica como en tiempo real, ha sido de gran valor para el sector.

A continuación, se muestra un caso de uso muy frecuente de solución de inteligencia de negocio para el sector de automoción.

#### 3.1.2. Caso

Una empresa líder del sector de alquiler y administración de vehículos, con una de las mayores flotas de vehículos del mundo, solicita una solución de inteligencia de negocio que permita a la compañía optimizar la gestión de sus flotas, así como mejorar la calidad del servicio y monetización de la información.

El objetivo era obtener una solución integral que permitiera a la compañía lo siguiente:

1

Recoger indicadores que representasen la cantidad y calidad de servicio a sus clientes.

2

Tomar decisiones sobre el estado y redistribución de flota de forma geográfica.

3

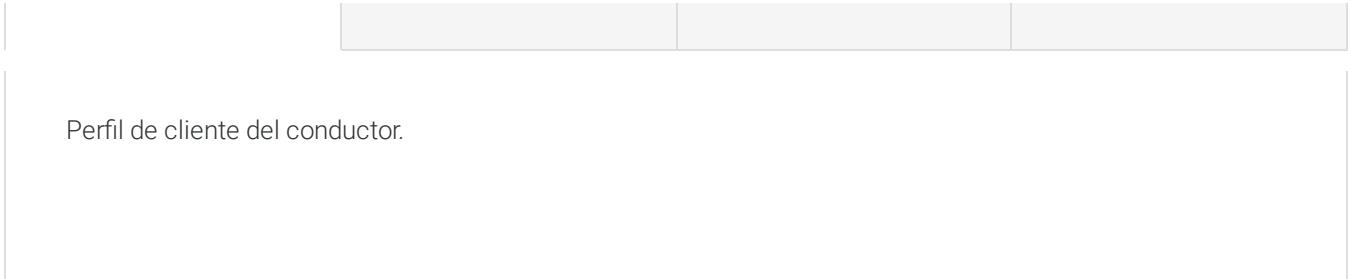
Perfiles de conducción de los usuarios y asignación optimizada de vehículos.

El principal reto de la compañía era ser capaces de recopilar la información de todas las fuentes e integrarla de forma automática, de manera que pudiera ser analizada conjuntamente y en tiempo real.

Como resultado, la solución y arquitectura tecnológica implementada se basaba en una sólida arquitectura con una fuerte capa de integración de fuentes, basada en procesos ETL y tecnología *big data* para *streaming*. Con ello, la solución integra principalmente tres fuentes de información:

- Datos sobre los clientes a partir de sus operacionales.
- Datos de la telemetría y sensores de los vehículos.
- Datos de fuentes externas, como meteorología.

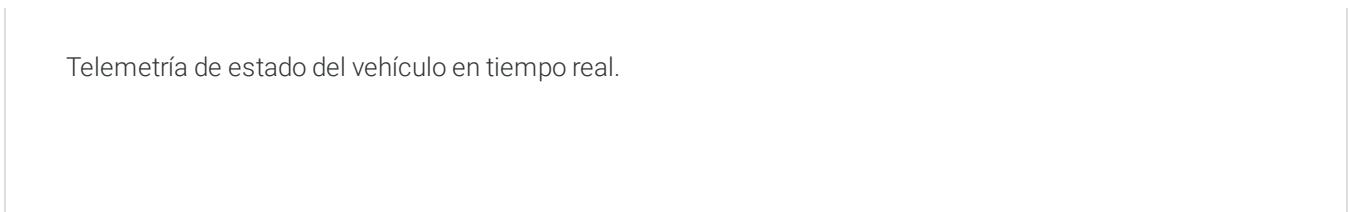
Y, con ello, obtener en tiempo real:



Perfil de cliente del conductor.



Perfil de conducción del conductor.



Telemetría de estado del vehículo en tiempo real.



Perfil de flota.

## CONTINUAR

Con la generación de esta información, la solución pudo proporcionar un CAU en tiempo real para la gestión de incidencias de conductores, así como un cuadro de mando integral de gestión de flota y perfiles de conducción de los usuarios, que permitieron la monetización de dicha información, siendo de especial interés para las aseguradoras.

Este es un ejemplo básico de caso de uso de una solución de inteligencia de negocio, integrada con otras tecnologías para el sector del automóvil.

**i** Este sector es uno de los que más invierte en sistemas de soporte para la toma de decisiones, ya sea de negocio o autónomas basadas en inteligencia artificial (IA). De hecho, los medios actuales y futuros de transporte autónomo se presentan en diversas formas, como automóviles, camiones, vehículos aéreos no tripulados (drones) o maquinaria agrícola.

El ejemplo de vehículos autónomos es particularmente relevante debido a su impacto en la sociedad en general. La conducción autónoma mantiene la promesa de una experiencia de movilidad más fluida, segura y cómoda. La industria del automóvil está en un viaje continuo desde la conducción asistida a la conducción autónoma.

Hoy en día, la mayoría de los sistemas avanzados de asistencia de controladores (ADAS), como el reconocimiento de peatones, aún se realiza con programación basada en reglas. Sin embargo, construir y mantener esos sistemas es complejo, y más en los sistemas y tecnologías tradicionales de procesamiento en tiempo real. La cantidad de situaciones que deben cubrirse es prácticamente indefinida, dada la gran variedad y diversidad de escenarios de tráfico posibles. Por lo tanto, definir un conjunto completo de reglas no solo es poco práctico, sino más bien imposible. En la actualidad, el procesamiento de imágenes en tiempo real es esencial para este sector y, con las tecnologías hasta ahora usadas, no era suficiente. Para

completar el camino hacia decisiones verdaderamente autónomas, el uso de enfoques modernos de IA se convertirá en un requisito previo.

**CONTINUAR**

## **3.2. Sector de recambios, consumibles, logística**

### **3.2.1. Caso**

Una empresa mayorista de referencia en la distribución de neumáticos, que pertenece a un grupo empresarial con más de 72 años de experiencia en el gremio, desea optimizar el proceso de pedidos con sus proveedores, ajustando el número de unidades solicitadas. La empresa no contaba inicialmente con una solución analítica, por lo que la nueva plataforma debería dar respuesta al resto de necesidades analíticas de la compañía, incluyendo las relaciones con sus colaboradores.

### **3.2.2. Solución**

Se optó por la solución Pentaho BA, una suite completa *open source* que cubría todas las necesidades del proyecto, incluyendo la integración con las librerías de R para los procesos *machine learning*.

Se implementó un repositorio central de datos que daba respuesta a los diferentes ámbitos de conocimiento, proporcionando indicadores de compras, ventas, márgenes, stock. Además, se implementaron procesos *machine learning* para la optimización de stock, en función de diferentes variables del sector: tiempo de recepción, nivel de servicio, estacionalidad...

Dispone de una plataforma BI que proporciona indicadores, cuadros de mando, informes y vistas OLAP de forma flexible y satisfactoria, gracias a tecnologías y metodologías de **open source**, como Pentaho BA, PostgreSQL o R.

## Beneficios

Un sistema capaz de optimizar sus pedidos a proveedores, estimando las ventas futuras y realizando propuestas de pedidos.

### 3.3. Sector industrial

#### 3.3.1. Contexto/caso

El mantenimiento predictivo tiene como objetivo mejorar la productividad de los activos mediante el uso de datos, para anticipar las averías de la máquina.

**i** Un método bien establecido y relativamente simple para reconocer fallos desde el principio es el monitoreo del estado o condición de los elementos de interés. La complejidad de la predicción del fallo a menudo se debe a la enorme cantidad de posibles factores o variables de influencia. Las fuentes de datos pueden ser múltiples y dependen del escenario.

Por ejemplo, en motores, cajas de engranajes o aire acondicionado, el análisis del sonido puede detectar una anomalía en el funcionamiento del dispositivo. Y en interruptores, máquinas y robots, las vibraciones se

pueden medir y usar para detectar errores.

Dado que los nuevos sensores y dispositivos IoT se pueden integrar en procesos de producción y operaciones, la disponibilidad de datos aumenta drásticamente. Los algoritmos basados en inteligencia artificial son capaces de reconocer errores y diferenciar el ruido de la información importante para predeterminar averías y guiar futuras decisiones.

### 3.3.2. Solución

Se trata de sistemas avanzados de inteligencia de negocio integrados con tecnologías *big data* y *analytics* que buscan como objetivo la optimización de procesos y, por tanto, el ahorro de costes.

Según el punto de partida y el nivel de redundancia, la disponibilidad, a veces, puede aumentar en más del 20 %. Los costes de inspección pueden reducirse hasta en un 25 % y es posible una reducción general de hasta el 10 % de los costes anuales de mantenimiento.

Para ello, se procesa la información de una gran variedad de fuentes de datos, más allá de las salidas de los sensores, tales como registros de mantenimiento, medición de calidad de salidas de máquinas y, si corresponde, fuentes de datos externas, como, por ejemplo, datos meteorológicos, lo que permite la predicción de eventos que antes no era posible modelar.

Junto con este procesado conjunto, se hace uso del modelado de algoritmos y técnicas de aprendizaje automático, que son los que, en último término, identifican y predicen la posibilidad y probabilidad del próximo fallo.

Como resultado, se tendrá una solución que presenta, a operadores y gestores, información en tiempo real y estimaciones/predicciones, mediante cuadros de mando capaces de optimizar el mantenimiento y la vida útil de las máquinas y componentes.

**CONTINUAR**

## 3.4. Sector de la salud

La riqueza de los datos disponibles en la atención médica continuará aumentando. Esto, junto con una demanda creciente de una mejor calidad en la atención al paciente, requiere la mejora de *healthcare business intelligence*, lo que se puede abordar mejor al considerar las soluciones tecnológicas de inteligencia de negocio de adquisición, almacenamiento, interpretación y evaluación de datos.

### 3.4.1. Contexto

#### Datos heterogéneos en la asistencia sanitaria

La atención médica se enfrenta a una gran heterogeneidad en los datos, ya que provienen de diferentes fuentes, estructura y naturaleza compleja. Así, la información de salud se presenta en distintos formatos: numéricos, textuales, digitales, imágenes, vídeos, multimedia, etc.

El registro electrónico de salud contiene cientos de filas de datos textuales y numéricos correspondientes a datos demográficos del paciente, notas de progreso, signos vitales, historias médicas, diagnósticos, imágenes de radiología, medicamentos, resultados de laboratorio y prueba, etc.

En la información del sector de la salud, los datos están **estructurados** y **no estructurados**. Los datos estructurados se refieren a los datos demográficos del laboratorio o del paciente que son consistentes y se almacenan en un formato predefinido. Los datos no estructurados no son uniformes y pueden ser de gran valor para analizar los datos del paciente. Estos incluyen notas clínicas, dictados de voz de audio, mensajes de correo electrónico y archivos adjuntos, mensajes de texto, vídeo en línea y transcripciones tipeadas. La presencia de datos estructurados y no estructurados hace que el conjunto de datos de la asistencia sanitaria se procese y aumente conforme se incrementa el número de variables.

## **Uso del sensor en la asistencia sanitaria**

La inteligencia empresarial en la asistencia sanitaria proporciona una amplia gama de análisis para mejorar el proceso de toma de decisiones relacionado con el paciente y el rendimiento de muchas áreas funcionales, incluidas la planificación de recursos, la prestación de atención, la contabilidad de los pacientes y el ciclo de financiación y de ingresos.

El avance tecnológico ha permitido un mayor acceso a los datos. El uso de sensores en este sector produce grandes volúmenes de datos de forma continua a lo largo del tiempo. Se puede ver una aplicación en la unidad de cuidados intensivos (UCI), donde se utilizan sensores para monitorear el estado actual de los pacientes, que incluyen electrocardiograma (ECG), electroencefalograma (EEG), monitores de presión arterial y monitores respiratorios.

### **3.4.2. Caso de estudio de la UCI**

#### **Ejemplo**

Las unidades de cuidados intensivos tienen un entorno de datos múltiples: fuentes de datos continuas se originan desde distintos dispositivos médicos: electroencefalograma, monitores de cabecera, monitor de oxígeno en tejido cerebral, catéteres de centro venoso (CVC), sistemas de información clínica y ventiladores, lo que se convierte en varios kilobits de datos del paciente por segundo.

Un paciente en estado de salud grave, a menudo, es controlado por una gran cantidad de sensores corporales conectados a dispositivos de

monitorización, que producen grandes volúmenes de datos fisiológicos, junto con datos clínicos exhaustivos y detallados y tendencias, minuto a minuto, del paciente. Otra monitorización del paciente en la UCI implica que el sistema genere algunas alertas o alarmas cuando el estado fisiológico del paciente muestra la detección de anomalías relevantes o cambios en la condición del paciente. Los médicos usan ciertos umbrales, que, cuando se sobrepasan, se activa la alarma. Sin embargo, tales esquemas de alerta simples pueden dar como resultado un gran número de falsas alarmas.

Es crucial que el paciente reciba la asistencia médica lo antes posible. El avance reciente en la tecnología de sensores ayuda a lograr este objetivo, monitoreando parámetros cruciales del paciente, como latidos cardíacos, temperatura corporal, monitoreo de sangre, EEG, etc.

### **3.4.3. Solución/enfoque**

En el escenario dado, la inteligencia de negocio puede desempeñar un papel clave, llevando a cabo el análisis predictivo, lo que implica integrar los datos del sensor con los datos tradicionales disponibles.

Los datos procesados que proporcionan las estadísticas vitales del paciente pueden transmitirse al monitor del enfermero que viaja con el paciente. Además, esta información puede usarse para tomar medidas apropiadas cuando el paciente llegue al hospital.

## **3.5. Sector hotelero**

La riqueza de los datos disponibles en los centros hoteleros es inmensa. Esto, junto con una demanda creciente de una mejor calidad en la atención al cliente, requiere la mejora de *business intelligence* dentro del sector, lo que se puede abordar mejor al considerar las soluciones tecnológicas de inteligencia de negocio de adquisición, almacenamiento, interpretación y evaluación de datos.

### 3.5.1. Contexto

#### Datos heterogéneos en los centros hoteleros

Los centros hoteleros tienen un gran volumen y variedad de datos que provienen de diferentes fuentes con estructuras y naturaleza diversas.

El sistema de reservas y registro de llegadas y salidas de los hoteles contiene muchísima información en formato texto y numérico, referente al cliente, la reserva y la estancia.

En la información del sector hotelero, los datos están **estructurados y no estructurados**. Los datos estructurados se refieren a los datos típicos del cliente y de sistemas operacionales de reservas, llegadas y estancias. Los datos no estructurados no son uniformes y pueden ser de gran valor para analizar los datos del cliente. Estos incluyen navegación por la web, ciertos datos de interacciones con sistemas del hotel, como el wifi, e incluso datos de sensores que permiten registrar la ocupación de estancias como salones de cenas, etc. La presencia de datos estructurados y no estructurados hace que el conjunto de datos de los hoteles se procese y aumente conforme se incrementa el número de variables.

#### Uso del sensor en los centros hoteleros

La inteligencia empresarial en los centros hoteleros proporciona una amplia gama de datos para realizar análisis avanzados para mejorar el proceso de toma de decisiones relacionado con el cliente y el rendimiento de ciertos sectores del hotel, incluida la planificación de recursos, la contabilidad de los clientes, la planificación de ocupación, la gestión de compras, sistemas de puntos para fidelización, etc.

El uso de sensores en este sector permite evaluar la ocupación de las estancias, los caminos que suelen tomar los clientes, para poder, por ejemplo, decidir dónde establecer una campaña de marketing, o cómo potenciar otros servicios del hotel.

### **3.5.2. Caso de estudio del *forecast* de ocupación**

Para los centros hoteleros es crucial su previsión de ocupación, ya que según este dato se ajustarán las contrataciones de personal y los aprovisionamientos realizados, entre otras cosas. Se entiende como *forecast* o pronóstico el proceso de estimación en situaciones de incertidumbre.

Esta previsión se obtiene mediante una mezcla entre experiencias pasadas, promociones realizadas y las tendencias del propio año. Con este dato se pueden gestionar y anticipar los gastos que va a tener el centro hotelero y mejorar las condiciones de compra.

### **3.5.3. Solución/enfoque**

En el escenario dado, la inteligencia de negocio puede desempeñar un papel clave, llevando a cabo el análisis predictivo, realizando una combinación de datos pasados, variables que incrementen la ocupación como las promociones y otras variables macroeconómicas y microeconómicas.

Los datos procesados permitirán saber la previsión de ocupación hasta cierto intervalo de tiempo, teniendo claro que, a mayor lejanía en el tiempo, mayor riesgo tiene la predicción.

## IV. Herramientas de inteligencia de negocio

---

En esta sección se detallarán las principales herramientas de BI que existen en el mercado, aunque en módulos anteriores se ha hablado y visto en profundidad varias de ellas, como Tableau y PowerBI.

Para ello, se va a profundizar en el cuadrante de Gartner de las herramientas de BI.

Debido al crecimiento de las tecnologías, las grandes empresas se hacen una serie de preguntas:

- ¿Qué tendencias tecnológicas están marcando actualmente la diferencia?
- ¿Qué producto o solución de los diferentes fabricantes lidera el mercado?
- ¿Cuáles son los mejores proveedores para el negocio?

Gartner elabora un ranking de los fabricantes con mejores soluciones y productos. Este consta de dos ejes explicados a continuación:

1

2

En el **eje X**, Gartner define la categoría “integridad de visión”, la cual representa el conocimiento de los proveedores sobre cómo se puede aprovechar el momento actual del mercado para generar valor, tanto para sus clientes como para ellos mismos.

1

2

En el **eje Y** se encuentra la “capacidad de ejecutar”, la cual mide la habilidad de los proveedores para ejecutar con éxito su particular visión del mercado.

Ambas divisiones fragmentan el cuadrante en cuatro sectores, donde aparecerán las principales compañías de cada competencia en función de su tipología y la de sus productos: líderes (*leaders*), retadores o aspirantes (*challengers*), visionarios (*visionaries*) y jugadores de nicho (*niche players*). En la figura 1 se muestra la imagen más característica y representativa del popular cuadrante mágico de Gartner.

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



Figura 1. Cuadrante de Gartner.

Fuente: Gartner

Se van a presentar aquellas en las que no se ha profundizado anteriormente:

Qlik

Según se define en su página web, es una plataforma completa para el descubrimiento. No visualiza datos, sino que los almacena y procesa, al ser una plataforma completa de BI.

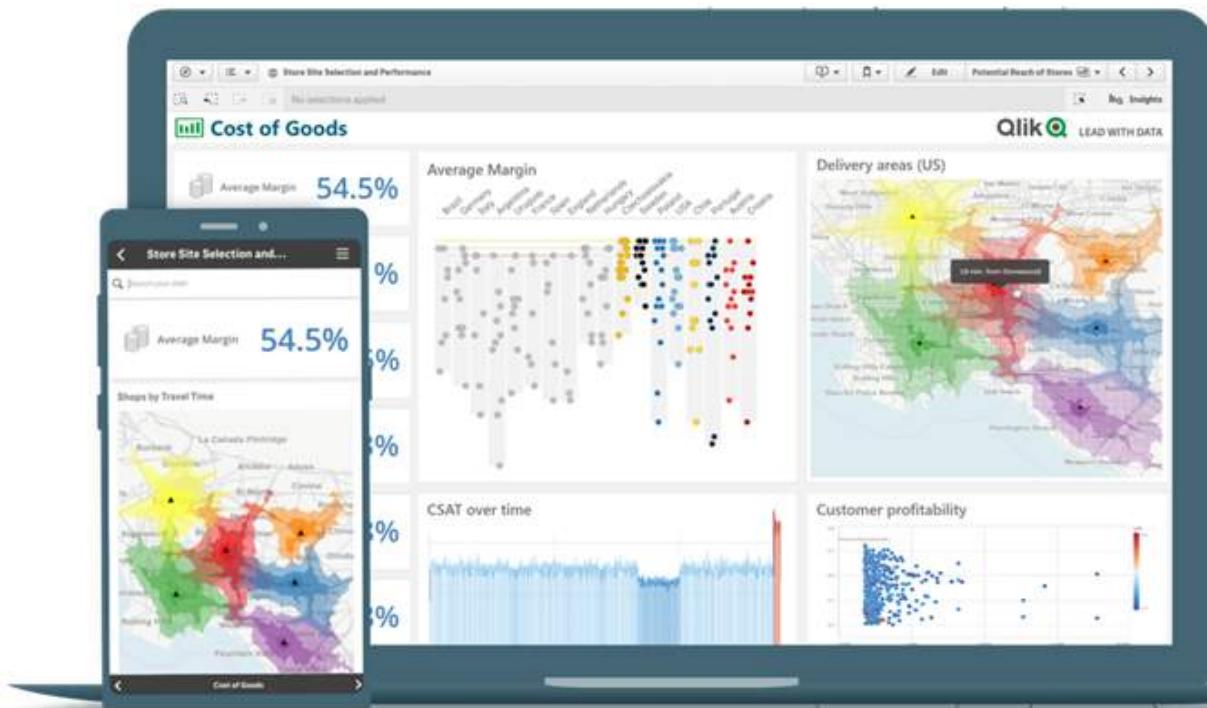
Qlik combina IA con la interacción humana, proporcionando una inteligencia aumentada sobre los datos, mediante la indexación asociativa que genera conocimiento sobre los datos para permitir ser explorados por los usuarios. Está [disponible aquí](#).

