

Sistema de inteligencia de negocio para el análisis de los tratamientos de reducción del colesterol

Fernando Mantilla Gómez

Máster en Ingeniería Informática

Área de Business Intelligence

Nombre Consultor/a: David Amorós Alcaraz

Nombre Profesor/a responsable de la asignatura: Ferran Prados Carrasco

Fecha Entrega: 01/2019



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Sistema de inteligencia de negocio para el análisis de los tratamientos de reducción del colesterol</i>
Nombre del autor:	<i>Fernando Mantilla Gómez</i>
Nombre del consultor/a:	<i>David Amorós Alcaraz</i>
Nombre del PRA:	<i>Ferran Prados Carrasco</i>
Fecha de entrega (mm/aaaa):	01/2019
Titulación::	<i>Master en Ingeniería Informática</i>
Área del Trabajo Final:	<i>El nombre de la asignatura de TF</i>
Idioma del trabajo:	<i>Español (castellano)</i>
Palabras clave	<i>Bussiness Intelligence, OLAP, ETL</i>
<p>Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.</i></p>	
<p>Crear un entorno Business Intelligence (BI) que posibilite el análisis de la información generada durante un experimento que tiene como finalidad comprobar la eficacia de los diferentes tratamientos que tienen como objetivo la reducción de los niveles de colesterol.</p> <p>La información se analizará en forma de cubos de datos lo que permitirá un análisis mucho más cuidadoso y la extracción de conclusiones mucho más elaboradas.</p> <p>Para el desarrollo del proyecto se utilizará una metodología para el desarrollo de proyectos de Business Intelligence que facilite en análisis, diseño e implementación de la solución.</p> <p>Se utiliza Pentaho como solución de Business Intelligence la creación de los procesos ETL, la generación de cubos y la creación del cuadro de mandos.</p> <p>El resultado final es la implementación de un sistema Business Intelligence que facilita la adquisición, el almacenamiento y la explotación de datos asociados a pacientes a los que se los ha diagnosticado los niveles de colesterol, permitiendo a los usuarios comprobar la eficacia de los tratamientos y realizar filtros por zona geográfica y hábitos alimenticios entre otros.</p>	
<p>Abstract (in English, 250 words or less):</p>	
<p>Create a Business Intelligence (BI) environment that enables the analysis of the information generated during an experiment that aims to verify the effectiveness</p>	

of different treatments that aim to reduce cholesterol levels.

The information will be analyzed in the form of cubes of data which will allow a much more careful analysis and the extraction of much more elaborate conclusions.

For the development of the project, a methodology will be used for the development of Business Intelligence projects that facilitate analysis, design and implementation of the solution.

Pentaho is used as a Business Intelligence solution for the creation of ETL processes, the generation of cubes and the creation of the scorecard.

The final result is the implementation of a Business Intelligence system that facilitates the acquisition, storage and exploitation of data associated with patients who have been diagnosed with cholesterol levels, allowing users to check the effectiveness of treatments and perform filters by geographical area and eating habits among others.

Índice

1. Introducción.....	1
1.1 Contexto y justificación del Trabajo	1
1.2 Objetivos del Trabajo	1
1.3 Estado del arte	2
1.4 Enfoque y método seguido	2
1.5 Planificación del Trabajo	3
1.6 Breve resumen de productos obtenidos	5
1.7 Breve descripción de los otros capítulos de la memoria	5
2. Resto de capítulos.....	6
2.1 Etapa de justificación y planteamiento	6
2.2 Etapa de análisis	9
2.2.1 Definición de los requerimientos	9
2.2.2 Selección del entorno tecnológico	10
2.2.3 Análisis de datos	12
2.2.4 Análisis repositorio de Meta Data	13
2.3 Etapa de diseño	13
2.3.1 Diseño de la base de datos. DataWareHouse	13
2.3.2 Diseño ETL	14
2.4 Etapa de construcción	15
2.4.1 Instalación Pentaho	15
2.4.2 Desarrollo ETL	15
2.4.3 Desarrollo de la aplicación. Gestión de información. Cubo OLAP ..	16
2.4.5 Visualización de la información	17
2.4.5.1 Informes	17
2.4.5.2 Análisis OLAP	23
2.4.5.1 Cuadro de mandos	27
3. Conclusiones.....	28
4. Glosario	32
5. Bibliografía	33
6. Anexos	34
6.1 Manual de instalación	34

Lista de figuras

Figura 1. Planificación del Trabajo. Diagrama Gant	8
Figura 2. Diseño Base de Datos.....	14
Figura 3. Diagrama ETL.....	15
Figura 4. Cubo OLAP	16
Figura 5. Informe evolución tratamiento por trimestre	18
Figura 6. Informe evolución tratamiento por hábito y trimestre. I.....	20
Figura 7. Informe evolución tratamiento por hábito y trimestre. II.....	21
Figura 8. Informe evolución tratamiento por comunidad	23
Figura 9. Saiku tratamientos por hábito y trimestre	24
Figura 10. Saiku grafico tratamientos por hábito y trimestre	24
Figura 11. Saiku tratamiento por zona geográfica.....	25
Figura 12. Saiku gráfico tratamiento por zona geográfica	25
Figura 13. Saiku tratamiento por hábito y comunidad	26
Figura 14. Saiku gráfico tratamiento por hábito y comunidad	26
Figura 15. Tratamiento por paciente	26
Figura 16. Cuadro de mandos.....	27
Figura 17. Evolución tratamiento por paciente	28
Figura 18. Evolución tratamiento por trimestre.....	29
Figura 19. Relación Tratamiento-Hábito.....	29
Figura 20. Gráfico tratamiento Natural por CCAA	30
Figura 21. Gráfico tratamiento Homeopathic por CCAA.....	30

1. Introducción

1.1 Contexto y justificación del Trabajo

Suponemos que este trabajo fin de máster es para una empresa farmacéutica que está probando diferentes tratamientos para reducir los niveles de colesterol.

El colesterol es un lípido (grasa) que está relacionado con enfermedades graves, como las cardiopatías, la angina de pecho y los accidentes cerebrovasculares.

La empresa farmacéutica tiene diversos tratamientos que producen unos resultados en los experimentos realizados. Estos resultados están relacionados con diversos parámetros de los experimentos.

Actualmente para determinar si un tratamiento es mejor que otro deben compararlos manualmente, con la dificultad que eso entraña puesto que un tratamiento puede ser mejor dependiendo en un tipo de persona por sus características biológicas y peor en otra persona con otras características. A parte de esta dificultad también se añade el posible error humano y el tiempo que se tarda en realizar esa comprobación.

Es importante determinar cuál es el mejor tratamiento puesto que si se elige un tratamiento que no tiene los resultados esperados las pérdidas pueden ser millonarias.

El problema que se pretende resolver es ayudar a determinar cuáles de los diferentes tratamientos son más efectivos para reducir el nivel de colesterol. Se pretende que se haga de manera automática mostrando informes que ayuden a tomar la decisión.

1.2 Objetivos del Trabajo

El objetivo de este trabajo es el diseño e implementación de un sistema de Business Intelligence que facilite la adquisición, el almacenamiento y la explotación de datos asociados a pacientes a los que se les ha diagnosticado los niveles de colesterol (LDL). Este sistema BI debe ayudar a resolver preguntas como las siguientes:

- ¿Cuál es la relación entre los diferentes tratamientos y la evolución de los pacientes?
- ¿Existen terapias más eficaces?
- ¿Ha influido en el resultado, los hábitos de los pacientes?
- ¿La evolución a lo largo del tiempo, por un mismo tratamiento, dependen de algún factor como los hábitos?

- ¿Hay diferencias en el resultado de un tratamiento según el lugar geográfico del paciente?
- ¿Hay algún periodo del año donde el tratamiento sea más o menos efectivo?

1.3 Estado del arte

Una vez planteados los objetivos y requerimientos del proyecto, se ha realizado una búsqueda de soluciones y aplicaciones del mercado que puedan servir para cubrir las necesidades planteadas.

En el mercado se han encontrado plataformas ya implementadas como **SAIMA SOLUTIONS**, que ofrece un servicio de Business Intelligence para hospitales que permite mejorar su eficiencia y rentabilidad. Está basada en **IBM Cognos Analíticos**, que es un producto BI con licenciamiento de pago de IBM. El problema de esta solución es que está centrada en hospitales por lo que tiene muchas funcionalidades que no son necesarias como la gestión de citas, pacientes o diagnósticos que encarecen el producto y no aportan nada a los objetivos.

Otra aplicación es **Mural Med**, que al igual que la anterior tiene una plataforma de Business Intelligence que proporciona información sobre multitud de servicios como receta electrónica, expedientes clínicos y tratamientos.

Estas soluciones tienen el hándicap de que no son específicas para los objetivos que perseguimos por lo que su coste se incrementa sustancialmente al tener más funcionalidades de las deseadas.

Por este motivo, para afrontar este proyecto se decide implementar una solución Business Intelligence concreta utilizando un software específico para este tipo de proyectos.

1.4 Enfoque y método seguido

Para conseguir la resolución del problema lo que se plantea es la instalación, configuración y parametrización de una solución de Business Intelligence basada en Open Source.

Lo que se pretende es que esta herramienta importe toda la información posible y en base a unos algoritmos muestre informes que ayuden a tomar la decisión sobre qué tratamiento es mejor. Es decir, que ayude a resolver las cuestiones anteriormente planteadas.

En concreto lo que se pretende es:

- Diseñar un almacén de datos (Data Warehouse) que permita almacenar la información adquirida de los diferentes orígenes de datos. Teniendo

en cuenta que tendremos un conjunto de pacientes que han sido sometidos, por grupos, a diferentes tratamientos.