Qlik es una plataforma de BI que se compone las siguientes herramientas:

- Qlik:
- Analítica:
  - Qlik Sense.
  - QlikView.
- Plataforma de desarrollo:
  - Qlik Core.
- Integración de datos:
  - Attunity Replicate.
  - Attunity Compose.
  - Attunity Enterprise Manager.
  - Attunity Gold Client.
  - Qlik Data Catalyst.

**Figura 2.** Qlik.

Fuente: Qlik® ([wwwqlik.com](http://wwwqlik.com)).



## MicroStrategy

Es una compañía que ofrece **software** ROLAP y de **reporting** como solución BI para empresas. Su **software** permite crear informes y análisis de datos almacenados en una base de datos relacional y otras fuentes.

Visual Insight es un producto que permite explorar los datos visualmente para descubrir ideas de negocio, analizar los datos importantes almacenados en Hadoop y otros motores, orientados a **big data** y mejorar la toma de decisiones.

Microstrategy cuenta con aplicaciones móviles como Microstrategy Mobile, plataforma que permite construir aplicaciones móviles que proporcionan funcionalidades de inteligencia de negocio en teléfonos móviles y tabletas.

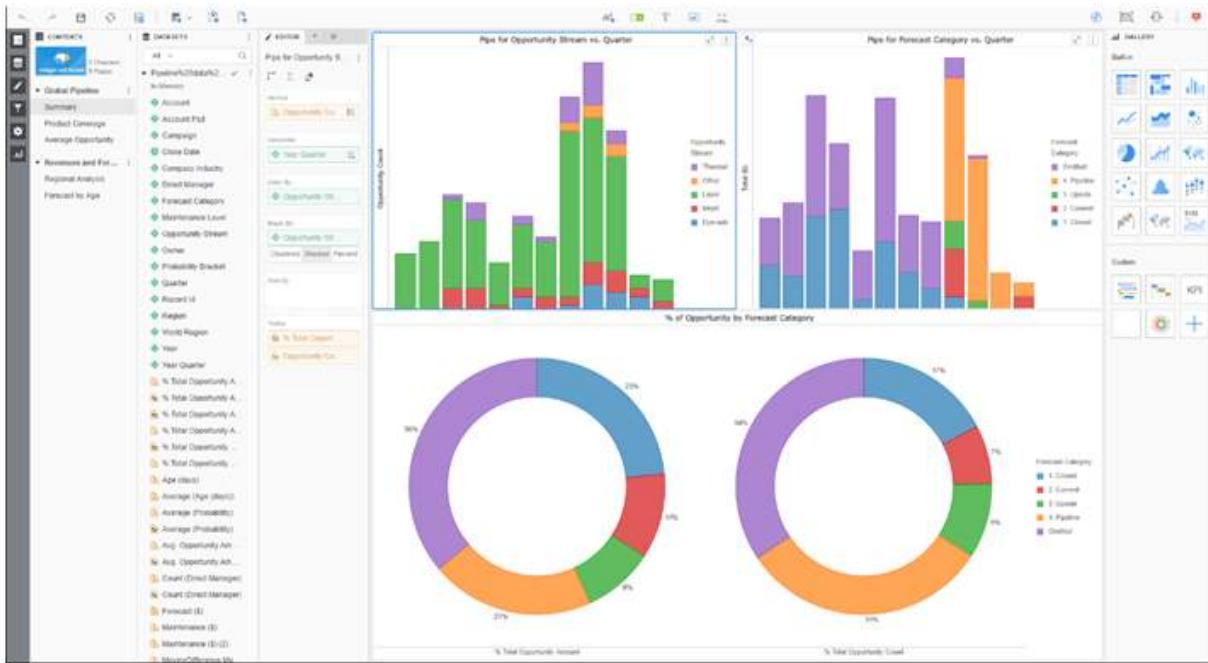
MicroStrategy Cloud proporciona un fácil acceso a una plataforma de inteligencia de negocio que permite implementar aplicaciones de forma sencilla reduciendo drásticamente riesgos y costes.

Como resultado de la consolidación en la industria de negocios inteligentes, MicroStrategy es uno de los pocos proveedores independientes de **software** BI.

Está [disponible aquí](#).

**Figura 3.** MicroStrategy.

Fuente: [MicroStrategy®](#).



## Tibco

Spotfire es la solución de descubrimiento de datos y BI de autoservicio de Tibco. Está diseñado para que el usuario obtenga datos y cree sus informes y procesos analíticos con independencia del departamento de IT.

Jaspersoft es una suite de **business intelligence** cuyo núcleo es **software open source**.

La plataforma de BI proporciona herramientas para la creación de una capa de metadatos sobre la que se crean informes de detalle, analíticos o cuadros de mando.

Se puede ver más en [este enlace](#).

## Sisense

Es un **software** de creación de informes de inteligencia de negocio que permite transformar fácilmente los datos en informes interactivos deslumbrantes.

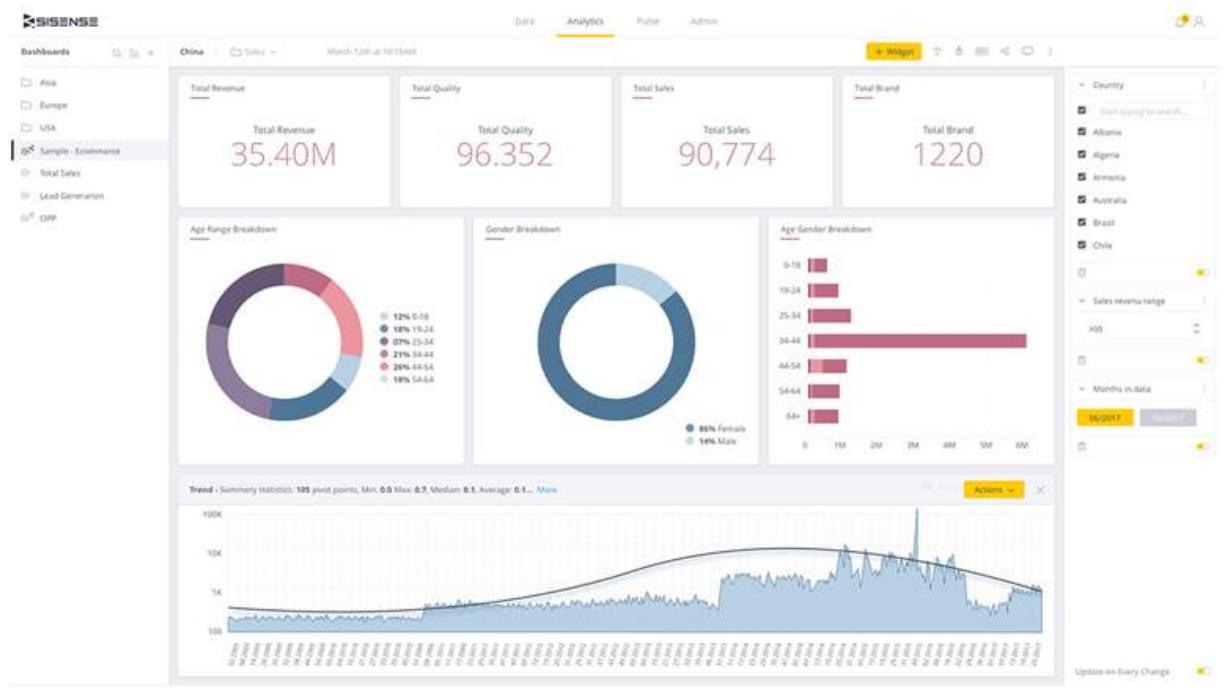
Con la solución integral de Sisense se puede generar conocimiento desde un conjunto de datos, con un coste en IT reducido. Cualquiera puede conectarse directamente a las bases de datos o aplicaciones,

mezclar múltiples fuentes de datos y visualizarlos en cuadros de mando o en informes en PDF.

Está disponible en [este enlace](#).

**Figura 4.** Sisense.

Fuente: [Sisense®](#).



## SAS

SAS Enterprise BI Server es una plataforma de **business intelligence** que proporciona acceso a informes vía web cuyo origen de datos pueden ser modelos relacionales o cubos OLAP de SAS.

SAS (anteriormente, Sistema de Análisis Estadístico) es un paquete de **software** estadístico desarrollado por SAS Institute para la gestión de datos, análisis avanzado, análisis multivariante, inteligencia empresarial y análisis predictivo. SAS se integra con Microsoft Office, incluyendo visualización de cuadros de mando y KPI en Outlook y Sharepoint, MS Excel, PowerPoint o Word.

El **software** complementario SAS Enterprise Guide proporciona un entorno de desarrollo de modelos analíticos de autoservicio para analistas y estadísticos.

La instalación habitual suele ser *on premise*, en servidores del cliente, aunque ofrece la posibilidad de realizar instalaciones en *cloud* para cualquier **software** de SAS, de manera que el servidor de BI también se podría alojar en la nube de SAS.

Está disponible en [este enlace](#).

**Figura 5.** SAS.  
Fuente: [Data Prix®](#).



# V. Herramientas de inteligencia de negocio: Pentaho Business Analytics

---

## 5.1. Iniciar el servidor de PENTAHO

Para iniciar el servidor de Pentaho Business Analytics en la máquina virtual, hay que acceder a:

`/home/pentaho/Escritorio/pentaho-server/`

Se hace clic en botón derecho y se abre un terminal en la ruta.

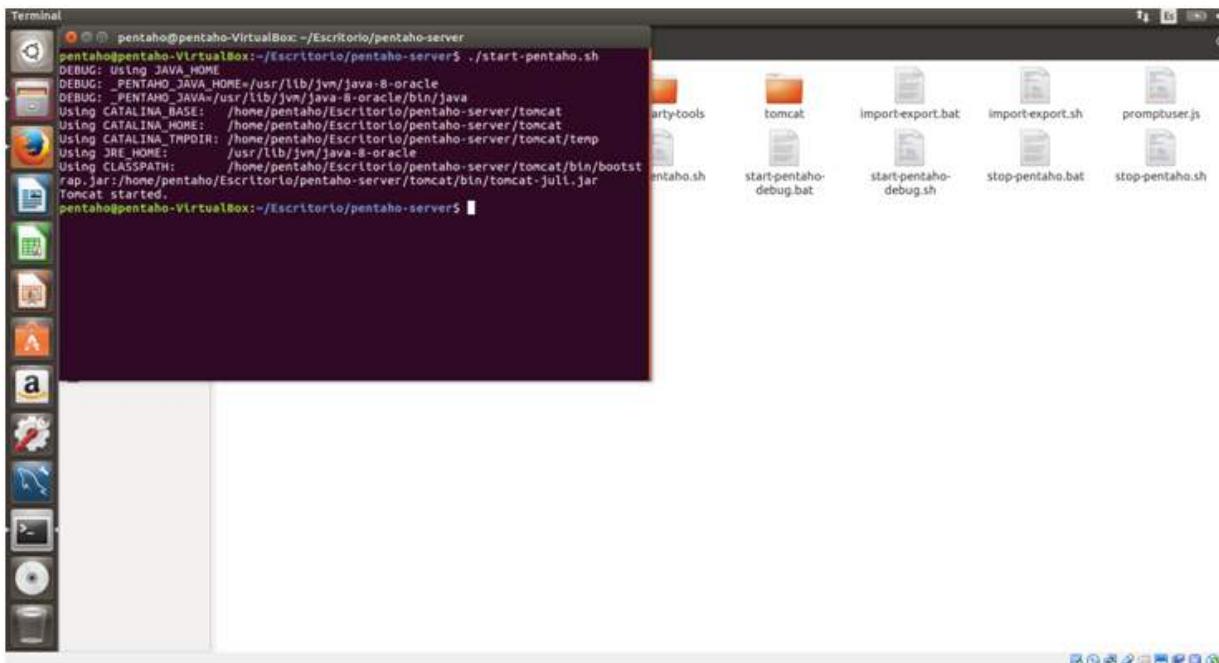


**Figura 6.** PBA.

Fuente: elaboración propia.

---

Ahora, se ejecuta "start-pentaho.sh" escribiendo `./start-pentaho.sh` en el terminal.



**Figura 7. PBA.**

Fuente: elaboración propia.

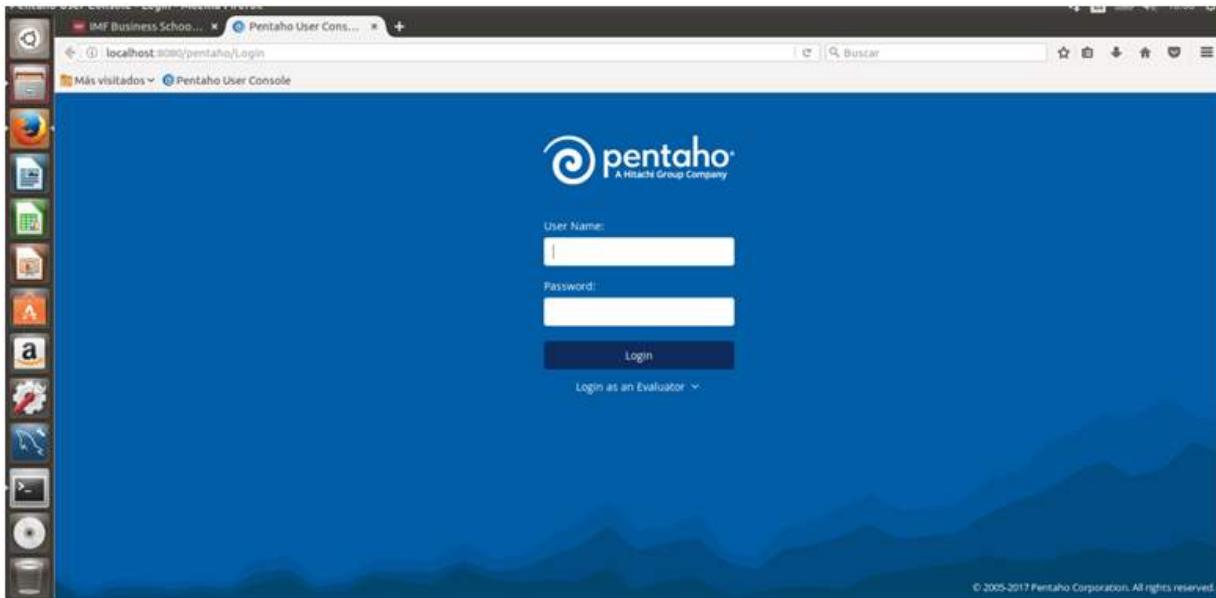
**CONTINUAR**

## 5.2. Acceso a la consola de usuario Pentaho (PUC)

Iniciar sesión en la PUC por primera vez es fácil siguiendo los pasos que se describen a continuación.

Hay que iniciar cualquier navegador web e ingresar la URL del servidor Pentaho: **localhost:8080**. También se puede acceder mediante el acceso directo de marcadores.

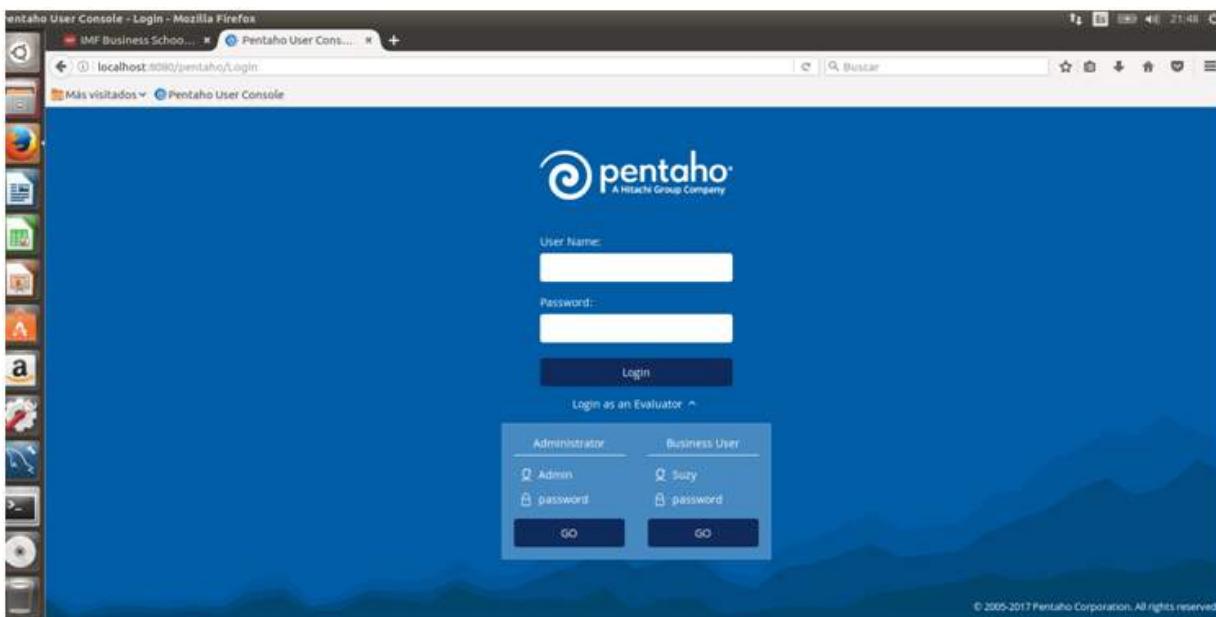
Como se puede ver, la página carga una pantalla introductoria con una sección de inicio de sesión:



**Figura 8.** PBA.

Fuente: elaboración propia.

Hay que introducir el nombre de usuario y contraseña y hacer clic en *Iniciar sesión* o usar la opción *Iniciar sesión como evaluador*.



**Figura 9.** PBA.

Fuente: elaboración propia.

## 5.3. Rápido tour por la consola de usuario Pentaho (PUC)

### 5.3.1. Home

Lo primero que el alumno verá después de iniciar sesión en la consola es la página de inicio, que sirve como punto de partida para la PUC y todas las tareas que realiza con ella.



Figura 10. PBA.

Fuente: elaboración propia.

Item	Nombre	Función

1	<i>Home view</i>	El menú desplegable de inicio permite ir fácilmente de página a página o regresar a la página de inicio.
2	<i>Getting Started</i>	Opciones recomendadas para comenzar a trabajar con Pentaho.
3	<i>Current User and Log Out</i>	Muestra el nombre de la persona que actualmente inició sesión en la consola de usuario. Al hacer clic en la flecha que está al lado del nombre, se puede cerrar la sesión de la consola de usuario.
4	<i>Browse Files</i>	Lleva a un explorador de ficheros, donde se pueden ubicar los archivos usando los paneles <i>Explorar</i> y <i>Archivos</i> , y administrarlos con el panel <i>Acciones</i> .  Todos los archivos que se abran aparecen en una nueva ventana.
5	<i>Create New</i>	Da opciones para crear un nuevo informe de análisis, informe interactivo o cuadro de mandos.  También se puede crear un nuevo origen de datos, si se tienen permisos suficientes.

6	<i>Documentation</i>	Lleva a la documentación oficial de Pentaho, que se abre en una nueva ventana o pestaña.
7	<i>Recents, Favorites</i>	Muestra una lista de los archivos abiertos más recientemente. Al hacer clic en la estrella junto a un archivo recientemente abierto, se agrega a la lista de favoritos.