- Implementar este almacén de datos y programar los procesos ETL (siglas en inglés de extracción, transformación y carga) que permitan alimentar el Data Warehouse a partir de los ficheros base facilidades.
- Analizar las diferentes plataformas BI Open Source disponibles al mercado que nos permitirían explotar la información almacenada.
- Seleccionar e implantar una de estas herramientas Open Source de tal forma que se disponga de una capa de software por el análisis de la información.
- Para el desarrollo del proyecto se seguirá una metodología específica para proyectos de Business Intelligence que contempla las mismas etapas para los proyectos comunes de ingeniería (Justificación, Planteamiento, Análisis, Diseño y Construcción)

Utilizar una herramienta Open Source de BI es la mejor solución ya que están muy probadas y adaptadas a resolver este tipo de problemas. Realizar una aplicación a medida no tiene sentido por el coste y el tiempo de implementación.

1.5 Planificación del Trabajo

Los recursos necesarios para realizar el trabajo serían:

ROL	TAREAS
Jefe de proyecto	<ul style="list-style-type: none"> • Gestión del proyecto • Seguimiento del proyecto • Selección de la herramienta
Consultor	<ul style="list-style-type: none"> • Seguimiento del trabajo • Resolución de dudas • Revisión del trabajo
Analista y consultor BI	<ul style="list-style-type: none"> • Análisis de la solución • Diseño e implementación del DWH • Instalación de la herramienta BI • Definición de cubos • Procesos ETL

Las tareas a realizar en el proyecto serían

Nombre de tarea	Duración	Comienzo	Fin
TFM	78 días	jue 20/09/18	lun 07/01/19
Gestión y control del proyecto	78 días	jue 20/09/18	lun 07/01/19
Actividades gestión y control	75 días	lun 24/09/18	vie 04/01/19
PEC1	8 días	jue 20/09/18	lun 01/10/18
PEC2	25 días	mar 02/10/18	lun 05/11/18
PEC3	20 días	mar 06/11/18	lun 03/12/18
PEC4	25 días	mar 04/12/18	lun 07/01/19
Elaboración memoria	70 días	vie 21/09/18	jue 27/12/18
Justificación y planteamiento	14 días	jue 20/09/18	mar 09/10/18
Justificación del proyecto	3 días	jue 20/09/18	lun 24/09/18
Metas y objetivos del proyecto	4 días	mar 25/09/18	vie 28/09/18
Alcance	3 días	lun 01/10/18	mié 03/10/18
Riesgos	4 días	jue 04/10/18	mar 09/10/18
Análisis	14 días	mar 09/10/18	vie 26/10/18
Definir requerimientos	4 días	mar 09/10/18	vie 12/10/18
Selección entorno tecnológico	4 días	lun 15/10/18	jue 18/10/18
Análisis de datos	4 días	vie 19/10/18	mié 24/10/18
Análisis repositorio MetaData	3 días	vie 19/10/18	mar 23/10/18
Diseño	11 días	vie 26/10/18	vie 09/11/18
Diseño BBDD	2 días	vie 26/10/18	lun 29/10/18
Diseño ETL	9 días	mar 30/10/18	vie 09/11/18
Construcción	26 días	vie 09/11/18	vie 14/12/18
Instalación herramienta	3 días	vie 09/11/18	mar 13/11/18

Desarrollo ETL	5 días	mié 14/11/18	mar 20/11/18
Cubo OLAP	5 días	mié 21/11/18	mar 27/11/18
Cuadro de mandos	11 días	mié 28/11/18	mié 12/12/18
Realización pruebas	3 días	mié 12/12/18	vie 14/12/18
Cierre proyecto	10 días	vie 14/12/18	jue 27/12/18
Elaboración presentación	5 días	vie 14/12/18	jue 20/12/18
Elaboración video	5 días	vie 21/12/18	jue 27/12/18

1.6 Breve resumen de productos obtenidos

El producto obtenido será la instalación, configuración y parametrización de una solución Business Intelligence Open Source que permita obtener datos de distintas fuentes y sacar informes que ayuden a la toma de decisión para determinar la eficacia de un tratamiento contra el colesterol.

1.7 Breve descripción de los otros capítulos de la memoria

Los capítulos se dividen según las etapas de justificación, análisis, diseño y construcción.

En el capítulo de justificación se definen los objetivos, metas, alcance y riesgos del proyecto.

En el capítulo de análisis se definen los requerimientos del proyecto, se estudian las diferentes soluciones tecnológicas y se hace el análisis de los datos.

En el capítulo de diseño se diseñan los procesos ETL y el esquema de Base de Datos.

En el capítulo de construcción se implementa la solución con los procesos ETL, el cubo y el cuadro de mandos.

2. Resto de capítulos

2.1 Etapa de justificación y planteamiento

2.1.1 Justificación del proyecto

Una empresa farmacéutica tiene diversos tratamientos para reducir el nivel de colesterol en seres humanos. Estos tratamientos están en fase de pruebas y producen unos resultados a partir de los experimentos realizados.

Los efectos del tratamiento dependen de varias variables o parámetros como el estilo de vida de los pacientes, el nivel de colesterol o la presión sanguínea.

Actualmente para determinar si un tratamiento es mejor que otro deben compararlos manualmente, con la dificultad que eso entraña puesto que un tratamiento puede ser mejor dependiendo en un tipo de persona por sus características biológicas y peor en otra persona con otras características diferentes. A parte de esta dificultad también se añade el posible error humano y el tiempo que se tarda en realizar esa comprobación.

Es importante determinar cuál es el mejor tratamiento puesto que si se elige un tratamiento que no tiene los resultados esperados las pérdidas pueden ser millonarias.

La empresa farmacéutica ha decidido desarrollar un proyecto Business Intelligence que les ayude a determinar cuáles de los diferentes tratamientos son más efectivos para reducir el nivel de colesterol, desechando aquellos que no funcionan. Se pretende que se haga de manera automática mostrando informes que ayuden a tomar la decisión. Este proyecto les ayudará a reducir las pérdidas al no tener que destinar presupuestos a tratamientos que no son efectivos.

2.1.2 Metas y objetivos del proyecto

La principal meta u objetivo del proyecto es crear un producto que ayude a determinar cuáles de los tratamientos de colesterol son mejores y la relación que pueden tener los tratamientos con la zona geográfica del cliente, sus hábitos o la duración del tratamiento.

Esto ayudará a personalizar el tratamiento en relación a la casuística de cada paciente.

2.1.3 Alcance

El producto obtenido será la instalación, configuración y parametrización de una solución Business Intelligence Open Source que permite obtener datos de

distintas fuentes y sacar informes que ayuden a la toma de decisión para determinar la eficacia de un tratamiento contra el colesterol.

El sistema debe ayudar a responder preguntas como las siguientes:

- ¿Cuál es la relación entre los diferentes tratamientos y la evolución de los pacientes?
- ¿Existen terapias más eficaces?
- ¿Ha influido en el resultado, los hábitos de los pacientes?
- ¿La evolución a lo largo del tiempo, por un mismo tratamiento, dependen de algún factor como los hábitos?
- ¿Hay diferencias en el resultado de un tratamiento según el lugar geográfico del paciente?
- ¿Hay algún periodo del año donde el tratamiento sea más o menos efectivo?

2.1.4 Riesgos del proyecto

Los posibles riesgos a los que nos enfrentamos en el desarrollo del proyecto son:

Id	Riesgo	Descripción	Respuesta	Probabilidad
1	Falta de experiencia	El equipo de desarrollo no tiene experiencia en proyectos BI	Contratar a un experto en BI	Baja
2	Falta de participación	El cliente o los usuarios no se involucran en el proyecto	Realizar reuniones de seguimiento	Baja
3	Cambios de requisitos	El cliente cambia constantemente los requisitos	Aprobar el análisis de requisitos	Baja
4	Integración de datos	Se desconocen todas las fuentes de datos o no se pueden integrar	Realizar un estudio previo para conocer todas las fuentes	Media
5	El tiempo y el alcance es irreal	El tiempo del proyecto es demasiado corto para el alcance planteado.	Acotar el alcance del proyecto	Media

2.1.5 Planificación del proyecto

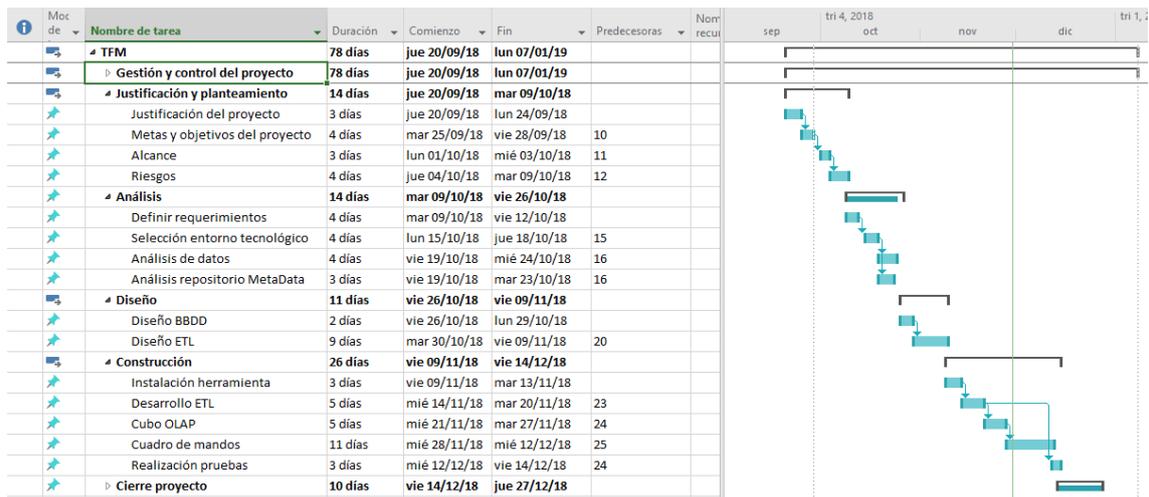


Figura 1. Planificación del Trabajo. Diagrama Gant

Nombre de tarea	Duración	Comienzo	Fin
TFM	78 días	jue 20/09/18	lun 07/01/19
Gestión y control del proyecto	78 días	jue 20/09/18	lun 07/01/19
Actividades gestión y control	75 días	lun 24/09/18	vie 04/01/19
PEC1	8 días	jue 20/09/18	lun 01/10/18
PEC2	25 días	mar 02/10/18	lun 05/11/18
PEC3	20 días	mar 06/11/18	lun 03/12/18
PEC4	25 días	mar 04/12/18	lun 07/01/19
Elaboración memoria	70 días	vie 21/09/18	jue 27/12/18
Justificación y planteamiento	14 días	jue 20/09/18	mar 09/10/18
Justificación del proyecto	3 días	jue 20/09/18	lun 24/09/18
Metas y objetivos del proyecto	4 días	mar 25/09/18	vie 28/09/18
Alcance	3 días	lun 01/10/18	mié 03/10/18
Riesgos	4 días	jue 04/10/18	mar 09/10/18
Análisis	14 días	mar 09/10/18	vie 26/10/18
Definir requerimientos	4 días	mar 09/10/18	vie 12/10/18
Selección entorno tecnológico	4 días	lun 15/10/18	jue 18/10/18