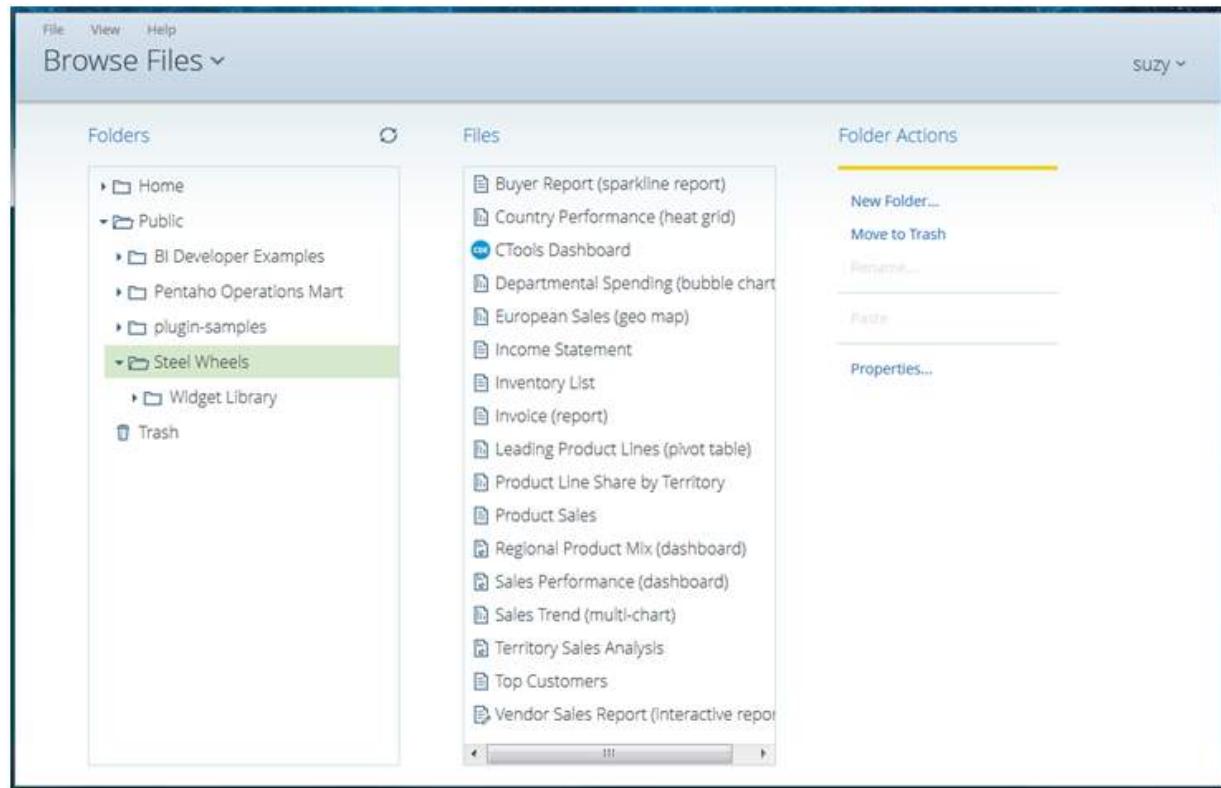
**Tabla 1.** Elementos fundamentales del *home*.

Fuente: elaboración propia.

**CONTINUAR**

### 5.3.2. *Browse Files* o explorador de ficheros

El explorador de archivos ayuda a mantener los archivos organizados y hace que sea más fácil encontrarlos y trabajar con ellos:

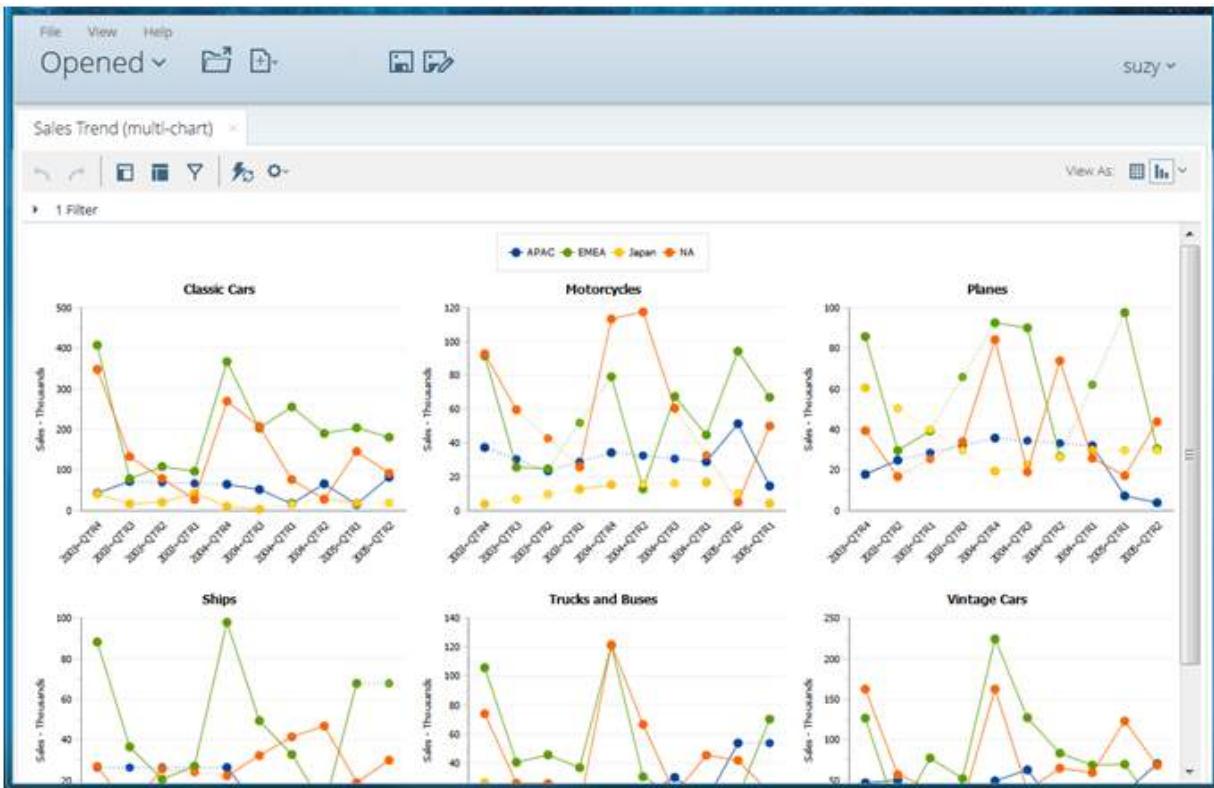


**Figura 11.** PBA.

Fuente: elaboración propia.

### 5.3.3. Opened

La opción *Opened* o abierto se activa después de abrir un archivo desde el explorador de archivos y proporciona un espacio simple para trabajar con dichos archivos. A continuación, se muestra un ejemplo:



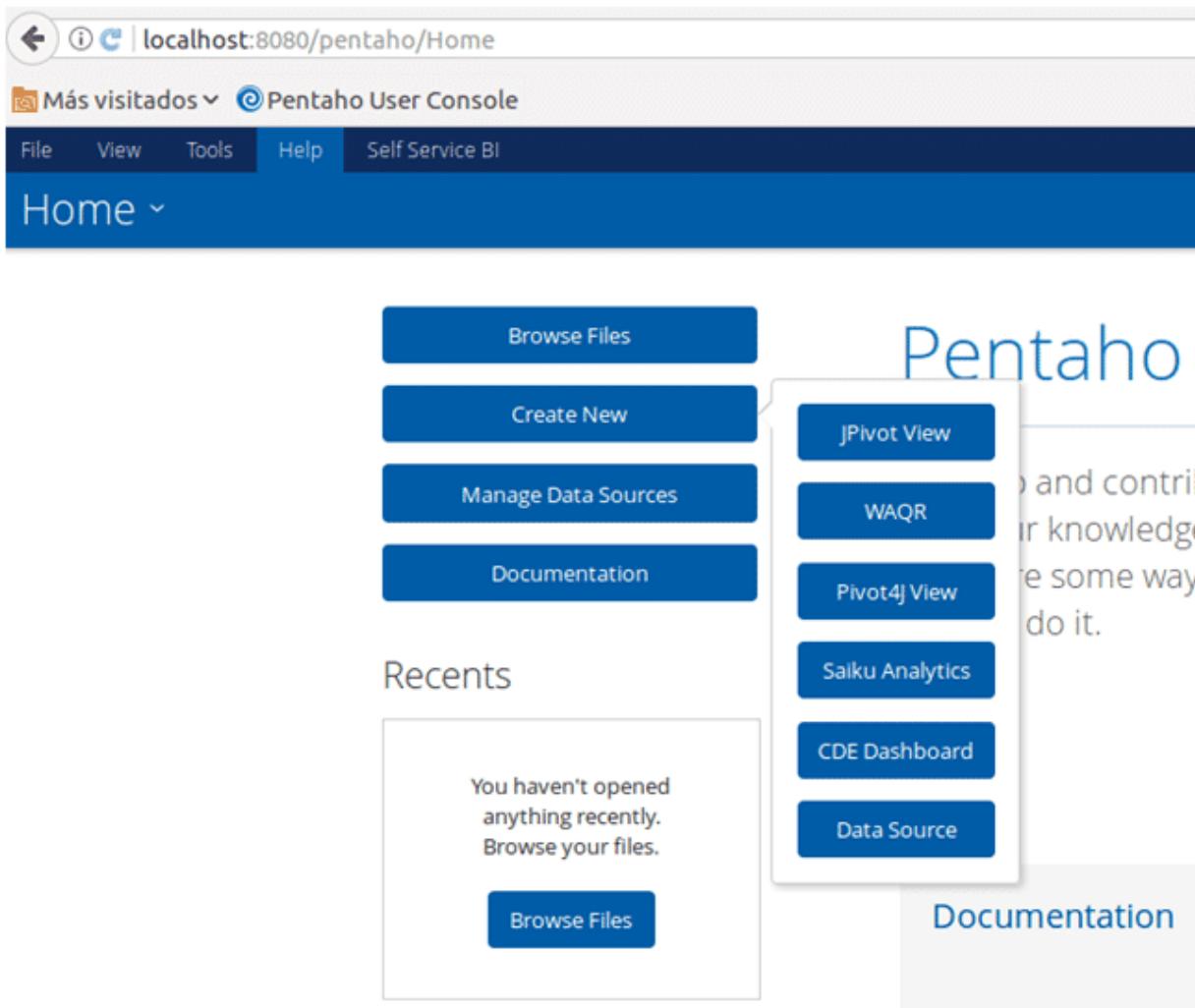
**Figura 12.** PBA.

Fuente: elaboración propia.

### 5.3.4. Herramientas Pentaho Community Edition

La versión Community de Pentaho incluye un grupo de herramientas básicas para realizar informes, cuadros de mando y vistas dinámicas OLAP. Aunque no son tan avanzadas como las que se pueden encontrar en la versión Enterprise, permiten al usuario diseñar y crear elementos básicos de análisis.

Las distintas herramientas que se pueden usar en la versión Community son accesibles cuando se crea un nuevo elemento:



**Figura 13.** PBA.

Fuente: elaboración propia.

A continuación, se resumen, en una tabla, las herramientas principales de la versión Community y las recomendaciones de cuándo usar cada una de ellas:

Explore Considerations	Opciones		
	<i>Reporting ad hoc con WAQR.</i>	Visores OLAP (Saiku, Pivot4j,	<i>PIN Board para dashboards.</i>

		Jpivot).	
<b>Resumen</b>	Interfaz de diseño basada en web que se utiliza para crear informes operativos sencillos, como <i>ad hoc</i> , sin depender de los desarrolladores de TI o de informes.	Herramienta intuitiva de visualización analítica sobre modelos multidimensionales y OLAP que permite navegar por la información.	Permite a los usuarios crear <i>dashboards</i> con poca o ninguna capacitación. Contiene varios informes diferentes reunidos dentro de una pantalla.
<b>Grado de especialización y conocimiento</b>	Conocimiento de funciones básicas del PC, como sistemas operativos y navegadores web.	Conocimiento de funciones básicas del PC, como sistemas operativos y navegadores web.	Conocimiento de funciones básicas del PC, como sistemas operativos y navegadores web.
<b>Recomendación</b>	Es recomendable usar informes interactivos si se desea crear un informe tabular que responda a una pregunta comercial inmediata. Resulta muy profesional, se puede imprimir rápidamente y brinda un control significativo sobre los elementos de formato, como fuentes, ancho o clasificación de columna, colores de fondo y más.	Es recomendable usar los informes OLAP si se desea agregar datos rápidamente mientras se exploran visualmente los datos. Realiza una clasificación avanzada y el filtrado de los datos. También es recomendable si se desea ver las visualizaciones de gráficos, que incluyen una iluminación de parada detallada.	Es recomendable usar el diseñador de cuadros de mando si se desea crear una interfaz para ver muchos informes diferentes a la vez. También para ver gráficos y tablas dinámicas dentro de un espacio mientras se crean informes en otro.

**Tabla 2.** Herramientas de exploración de Community Edition.

Fuente: elaboración propia.

**CONTINUAR**

### 5.3.5. *Schedules* (programaciones)

Permite hacer seguimiento de las programaciones de informes que se hayan creado.

Todos los informes programados activos aparecen en la lista de programas, a los que se puede acceder haciendo clic en el menú desplegable *Inicio* y luego en el enlace *Horarios*, en la esquina superior izquierda de la PUC.

La lista de programas muestra qué informes están programados para ejecutarse, el patrón de recurrencia para el programa, cuándo fue la última vez que se ejecutó, cuándo se configura para ejecutarse nuevamente y el estado actual de la programación. Se pueden editar y mantener cada una de las programaciones utilizando los controles que se encuentran en la parte superior de la lista de horarios, en el extremo derecho de la barra de herramientas:

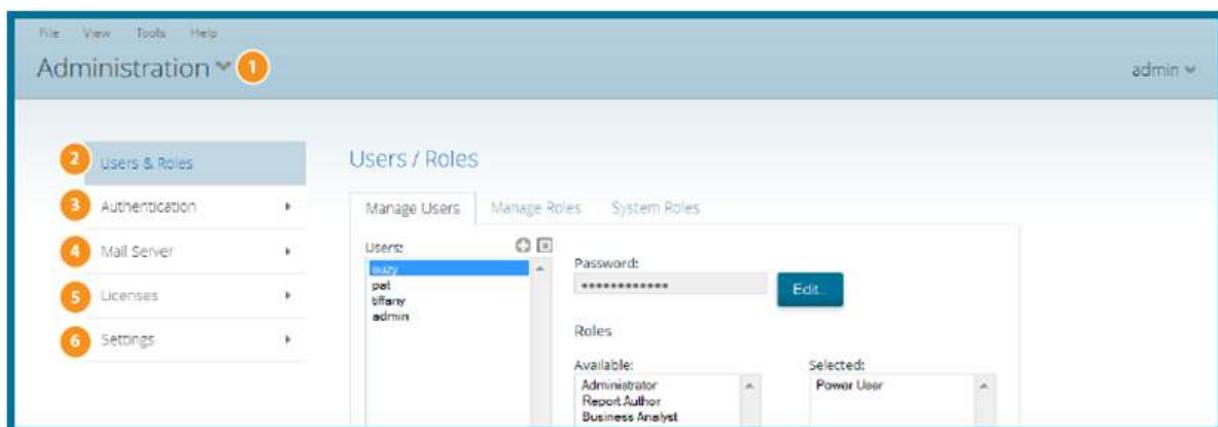
Schedule Name	Repeats	Source File	Output Location	Last Run	Next Run	Status
Inventory List (report)	The first Sunday of every month at 17:00:00	/public/Steel Wheels/Inventory List (report)	/home/suzy	-	2015 Feb 1 17:00:00	Normal
Leading Product Lines (pivot table)	Every Friday at 23:00:00	/public/Steel Wheels/Leading Product Lines (pivot table)	/home/suzy	-	2015 Jan 16 23:00:00	Normal
European Sales (geo map)	Every Sunday at 23:00:00	/public/Steel Wheels/European Sales (geo map)	/home/suzy	-	2015 Jan 18 23:00:00	Normal

**Figura 14.** PBA.  
Fuente: elaboración propia.

### 5.3.6. Administración

En la PUC, la administración puede ser utilizada por usuarios que hayan iniciado sesión y que tengan asignada una función que tenga permisos para administrar la seguridad y que, a su vez, puedan realizar tareas de configuración y mantenimiento del sistema.

La administración aparece en el menú desplegable de la izquierda en la página de inicio. Se puede hacer clic para mostrar los elementos del menú específicos para la administración del servidor Pentaho. Si no se tienen privilegios de administración, no aparece dentro de la consola:



**Figura 15.** PBA.  
Fuente: elaboración propia.

Item	Nombre	Función
1	<i>Administration</i>	La perspectiva de administración permite configurar usuarios,

			configurar el servidor de correo, cambiar la configuración de autenticación en el servidor Pentaho e instalar licencias de software para Pentaho.
2	<i>Users &amp; Roles</i>	Permite administrar los usuarios o roles de Pentaho para el servidor Pentaho. Permite administrar los usuarios o roles de Pentaho para el servidor Pentaho.	
3	<i>Authentication</i>	Opción para integrar con LDAP/Active Directory.	
4	<i>Mail Server</i>	Configura el servidor de correo saliente y la cuenta utilizada para enviar informes por correo electrónico.	
5	<i>Licenses</i>	Administra la configuración para eliminar archivos generados antiguos, ya sea manualmente o mediante la creación de una programación para su eliminación.	
6	<i>Settings</i>	Lleva a la documentación oficial de Pentaho, que se abre en una nueva ventana o pestaña.	

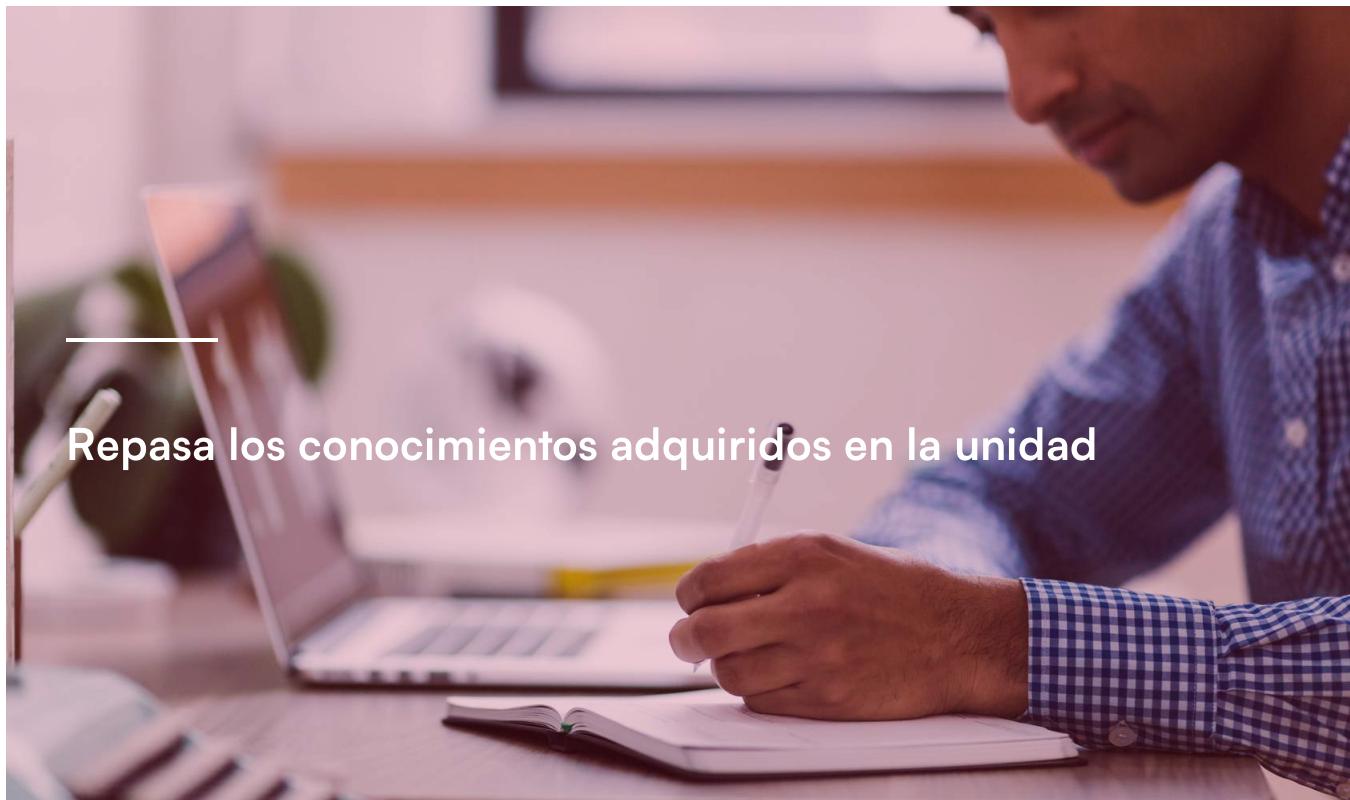
**Tabla 3.** Funciones de administración en Community Edition.