Análisis de datos	4 días	vie 19/10/18	mié 24/10/18
Análisis repositorio MetaData	3 días	vie 19/10/18	mar 23/10/18
Diseño	11 días	vie 26/10/18	vie 09/11/18
Diseño BBDD	2 días	vie 26/10/18	lun 29/10/18
Diseño ETL	9 días	mar 30/10/18	vie 09/11/18
Construcción	26 días	vie 09/11/18	vie 14/12/18
Instalación herramienta	3 días	vie 09/11/18	mar 13/11/18
Desarrollo ETL	5 días	mié 14/11/18	mar 20/11/18
Cubo OLAP	5 días	mié 21/11/18	mar 27/11/18
Cuadro de mandos	11 días	mié 28/11/18	mié 12/12/18
Realización pruebas	3 días	mié 12/12/18	vie 14/12/18
Cierre proyecto	10 días	vie 14/12/18	jue 27/12/18
Elaboración presentación	5 días	vie 14/12/18	jue 20/12/18
Elaboración video	5 días	vie 21/12/18	jue 27/12/18

2.2 Etapa de análisis

2.2.1 Definición de los requerimientos

El sistema Business Intelligence que vamos a implementar debe responder a las siguientes cuestiones:

- ¿Cuál es la relación entre los diferentes tratamientos y la evolución de los pacientes?
- ¿Existen terapias más eficaces?
- ¿Ha influido en el resultado, los hábitos de los pacientes?
- ¿La evolución a lo largo del tiempo, por un mismo tratamiento, dependen de algún factor como los hábitos?
- ¿Hay diferencias en el resultado de un tratamiento según el lugar geográfico del paciente?

- ¿Hay algún periodo del año donde el tratamiento sea más o menos efectivo?

A raíz de estas preguntas, se concluye que existen cinco dimensiones: el tratamiento, los pacientes, los hábitos, la zona geográfica y el tiempo.

De la misma manera se definen las siguientes medidas: el nivel de colesterol, y la presión sanguínea.

Existirá también una tabla de hechos que almacena la información de la evolución del paciente.

A la hora de definir el cubo OLAP optamos por un modelo de datos estrella ya que nos garantiza un modelo más simple con mayor velocidad en los análisis multidimensionales. A esto hay que añadir que para el usuario final este modelo es más sencillo de entender permitiéndole realizar las consultas que desee de una forma sencilla.

La solución Business Intelligence a implantar debe ser open source.

2.2.2 Selección del entorno tecnológico

Una vez definidos los requisitos del proyecto se procede al estudio de los diferentes entornos tecnológicos que nos permitan tener una solución Business Intelligence que dé respuesta a las preguntas planteadas.

Los requisitos que debe tener la herramienta Business Intelligence son:

- Open Source.
- Disponibilidad de entorno gráfico.
- Multiplataforma
- Debe ser un sistema intuitivo que permita a los usuarios crear sus propios informes. Es decir, debe incluir cuadros de mando.
- Debe poder importar datos de diferentes tipos de fuentes
- Debe ser una solución integrada que permita la gestión de procesos ETL, creación de informes interactivos y análisis multidimensionales de información (OLAP), minería de datos y cuadros de mando.
- La herramienta debe ser gratuita y debe disponer de una amplia comunidad de desarrolladores y de soporte tanto en español como en inglés.

Con los requisitos anteriormente definidos, vamos a estudiar las principales alternativas existentes en el mercado para quedarnos con la que consideremos más apropiada para nuestro proyecto.

BIRT

Es un proyecto Open Source cuya finalidad principal es la de crear informes de inteligencia de negocio. Permite procesamiento multidimensional analítico (OLAP).

Puede integrarse con otras aplicaciones de software empresarial y está basado en Java, por lo que es multiplataforma.

No tiene un módulo específico para los procesos ETL ni minería de datos.

Jasper Reports

Se trata de un servidor de informes con interfaz gráfica basada en Java que se pueden utilizar independientemente o integrar en cualquier otra aplicación o web.

Es multiplataforma y dispone de módulos para gestionar los procesos ETL y utilizar datos de distintas fuentes.

Permite crear esquemas OLAP y cuadros de mando.

KNIME

Es una plataforma de minería de datos con entorno gráfico programada en Java por lo que es multiplataforma.

La generación de informes la realiza integrándose con la plataforma BIRT.

Disponen de diferentes nodos o módulos para realizar minería de datos, procesos ETL y análisis de datos.

Es una herramienta de software libre y gratuita.

Pentaho

Es una plataforma de código abierto líder mundial en Business Intelligence. Dispone de diversos módulos que proporcionan herramientas para los procesos ETL, minería de datos, OLAP y cuadros de mando avanzado.

Dispone de entorno gráfico y tiene una distribución gratuita y otra de pago.

Tiene una amplia comunidad de desarrolladores tanto en español como en inglés.

Knowage

Es una suite de código abierto que proporciona una solución para los procesos de inteligencia de negocio.

Dispone de una versión Community que es gratuita y una edición Enterprise de pago.

Al igual que las anteriores tiene una interfaz gráfica y dispone de diferentes módulos para el análisis multidimensional (OLAP), procesos ETL, minería de datos e informes y cuadros de mando.

Solución seleccionada

Una vez revisadas las características de las anteriores soluciones vemos que tienen unas funcionalidades similares y que se adecúan a lo que buscamos y al alcance del proyecto. Menos BIRT, que parece que está más orientada a los informes, vemos que podríamos utilizar cualquiera de las otras soluciones.

Finalmente nos hemos decidido por Pentaho porque es una de las soluciones líder en el mercado y porque tiene una gran comunidad de desarrolladores detrás.

2.2.3 Análisis de datos

La fuente con los distintos datos sobre los que debemos trabajar se encuentran en cuatro hojas de un libro Excel:

- **Treatment:** se almacenan los distintos tratamientos.
 - **Terapias farmacológicas:** son aquellas que se tratan con medicamentos farmacéuticos que han superado todas fases clínicas establecidas.
 - **Terapias naturales:** son aquellas terapias que usan productos naturales (raíces, entonces, etc.).
 - **Terapias homeopáticas:** aquellas que usan productos homeopáticos
- **Patients:** se guardan los datos de los pacientes. Su comunidad, ciudad, edad, género y tratamiento que se le aplica.
- **Habits:** son los hábitos de cada paciente a lo largo del tratamiento. Hay un registro por semana y se indica para cada paciente la actividad realizada y la dieta seguida.
- **Indicators:** son los indicadores de cada paciente para cada semana. Estos indicadores son el nivel de colesterol y la tensión arterial.

2.2.4 Análisis repositorio de Meta Data

Para almacenar los datos capturados de las distintas fuentes con el fin de que el sistema BI pueda trabajar con ellos, se debe utilizar un sistema gestor de base de datos.

Este sistema gestor de base de datos también tiene que ser open source y con una amplia comunidad de desarrolladores detrás.

Los sistemas gestores de base de datos más utilizadas con estas características, y con mayor aceptación en el mercado son MySQL y PostgreSQL.

Las características de ambos sistemas son muy similares por lo que nos decantamos por **PostgreSQL**.

2.3 Etapa de diseño

2.3.1 Diseño de la base de datos. DataWareHouse

El primer paso es definir el esquema de nuestro DataWareHouse. Una vez analizada la información sobre la que debemos de trabajar, se llega a la conclusión de que es necesaria una tabla de hechos que almacena la información de la evolución del paciente, y cinco tablas de dimensiones: el tratamiento, los pacientes, los hábitos, la zona geográfica y el tiempo.

Se opta por un esquema en estrella puesto que tiene una implementación más sencilla y para el usuario es más fácil de entender.

Por lo tanto, el diseño de la base de datos sería:

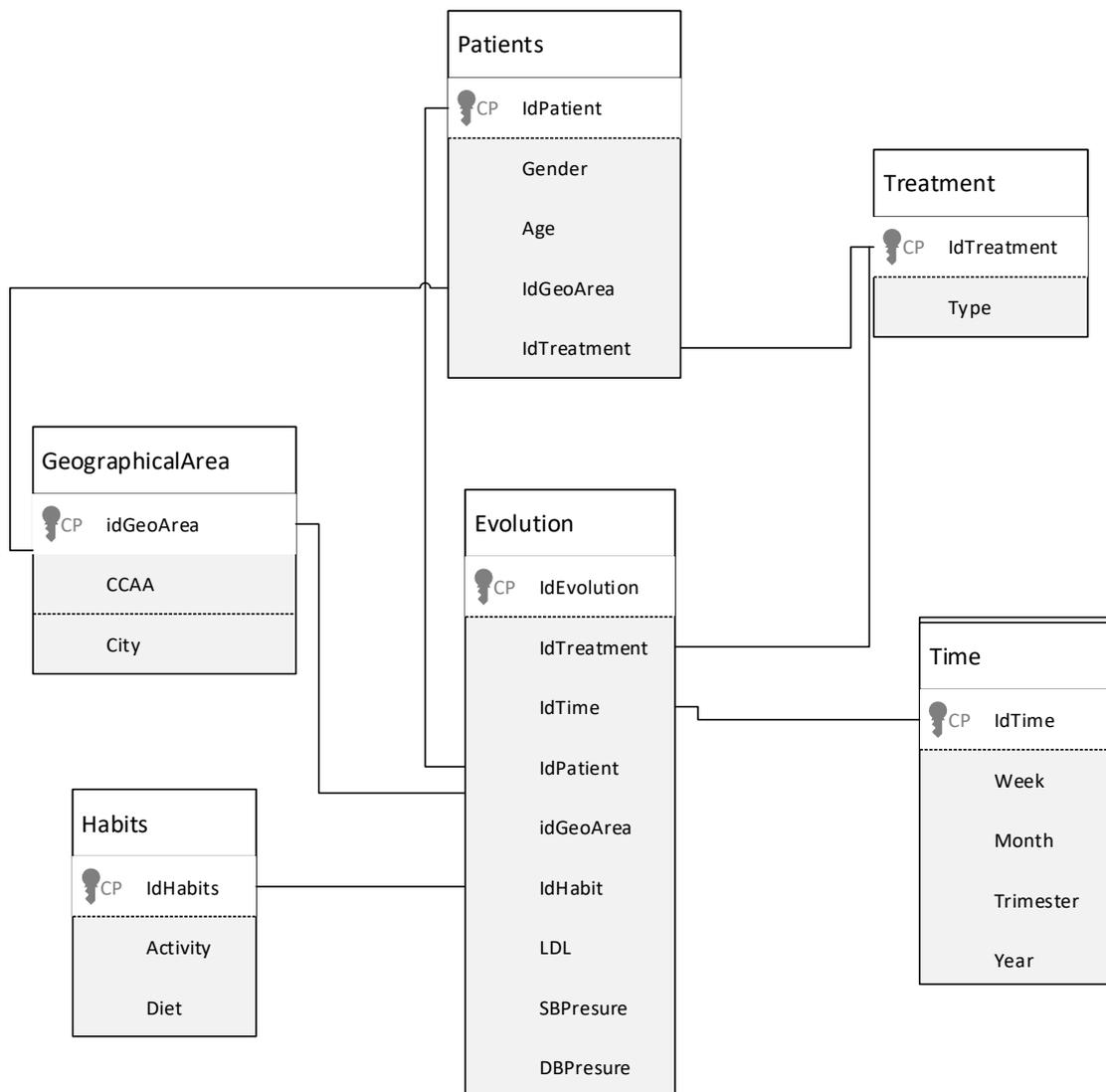


Figura 2. Diseño Base de Datos

2.3.2 Diseño ETL

Para capturar la información de la fuente de datos y almacenarla en nuestra base de datos (DataWareHouse) se crearán los siguientes procesos ETL:

- **Treatment:** su objetivo es obtener los datos de la hoja treatment del fichero Excel y almacenarlo en la tabla treatment
- **Geografical Area:** igual que los dos anteriores, toma los datos de la hoja patient del fichero Excel y almacenarlo en la tabla geograficalarea. Como puede estar repetido, se debe incluir una transformación para que incluya solo registros únicos
- **Patient:** almacena los datos de los pacientes en la tabla patient. Se debe incluir transformaciones de búsqueda para que obtenga las claves primarias de la tabla geografical área.

- **Habits:** su objetivo es obtener los datos de la hoja habits del fichero Excel y almacenarlo en la tabla habits. Debe incluir registros únicos.
- **Time:** con el siguiente es el proceso ETL más complicado. Guarda los registros del tiempo en formato fecha, semana, mes, año y trimestre. Por lo que se deben realizar transformaciones para obtener dichos datos.
- **Evolution:** inserta los registros en la tabla evolution, por lo que debe obtener datos de casi todas las dimensiones. Como los datos de los hábitos están en otra hoja Excel, se debe realizar una transformación para que relacione los registros semanales de los hábitos, con los indicadores.

2.4 Etapa de construcción

2.4.1 Instalación Pentaho

La versión de Pentaho que se va a utilizar es la Community Edition. Los diferentes módulos de Pentaho nos los descargamos de la página oficial, que es: <https://community.hitachivantara.com/docs/DOC-1009931-downloads>

2.4.2 Desarrollo ETL

Utilizamos el módulo Pentaho Data Integration para realizar los procesos ETL antes definidos. Este módulo nos permite conectarnos a nuestra base de datos PostgreSQL e insertar los registros obtenidos del fichero Excel.

La siguiente imagen muestra los procesos ETL en Pentaho.

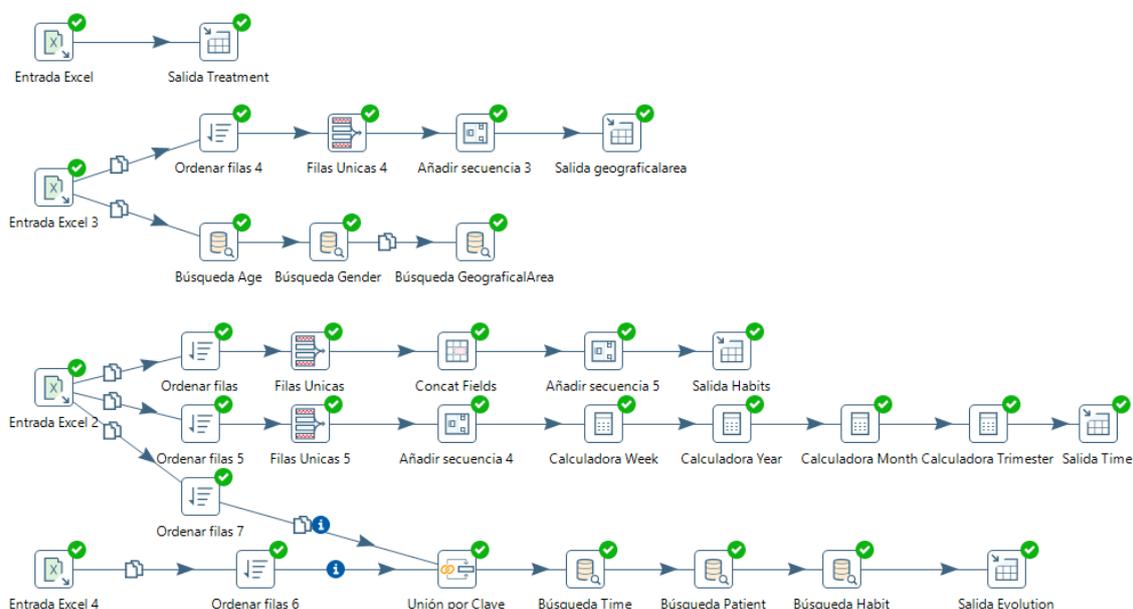


Figura 3. Diagrama ETL

2.4.3 Desarrollo de la aplicación. Gestión de la información. Cubo OLAP

Para definir los cubos OLAP utilizamos la herramienta Schema Workbench de Pentaho.

Con esta herramienta creamos las cinco dimensiones: treatment, time, geogaficalarea, habit y patient.

A continuación, creamos el cubo Evolucion, que contiene la tabla Evolución con todos los registros del tratamiento. Tiene las cinco dimensiones antes mencionadas y las medidas serán: la media (avg) del campo ldl, la media del campo sbpreasure y la media del campo ldpreasure.