Fuente: elaboración propia.



## VI. Resumen

---



**Repasa los conocimientos adquiridos en la unidad**

En esta unidad se han mostrado ejemplos de algunos casos de aplicación de inteligencia de negocio para que el alumno pueda familiarizarse con la puesta en práctica de este tipo de tecnología y soluciones.

Se han mostrado casos de uso reales de aplicación de sistemas de inteligencia de negocio de varios sectores, como el sector de la automoción, recambios, industrial, salud y hotelero.

En el caso del sector de la automoción, se refiere a una empresa de alquiler y administración de vehículos que solicita una solución de inteligencia de negocio que permita a la compañía optimizar la gestión de sus flotas, mejorar la calidad del servicio y monetización de la información. Con unos objetivos tan claros, el reto es recopilar la información de forma automática y en tiempo real, por lo que se usaron tecnologías *big data* para realizar el *streaming*.

El caso de uso industrial es la búsqueda de la mejora del rendimiento mediante el mantenimiento proactivo y predictivo, anticipando las averías de la máquina. La monitorización del estado o condición de los elementos de interés de la cadena de producción es clave en este sentido. La complejidad de la predicción del fallo a menudo se debe a la enorme cantidad de posibles factores o variables de influencia. Las fuentes de datos pueden ser múltiples y dependen del escenario.

El sector de la salud y el de la farmacia son y serán en el futuro grandes demandantes de soluciones *big data* y *business intelligence*, por su gran variedad de datos, volumetría y el valor intrínseco de estos. En el caso de uso se habla desde el punto de vista de la asistencia sanitaria y la monitorización de los pacientes, pero en este ámbito entran en juego desde la buena gestión de los recursos hasta la predicción de enfermedades.

Por último, el sector hotelero, uno de los motores económicos más importantes para muchos países. En este caso, la aplicación de técnicas de *data analytics* es tan extensa como variopintos son sus datos. El caso de uso se centra en un *forecast*, herramienta que se usa en cualquier sector, pero en el caso hotelero se centra en la predicción de la ocupación del año en curso.

También se han detallado las herramientas de inteligencia de negocio más comunes en el mercado actual, acorde con el diagrama de Gartner. Entre ellas, Qlik, MicroStrategy, Tibco, Sisense y SAS.

Del mismo modo, se ha realizado una introducción a la solución Pentaho Business Analytics y se ha llevado a cabo una primera toma de contacto para que el alumno asimile los conceptos y tecnologías que se han ido desarrollando a lo largo de las unidades de este módulo.

En el caso práctico que se expone más adelante, el alumno podrá llevar a la práctica los conceptos adquiridos.



Además, como complemento a la unidad, se ofrece a los estudiantes la posibilidad de acceder de forma gratuita a los materiales y laboratorios prácticos de la herramienta Qlik Sense. Para registrarse, se deberán seguir los pasos que se indican en el siguiente anexo:



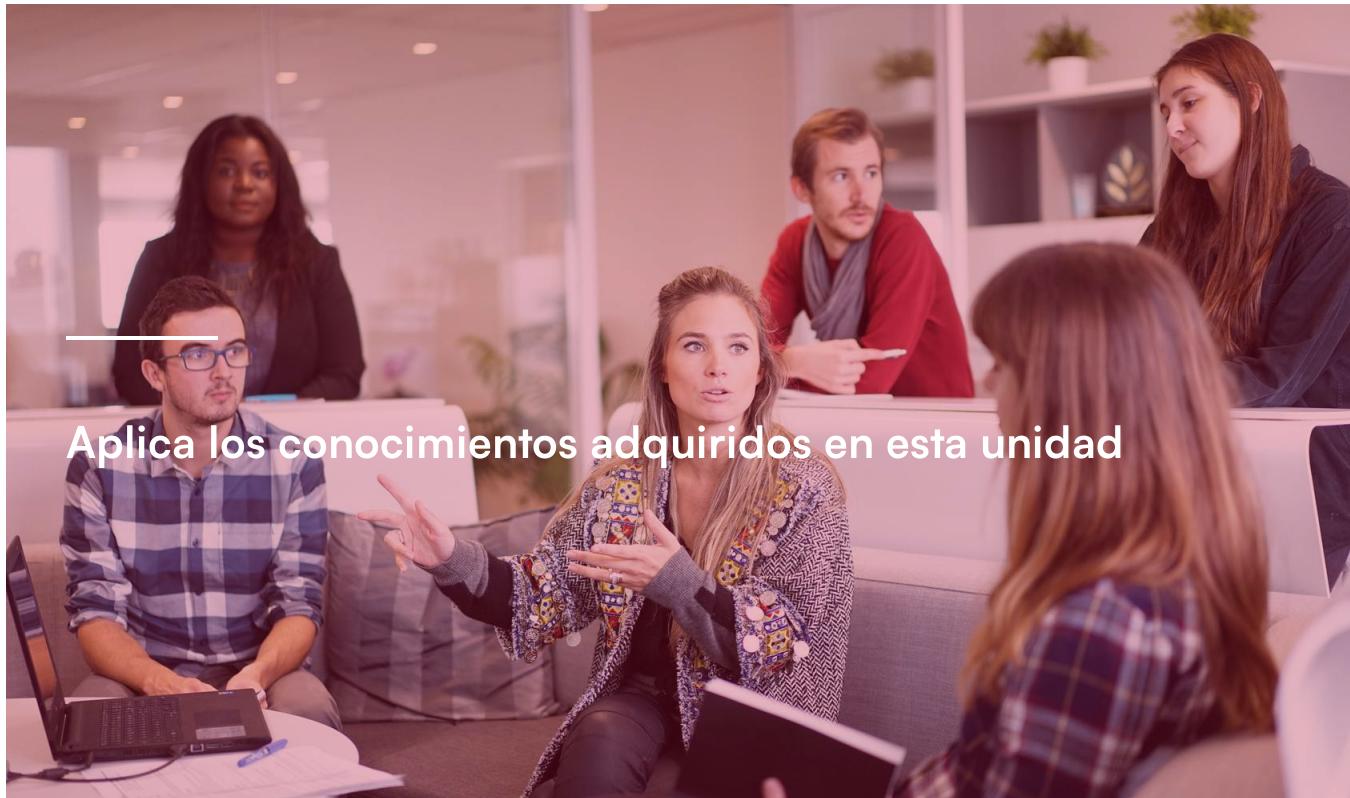
**Anexo\_QlikSense.pdf**

320.5 KB



## VII. Caso práctico con solución

---



**Aplica los conocimientos adquiridos en esta unidad**

### SE PIDE

Haciendo uso de Pentaho Business Analytics, el departamento de ventas de una compañía, realizar un modelo multidimensional OLAP de prueba, para poder realizar un análisis de la información departamental.

Para ello, el departamento de ventas facilita un fichero fuente de datos, a partir del cual hay que diseñar el modelo multidimensional que implementar:

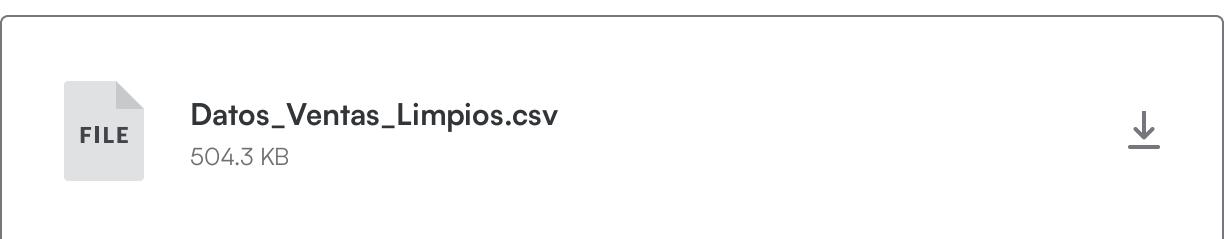
- Se puede descargar dicho fichero de datos en el siguiente enlace: [Datos\\_Ventas\\_Limpios.csv](#)
- Este fichero se adjunta también en la máquina virtual.

El objetivo es generar una solución que proporcione al equipo de ventas la capacidad de analizar sus datos mediante informes dinámicos OLAP.

**Figura 17.** Fuente de datos para realizar un análisis de la información departamental.

Fuente: elaboración propia.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	DE
1	ORDERNU	QUANTITY	PRICEACI	ORDERLIN	SALES	ORDERDATE	STATUS	QTR_ID	MONTH_ID	YEAR_ID	PRODUCTC	MSRP	PRODUCTCUSTOME	PHONE	ADDRESS1	ADDRESS2	CITY	STATE	POSTALC	COUNTRY	TERRITORI	CONTACTC	CONTACTN	CONTACTF	DE	
2	10107	30	95,7	2	2871	02/24/2003	Shipped	1	2	2003	Motorcycl	95	S10_1678	Land of Tc	2,13E+09	897 Long Airport Ave NYC	NY		10022	USA	NA	Yu	Kwai	Sri		
3	10121	34	81,35	5	2765,9	05/07/2003	Shipped	2	5	2003	Motorcycl	95	S10_1678	Reims Col 26.47.155159 rue de l'Abbaye	Reims			51100	France	EMEA	Henriot	Paul	Sri			
4	10134	41	94,74	2	3884,34	07/01/2003	Shipped	3	7	2003	Motorcycl	95	S10_1678	Lyon Souv +33 1 46 6,27 rue du Colonel Pi	Paris			75508	France	EMEA	Da Cunha	Daniel	M			
5	10145	45	83,26	6	3746,7	08/25/2003	Shipped	3	8	2003	Motorcycl	95	S10_1678	ToysAGrov 6,27E+09 78934 Hillside Dr.	Pasadena CA			90003	USA	NA	Young	Julie	M			
6	10159	49	100	14	5205,27	10/10/2003	Shipped	4	10	2003	Motorcycl	95	S10_1678	Corporate 6,51E+09 7734 Strong St.	San Franci CA				USA	NA	Brown	Julie	M			
7	10168	36	96,66	1	3479,76	10/28/2003	Shipped	4	10	2003	Motorcycl	95	S10_1678	Technics S 6,51E+09 9408 Furth Circle	Burlingam CA			94217	USA	NA	Hirano	Juri	M			
8	10180	29	86,13	9	2497,77	11/11/2003	Shipped	4	11	2003	Motorcycl	95	S10_1678	Daedalos 20,16.1551180, chausse de Tou Lille			59000	France	EMEA	Rance	Martine	Sri				
9	10184	48	100	1	5512,32	11/18/2003	Shipped	4	11	2003	Motorcycl	95	S10_1678	Herku Gi 47 2267 9 Dammaren 121, PR 74 Bergen			N 5804	Norway	EMEA	Oeztan	Veysel	M				
10	10201	22	98,57	2	2168,54	12/01/2003	Shipped	4	12	2003	Motorcycl	95	S10_1678	Mini Whe 6,51E+09 5557 North Pendale	San Franci CA				USA	NA	Murphy	Julie	Sri			
11	10211	41	100	14	4708,44	01/15/2004	Shipped	1	1	2004	Motorcycl	95	S10_1678	Auto Can(1)47.55.25, rue Lauriston	Paris			75016	France	EMEA	Perrier	Dominiqu	M			
12	10223	37	100	1	3965,66	02/20/2004	Shipped	1	2	2004	Motorcycl	95	S10_1678	Australiar 03 9520 45 6365 St Klc Level 3	Melbourn Victoria			3004	Australia	APAC	Ferguson	Peter	M			
13	10237	23	100	7	2333,12	04/05/2004	Shipped	2	4	2004	Motorcycl	95	S10_1678	Vitachron 2,13E+09 2678 King Suite 101	NYC NY			10022	USA	NA	Frick	Michael	Sri			
14	10251	28	100	2	3188,64	05/18/2004	Shipped	2	5	2004	Motorcycl	95	S10_1678	Tekni Coll 2,02E+09 7476 Moss Rd.	Newark NJ			94019	USA	NA	Brown	William	M			
15	10263	34	100	2	3676,76	06/28/2004	Shipped	2	6	2004	Motorcycl	95	S10_1678	Gift Depo! 2,04E+09 25593 South Bay Ln.	Bridgewat CT			97562	USA	NA	King	Julie	M			
16	10275	45	92,83	1	4177,35	07/23/2004	Shipped	3	7	2004	Motorcycl	95	S10_1678	La Rochell40,67.855,67, rue des Cinquant Nantes			44000	France	EMEA	Labrune	Janine	M				
17	10285	36	100	6	4099,68	08/27/2004	Shipped	3	8	2004	Motorcycl	95	S10_1678	Marta's Re 6,18E+09 39323 Spinaker Dr.	Cambridg MA			51247	USA	NA	Hernande	Marta	M			
18	10299	23	100	9	2597,39	09/30/2004	Shipped	3	9	2004	Motorcycl	95	S10_1678	Toys of Fii 90-224 851 Keskuskatu 45	Helsinki			21240	Finland	EMEA	Karttunen	Matti	Sri			
19	10309	41	100	5	4394,38	10/15/2004	Shipped	4	10	2004	Motorcycl	95	S10_1678	Baan Mir 07-98 9555 Erling Skakkes gate 7 Stavern			4110	Norway	EMEA	Bergulffse	Jonas	M				
20	10318	46	94,74	1	4358,04	11/02/2004	Shipped	4	11	2004	Motorcycl	95	S10_1678	Diecast Cl 2,16E+09 7586 Pompton St.	Allentown PA			70267	USA	NA	Yu	Kyung	M			
21	10329	42	100	1	4396,14	11/15/2004	Shipped	4	11	2004	Motorcycl	95	S10_1678	Land of Tc 2,13E+09 897 Long Airport Ave NYC	NY			10022	USA	NA	Yu	Kwai	M			
22	10341	41	100	9	7737,93	11/24/2004	Shipped	4	11	2004	Motorcycl	95	S10_1678	Salzburg 6562-9555 Geisweg 14	Salzburg			5020	Austria	EMEA	Pipps	Georg	La			
23	10361	20	72,55	13	1451	12/17/2004	Shipped	4	12	2004	Motorcycl	95	S10_1678	Souvenir+61 2 9495 Monitor N Level 6	Chatswo NSW			2067	Australia	APAC	Huxley	Adrian	Sri			
24	10375	21	34,91	12	733,11	02/03/2005	Shipped	1	2	2005	Motorcycl	95	S10_1678	La Rochell40,67.855,67, rue des Cinquant Nantes			44000	France	EMEA	Labrune	Janine	Sri				
25	10388	42	76,36	4	3207,12	03/03/2005	Shipped	1	3	2005	Motorcycl	95	S10_1678	FunGiftId 5,09E+09 1785 First Street	New Beaf MA			50553	USA	NA	Benitez	Violeta	M			
26	10403	24	100	7	2434,56	04/08/2005	Shipped	2	4	2005	Motorcycl	95	S10_1678	UK Collect(171) 555- Berkeley Gardens 12 Liverpool			WX1 BLT	UK	EMEA	Devon	Elizabeth	Sri				
27	10417	66	100	2	7516,08	05/13/2005	Disputed	2	5	2005	Motorcycl	95	S10_1678	Euro Shop (91) 555 9 C/ Moratalzar, 86	Madrid			28034	Spain	EMEA	Freyne	Diego	La			
28	10103	26	100	11	5404,62	01/29/2003	Shipped	1	1	2003	Classic Cai	214	S10_1349	Baan Mir 07-98 9555 Erling Skakkes gate 7 Stavern			4110	Norway	EMEA	Bergulffse	Jonas	M				
29	10112	29	100	1	7209,11	03/24/2003	Shipped	1	3	2003	Classic Cai	214	S10_1349	Volvo Mo 0921-12 3 Berguvv-gen 8	Lule			5-558,22	Sweden	EMEA	Berglund	Christina	La			



## VER SOLUCIÓN

### SOLUCIÓN

Para plantear la solución de este caso práctico, primero hay que hacer un rápido análisis del enunciado. Es evidente que es bastante abierto, por lo que podrían admitirse varias soluciones.

Para realizar este rápido análisis, será conveniente hacer uso de conceptos explicados en unidades anteriores.

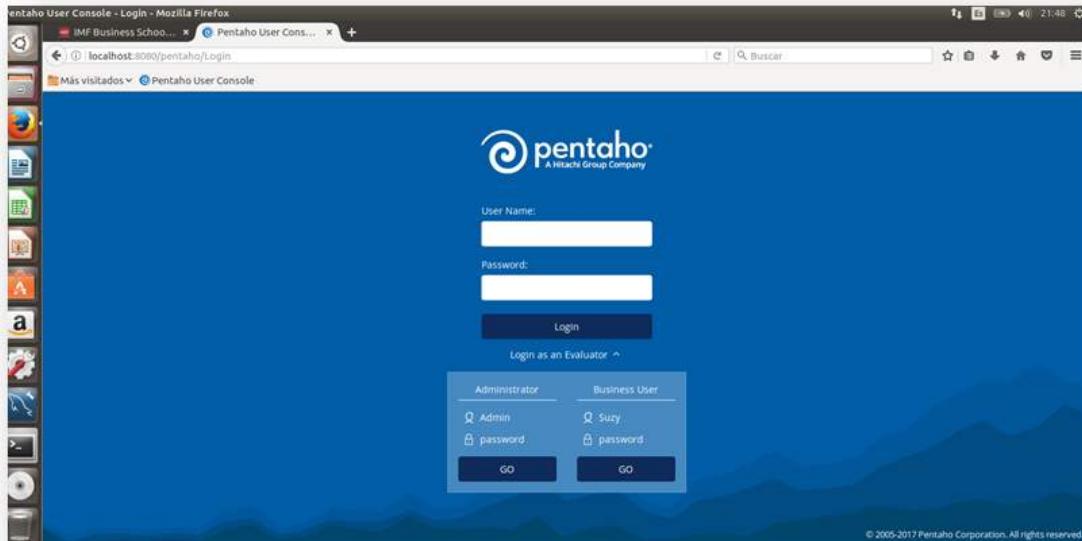
En primer lugar, el enunciado solicita hacer uso de la solución Pentaho Business Analytics, que es una solución HOLAP, aunque aquí se utilizará como ROLAP, u OLAP sobre relacional, lo que implica que se debe usar una base de datos relacional para implementar la capa física y una capa de metadatos para implementar el modelo lógico multidimensional.

Del mismo modo, el enunciado indica que se trata de una prueba de concepto para determinar la viabilidad de este desarrollo, lo que implica carácter local y temporal del desarrollo. Con este requerimiento, es recomendable plantear una estrategia de diseño inicial de Kimball o *bottom-up* que permita comenzar a nivel local o departamental, para escalar al resto de la organización.