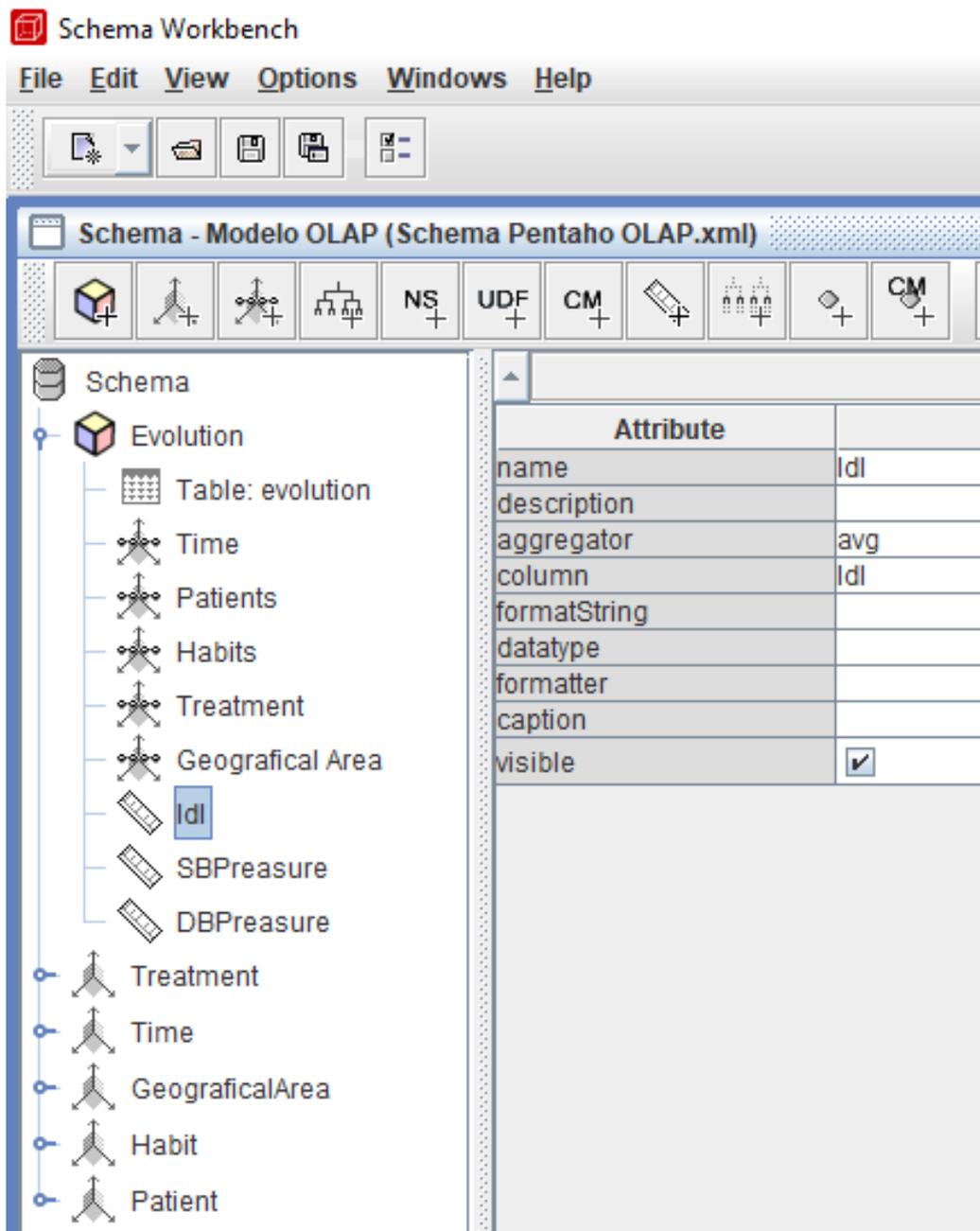


Figura 4. Cubo OLAP

2.4.5 Visualización de la información

2.4.5.1 Informes

Se realizan los siguientes informes con la herramienta Pentaho Report Designer

Informe Tratamiento por Trimestre

Este informe muestra la evolución de los distintos tratamientos por cada trimestre. Es muy útil para que los usuarios puedan ver de manera rápida cuál de los distintos tratamientos es mejor.

La consulta que se realiza es:

```
SELECT
    "datawarehouse"."treatment"."type",
    "datawarehouse"."time"."trimester",
    round(avg("datawarehouse"."evolution"."ldl"),2)
FROM
    "datawarehouse"."time"      INNER JOIN      "datawarehouse"."evolution"      ON
    "datawarehouse"."time"."idtime" = "datawarehouse"."evolution"."idtime"
    INNER JOIN "datawarehouse"."treatment" ON "datawarehouse"."evolution"."idtreatment" =
    "datawarehouse"."treatment"."idtreatment"
group by
    "datawarehouse"."treatment"."type",
    "datawarehouse"."time"."trimester"
order by "datawarehouse"."treatment"."type" asc, "datawarehouse"."time"."trimester" asc
```

A continuación, se puede ver el resultado de la consulta con su gráfico correspondiente

Evolución de los distintos tratamientos por trimestre

TratamientoHOMEOPATHIC

Trimestre	Media LDL
1	326
2	327
3	330
4	331

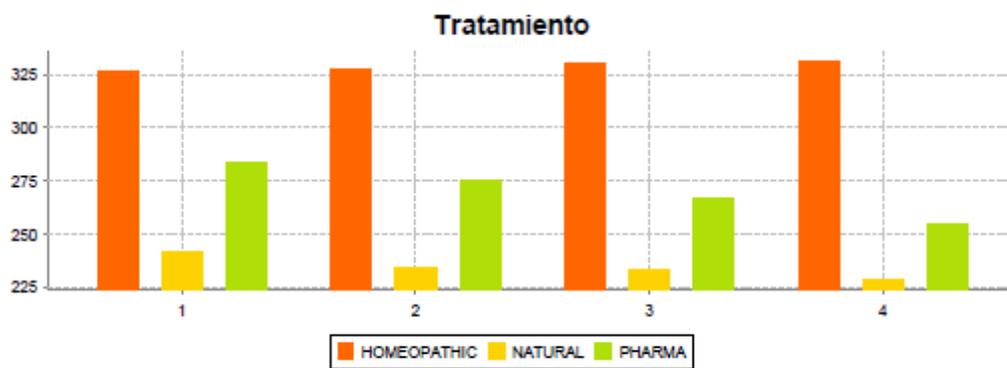
TratamientoNATURAL

Trimestre	Media LDL
1	241
2	234
3	234
4	229

TratamientoPHARMA

Trimestre	Media LDL
1	283
2	275
3	267
4	254

Report Footer



Sat Nov 24 13:24:55 CET 2018

Figura 5. Informe evolución tratamiento por trimestre

Informe Tratamiento por Trimestre y Hábito

El objetivo de este informe es mostrar la evolución de cada tratamiento en relación al hábito seguido.

La consulta realizada es:

```
SELECT
  round(avg("datawarehouse"."evolution"."ldl"),2),
  "datawarehouse"."habits"."name",
  "datawarehouse"."treatment"."type",
```

```

"datawarehouse"."time"."trimester"
FROM
"datawarehouse"."time" INNER JOIN "datawarehouse"."evolution" ON
"datawarehouse"."time"."idtime" = "datawarehouse"."evolution"."idtime"
INNER JOIN "datawarehouse"."treatment" ON "datawarehouse"."evolution"."idtreatment" =
"datawarehouse"."treatment"."idtreatment"
INNER JOIN "datawarehouse"."habits" ON "datawarehouse"."evolution"."idhabit" =
"datawarehouse"."habits"."idhabits"
group by
"datawarehouse"."habits"."name",
"datawarehouse"."treatment"."type",
"datawarehouse"."time"."trimester"
order by "datawarehouse"."habits"."name" asc, "datawarehouse"."treatment"."type" asc,
"datawarehouse"."time"."trimester" asc

```

El resultado del informe es:

Evolución del tratamiento por Hábito

Habito HEALTHY-FAT

Tratamiento	Trimestre	Media LDL
HOMEOPATHIC	1	312
HOMEOPATHIC	2	316
HOMEOPATHIC	3	321
HOMEOPATHIC	4	324
NATURAL	1	266
NATURAL	2	245

Habito HEALTHY-MEDITERRANEAN

Tratamiento	Trimestre	Media LDL
HOMEOPATHIC	1	336
HOMEOPATHIC	2	324
HOMEOPATHIC	3	329
HOMEOPATHIC	4	333
NATURAL	1	269
NATURAL	2	254
NATURAL	3	253
NATURAL	4	243
PHARMA	2	314
PHARMA	3	304
PHARMA	4	285

Habito HEALTHY-VEGETARIAN

Tratamiento	Trimestre	Media LDL
HOMEOPATHIC	1	348
HOMEOPATHIC	2	342
HOMEOPATHIC	3	339
HOMEOPATHIC	4	337
NATURAL	4	245
PHARMA	4	274

Habito NORMAL-FAT

Tratamiento	Trimestre	Media LDL
-------------	-----------	-----------

Figura 6. Informe evolución tratamiento por hábito y trimestre. I

Tratamiento	Trimestre	Media LDL
HOMEOPATHIC	1	311
HOMEOPATHIC	2	311
HOMEOPATHIC	3	322
NATURAL	1	219
NATURAL	2	214
NATURAL	3	215
NATURAL	4	215
PHARMA	1	319
PHARMA	2	306
PHARMA	3	302

Habito NORMAL-MEDITERRANEAN

Tratamiento	Trimestre	Media LDL
HOMEOPATHIC	2	341
HOMEOPATHIC	4	337
NATURAL	1	240
NATURAL	2	257
NATURAL	3	232
NATURAL	4	212
PHARMA	1	315
PHARMA	2	313
PHARMA	3	287
PHARMA	4	275

Habito NORMAL-VEGETARIAN

Tratamiento	Trimestre	Media LDL
HOMEOPATHIC	1	348
HOMEOPATHIC	2	342
HOMEOPATHIC	3	342
PHARMA	4	264

Habito SEDENTARIAN-FAT

Tratamiento	Trimestre	Media LDL
NATURAL	1	214
NATURAL	2	210
NATURAL	3	212
PHARMA	1	259
PHARMA	2	263
PHARMA	3	230

Habito SEDENTARIAN-MEDITERRANEAN

Tratamiento	Trimestre	Media LDL
PHARMA	1	260
PHARMA	2	238
PHARMA	3	233
PHARMA	4	223

Habito SEDENTARIAN-VEGETARIAN

Tratamiento	Trimestre	Media LDL
PHARMA	4	229

Figura 7. Informe evolución tratamiento por hábito y trimestre. II

Informe Evolución Tratamiento por Área Geográfica

Este informe muestra la evolución de los distintos tratamientos divididos por Comunidades

La consulta es:

SELECT

```
round(avg("datawarehouse"."evolution"."idl"),2),  
"datawarehouse"."treatment"."type",  
"datawarehouse"."geographicalarea"."caa",  
"datawarehouse"."time"."trimester"
```

FROM

```
"datawarehouse"."treatment" INNER JOIN "datawarehouse"."evolution" ON  
"datawarehouse"."treatment"."idtreatment" = "datawarehouse"."evolution"."idtreatment"  
INNER JOIN "datawarehouse"."geographicalarea" ON  
"datawarehouse"."evolution"."idgeograficalarea" =  
"datawarehouse"."geographicalarea"."idgeoarea"  
INNER JOIN "datawarehouse"."time" ON "datawarehouse"."evolution"."idtime" =  
"datawarehouse"."time"."idtime"
```

GROUP BY

```
"datawarehouse"."treatment"."type",  
"datawarehouse"."geographicalarea"."caa",  
"datawarehouse"."time"."trimester"
```

ORDER BY

```
"datawarehouse"."geographicalarea"."caa" ASC,  
"datawarehouse"."treatment"."type" ASC,  
"datawarehouse"."time"."trimester" ASC
```

El resultado del informe se puede ver a continuación:

Evolución del tratamiento por Comunidad

Comunidad CATALUNYA		
Tratamiento	Trimestre	Media LDL
NATURAL	1	289
NATURAL	2	254
NATURAL	3	253
NATURAL	4	243
PHARMA	1	244
PHARMA	2	238
PHARMA	3	232
PHARMA	4	223

Comunidad EUSKADI		
Tratamiento	Trimestre	Media LDL
HOMEOPATHIC	1	305
HOMEOPATHIC	2	313
HOMEOPATHIC	3	321
HOMEOPATHIC	4	325
NATURAL	1	214
NATURAL	2	214
NATURAL	3	214
NATURAL	4	215

Comunidad MADRID		
Tratamiento	Trimestre	Media LDL
HOMEOPATHIC	1	348
HOMEOPATHIC	2	342
HOMEOPATHIC	3	340
HOMEOPATHIC	4	338
PHARMA	1	323
PHARMA	2	313
PHARMA	3	301
PHARMA	4	285

Figura 8. Informe evolución tratamiento por comunidad

2.4.5.2 Análisis OLAP

Con la herramienta Saiku Analytics permitimos al usuario el análisis de la información filtrando por los campos que desee.

En la siguiente imagen se puede ver el resultado de los tratamientos por hábitos en cada trimestre

Habit	treatment	PHARMA				NATURAL				HOMEOPATHIC			
		1	2	3	4	1	2	3	4	1	2	3	4
	Trimester	ldl	ldl	ldl	ldl	ldl	ldl	ldl	ldl	ldl	ldl	ldl	ldl
HEALTHY-FAT		-	-	-	-	266	245	-	-	312	315,583	321,455	324,455
HEALTHY-MEDITERRANEAN		-	314	304	265	269	254,455	253,364	242,909	336	324	329	333
HEALTHY-VEGETARIAN		-	-	-	274	-	-	-	245	347,833	341,857	338,889	337,333
NORMAL-FAT		319	306	302	-	219,273	214,364	215,1	215,273	311,143	311	322	-
NORMAL-MEDITERRANEAN		315,1	313	287,273	275,455	240	257	232	211,5	-	341	-	337
NORMAL-VEGETARIAN		-	-	-	264	-	-	-	-	347,5	341,667	342	-
SEDENTARIAN-FAT		259,2	262,667	230	-	214	210	212	-	-	-	-	-
SEDENTARIAN-MEDITERRANEAN		260	238	232,6	222,667	-	-	-	-	-	-	-	-
SEDENTARIAN-VEGETARIAN		-	-	-	229	-	-	-	-	-	-	-	-

Figura 9. Saiku tratamientos por hábito y trimestre

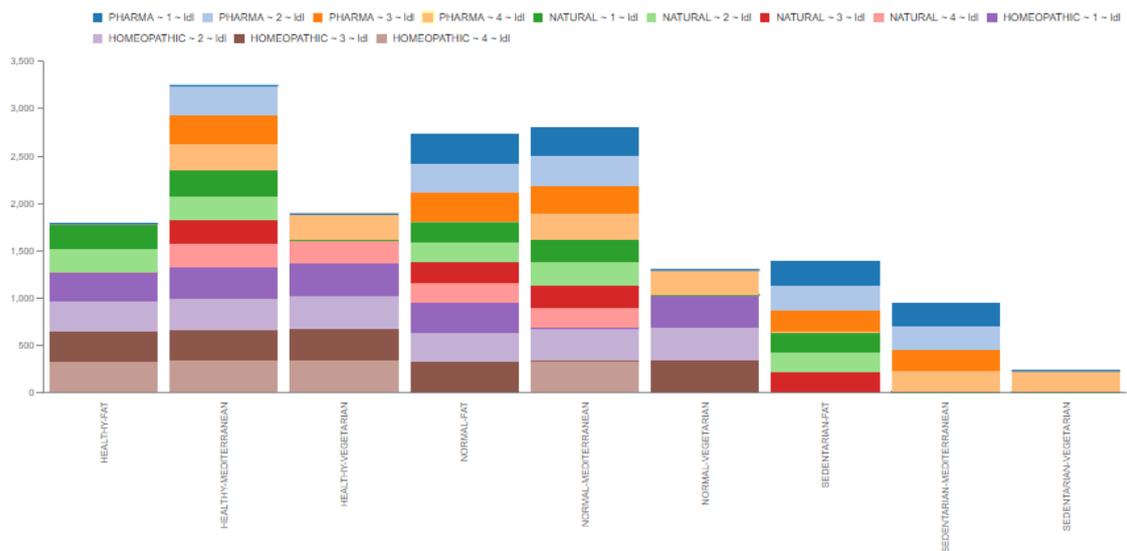


Figura 10. Saiku grafico tratamientos por hábito y trimestre

Y en la siguiente se pueden ver los resultados por tratamiento, trimestre y zona geográfica

Medidas

Idi

Columnas

Treatment

treatment

GeograficalArea

CCAA

Filas

Time

Trimester

treatment	PHARMA		NATURAL		HOMEOPATHIC		
	CCAA	CATALUNYA	MADRID	CATALUNYA	EUSKADI	EUSKADI	MADRID
Trimester	Idi	Idi	Idi	Idi	Idi	Idi	
1		243,538	323,077	268,848	213,923	304,769	348,154
2		237,846	312,615	253,923	213,692	312,538	341,692
3		231,615	301,462	253,154	214,385	321,308	339,615
4		223,462	285,308	243,231	214,692	324,538	337,538

Figura 11. Saiku tratamiento por zona geográfica

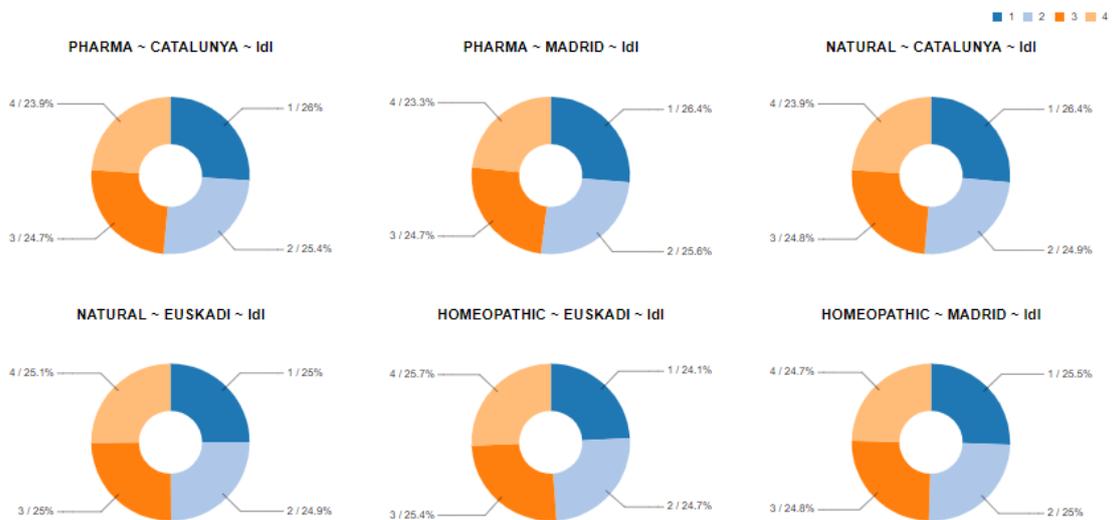


Figura 12. Saiku gráfico tratamiento por zona geográfica

Y en las siguientes imágenes se pueden ver los datos por tratamiento, hábito, comunidad y trimestre

		CCAA				CATALUNYA				EUSKADI				MADRID			
		Trimester				1	2	3	4	1	2	3	4	1	2	3	4
treatment	Habit	Idl	Idl	Idl	Idl	Idl	Idl	Idl	Idl	Idl	Idl	Idl	Idl	Idl	Idl	Idl	Idl
PHARMA	HEALTHY-MEDITERRANEAN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	314	304
	HEALTHY-VEGETARIAN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	274
	NORMAL-FAT	-	-	-	-	-	-	-	-	-	-	-	-	319	306	302	-
	NORMAL-MEDITERRANEAN	247	-	227,5	222	-	-	-	-	-	-	-	-	322,667	313	300,556	287,333
	NORMAL-VEGETARIAN	-	-	-	-	-	-	228	-	-	-	-	-	-	-	-	282
	SEDENTARIAN-FAT	241,5	237	230	-	-	-	-	-	-	-	-	-	330	314	-	-
	SEDENTARIAN-MEDITERRANEAN	244,125	238	232,6	222,667	-	-	-	-	-	-	-	-	323,5	-	-	-
	SEDENTARIAN-VEGETARIAN	-	-	-	229	-	-	-	-	-	-	-	-	-	-	-	-
NATURAL	HEALTHY-FAT	266	245	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	HEALTHY-MEDITERRANEAN	269	254,455	253,364	242,909	-	-	-	-	-	-	-	-	-	-	-	-
	HEALTHY-VEGETARIAN	-	-	-	245	-	-	-	-	-	-	-	-	-	-	-	-
	NORMAL-FAT	271	-	-	-	214,1	214,384	215,1	215,273	-	-	-	-	-	-	-	-
	NORMAL-MEDITERRANEAN	268	257	252	-	212	-	212	211,5	-	-	-	-	-	-	-	-
	SEDENTARIAN-FAT	-	-	-	-	214	210	212	-	-	-	-	-	-	-	-	-
	SEDENTARIAN-MEDITERRANEAN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	SEDENTARIAN-VEGETARIAN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HOMEOPATHIC	HEALTHY-FAT	-	-	-	-	305,5	313	321,455	324,455	351	344	-	-	-	-	-	-
	HEALTHY-MEDITERRANEAN	-	-	-	-	305	309	319	325	346,333	339	339	338,333	-	-	-	-
	HEALTHY-VEGETARIAN	-	-	-	-	-	-	-	-	347,833	341,857	338,889	337,333	-	-	-	-
	NORMAL-FAT	-	-	-	-	304	311	322	-	354	-	-	-	-	-	-	-
	NORMAL-MEDITERRANEAN	-	-	-	-	-	-	-	-	-	-	341	337	-	-	-	-
	NORMAL-VEGETARIAN	-	-	-	-	-	-	-	-	-	347,5	341,667	342	-	-	-	-
	SEDENTARIAN-FAT	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	SEDENTARIAN-MEDITERRANEAN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figura 13. Saiku tratamiento por hábito y comunidad