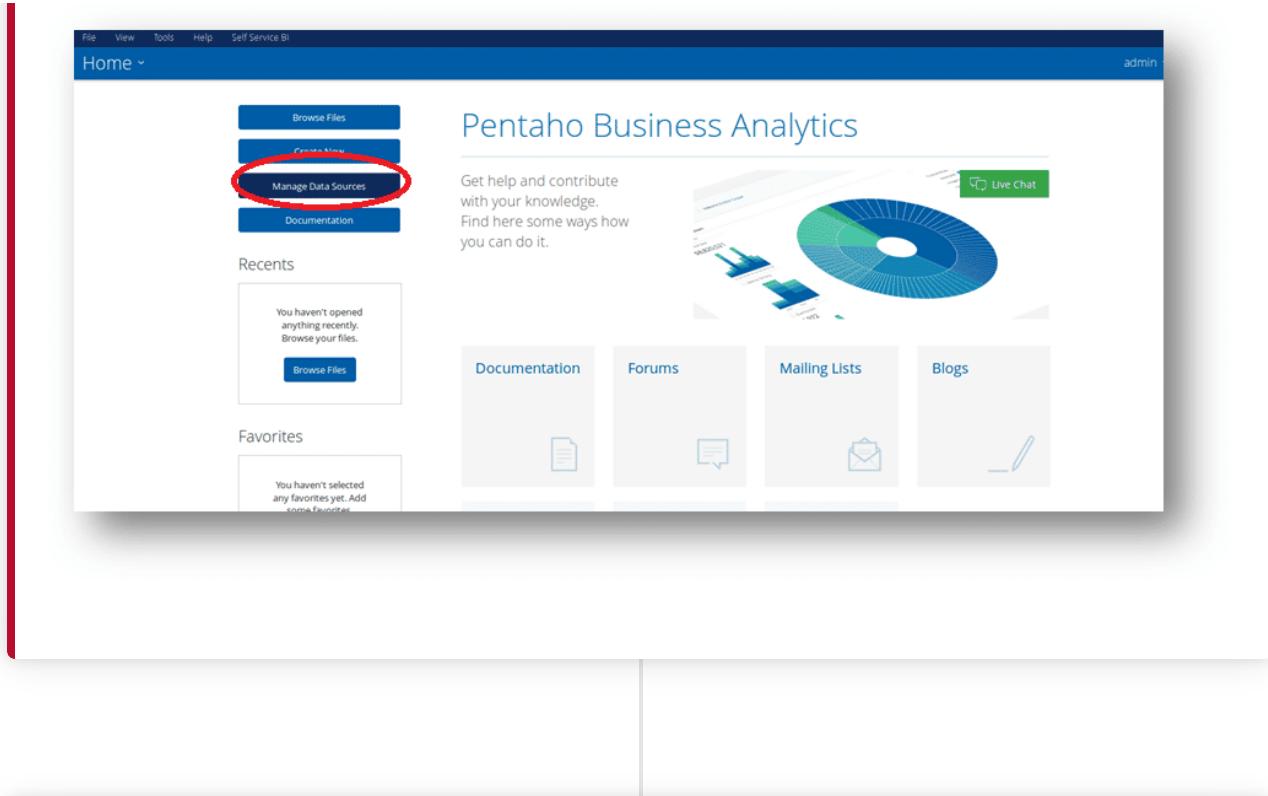
Sin embargo, el carácter de prueba que se señala en el enunciado hace que sea más conveniente utilizar herramientas ágiles de modelado, orientadas a obtener un resultado final que el usuario pueda explotar y evaluar, que el método tradicional de diseño e implementación, que puede ser más profundo, pero también más lento. Es por ello que, para este caso, en vez de crear un proceso ETL específico para la fuente de datos identificada, y generar un almacén de datos físico sobre la base de datos relacional, se realizará una implementación ágil del modelo lógico multidimensional, que se encuentra por encima del modelo físico, obviando, por tanto, este modelo físico y haciendo que Pentaho cargue en memoria los datos directamente.

Para realizar el modelo multidimensional, hay que acceder a la máquina virtual.

Para acceder a la PUC de Pentaho Business Analytics, hay que asegurarse de que está el servicio arrancado haciendo doble clic en “**start-pentaho.batch**”, que se encuentra en el escritorio. En la máquina virtual, hay que acceder a la URL **localhost:8080** del navegador Edge, y loguearse como usuario *Admin*.



Una vez en la PUC de Pentaho, para crear el modelo multidimensional, hay que identificar la nueva fuente de datos, en este caso, el fichero. Para ello, hay que acceder a la opción del menú lateral **Manage Data Sources**.



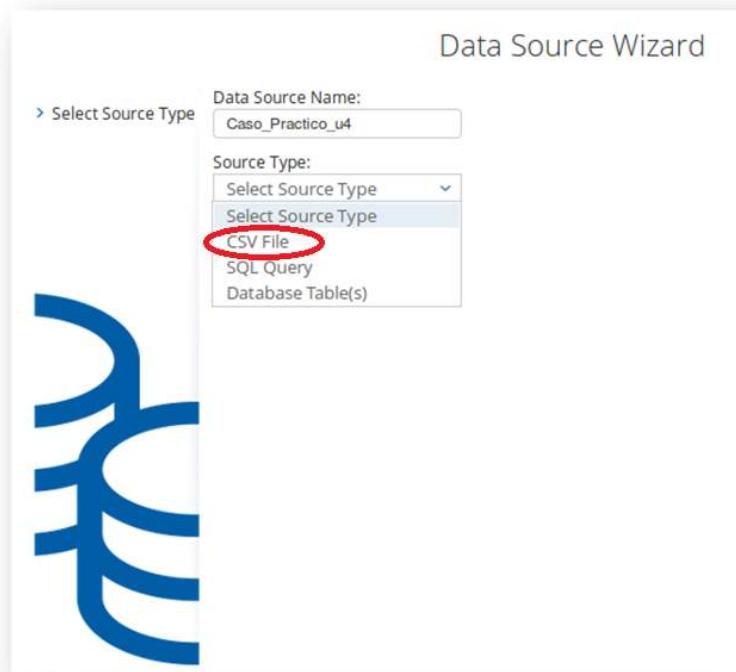
Ahora, se debe seleccionar la opción **New Data Source** en el nuevo menú que aparece.

The screenshot shows a modal dialog box titled "Manage Data Sources". At the top right of the dialog is a blue button labeled "New Data Source" which is circled in red. Below this button is a table with two columns: "Datasource" and "Type". There is one entry in the table: "MySql Local" under "Datasource" and "JDBC" under "Type". At the bottom right of the dialog is a blue "Close" button.

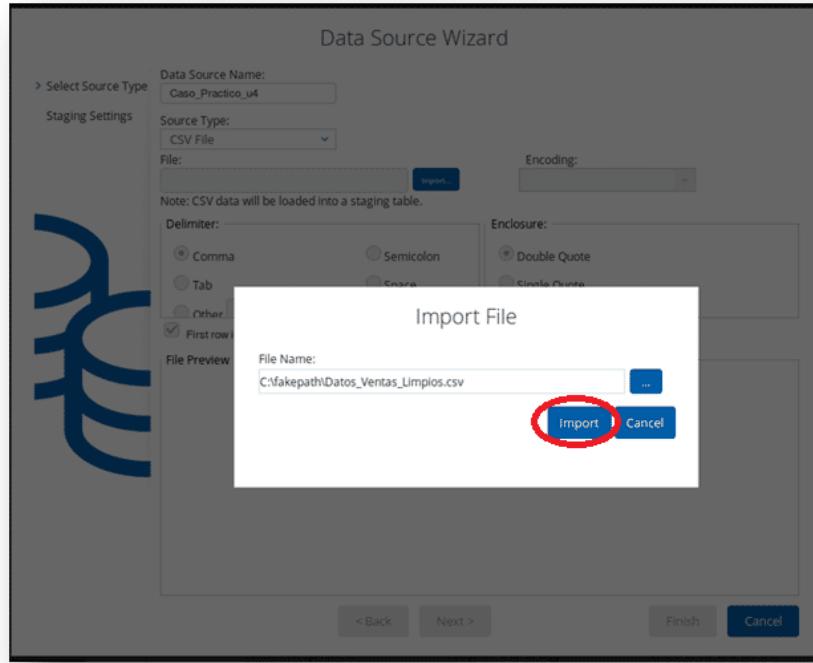
Datasource	Type
MySql Local	JDBC

Esta opción lanzará un proceso a modo de Wizard que permitirá crear e identificar la fuente de datos y el modelo que se quiere implementar.

El primer paso es seleccionar la fuente de datos, en este caso, un fichero CSV, por lo que hay que seleccionar un nombre para esta nueva fuente y el tipo de fuente (CSV, en este caso).

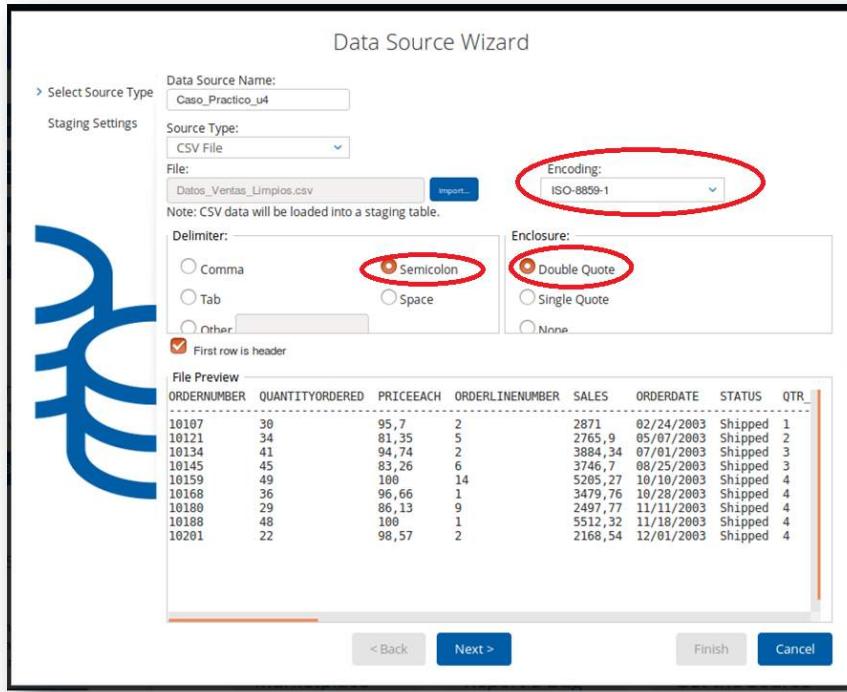


Una vez seleccionado el tipo de *data source*, se podrá importar el fichero CSV de la ruta donde esté ubicado.



Una vez importado, será necesario configurar tanto el *encoding* usado por el fichero como el delimitador de campos usados y de texto.

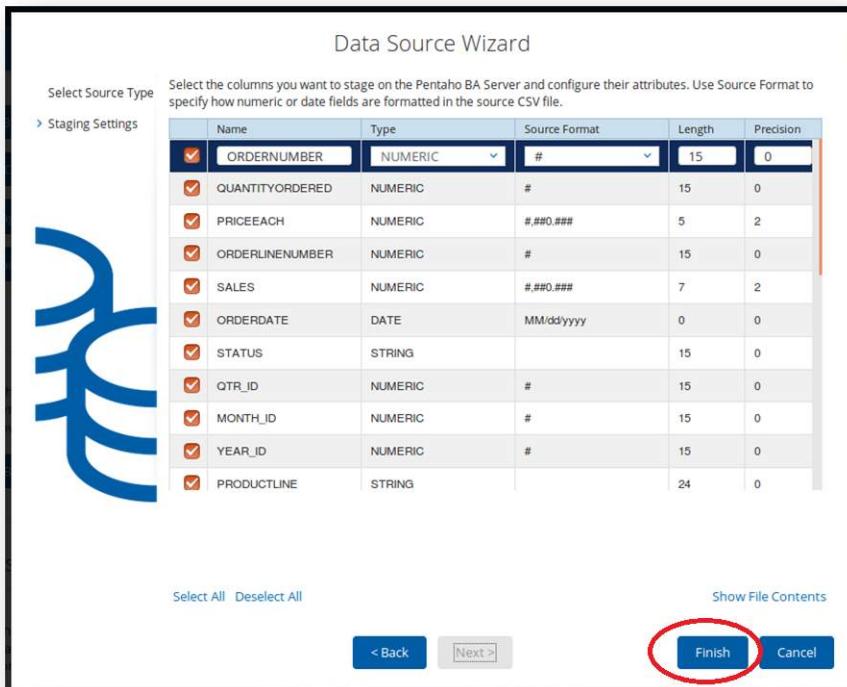
Se podrán ver en todo momento los datos que se van recuperando en la parte inferior de la pantalla del proceso.



Es importante asegurarse de que los datos se previsualizan bien y usar los separadores adecuados.

Tras esto, hay que hacer clic sobre *Next* para pasar a la siguiente fase.

A continuación, Pentaho proporciona un listado de todos los nombres de campos con una descomposición de tipos, longitud y formato para cada uno de ellos, y la posibilidad de modificarlos o ajustarlos al criterio propio.



Cuando se ha finalizado, hay que clicar sobre el botón *Finish* para terminar con la importación de la nueva fuente de datos.

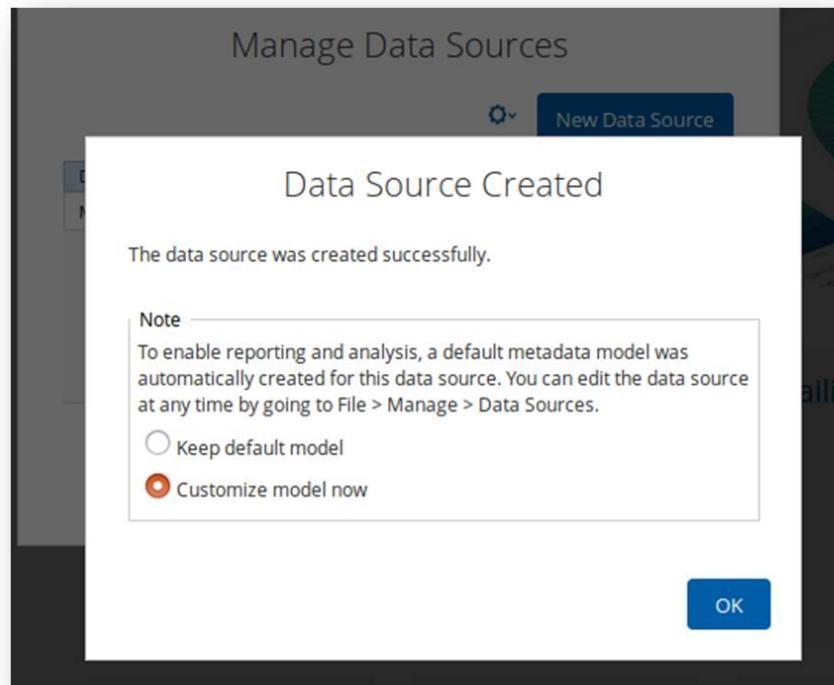
Este último paso dará lugar a la carga automática en memoria de los datos, para su posterior explotación, y dará comienzo el proceso de creación del modelo multidimensional.

En este proceso, Pentaho ha creado una tabla de datos relacional que conformará el modelo físico del nuevo modelo: se trata de la capa física, pero, como ya es sabido, se necesita, además, una capa lógica,

donde se define el modelo multidimensional, que es lo que permite traducir/ejecutar un motor OLAP

multidimensional de Pentaho (Mondrian) sobre una base de datos relacional.

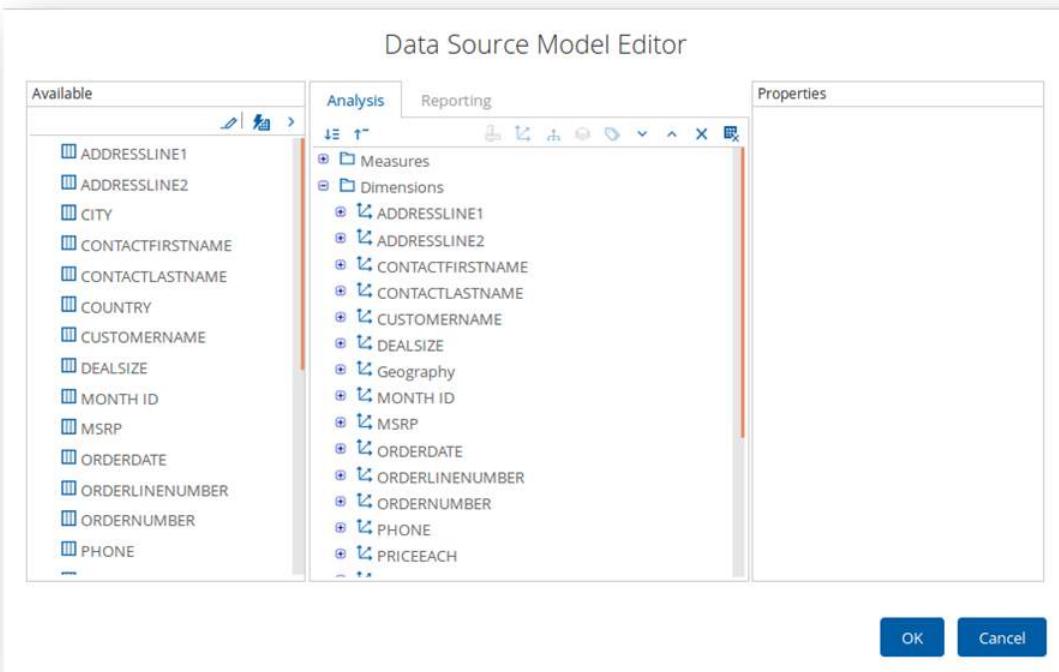
Wizard de Pentaho lanza automáticamente la fase de creación del modelo multidimensional. Y para ello da la opción de crear uno por defecto automático o customizarlo/crearlo uno mismo. Se elegirá esta segunda opción para ver qué es lo que realmente se está haciendo.



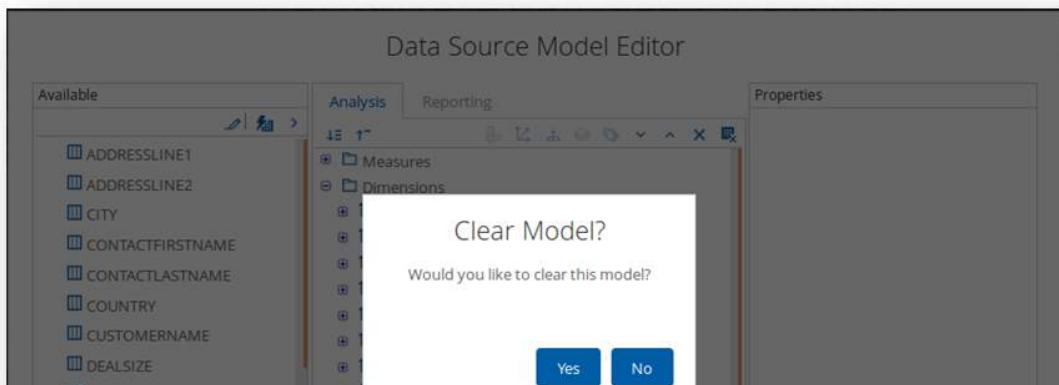
Una vez que se ha indicado que se desea crear un modelo propio, hay que acceder a un editor ligero de creación de cubos OLAP.

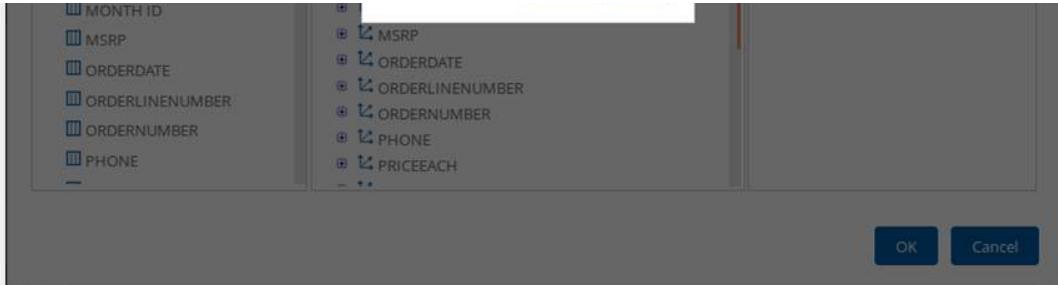
En una primera aproximación, Pentaho propone el modelo más sencillo de cubo, que sería el que considera cada campo de tipo texto como dimensión, y cada campo de tipo numérico como métricas.

Sin embargo, esto no es lo que aquí se está buscando, por lo que habrá que rehacerlo desde el principio. Para ello, se borra la propuesta por defecto clicando sobre el botón *Clear model*.



Se acepta. Esto borrará todo el modelo inicial.



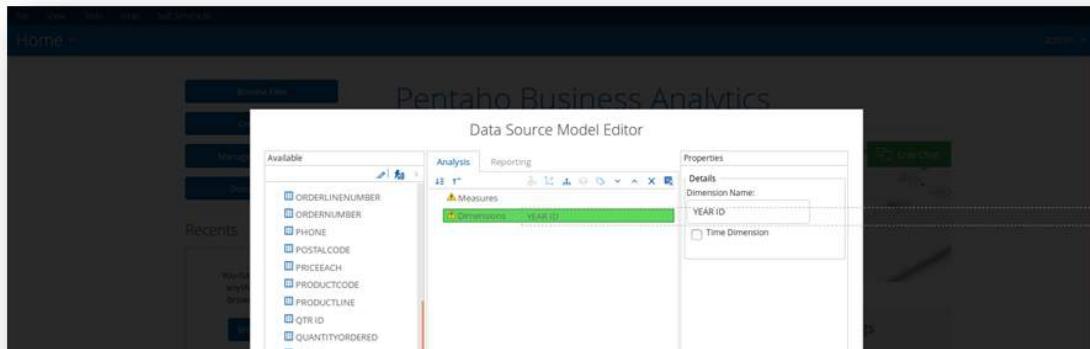


Como se ha visto en las unidades anteriores de este módulo, para crear el modelo lógico, es necesario identificar y definir las dimensiones, jerarquías, niveles, atributos, métricas y agregaciones.