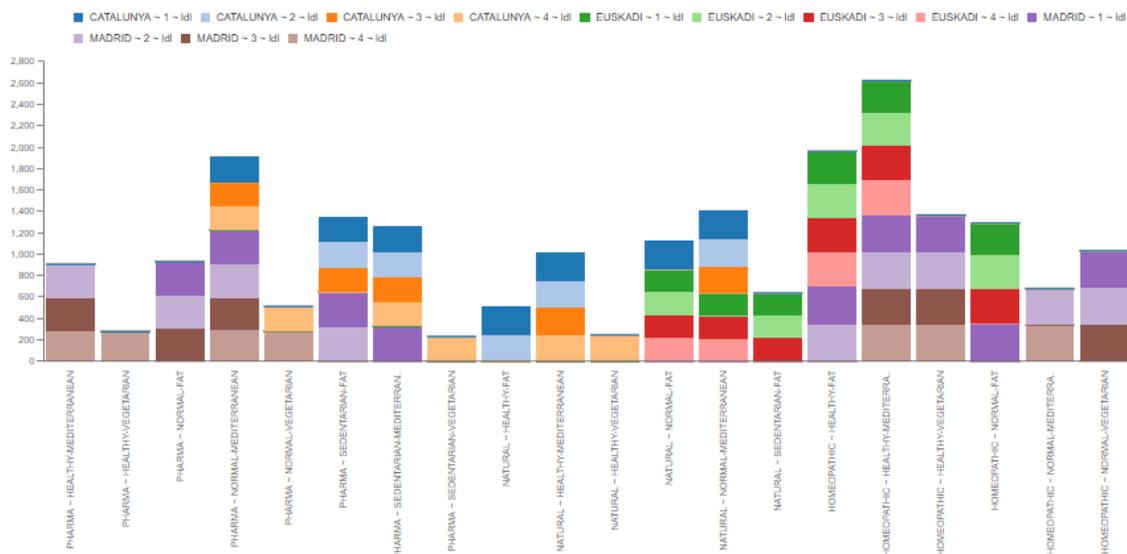


Figura 14. Saiku gráfico tratamiento por hábito y comunidad

Por último, en el siguiente análisis se puede ver la evolución de cada tratamiento en los pacientes

treatment	PHARMA				NATURAL				HOMEOPATHIC			
Trimester	1	2	3	4	1	2	3	4	1	2	3	4
Patient	Idl	Idl	Idl	Idl								
P1	243,538	237,846	231,615	223,462	-	-	-	-	-	-	-	-
P2	-	-	-	-	268,846	253,923	253,154	243,231	-	-	-	-
P3	-	-	-	-	-	-	-	-	304,769	312,538	321,308	324,538
P4	-	-	-	-	213,923	213,692	214,385	214,692	-	-	-	-
P5	323,077	312,615	301,462	285,308	-	-	-	-	-	-	-	-
P6	-	-	-	-	-	-	-	-	348,154	341,692	339,615	337,538

Figura 15. Tratamiento por paciente

2.4.5.1 Cuadro de mandos

Se ha creado el siguiente cuadro de mando que muestre la información de una manera visual que permita al usuario mostrarla de la manera que desee.

El cuadro de mando se crea con el componente CDE Dashboard de Pentaho.

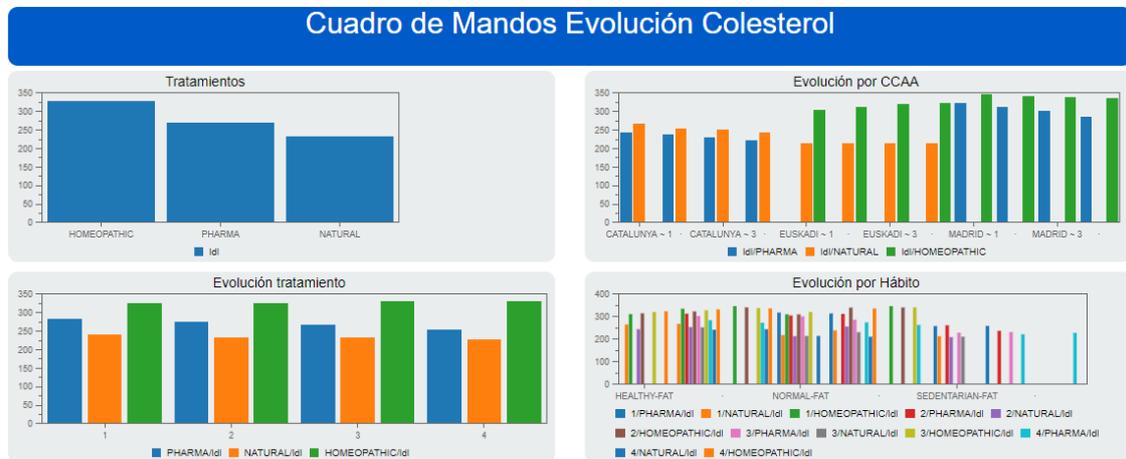


Figura 16. Cuadro de mandos

En el panel superior izquierdo se encuentra el valor medio de LDL de cada tratamiento para todo el periodo consultado.

En el panel superior derecho se puede ver el valor medio de LDL de cada tratamiento en cada comunidad por cada trimestre.

El panel inferior izquierdo muestra la evolución de los distintos tratamientos por cada trimestre.

Y el panel superior derecho muestra para cada hábito y tratamiento la evolución en cada trimestre.

3. Conclusiones

A la hora de elegir el tema sobre el que desarrollar el trabajo fin de master me decanté por el área de business Intelligence porque no conocía nada de este tema, pero siempre me pareció muy interesante por la utilidad que tiene en todos los ámbitos académicos, profesionales y sociales. Es una herramienta muy potente que nos permite obtener información de diferentes fuentes de datos con el fin de procesarla de manera conjunta para ayudar a tomar decisiones empresariales o de negocio.

A nivel personal los objetivos se han cumplido satisfactoriamente puesto que nos ha permitido conocer la potencia de este tipo de soluciones adquiriendo el conocimiento suficiente para poder afrontar proyectos similares.

Al inicio del proyecto se plantearon una serie de objetivos que básicamente consistían en proporcionar una solución business Intelligence a una empresa farmacéutica con el fin de que pueda determinar que tratamiento está funcionando mejor o si en el resultado del tratamiento influye la comunidad autónoma donde vive el paciente o sus hábitos alimenticios.

Para determinar si se han cumplido los objetivos propuestos, vamos a ver si la solución propuesta puede responder las preguntas planteadas en los objetivos de una forma fácil y rápida.

¿Cuál es la relación entre los diferentes tratamientos y la evolución de los pacientes?

Tal y como se puede ver en la siguiente figura, los dos pacientes con el tratamiento PHARMA han conseguido reducir el nivel de colesterol a lo largo del tiempo. En cambio, para los tratamientos NATURAL y HOMEOPATHIC solo uno de los dos pacientes que probaron dicho tratamiento han conseguido una reducción del nivel de colesterol.

Esto significa que para los tratamientos NATURAL y HOMEOPATHIC influyen otros factores, aparte del tratamiento, para conseguir una reducción del nivel de colesterol.

treatment	PHARMA				NATURAL				HOMEOPATHIC			
Trimester	1	2	3	4	1	2	3	4	1	2	3	4
Patient	ldl	ldl	ldl	ldl								
P1	243,538	237,846	231,615	223,462	-	-	-	-	-	-	-	-
P2	-	-	-	-	268,846	253,923	253,154	243,231	-	-	-	-
P3	-	-	-	-	-	-	-	-	304,769	312,538	321,308	324,538
P4	-	-	-	-	213,923	213,692	214,385	214,692	-	-	-	-
P5	323,077	312,615	301,462	285,308	-	-	-	-	-	-	-	-
P6	-	-	-	-	-	-	-	-	348,154	341,692	339,615	337,538

Figura 17. Evolución tratamiento por paciente

¿Existen terapias más eficaces?

Sí, el tratamiento PHARMA es el más efectivo, ya que como se puede observar en la siguiente imagen y en la figura 5, es el que más ha conseguido reducir el colesterol a lo largo del tratamiento.

En cambio, el tratamiento HOMEOPATHIC no ha conseguido ninguna reducción del colesterol.

treatment	PHARMA	NATURAL	HOMEOPATHIC
Trimester	ldl	ldl	ldl
1	283,308	241,385	326,462
2	275,231	233,808	327,115
3	266,538	233,769	330,462
4	254,385	228,962	331,038

Figura 18. Evolución tratamiento por trimestre

¿Ha influido en el resultado, los hábitos de los pacientes?

Si observamos la figura 9 y en la siguiente imagen, que nos muestra la evolución del tratamiento según el hábito del paciente, podemos observar que los pacientes que tienen el hábito HEALTHY-VEGETARIAN consiguen una reducción de los niveles de colesterol con el tratamiento HOMEOPATHIC, que como vimos en el apartado anterior no consigue ninguna reducción de colesterol, lo que se puede comprobar en la figura 9 para el resto de hábitos.

Para el tratamiento NATURAL, se puede observar que funciona bien si la actividad es HEALTHY, en cambio para las actividades NORMAL y SEDENTARIAN no obtiene buenos resultados.

En cambio, el tratamiento PHARMA obtiene buenos resultados independientemente del hábito seguido.

Por lo que se puede concluir que para el tratamiento NATURAL y HOMEOPATHIC sí que influye el hábito, pero para el tratamiento PHARMA no.

treatment	PHARMA				NATURAL				HOMEOPATHIC			
	1	2	3	4	1	2	3	4	1	2	3	4
Trimester	ldl	ldl	ldl	ldl	ldl	ldl	ldl	ldl	ldl	ldl	ldl	ldl
HEALTHY-FAT	-	-	-	-	266	245	-	-	312	315,583	321,455	324,455
HEALTHY-MEDITERRANEAN	-	314	304	285	269	254,455	253,364	242,909	336	324	329	333
HEALTHY-VEGETARIAN	-	-	-	274	-	-	-	245	347,833	341,857	338,889	337,333
NORMAL-FAT	319	306	302	-	219,273	214,364	215,1	215,273	311,143	311	322	-
NORMAL-MEDITERRANEAN	315,1	313	287,273	275,455	240	257	232	211,5	-	341	-	337
NORMAL-VEGETARIAN	-	-	-	264	-	-	-	-	347,5	341,667	342	-
SEDENTARIAN-FAT	259,2	262,667	230	-	214	210	212	-	-	-	-	-
SEDENTARIAN-MEDITERRANEAN	260	238	232,6	222,667	-	-	-	-	-	-	-	-
SEDENTARIAN-VEGETARIAN	-	-	-	229	-	-	-	-	-	-	-	-

Figura 19. Relación Tratamiento-Hábito

¿La evolución a lo largo del tiempo, por un mismo tratamiento, dependen de algún factor como los hábitos?

Sí, para el tratamiento NATURAL si se tiene un hábito con actividad HEALTHY se consigue una reducción de los niveles de colesterol. Y para el tratamiento HOMEOPATHIC solo se obtiene un buen resultado si el hábito es HEALTHY-VEGETARIAN

¿Hay diferencias en el resultado de un tratamiento según el lugar geográfico del paciente?

Pues sí, si nos fijamos en la siguiente imagen del cuadro de mando (y en la figura 11), podemos observar cómo para Cataluña el tratamiento Natural ha conseguido una reducción del colesterol, en cambio ese mismo tratamiento para Euskadi no ha conseguido reducirlo.

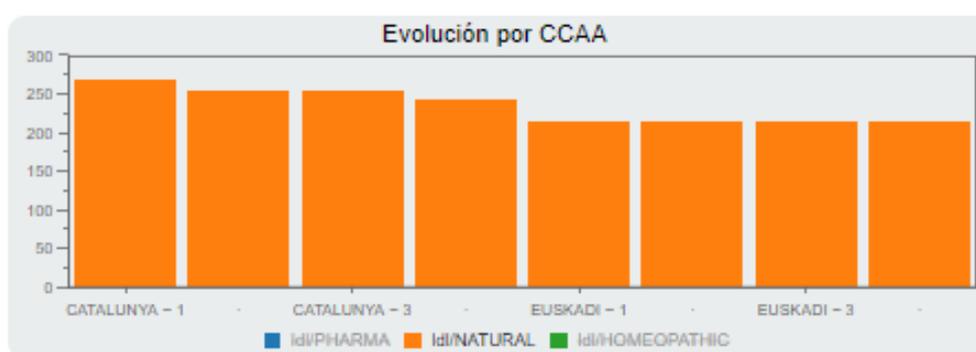


Figura 20. Gráfico tratamiento Natural por CCAA

Lo mismo ocurre para el tratamiento Homeopathic, que en Euskadi tiene un aumento de colesterol a lo largo del tratamiento, en cambio en Madrid obtiene una ligera mejoría a lo largo de los trimestres del tratamiento.

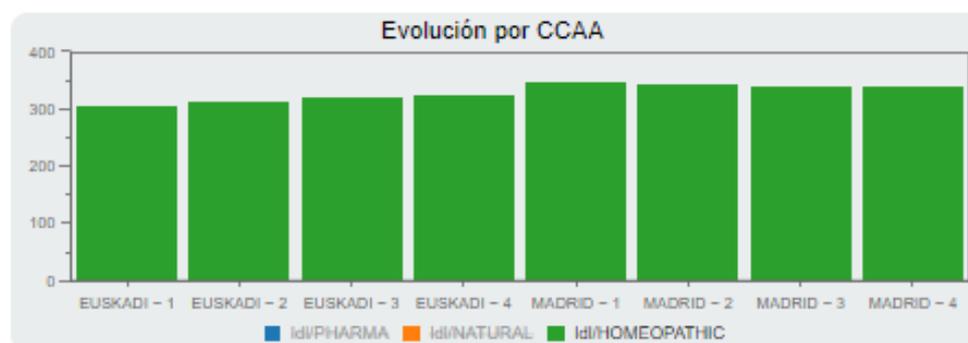


Figura 21. Gráfico tratamiento Homeopathic por CCAA

¿Hay algún periodo del año donde el tratamiento sea más o menos efectivo?

Sí, si observamos la figura 18, podemos ver que para el tratamiento NATURAL y HOMEOPATHIC el tercer trimestre no consigue mejorar los resultados del trimestre anterior.

A la vista de las respuestas a las preguntas, se puede afirmar que se han alcanzado todos los objetivos planteados puesto que se ha terminado el producto, proporcionando una solución al cliente que le permite responder a las preguntas planteadas en los objetivos de una forma rápida, visual y fácil.

El único problema que nos encontramos a la hora de realizar el trabajo ha sido la creación del cuadro de mandos en la herramienta Pentaho. El problema se debió a que hay un bug en la versión 8.1 pentaho server que hacía que no se mostrasen los gráficos del cuadro de mando. En diciembre de 2018 salió la versión 8.2 de pentaho server que ya no tenía este bug y es la que he usado finalmente.

Una línea de trabajo futuro sería aplicar técnicas de minería de datos para conocer qué tipos de datos influyen en los resultados de los tratamientos, con el fin de eliminar aquellos datos que no tienen ninguna correlación.

4. Glosario

Business Intelligence: se trata de un software que permite transformar todos los datos de una compañía en información y conocimiento útil.

Proceso ETL: es el proceso que nos permite obtener datos de múltiples fuentes, transformarlos y cargarlos en una base de datos.

Cubo OLAP: se trata de una base de datos multidimensional para el almacenamiento físico de los datos en un vector multidimensional.

Pentaho: conjunto de programas libres para soluciones business Intelligence.

Data Warehouse: es la base de datos donde se almacena de forma unificada los datos de las diferentes fuentes.

Cuadro de mandos: herramienta gráfica para mostrar la información de forma sencilla y útil.

5. Bibliografía

<http://www.stratebi.com>. Fecha 07/10/2018

<https://blog.capterra.com/top-8-free-and-open-source-business-intelligence-software/>. Fecha 06/10/2018

SAIMA SOLUTIONS: <https://saimasolutions.com/bi-hospital/>. Fecha 07/10/2018

BIRT: <http://www.eclipse.org/birt/>. Fecha 07/10/2018

Pentaho: <http://www.pentaho.org>. Fecha 07/10/2018

SpagoBI: : <http://www.spagoworld.org>. Fecha 07/10/2018

Jasper Reports: <http://community.jaspersoft.com/>. Fecha 07/10/2018

KNIME: <https://www.knime.com/>. Fecha 07/10/2018

Saiku: <https://www.meteorite.bi/products/saiku>. Fecha 12/11/2018

6. Anexos

6.1 Manual de instalación

1. Instalación sistema de gestión de base de datos.

El sistema de gestión de base de datos que se utiliza es Postgres, por lo que el primer paso es descargar e instalar PostgreSQL. La url de descarga es <https://www.postgresql.org/download/>

2. Configuración de la base de datos.

Se tiene que crear una base de datos llamada bi_datawarehouse con un esquema llamado datawarehouse. En dicho esquema estarán las tablas que contendrán la información.

Los scripts que crean esta estructura son:

```
CREATE DATABASE bi_datawarehouse
WITH
OWNER = postgres
ENCODING = 'UTF8'
CONNECTION LIMIT = -1;
```

```
CREATE SCHEMA datawarehouse
AUTHORIZATION postgres;
```

A continuación, se debe ejecutar el script scriptBBDD.sql dentro del esquema datawarehouse para que cree la estructura de tablas.

3. Ejecución procesos ETL.

El software que realiza la ejecución de los procesos ETL es el Data Integration de Pentaho.

Por lo tanto, el primer paso es descargarlo de <https://community.hitachivantara.com/docs/DOC-1009931-downloads>

Una vez descargado, se debe abrir ejecutando el fichero spoon.bat.

Posteriormente se debe crear una conexión al esquema de la base de datos creada anteriormente

A continuación, se debe abrir y ejecutar el fichero "ETL pentaho.ktr". Esta ejecución insertará en las tablas todos los datos que se encuentran en el Excel

DATASHEETCOLOCT18.xls. Por lo tanto, en las entradas del proceso ETL se debe indicar la ruta donde esté ubicado el Excel antes de proceder a la ejecución del fichero “ETL pentaho.ktr”

4. Configuración Pentaho Server

El software de Business Intelligence que nos va a mostrar toda la información de forma gráfica y agrupada es Pentaho Server en su versión Community Edition.

La versión que vamos a utilizar es la 8.2, llamada Pentaho Server CE, que se debe <https://sourceforge.net/projects/pentaho/files/Pentaho%208.2/server/> de

Este software ejecuta un Tomcat, por lo que en Windows se debe crear una variable del sistema llamada JAVA_HOME, que apunte al directorio donde esté instalado el JDK de Java.

Posteriormente, el primer paso es instalar Saiku. Para ello hay que descargarlo de <https://www.meteorite.bi/products/saiku>, obtener una licencia de <https://licensing.meteorite.bi/licenses> .

Una vez instalado saiku, hay que ejecutar el Pentaho Server y crear una conexión de base de datos al esquema PostgreSQL creado en el primer punto.

Posteriormente se debe importar el fichero “Schema Pentaho OLAP.xml”, que contiene el cubo OLAP. Esta importación se hace desde la opción Import Analysis.

Con esto ya estaría configurado Pentaho Server, por lo que el último paso es importar los análisis realizados con Saiku, los informes y el cuadro de mandos. Estos se encuentran en los ficheros:

- Informe Tratamiento por Trimestre y CCAA.ptr
- Informe Tratamiento por Trimestre y Habito.ptr
- Informe Tratamiento por Trimestre.ptr
- Tratamiento_habito_trimestre.saiku.zip
- Tratamiento_habito_CCAA.saiku.zip
- Tratamiento_CCAA.saiku.zip
- Tratamiento_Paciente.saiku.zip
- Cuadro_de_mandos.cda.zip
- Cuadro_de_mandos.cdfde.zip
- Cuadro_de_mandos.wcdf.zip