Hay que ir definiendo sobre Pentaho los datos de ventas de entradas de los que se dispone.

Como siempre, la primera dimensión que ha de identificarse es el *Tiempo* (o temporal). Esta dimensión siempre debe estar presente en todo cubo multidimensional y, por norma general, en todo almacén de datos o *data warehouse*. Es importante recordar que el *data warehouse* tiene una de característica que destaca sobre el resto de bases de datos: su cualidad de almacenar y mantener la información histórica. Este hecho hace que una de las principales capacidades del *data warehouse* sea la posibilidad del análisis de los datos a lo largo del tiempo, y es por ello que el eje temporal, o dimensión temporal, es el más relevante en cualquier modelo multidimensional.

Para crear la dimensión, hay que arrastrar el campo *YEAR\_ID* sobre *Dimensiones* y soltarlo:





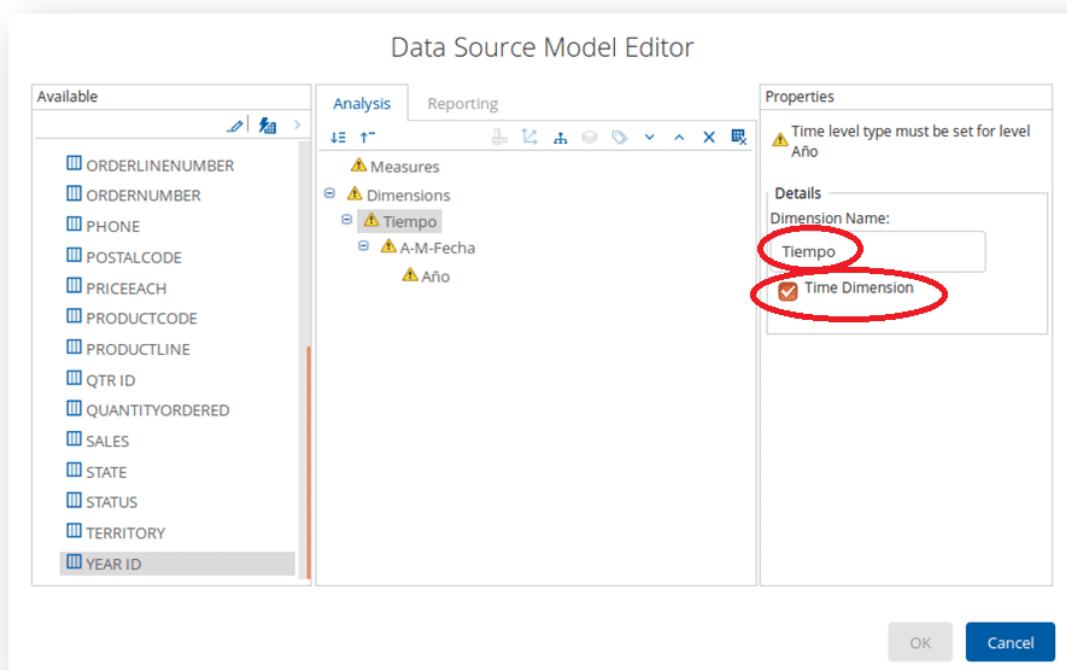
Esta acción creará por defecto una dimensión, una jerarquía y un primer nivel, todos ellos con el nombre del campo, por defecto.

Data Source Model Editor

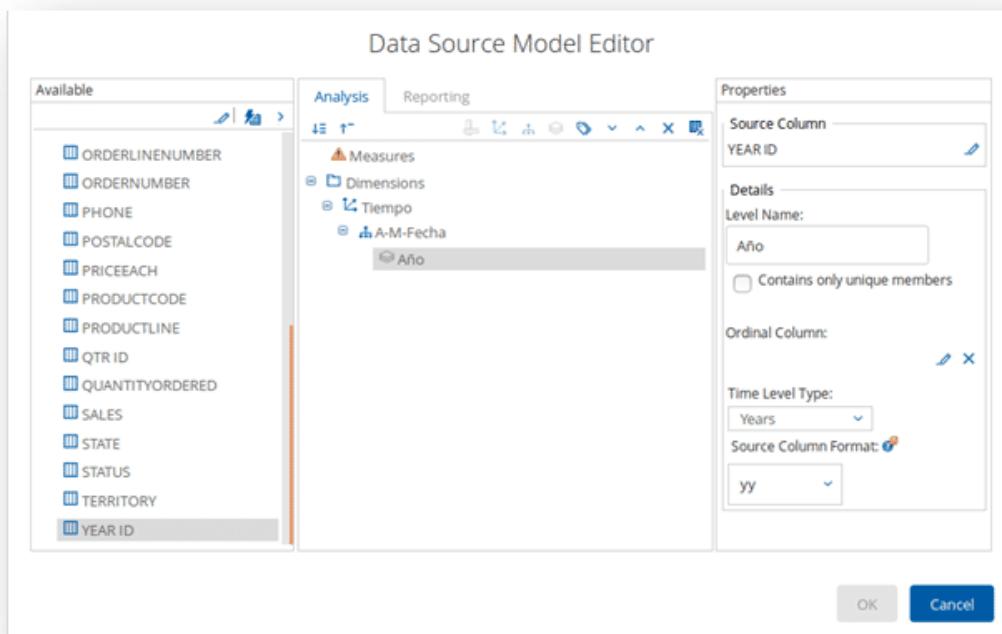
The screenshot shows the Data Source Model Editor window. On the left, the 'Available' list contains various fields: ORDERLINENUMBER, ORDERNUMBER, PHONE, POSTALCODE, PRICEEACH, PRODUCTCODE, PRODUCTLINE, QTR ID, QUANTITYORDERED, SALES, STATE, STATUS, TERRITORY, and YEAR ID. The 'YEAR ID' item is selected and highlighted with a red box. In the center, the 'Analysis' tab is active, showing the 'Measures' and 'Dimensions' sections. Under 'Dimensions', there is a tree view with 'YEAR ID' expanded, showing its first level. On the right, the 'Properties' panel has a 'Details' section with 'Dimension Name:' set to 'YEAR ID'. There is also a checked checkbox for 'Time Dimension'. At the bottom right are the 'OK' and 'Cancel' buttons.

A continuación, hay que particularizar la definición de la dimensión:

- Renombrar la dimensión como *Tiempo* e identificarla como dimensión temporal.
- Renombrar la jerarquía con un nombre indicativo. En este caso se desea crear una jerarquía de tres niveles: año, mes y fecha, por lo que se renombra como *A-M-Fecha*.
- Renombrar el nivel como *Año*.



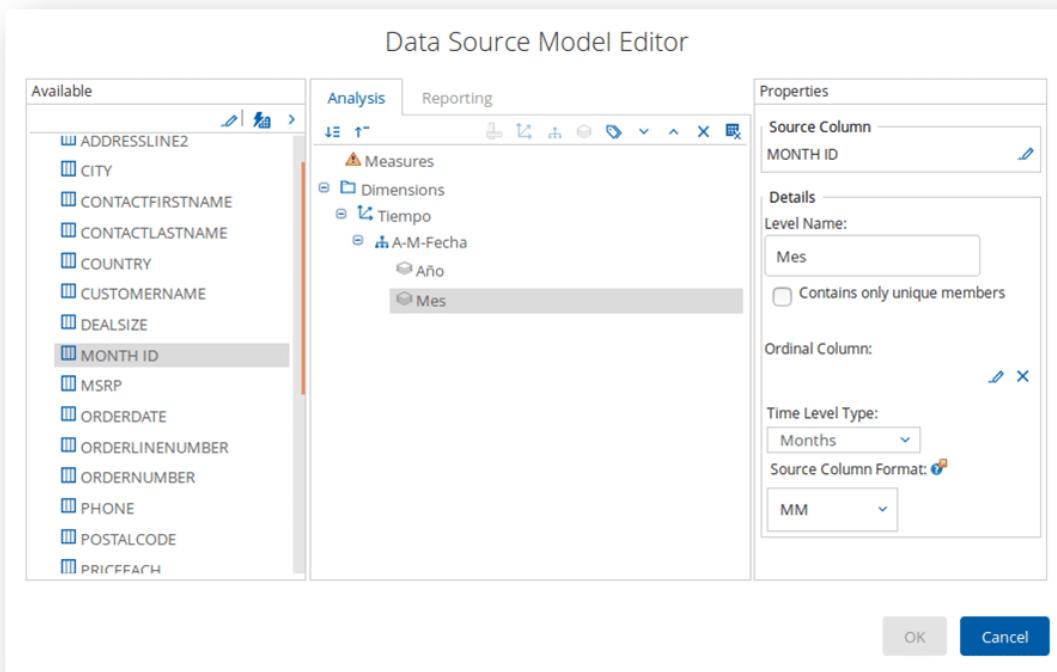
En el nivel *Año*, es importante también identificar el tipo de nivel temporal a Pentaho, así como el formato de columna.



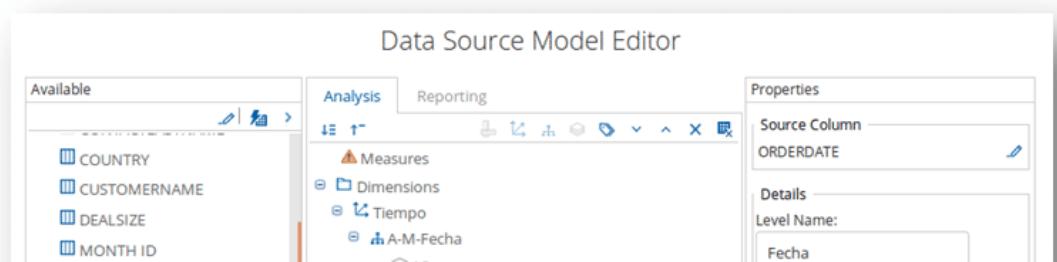
A continuación, para agregar un nuevo nivel a la jerarquía generada *A-M-Fecha*, se arrastra el campo *MONTH\_ID* y se deja caer sobre el nombre de la jerarquía. Como se puede apreciar en la siguiente

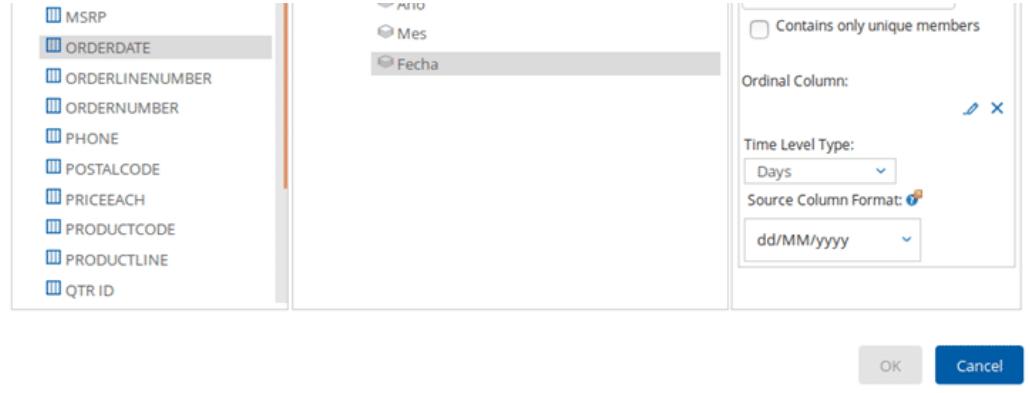
captura de pantalla, se creará un nuevo nivel, a la misma altura de Año, pero con un nivel inferior. Es importante guardar el orden descendente de los niveles. No sería correcto definir primero los meses y por debajo los años.

Se configuran el resto de opciones como se ha hecho con los años:



Para incluir el último nivel de esta jerarquía, se repite el proceso. Se arrastra el campo a la jerarquía, se renombra y se configura.





El modelo ya dispone de una dimensión completamente funcional: la dimensión *Tiempo*. Sin embargo, el cubo o modelo no es completo, pues, como mínimo, es necesario definir una métrica.

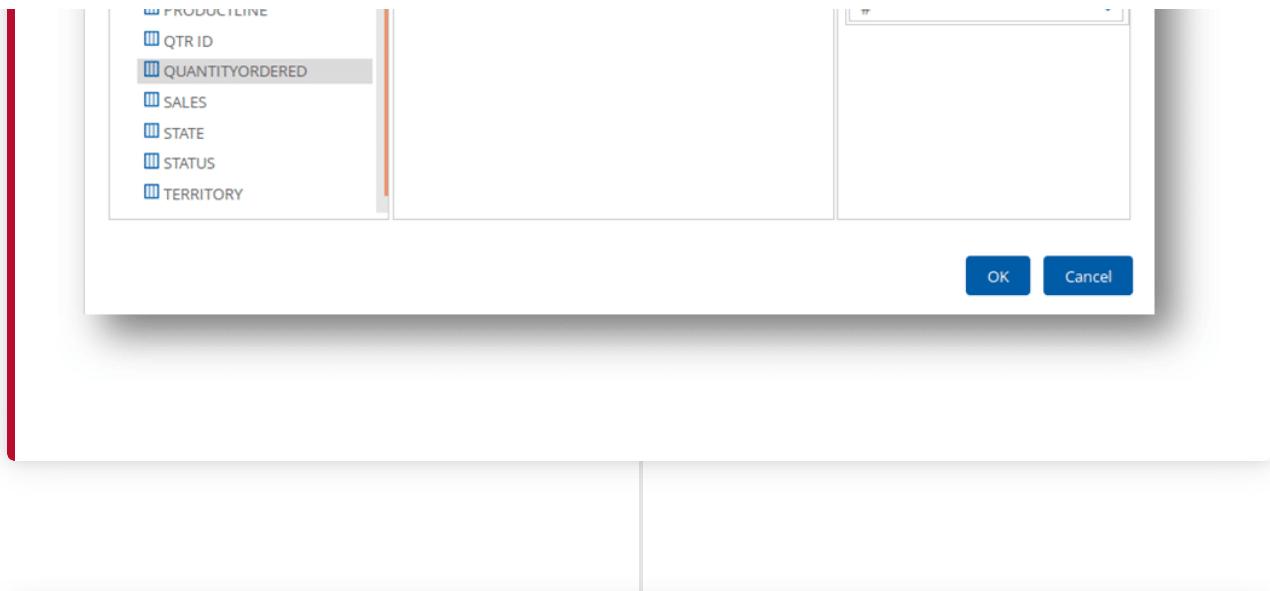
Para definir las métricas directas hay que recordar que estas están compuestas por un campo/valor y una función de agregación que indica al motor OLAP de Pentaho cómo debe agregar la información cuando se trabaja a niveles superiores de información.

Para crear la métrica, en este caso basta con identificar el campo que contiene los datos de interés, arrastrarlos y soltarlos sobre el grupo de métricas. Por ejemplo, el campo *QUANTITYORDERED*.

En este caso, se renombra como *Cantidad Pedida* y se le asigna una función de agregación *SUM*, de suma, que será la función que se aplique para agregar a niveles superiores.

**Data Source Model Editor**

Available	Analysis	Properties
ORDERDATE ORDERLINENUMBER ORDERNUMBER PHONE POSTALCODE PRICEEACH PRODUCTCODE PRODUCTLINE	<b>Analysis</b> Measures: Cantidad Pedida Dimensions: Tiempo A-M-Fecha: Año, Mes, Fecha	Source Column: QUANTITYORDERED Details: Display Name: Cantidad Pedida Default Aggregation: SUM Format: #

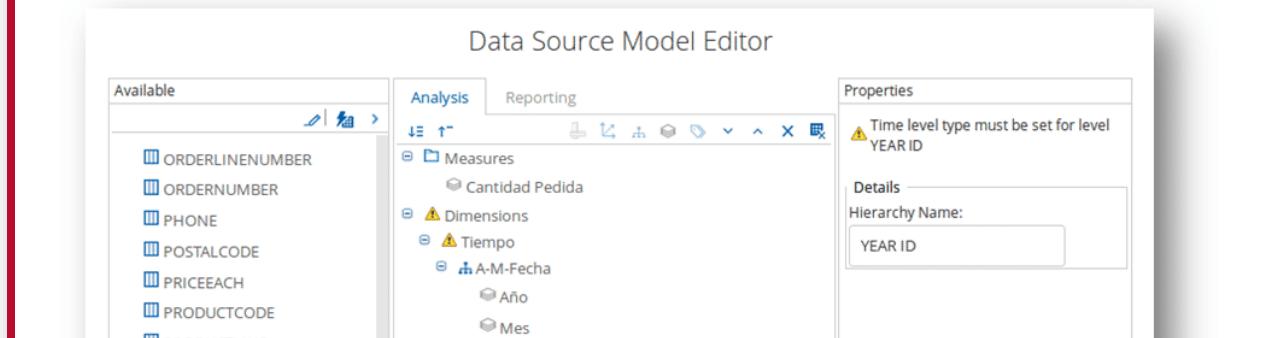


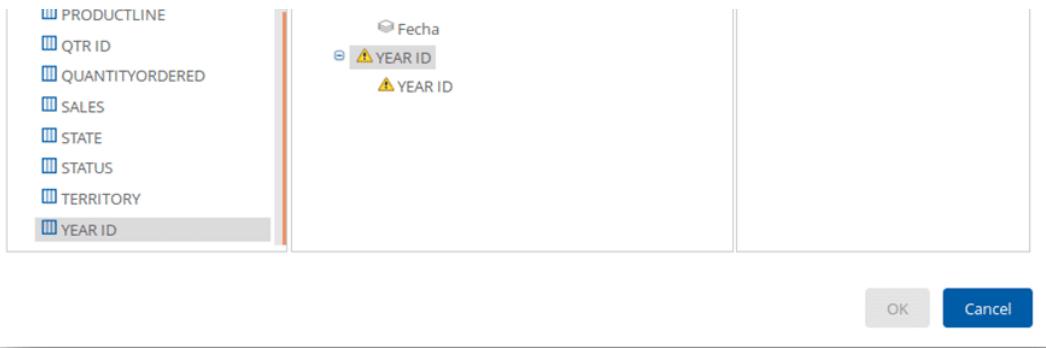
El modelo ya es funcional, pero se trata de un modelo mínimo (1 dimensión + 1 métrica): su capacidad de análisis sería muy limitada, prácticamente, lo único que podría mostrar sería el comportamiento de los pedidos a lo largo del tiempo.

Así pues, a continuación, se mostrará cómo enriquecer el modelo para ganar en capacidad de análisis.

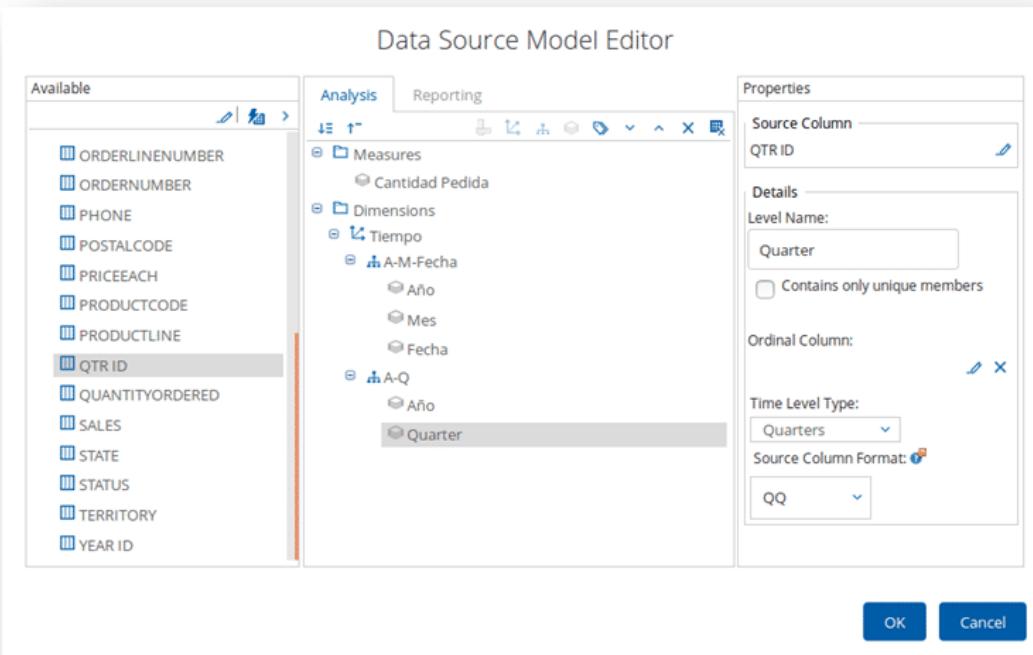
Hay que generar una nueva jerarquía en la dimensión *Tiempo*. Esto permitirá que los usuarios puedan navegar por la información de diferente manera y, por tanto, también podrán analizarla de distintos modos.

Téngase presente que una dimensión puede tener una o más jerarquías. En este caso, lo que se desea es crear una jerarquía que solo contenga dos niveles: años y trimestres. Para crear esta segunda jerarquía en la dimensión *Tiempo*, hay que arrastrar el campo *YEAR\_ID*, como ya se hizo, pero, esta vez, dejándolo caer sobre el nombre de la dimensión. La siguiente captura de pantalla muestra el resultado:



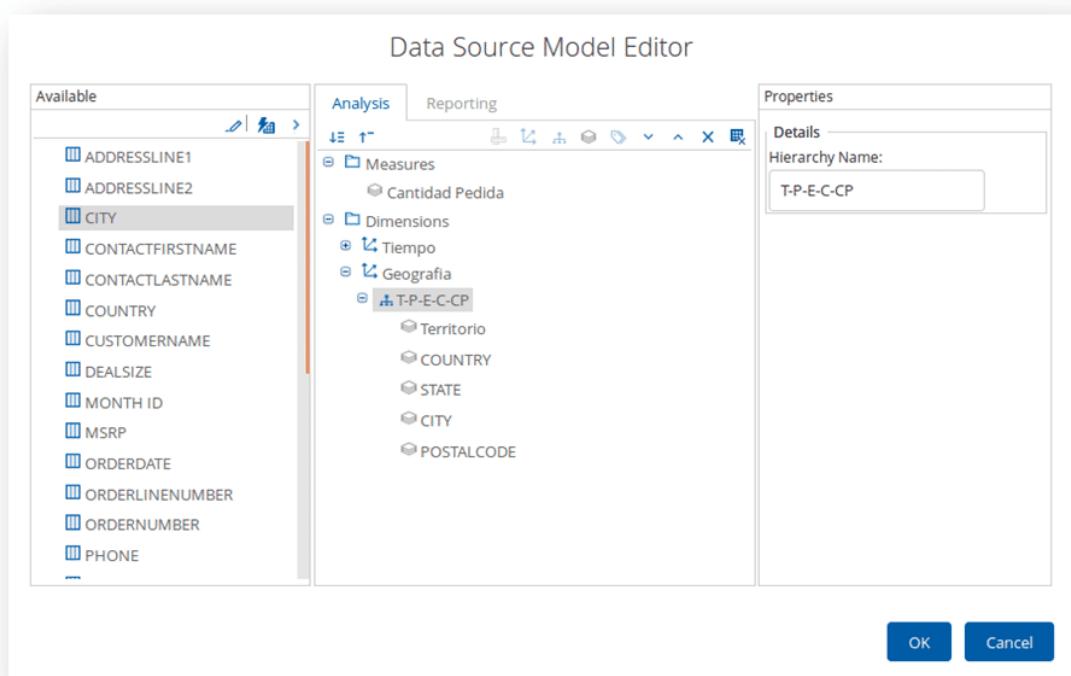


A continuación, hay que renombrar la jerarquía, esta vez como *A-Q*, y configurar los niveles: tanto el nivel año ya agregado, como el nivel *QTR ID*, que también se arrastra:

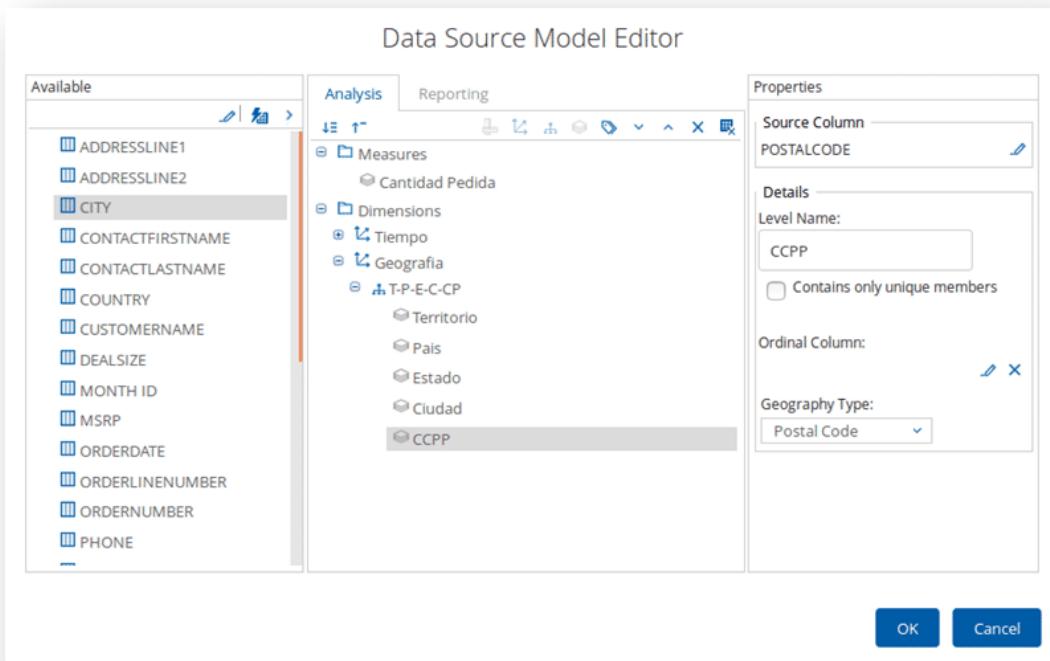


Hay que repetir este proceso con el resto de las dimensiones que quieran crearse. La siguiente dimensión que se propone es la geográfica.

Para ello, se arrastran los campos, se renombran y configuran, tal y como se muestra en la siguiente imagen:

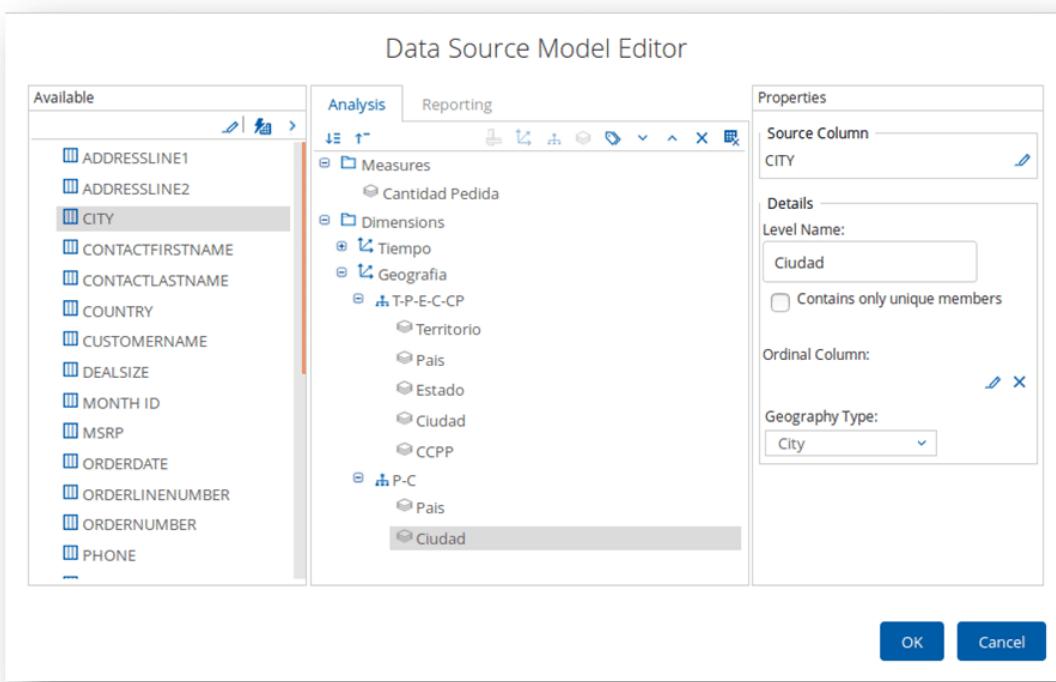


Para ello, hay que crear una jerarquía de cinco niveles: desde *Territorio* hasta *Código postal*.

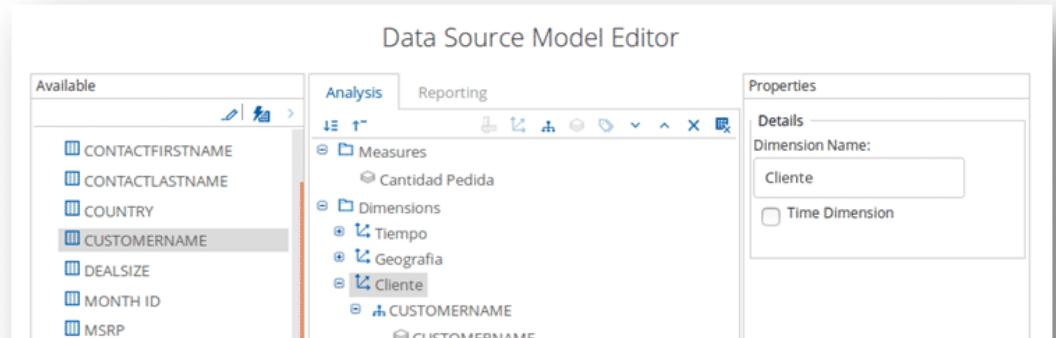


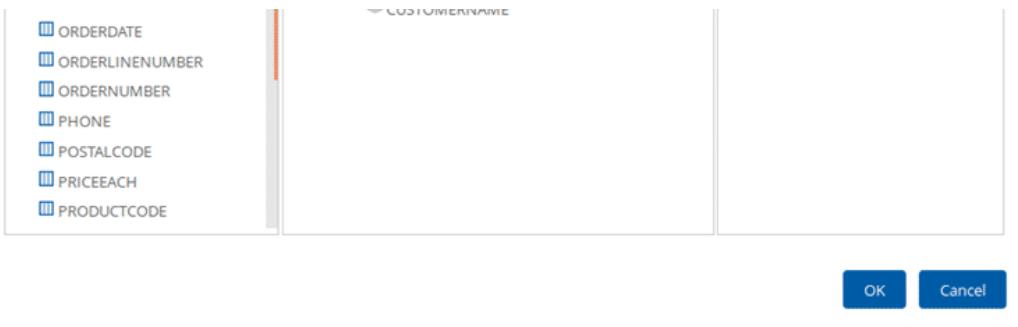
Una vez configurados y renombrados todos los niveles, se crea una segunda jerarquía, como, por ejemplo, *País* y *Ciudad*, con el objetivo de que el usuario pueda analizar más rápidamente sin tener que bajar tantos niveles.

El proceso de diseño es idéntico al mencionado en párrafos anteriores. Únicamente es necesario particularizar nombre y campos.

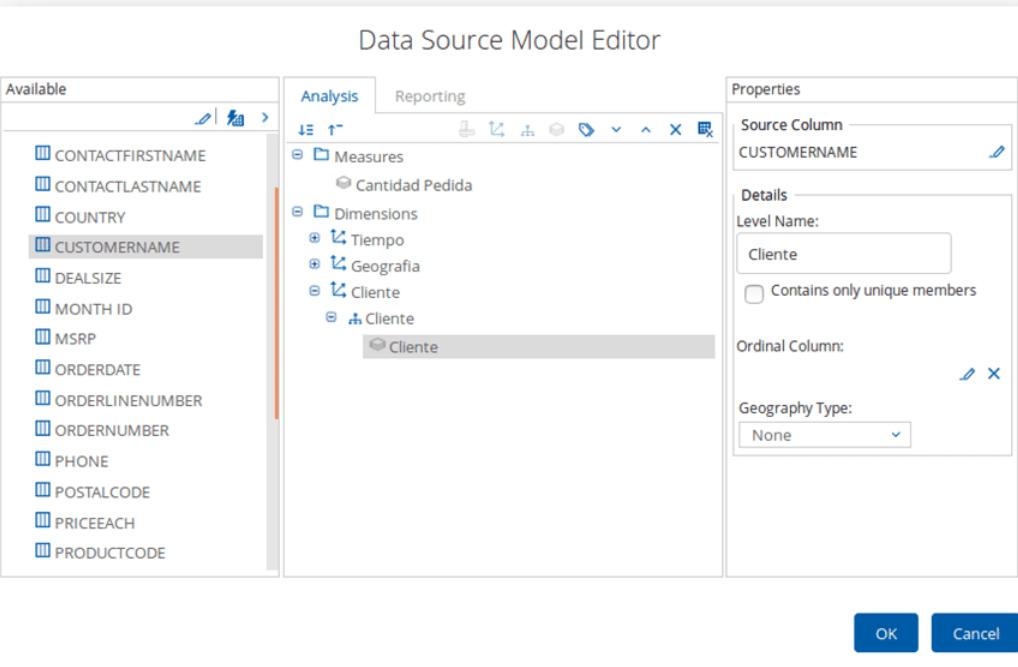


La siguiente dimensión que debe crearse es *Cliente*, a partir del campo *CUSTOMERNAME*:

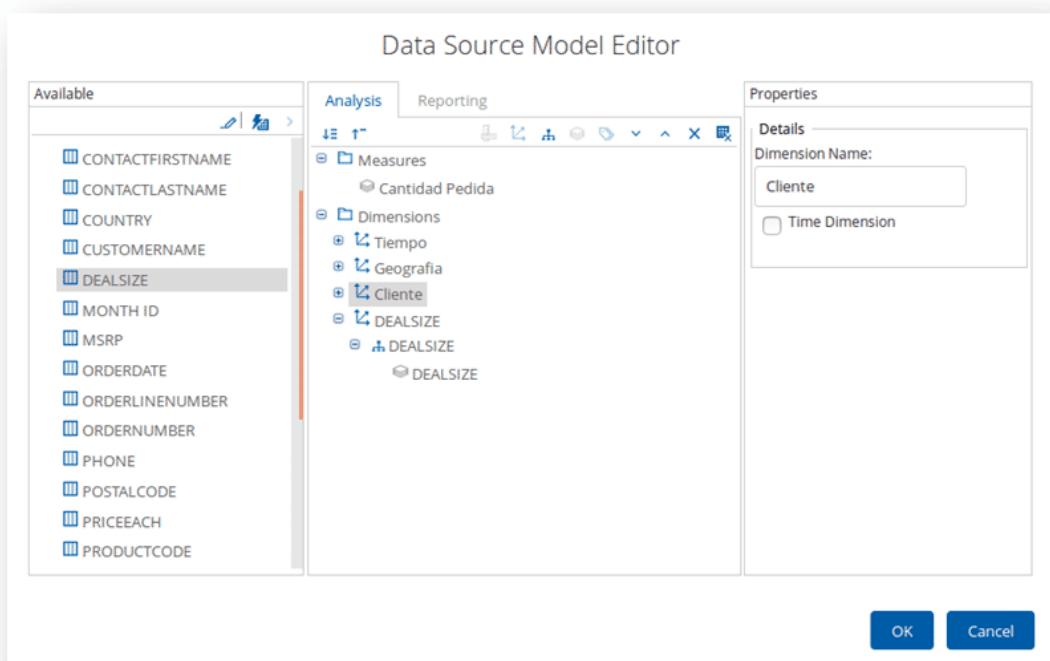




A continuación, se repite el proceso para editar y modificar el nombre de dimensión, jerarquía y nivel:



Se repite también el proceso para la dimensión *Tamaño Acuerdo*, a partir de la columna *DEALSIZE*:



Se repite también el proceso para editar y modificar el nombre de dimensión, jerarquía y nivel:

Data Source Model Editor

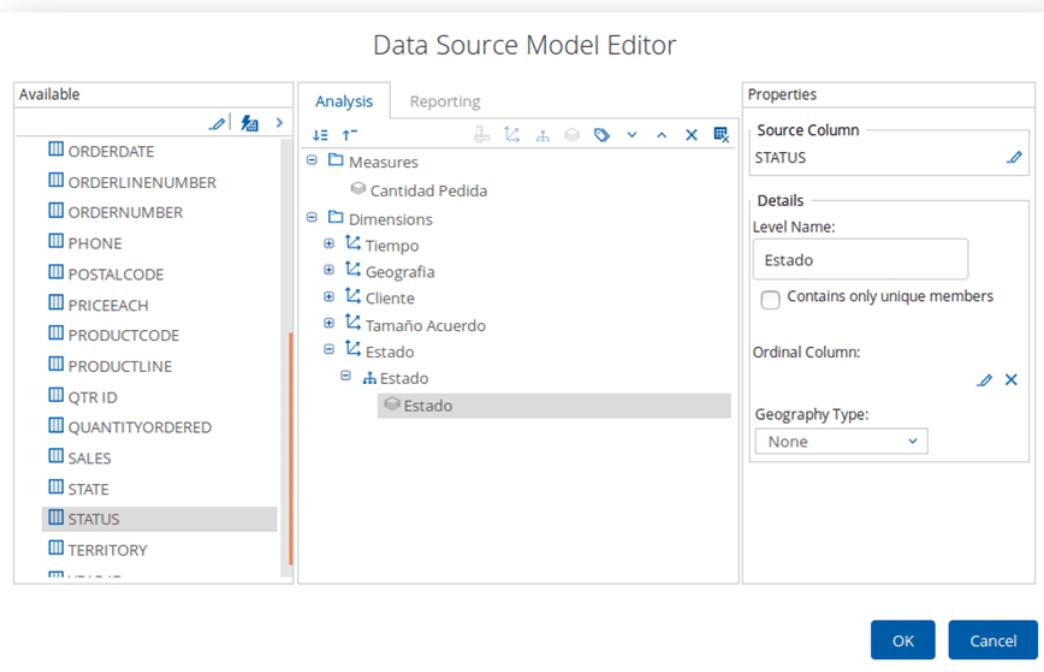
The screenshot shows the Data Source Model Editor interface. On the left, the 'Available' pane lists various columns: CONTACTFIRSTNAME, CONTACTLASTNAME, COUNTRY, CUSTOMERNAME, DEALSIZE, MONTH ID, MSRP, ORDERDATE, ORDERLINENUMBER, ORDERNUMBER, PHONE, POSTALCODE, PRICEEACH, and PRODUCTCODE. In the center, the 'Analysis' tab is selected, showing the 'Measures' and 'Dimensions' hierarchies. Under 'Dimensions', 'Tamaño Acuerdo' is expanded, and 'Tamaño Acuerdo' is selected. On the right, the 'Properties' pane is open, showing the 'Source Column' as 'DEALSIZE', 'Level Name' as 'Tamaño Acuerdo', and 'Geography Type' as 'None'. There is also a checkbox for 'Contains only unique members' which is unchecked. At the bottom right are 'OK' and 'Cancel' buttons.

Se repite el proceso para la dimensión *Estado*, a partir de la columna *STATUS*:

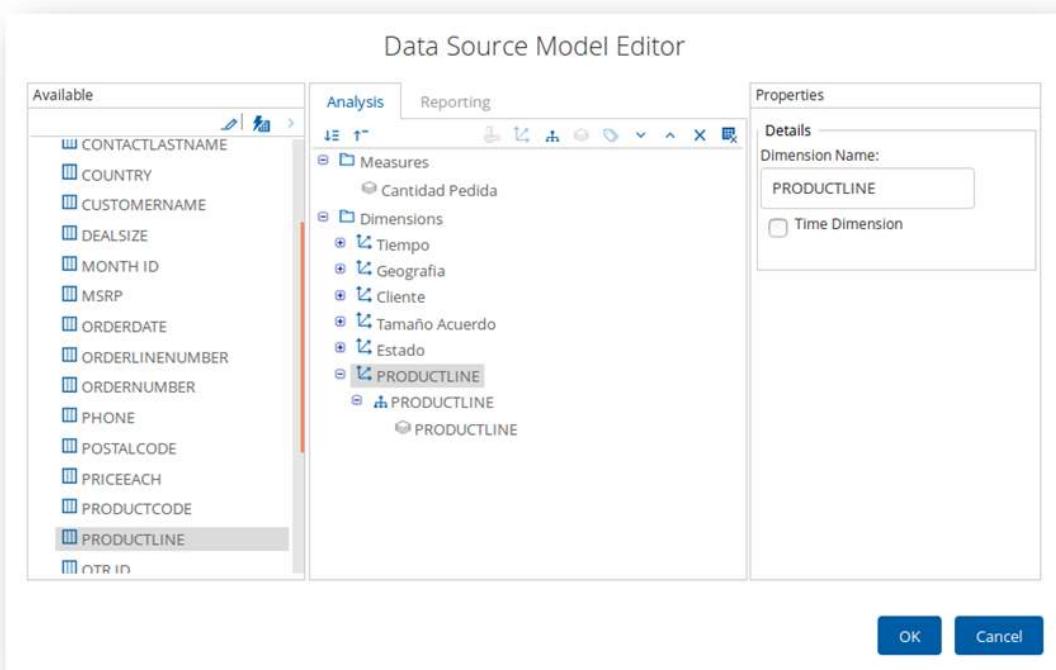
Data Source Model Editor

The screenshot shows the Data Source Model Editor interface. On the left, the 'Available' pane lists various columns: ORDERDATE, ORDERLINENUMBER, ORDERNUMBER, PHONE, POSTALCODE, PRICEEACH, PRODUCTCODE, PRODUCTLINE, QTR ID, QUANTITYORDERED, SALES, STATE, STATUS, and TERRITORY. In the center, the 'Analysis' tab is selected, showing the 'Measures' and 'Dimensions' hierarchies. Under 'Dimensions', 'Tamaño Acuerdo' is expanded, and 'STATUS' is selected. On the right, the 'Properties' pane is open, showing the 'Dimension Name' as 'Tamaño Acuerdo' and 'Time Dimension' checked. At the bottom right are 'OK' and 'Cancel' buttons.

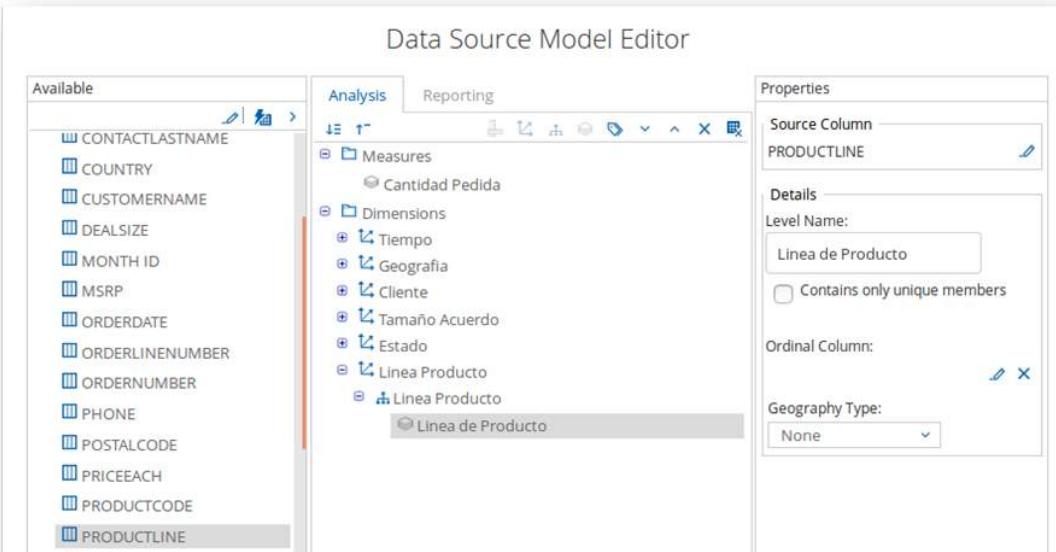
Se repite el proceso para editar y modificar el nombre de dimensión, jerarquía y nivel:

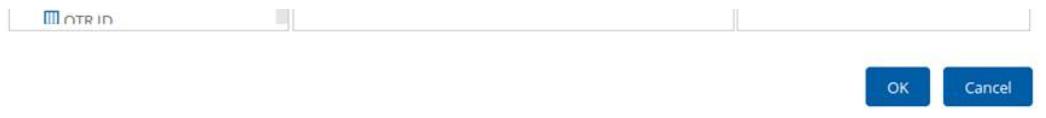


Y, por último, se añade una última dimensión, *Línea de Producto*, a partir de la columna *PRODUCTLINE*:



Se repite el proceso para editar y modificar el nombre de dimensión, jerarquía y nivel:





El modelo ya cuenta con seis dimensiones para poder analizar la información. Se añadirá una métrica más, que en este caso sería la métrica por defecto o básica: las ventas.

Para ello, y como ya se vio con la métrica anterior, se crea la métrica directa *Ventas*, a partir del campo *SALES*, y se define como función de agregación la suma, *SUM*:

Data Source Model Editor

Available

- ORDERLINENUMBER
- ORDERNRNUMBER
- PHONE
- POSTALCODE
- PRICEEACH
- PRODUCTCODE
- PRODUCTLINE
- QTR ID
- QUANTITYORDERED
- SALES**
- STATE
- STATUS
- TERRITORY
- YEAR ID

Analysis Reporting

Measures

- Cantidad Pedida
- SALES**

Dimensions

- Tiempo
- Geografía
- Cliente
- Tamaño Acuerdo
- Estado
- Línea Producto

Properties

Source Column: SALES

Details

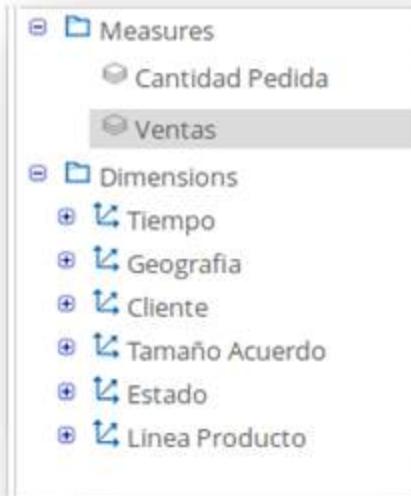
Display Name: SALES

Default Aggregation: SUM

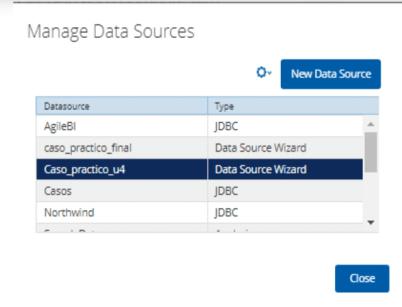
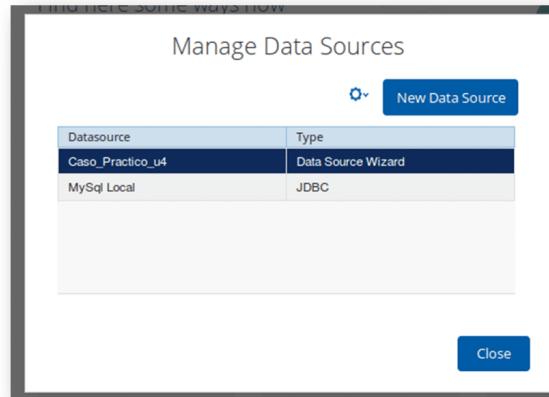
Format: #

OK Cancel

Por lo que el modelo quedaría del siguiente modo:



Una vez finalizado el modelo, hay que clicar sobre *OK* para que Pentaho guarde el modelo lógico o cubo en el repositorio y poder trabajar con él:



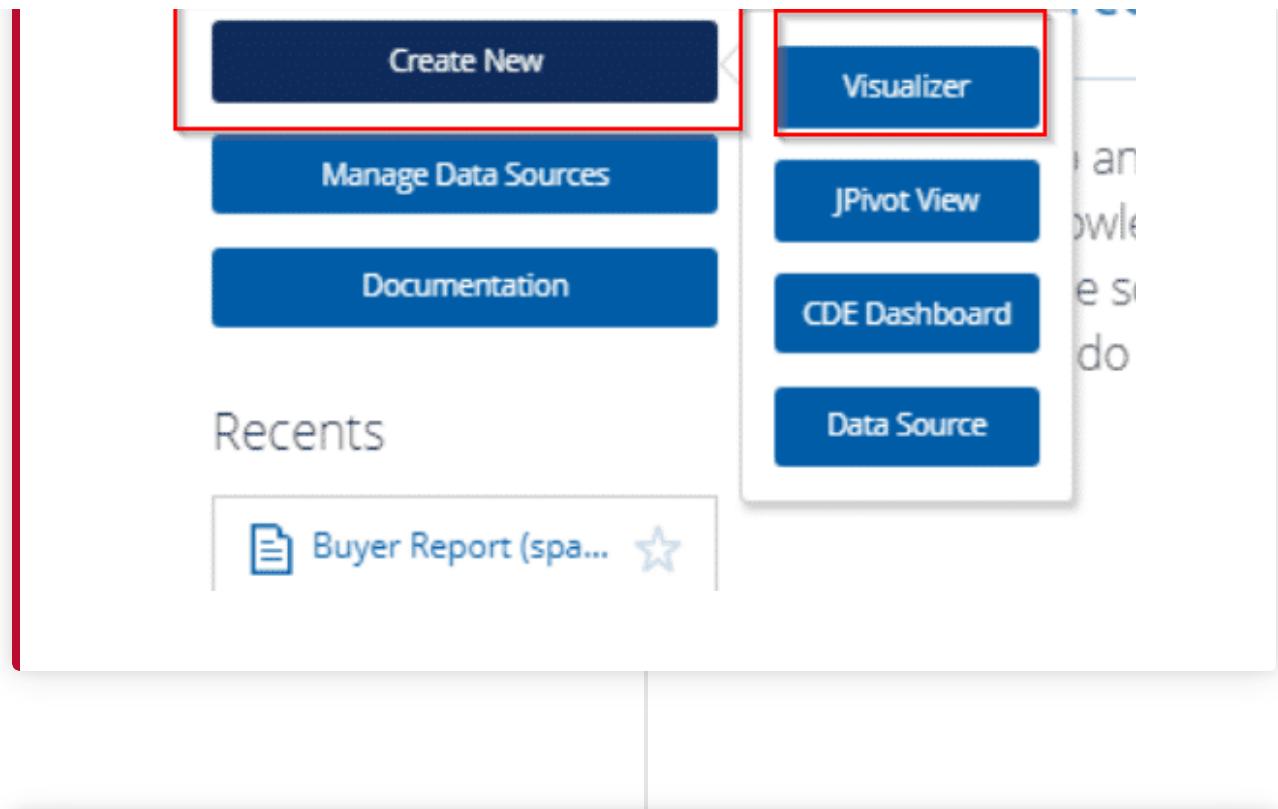
En este momento, el departamento de ventas ya tendría disponible un cubo multidimensional OLAP para poder hacer informes dinámicos OLAP y analizar la información, tal y como solicitaba el enunciado.

Para ver un ejemplo de cómo los usuarios utilizarían el modelo, a continuación, se creará un **informe dinámico**. Para ello, hay que clicar sobre la opción de crear nuevo contenido, *Create New*, y seleccionar un visor OLAP de los que hay disponibles en el entorno (Pivot4j view, Saiku Analytics, jPivot View).

Por ejemplo, se usará Visualizer:

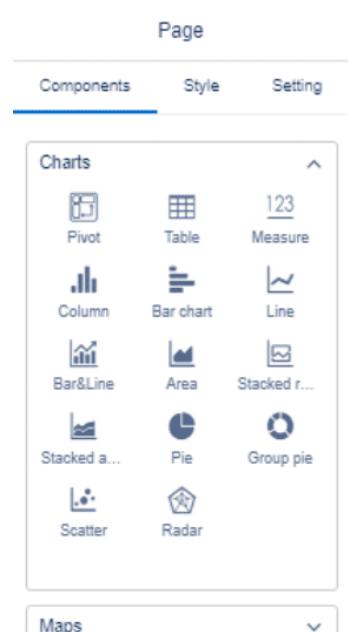
Browse Files

Penta



Se abrirá la opción de visualizar, desde donde se podrán configurar una serie de visualizaciones desde distintas fuentes.

Así, se arrastran los objetos que se quieren configurar para la visualización, en este caso, una tabla cruzada y un gráfico de barras que luego se filtrarán para permitir el análisis.



En el caso de la tabla cruzada, hay que seleccionar la fuente de datos, filas y columnas que la componen, así como la métrica que se va a mostrar.

En el caso de la tabla cruzada, hay que seleccionar la fuente de datos, filas y columnas que la componen, así como la métrica que se va a mostrar.

YEAR ID	PRODUCTLINE-YEAR ID, TERRITORY Ventas								
	2003				2004				
TERRITORY	APAC	EMEA	Japan	NA	APAC	EMEA	Japan	NA	APAC
	PRODUCTLINE	Ventas	Ventas	Ventas	Ventas	Ventas	Ventas	Ventas	Ventas
Classic Cars	85399	651273	120998	587428	127960	1011184	42071	581043	31410
Motorcycles	42353	137795	12839	178109	33124	204042	31958	291421	14492
Planes	22347	138987	2007	90018	41425	209128	49177	202942	11082
Ships		172428	14158	58238	1089	188992	10453	142904	3070
Trains	1681	20938	13279	28304		90973		25581	
Trucks and Buses	11298	228699	44498	135938	80834	182748	13349	252572	29917
Vintage Cars	90067	261595	17599	281727	61076	504052	21471	324815	57709

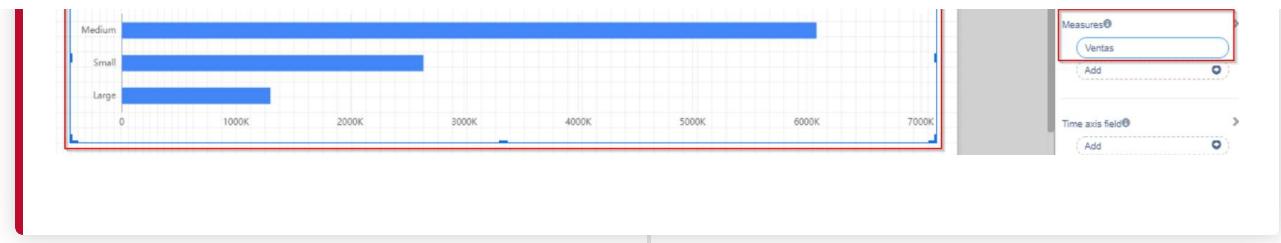
Analysis mode: test  
Rows: PRODUCTLINE  
Columns: YEAR ID, TERRITORY  
Measures: Ventas

El gráfico de barras se configura de forma similar:

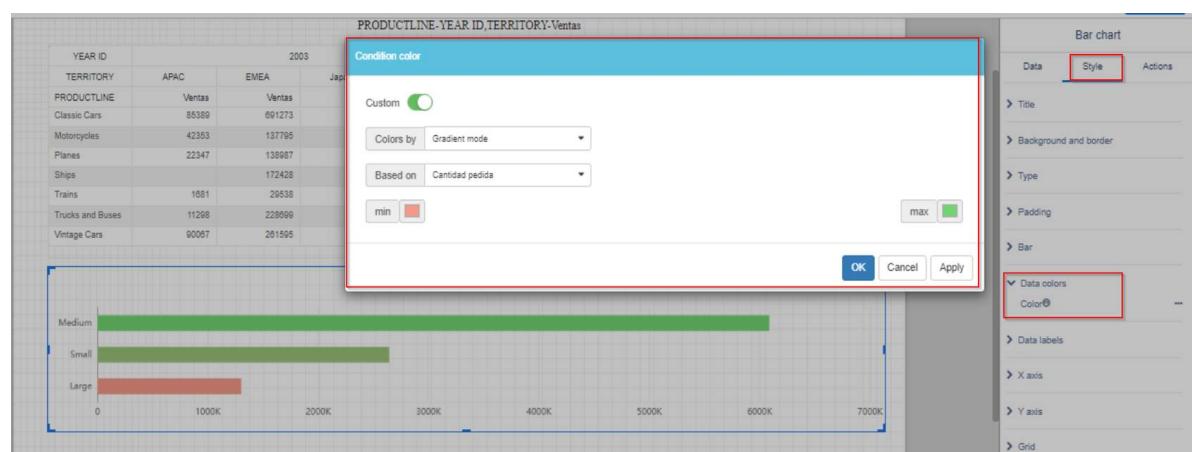
El gráfico de barras se configura de forma similar:

The screenshot shows a data visualization interface. On the right, a sidebar titled "Bar chart" contains tabs for "Data" (selected), "Style", and "Actions". Under "Analysis model", there is a dropdown set to "test". A red box highlights the "DEALSIZE" button in the "Actions" section. On the left, a grid table displays sales data for different products across four regions (APAC, EMEA, Japan, NA) for two years (2003, 2004). The table has "YEAR ID" and "TERRITORY" as columns. The data shows significant growth in all categories over the period.

PRODUCLINE-TEAK IN TERRITORI-VENTAS								
YEAR ID	2003				2004			
	TERRITORY	APAC	EMEA	Japan	NA	APAC	EMEA	Japan
PRODUCTLINE	Ventas	Ventas	Ventas	Ventas	Ventas	Ventas	Ventas	Ventas
Classic Cars	85389	691273	120698	587428	127980	1011184	42071	581043
Motorcycles	42353	137796	12639	178109	33124	204042	31969	281421
Planes	22347	136987	20697	90018	41425	209128	49177	202942
Ships		172428	14158	58238	1089	186992	10453	142904
Trains	1681	29538	13279	28304		90673		25551
Trucks and Buses	11298	228699	44498	135698	80834	182748	13349	252672
Vintage Cars	90067	261958	17599	281727	81078	504082	21471	324815



Se puede retocar el formato en la pestaña de *Style*.



Por último, se añaden los filtros.



The screenshot shows a data visualization tool's interface. On the left, there is a grid editor displaying a table with three columns: Japan, NA, and APAC. The data rows include: Japan Ventas 42071; NA Ventas 581043; APAC Ventas 31410; Japan 31959; NA 291421; APAC 14482; Japan 49177; NA 202942; APAC 11082; Japan 10453; NA 142904; APAC 3070; Japan 25551; NA 13349; APAC 29917; Japan 21471; NA 324815; APAC 57709. On the right, there is a sidebar with various styling options:

- Title
- Background and border
- List items
  - Font: sans-serif
  - Size: 12
  - Color: Black
  - Bold: B
  - Italic: I
- Grid line color: Black (unchecked)
- Others
  - Type: Dropdown
  - Show 'Search': On (checked)

Como resultado final, el departamento dispone de una solución de inteligencia de negocio para dar respuesta a las necesidades de prueba de los usuarios demandantes. Se ha creado en esta práctica un cubo en memoria temporal, para que se realicen los análisis y pruebas necesarios.

## VIII. Glosario

---



**El glosario contiene términos destacados para la comprensión de la unidad**

**ADAS** —

Sistemas avanzados de asistencia de controladores.

**CAU** —

Centro de atención al usuario.

**Dron** —

Vehículo aéreo no tripulado.

**IA** —

Inteligencia artificial, o inteligencia computacional; es la inteligencia exhibida por máquina.

**CVC** —

Catéteres de centro venoso.

**Healthcare** —

Conjunto de servicios que se proporcionan al individuo, con el fin de promover, proteger y restaurar su salud.

**Machine learning** —

El aprendizaje automático o ***machine learning*** es un subcampo de las ciencias de la computación y una rama de la inteligencia artificial cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender.

## PUC —

Consola de usuario de Pentaho.

## R —

Entorno y lenguaje de programación enfocado al análisis estadístico. Es una implementación de ***software*** libre del lenguaje S, pero con soporte de alcance estático.

## Sistema inteligente —

Programa de computación que cuenta con características y comportamientos similares a los de la inteligencia humana o animal, es decir, cuenta con la capacidad de decidir por sí mismo qué acciones realizar para alcanzar sus objetivos, basándose en sus percepciones, conocimientos y experiencias acumuladas.

## Telemetría —

Tecnología que permite la medición remota de magnitudes físicas y el posterior envío de la información hacia el operador del sistema.

**UCI**

—

Unidad de cuidados intensivos.

## IX. Anexos

---

Además, como complemento a la unidad, se ofrece a los estudiantes la posibilidad de acceder de forma gratuita a los materiales y laboratorios prácticos de la herramienta Qlik Sense. Para registrarse, se deberán seguir los pasos que se indican en el siguiente anexo:

