Universitat Oberta de Catalunya

uoc.edu

Sistema de inteligencia de negocio

Entorno para paciente con trastornos cognitivos

Estibaliz Sánchez Izquierdo

Máster en Ingeniería Informática

Business Intelligence

Ferran Prados Carrasco

David Amorós Alcaraz

Fecha Entrega 17/01/2021

FICHA DEL TRABAJO FINAL

Título del trabajo:	Sistema de inteligencia de negocio. Entorno para paciente con trastornos cognitivos
Nombre del autor:	*Estibaliz Sánchez Izquierdo*
Nombre del consultor/a:	**David Amorós Alcaraz**
Nombre del PRA:	**Ferran Prados Carrasco**
Fecha de entrega (mm/aaaa):	01/2021
Titulación:	*Máster en Ingeniería Informática*
Área del Trabajo Final:	*Business Intelligence*
Idioma del trabajo:	*Castellano*
Palabras clave	*Pentaho, Business Intelligence, Cognitive disorders*

Resumen del Trabajo:.

Este Trabajo de Fin de Máster se enmarca dentro de un proyecto que intenta desarrollar terapias que permitan entender y controlar enfermedades relacionadas con los trastornos cognitivos. Para ello, se ha hecho un estudio sobre veinte pacientes afectados por este tipo de enfermedades. Este estudio intenta relacionar los estados de ánimo y las actividades realizadas con la aparición de crisis agudas o empeoramientos temporales de los síntomas asociados a estas enfermedades. Poder extraer conclusiones sobre esta relación podría ayudar en la mejora de las condiciones de vida de enfermos con estos trastornos y, para ello, lo que se propone en este proyecto es el desarrollo de una herramienta de Business Intelligence.

Para la definición de los requisitos funcionales se han proporcionado una serie de preguntas a las que la herramienta desarrollada tiene que dar respuesta.

Con los requisitos funcionales definidos desde un comienzo, se define la estrategia tradicional de gestión de proyectos del PMBOK como método de desarrollo de la herramienta.

Asimismo, el resultado obtenido ha sido un panel de cuadros de mando donde se visualizan diferentes elementos que dan respuesta a las preguntas propuestas y que cargan los datos desde un almacén de datos implementado en MySQL. Para la carga del almacén de datos se han implementado procesos ETL mediante el *framework* Pentaho Data Integration.

Este proyecto es una buena forma de iniciarse en el ámbito de los sistemas de Business Intelligence.

Abstract:

This Master's Thesis is part of a project that intends to develop therapies that allow understanding and controlling diseases related to cognitive disorders. For this, a study has been conducted on twenty patients affected by these types of diseases. This study tries to relate the moods and activities carried out with the appearance of acute attacks or temporary worsening of the symptoms associated with these diseases. Being able to infer conclusions about this relationship could help improve the living conditions of patients with these disorders and, for this purpose, the development of a Business Intelligence tool has been proposed in this project.

For the definition of the functional requirements, a series of questions have been posed, which the developed tool must answer.

The traditional PMBOK project management strategy has been chosen as the tool's development method because the functional requirements are known from the beginning.

Likewise, the result obtained has been a dashboard panel where different elements are displayed that answer the proposed questions and that load the data from a data warehouse implemented in MySQL. To load the data warehouse, ETL processes have been implemented using the Pentaho Data Integration framework.

This project is a solid starting point in the field of Business Intelligence systems.

Índice

Lista de figuras

1 Introducción

1.1 Contexto y justificación del Trabajo

En la era digital que las empresas viven en la actualidad, tomar decisiones bien informadas es uno de los principales factores de diferenciación. Para ello, es esencial tener la información adecuada y un tiempo de respuesta corto para dar soporte a toda la gestión de las operaciones de la empresa de forma ágil y rápida.

Asimismo, las tecnologías de la información tienen un papel muy importante al permitir la recolecta, el almacenamiento y el procesamiento de datos generados por las operaciones de la empresa. Por ello, surgen conjuntos de procesos, aplicaciones y tecnologías que facilitan la obtención rápida y sencilla de datos provenientes de los sistemas de gestión empresarial y el análisis e interpretación de los datos para la toma de decisiones de la empresa. Esto es lo que en la actualidad se llama Inteligencia de Negocio o *Business Intelligence* (BI).

Las herramientas de BI por lo general muestran la información en forma de cuadros de mando que se pueden crear a partir de los datos internos o externos que se obtienen de la empresa, de tal forma que la información es presentada al usuario de manera ágil y accesible para que pueda realizar el análisis e interpretación correspondiente.

A pesar de que hoy en día este tipo de herramientas son más utilizadas en el ámbito de finanzas, ventas, logística, producción, etc. en el ámbito de la sanidad también son aplicables, al igual que lo es la minería de datos.

Este Trabajo de Fin de Máster se enmarca dentro de un proyecto que intenta desarrollar herramientas y terapias que permitan entender y controlar enfermedades relacionadas con los trastornos cognitivos. Para ello, se ha hecho un estudio sobre veinte pacientes afectados por este tipo de enfermedades. Este estudio intenta relacionar los estados de ánimo y las actividades realizadas con la aparición de crisis agudas o empeoramientos temporales de los síntomas asociados a estas enfermedades. Poder extraer conclusiones sobre esta relación podría ayudar en la mejora de las condiciones de vida de enfermos con estos trastornos y, para ello, lo que se propone es el desarrollo de una herramienta de BI.

1.1.1 Trastornos de déficit cognitivo

Conforme a la definición del Instituto Superior de Estudios Sociales y Sanitarios [1], los trastornos cognitivos alteran las funciones cognitivas de la persona que los padece, como pueden ser la memoria, el lenguaje, la atención,

la conducta, el aprendizaje o la orientación. Este tipo de trastornos suele darse en personas mayores, por lo que se debe trabajar para prevenir dicho deterioro cognitivo. Dentro de estos trastornos, se encuentran el delirium, la demencia o los trastornos amnésicos, que explicaremos a continuación:

- Delirium: Se trata del deterioro agudo y global de las funciones superiores. Su dato más característico es el deterioro del nivel de conciencia. Al principio sólo se detectan dificultades de atención, concentración y desorientación (temporal al inicio, luego espacial). Conforme se agrava, se desestructura el pensamiento y la percepción. En el delirium se diferencian dos patrones según la alteración de la conducta: agitado y estuporoso.

- Demencia: Es el síndrome caracterizado por el deterioro crónico y global de las denominadas funciones superiores. Lo normal en estos casos es un deterioro intelectual acompañado de alteraciones de la conducta y del estado de ánimo. Su prevalencia aumenta con la edad (de 65 a 70 años, 2%; >80 años, 20%), siendo la principal causa de incapacidad a largo plazo en la tercera edad. Suele iniciarse con el deterioro de la memoria y cambios de personalidad, sin que el paciente tenga conciencia de sus cambios que, con frecuencia, niega o disimula. La conducta se vuelve inapropiada y se pierde el interés por las cosas debido en gran parte a fuertes déficits de atención.

  Según la OMS, la demencia afecta a nivel mundial a unos 50 millones de personas, de las cuales alrededor del 60% viven en países de ingresos bajos y medios. Cada año se registran cerca de 10 millones de nuevos casos. Se prevé que el número total de personas con demencia alcance los 82 millones en 2030 y 152 millones en 2050. Buena parte de ese incremento puede achacarse al hecho de que en los países de ingresos bajos y medios el número de personas con demencia tenderá a aumentar cada vez más.

- Trastornos amnésicos: Es un deterioro específico en la memoria, normalmente de la memoria reciente. Los trastornos amnésicos más típicos son los siguientes:

  - Psicosis de Korsakoff: Trastorno de la memoria provocado por la deficiencia de vitamina B1. Afecta sobre todo a la memoria a corto plazo. Los pacientes que presentan este síndrome manifiestan por norma general dificultad al caminar y con el equilibrio, confusión, somnolencia, parálisis de algunos músculos oculares, neuropatía periférica, etc.

  - Traumatismos craneoencefálicos: Asociada a la amnesia retrógrada y anterógrada. Ambas se asocian con la intensidad del traumatismo. En él se asocian déficits cognitivos leves (deterioro de la atención o la memoria) con síntomas afectivos (ansiedad, labilidad emocional, tristeza), cambios de personalidad, cansancio, fatiga, cefalea, insomnio, inestabilidad.

- Amnesia global transitoria: Caracterizada por una pérdida brusca de la memoria reciente, provocándole un estado de desorientación y perplejidad al no poder retener información; el resto de la exploración es normal. El paciente conserva recuerdos lejanos (nombre, lugar de nacimiento); pero es incapaz de recordar cosas recientes a pesar de mantener un buen nivel de atención; es característico que el paciente repita de forma insistente la misma pregunta.

1.2 Objetivos del trabajo

Tal y como se ha mencionado en el anterior apartado, el objetivo general es obtener una herramienta que facilite al experto del dominio la extracción de conclusiones que relacionen los estados de ánimo y las actividades realizadas con la aparición de crisis agudas o empeoramientos temporales de los síntomas asociados a las enfermedades de trastorno cognitivo. Para ello, será indispensable que la herramienta contenga funcionalidades que den respuesta a las siguientes preguntas:

- ¿Cuál es la relación entre las actividades realizadas y los episodios de crisis graves?
- ¿Se puede establecer algún tipo de relación entre los valores de los diferentes estados de ánimo y los episodios de crisis?
- ¿Estas relaciones son iguales para cualquiera de las enfermedades o, en cambio, hay relaciones más acusadas por alguna de ellas?
- ¿Se puede establecer alguna relación en nivel geográfico, por ejemplo, entorno urbano o rural?
- ¿Cuál sido la evolución de los diferentes pacientes a lo largo del tiempo?
- ¿Se puede establecer alguna relación entre los episodios de crisis y el momento del día o de la semana o del año?
- ¿La realización de actividades físicas mejora o empeora el estado de ánimo de los pacientes?
- ¿Hay algún tipo de actividad que mejore el día a día de los pacientes?

Por tanto, el trabajo a realizar en este TFM 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 con enfermedades cognitivas provenientes de diferentes centros médicos. En concreto, los objetivos del trabajo son:

1. Diseñar un almacén de datos (*data warehouse*) que permita almacenar la información adquirida desde los diferentes orígenes de datos situados en cada centro médico.

2. Implementar este almacén de datos. Para ello, se puede utilizar un entorno relacional como MySql o MariaDB.

3. Utilizar un entorno BI (Power BI, Pentaho, etc) para la realización de los siguientes trabajos:

    a. Programar los procesos ETL (extracción, transformación y carga) que permitan alimentar el DW a partir de los ficheros base facilitados. Para ellos se podrá utilizar la herramienta *Data Integration* de Pentaho o crear procesos ad-hoc manualmente.

    b. Utilizar las utilidades propias según la herramienta para la definición de cubos.

    c. Creación de un cuadro de mando con capacidades multidimensionales para el análisis de la información y formulación de conclusiones.

1.3 Enfoque y método seguido

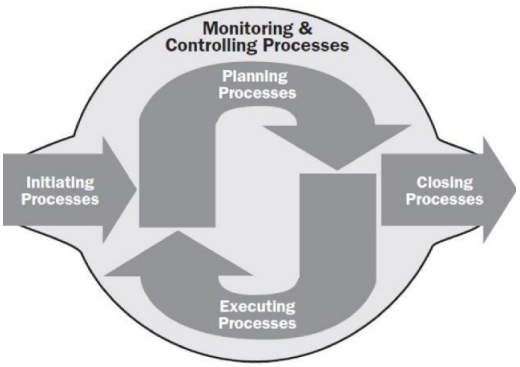


Imagen 1. Ciclo de vida del proyecto

En este caso, se desarrollará un producto nuevo, ya que no se tiene uno del que partir y, al igual que se suele hacer en la mayoría de los proyectos informáticos, la estrategia que se va a llevar a cabo para el desarrollo de este trabajo de fin de máster se basará en la guía PMBOK, donde la gestión de proyectos se organiza por fases.

Esta estrategia encaja bien con este proyecto porque el alcance está bien definido y, por tanto, los objetivos también. Esto permite poder tener una planificación desde el comienzo del proyecto, ya que a priori se conocen los requisitos funcionales mínimos y requerimiento de la herramienta a desarrollar.

1.4 Planificación del trabajo

El proyecto comienza el 17 de septiembre con la elección de la temática del TFM y su correspondiente planificación. Asimismo, finaliza el 1 de febrero con la evaluación del jurado.

A continuación, se muestra el diagrama de Gantt, donde se identifican las tareas a realizar y se identifican los principales hitos.

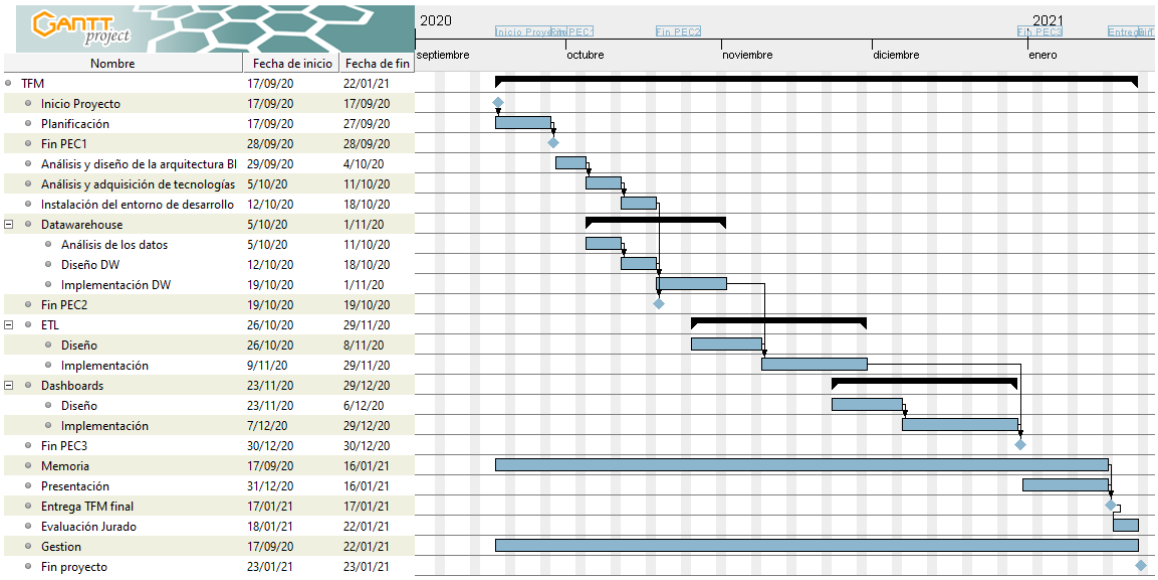


Imagen 2. Diagrama de Gantt del proyecto

1.4.1 Hitos

Un hito en gestión de proyectos es un momento específico, un punto de referencia, dentro del ciclo de vida de un proyecto que marca un evento importante para medir el progreso de un proyecto. Los hitos tienen una fecha fija pero no una duración. En este caso los hitos se han utilizado como señales para el inicio y fin del proyecto y como señales de las entregas de las PEC que, a su vez, sirven para la comprobación del cumplimiento de los objetivos de estos.

- Inicio del proyecto (17/09/2020).
- Fin PEC 1 (28/09/2020): Es la primera entrega del proyecto. En este se define el plan de trabajo.
- Fin PEC 2 (19/10/2020): Es la segunda entrega del proyecto. Tiene como objetivo el estudio del problema, la elección del entorno de desarrollo y el desarrollo del modelo de datos
- Fin PEC 3 (30/12/2020): Es la tercera entrega del proyecto. La memoria debe contener el detalle del trabajo realizado en el desarrollo de los procesos ETL y de la aplicación.
- Fin Entrega Final (17/01/2021): Es la última entrega del proyecto, donde se entrega la memoria del proyecto y la presentación.
- Fin del proyecto (22/01/2021).

1.5 Breve sumario de productos obtenidos

Este proyecto tiene como resultado una herramienta de análisis de trastornos cognitivos, donde de una forma visual, el experto de dominio puede

jugar con la parametrización de algunos factores que posiblemente estén relacionados con este tipo de enfermedad y analizar de una forma visual la relación entre estos factores.

Para poder entender mejor los requisitos funcionales de la herramienta, se ha realizado un análisis de las consultas que se proponen como objetivo para que la herramienta de respuesta a éstas. Todo ello ha sido documentado y relacionado con el modelo de datos.

Teniendo claro el objetivo, se ha generado un documento con el análisis de la fuente de datos, el fichero Excel. Asimismo, se ha diseñado e implementado un almacén de datos, donde la información se guarda adaptada a las necesidades de las visualizaciones tras haber realizado las transformaciones necesarias de los datos leídos del origen Excel.

Para poder adaptar los datos leídos a las necesidades de las funcionalidades y el modelo de datos, ha sido necesario diseñar e implementar un *workflow* con procesos ETL que se almacenan en ficheros legibles con Pentaho y un almacén de datos que se configura mediante un fichero SQL.

Por último, todo el sistema ha sido configurado tanto en un entorno Linux como en un entorno Windows por lo que se ha generado un manual de instalación.

1.6 Breve descripción de los otros capítulos de la memoria

A continuación, a modo resumen del documento, se describen los apartados que lo componen.

El apartado **Herramienta de análisis de trastornos cognitivos** describe la solución dada en el proyecto, tanto desde el punto de vista de la arquitectura como de su funcionalidad.

En el apartado **Tecnología** se detalla el análisis de las diferentes alternativas tecnológicas, así como la elección que se ha realizado de estas para este proyecto.

En el apartado **Fuente de datos** se describe y se analiza el documento Excel que este proyecto ha tenido como referencia de origen de datos.

En el apartado **Repositorio de información. Data Warehouse** se detalla el análisis de las consultas a las que la herramienta tiene que dar una respuesta enfocado al diseño del modelo de datos. Asimismo, se describe el modelo de datos diseñado para el data warehouse y su implementación.

En el apartado **Procesos ETL** se describe el *workflow* de procesos ETL diseñado, así como el detalle de cada uno de ellos.

En el apartado **Exploración de los datos. Cuadros de Mando** se describen las diferentes visualizaciones que se han implementado para dar respuesta a las consultas objeto de la herramienta.

En el apartado **Conclusiones** se recogen las reflexiones realizadas en cuanto a la experiencia y aprendizaje adquirido, las críticas en cuanto al cumplimiento de los objetivos planificados y la planificación y la metodología llevada a cabo. También se recogen las líneas futuras o mejoras que podrían llevar a cabo.

En el apartado **Glosario** se recogen las definiciones de los términos y acrónimos más relevantes utilizados dentro de este documento.

En el apartado **Bibliografía** se recogen aquellas referencias que se han tenido en cuenta durante el proyecto.

En el apartado **Anexo 1** se adjunta el manual de instalación del entorno.

En el apartado **Anexo 2** se adjunta el script SQL de instalación de la base de datos.

2 Herramienta de análisis de trastornos cognitivos

Tras analizar diferentes requisitos del sistema se ha diseñado una arquitectura para dar solución a la aplicación de BI que se plantea y en base a ello se ha realizado un análisis tecnológico. En los siguientes subapartados se describe la herramienta de análisis de trastornos cognitivos y la arquitectura sobre la que se despliega.

2.1 Descripción de la herramienta

A continuación, se describe la funcionalidad de la herramienta desarrollada como solución al sistema de análisis de trastornos cognitivos que se plantea en este proyecto.

- La visualización debe tener filtros de datos y cuadros de mandos que interaccionen con ello.
- Debe mostrar la relación entre actividades realizadas y los episodios.
- Debe mostrar la relación entre estados de ánimos y los episodios.
- Debe mostrar la relación entre actividades, episodios, horas dormidas y el tipo de desorden cognitivo.
- Debe mostrar cómo impacta el entorno geográfico en la enfermedad.
- Debe mostrar la evolución de los pacientes en el tiempo.
- Debe mostrar la evolución de los episodios en el tiempo.
- Debe mostrar la relación entre la actividad física y el estado de ánimo.
- Debe mostrar el impacto de las actividades en la evolución de los episodios.

2.2 Arquitectura del sistema

Cualquier aplicación o servicio se basa en una arquitectura específica, definida por los requisitos intrínsecos o por circunstancias externas, como pueden ser la utilización de recursos externos o requerimientos por parte de la empresa. Por eso, es importante identificar los requisitos tanto funcionales como no funcionales lo antes posible y diseñar una arquitectura válida para la solución

final. Además, la arquitectura es un factor decisivo en la elección de la tecnología a utilizar, por lo que se debe definir al comienzo del ciclo del proyecto.

El siguiente gráfico muestra de forma genérica la arquitectura que tiene la solución BI que se plantea.

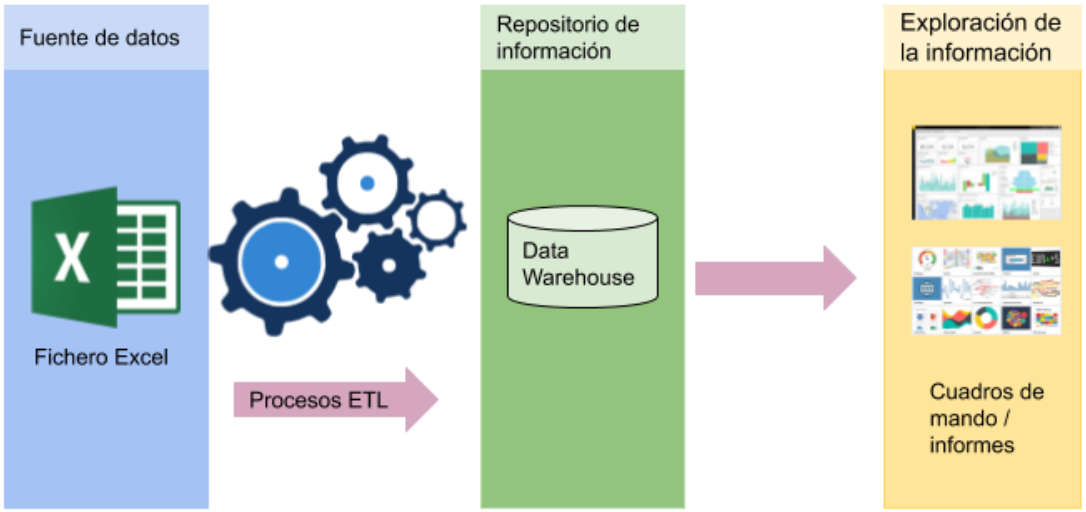


Imagen 3. Arquitectura del sistema BI

En él se distinguen los siguientes componentes:

- **Fuente de datos.** Este componente hace referencia a los distintos orígenes de datos sobre los que se empieza a montar el sistema BI. En este caso el origen de los datos es un fichero Excel, pero en otros contextos podrían ser múltiples.

- **Procesos ETL.** Sobre las fuentes de información se montan los procesos ETL (Extracción, Transformación y Carga) que recogen la información de las fuentes de datos origen, realizan las transformaciones oportunas y cargan la información en un nuevo repositorio de información adaptado para poder realizar sobre él la exploración de la información.

- **Repositorio de información.** Este componente hace referencia a los distintos elementos de almacenamiento de los datos transformados. El elemento más conocido y el que se ha utilizado en esta solución es el *data warehouse* (almacén de datos), que según la definición de William Harvey Inmon “es una colección de datos orientada al negocio, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia”. No obstante, en este componente se pueden utilizar conceptos tales como *data marts* o Cubos OLAP.

- **Exploración de la información.** Este componente hace referencia al conjunto de herramientas que permiten recuperar la información del

repositorio de información adaptada a las necesidades que se requieren.

3 Tecnología

Una de las partes más importantes del proyecto es la elección de la tecnología que va a dar soporte a la solución que se vaya a desarrollar. Generalmente, la adquisición tecnológica viene condicionada por la arquitectura del sistema y algunos requisitos no funcionales.

Para este proyecto en concreto, se ha identificado la necesidad de adquirir una tecnología para cada componente de la arquitectura:

- Procesos ETL. Se requiere de una tecnología que permita desarrollar los ETL bien mediante programación manual o, bien mediante una tecnología específica para ello.

- Repositorio de información. Se requiere una tecnología para el almacenamiento de los datos transformados.

- Exploración de los datos. Se requiere de una tecnología que permita leer los datos del repositorio de información y los visualice en cuadros de mandos.

No ha sido objetivo del proyecto realizar una comparativa exhaustiva entre las diferentes alternativas tecnológicas; sin embargo, a continuación, se señalan los aspectos más decisivos para su elección:

- Tecnología que se ajuste a la arquitectura definida y que sea instalable en *commodity hardware*.

- La experiencia previa tanto en la tecnología como en su lenguaje.

- Tecnología de actualidad o en auge.

- Soporte de la comunidad de desarrolladores.

- Tecnología de código abierto.

Teniendo todo esto en cuenta, a continuación, se describe el análisis tecnológico realizado y la decisión tomada respecto a la elección.

3.1.1 Almacenamiento de datos

Para el almacenamiento será necesario escoger un sistema de gestor de base de datos. En este caso los datos se encuentran bien estructurados y relacionados entre sí, sin necesidad de tener un esquema flexible, por tanto, en este proyecto se opta por un sistema relacional SQL.

Existen diferentes tecnologías que implementan bases de datos relacionales. No obstante, en este proyecto se tiene como requisito que sean de código libre y gratuitas. Por tanto, a continuación, se detallan las especificaciones de MySQL y PostgreSQL, los sistemas que en la actualidad están en auge.

- Características de MySQL
    - Base de datos relacional de objetos
    - No cumple del todo el estándar SQL
    - El sistema relacional más popular, por tanto, mayor soporte
- Características PostgreSQL
    - Base de datos relacional
    - Soporte para el estándar SQL
    - No es tan popular

Para este proyecto cualquiera de las tecnologías sería válida, por tanto, por la gran comunidad que tiene y el conocimiento previo que se tiene se ha escogido MySQL.

3.1.2 Tecnología BI

La analítica de datos es un tema de la actualidad en las empresas y, por ello, cada vez existen más tecnologías para dar soporte a las necesidades que se presentan en el ámbito empresarial. Incluso, existe un Cuadrante Mágico de Gartner para tecnologías de BI que permite a las empresas facilitar la elección éstas.

Imagen 4. Cuadrante Mágico de Gartner para tecnologías de BI 2020

Como se puede observar, en el año 2020 Tableu, Power BI (Microsoft) y Qlik fueron las herramientas comerciales más potentes en este ámbito. No obstante, ninguna de las tecnologías anteriormente mencionadas cumple con los requisitos que se han presentado y, por ello, se han identificado y analizado otras con el fin de cumplimentar los requisitos.

Apache Superset

Apache Superset, tecnología creada por la empresa AirBnb y adoptada por Apache, es una aplicación web de BI que proporciona una interfaz intuitiva para la exploración y visualización de conjuntos de datos y para crear paneles interactivos. [2]

- Fácil de crear cuadros de mando a golpe de *click*.
- Gran oferta de diferentes elementos de visualización.
- Compatible con la mayoría de las bases de datos SQL.
- Robusta para la gestión de grandes almacenes de datos implementados en Apache Druid.
- Contiene una capa semántica ligera, que permite controlar cómo se exponen las fuentes de datos al usuario mediante la definición de dimensiones y métricas.
- Control de usuarios mediante la integración con los principales *backends* de autentificación (base de datos, OpenId, LDAP, OAuth, REMOTE_USER, etc.).
- No hay un gran soporte por parte de la comunidad de desarrolladores.
- No es una tecnología suficientemente madura, ya que todavía se encuentra bajo la supervisión de Apache Incubator.

Pentaho

Pentaho se define a sí mismo como una plataforma de BI “orientada a la solución” y “centrada en procesos” que incluye todos los principales componentes requeridos para implementar soluciones basadas en procesos, tal como ha sido concebido desde el principio.

Las soluciones que Pentaho pretende ofrecer se componen fundamentalmente de una infraestructura de herramientas de análisis e informes integrado con un motor de *workflow* de procesos de negocio.

Las características principales de esta tecnología son:

- Pentaho Analysis Services (Mondrian). Es un servidor OLAP.

- Pentaho Data Integration: Es el módulo que permite definir procesos de transformación de datos (ETL).

- Pentaho Reporting. es un motor de presentación, capaz de generar informes programáticos sobre la base de un archivo de definición XML.

- Pentaho Data Mining. Es una envoltura alrededor del proyecto Weka. Es una suite de software que usa estrategias de aprendizaje de máquina, aprendizaje automático y minería de datos.

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

- Pentaho para Apache Hadoop. Es un conector de bajo nivel para facilitar el acceso a muy grandes volúmenes manejados en el proyecto Apache Hadoop. La Suite de Pentaho BI para Hadoop permite abordar los mayores desafíos que experimentan los usuarios de Hadoop.

- Tiene un foro propio para el soporte.

- Ofrece una versión *enterprise* y una *community*. La *enteprise* se obtiene mediante una suscripción anual y contiene funciones adicionales y soporte que no se encuentra en la edición *community*.

Knowage

Knowage es el nombre que ha adoptado el software de BI conocido como Spago BI. Es una plataforma que ofrece una suite de tecnologías para ofrecer soluciones para la presentación de informes, el análisis multidimensional (OLAP), la minería de datos (*data mining*), los tableros de mando (*dashboard*) y consultas ad-hoc.

Algunas de las características principales son las siguientes:

- Reportes para mostrar datos estructurados.

- Análisis OLAP para navegar por los datos.

- Gráficos para ofrecer vistas simples e intuitivas de la información.

- Dashboards, tablero de mandos para monitorear los KPI.

- Modelos KPI para construir y probar un modelo propio de supervisión de los Indicadores claves del desempeño.

- Procesos de *data mining* para descubrir la información oculta.

- Consola de supervisión en tiempo real, para monitorear las aplicaciones.

- Filtros Inteligentes, para la selección guiada de los datos.

- Procesos externos para la ejecución de los procesos capaces de interactuar con el sistema OLTP.

- Procesos ETL, para la recogida de datos de diferentes fuentes.

- No tiene gran soporte por parte de la comunidad de desarrolladores

- Permite la integración de soluciones propietarias (Business Objects o Microsoft Analysis Services, por ejemplo) para poder construir la mejor plataforma para un problema en particular.

En resumen, teniendo en cuenta las características principales de las tecnologías que se han analizado y que no se tiene conocimiento previo en ninguna de ellas, se ha optado por utilizar Pentaho por tener la capacidad de definir procesos de transformación de datos con Pentaho Data Integration, de crear visualizaciones con Pentaho Dashboard Editor y por el soporte de la comunidad de desarrolladores que hay en comparación con las demás tecnologías.

4 Fuente de datos

El proyecto que se encarga de desarrollar terapias para entender y controlar enfermedades relacionadas con los trastornos cognitivos ha hecho un estudio sobre veinte pacientes afectados por este tipo de enfermedad. Para ello, se ha recolectado diariamente información de cada paciente sobre la actividad realizada, el estado de ánimo tenido, las horas de sueño y los episodios sufridos. Todo ello se encuentra en un fichero Excel estructurado en diferentes hojas que, a continuación, se describen.

Asimismo, el registro realizado a los pacientes tiene una frecuencia diaria, por tanto, la granularidad mínima de los hechos es a nivel de día. En cuanto al rango cronológico de los datos, el fichero con el que se ha trabajado solo disponía de los datos correspondientes al año 2016. No obstante, hay que tener en cuenta que el sistema debe estar preparado para recoger datos de otro cualquier momento, ya sea anterior o posterior.

PATIENTS

En esta hoja se detalla información personal del paciente. Para ello se dispone de la siguiente información:

- Patient: identificador único del paciente. P1, P2, P3, Pn
- Cognitive disorder: tipo de desorden cognitivo que tiene el paciente. Los valores que puede tomar son DELIRIUM, AMNESIA o DEMENTIA.
- City: ciudad del paciente.
- Environment: tipo de entorno en el que vive el paciente. Los valores que puede tomar son URBAN, SEMIURBAN o RURAL.

HOURS SLEEP VALUES

En esta hoja se detalla el número de horas de sueño de cada paciente para cada día del año 2016.

- Fecha: Fecha del registro de sueño en formato dd/mm/aaaa.
- P1, P2, P3… Pn: Cada columna representa el número de horas de sueño para cada uno de los pacientes.

ACTIVITY VALUES

En esta hoja se detalla el tipo de actividad que ha llevado cada paciente cada día del año 2016.

- Fecha: Fecha del registro de la actividad realizada en formato dd/mm/aaaa.

- P1, P2, P3… Pn: Cada columna representa el tipo de actividad principal realizado por el paciente. Los valores que puede tomar el registro son NO ACTIVITY, RADIO/TV, SLEEP/SOFA, FAMILY, EXERCISE, READ/STUDY.

EPISODE VALUES

En esta hoja se detalla información sobre el tipo de episodio que ha sufrido cada paciente cada día del año 2016.

- Fecha: Fecha del registro del episodio en formato dd/mm/aaaa.

- P1, P2, P3… Pn: Cada columna representa el tipo de actividad principal realizado por el paciente. Los valores que puede tomar el registro son NO EPISODE, LIGHT, SEVERE y MODERATE.

MOOD VALUES

En esta hoja se detalla información sobre el estado de ánimo de cada paciente cada día del año 2016

- Fecha: Fecha del registro del estado de ánimo del paciente en formato fecha Excel. Esto será necesario cambiar para que tenga el mismo formato que el resto de las fechas.

- P1, P2, P3… Pn: Cada columna representa el estado de ánimo principal de cada paciente según sus familiares o cuidadores. Los valores que puede tomar el registro son SAD, NORMAL, HAPPY.

5 Repositorio de información. Data Warehouse

En el ámbito de los sistemas informáticos los *data warehouse* o los almacenes de datos hacen referencia a una colección de datos orientada a un determinado contexto, integrado, no volátil y variable en el tiempo y orientado a la toma de decisiones. Es un componente básico de los sistemas BI y se usa para crear informes o realizar análisis de datos. [3]

Esta colección de datos está formada por «dimensiones» y «medidas», entendiendo como dimensiones a aquellos elementos que participan en el análisis y medidas a los valores que se desean analizar.

En las bases de datos de los almacenes de datos los esquemas más utilizados son el esquema en estrella y, su versión compleja, esquema en copo de nieve. Estas estructuras son un tipo de modelo de datos que tienen una tabla de hechos que contiene los datos para el análisis, rodeada de las tablas de dimensiones.

A continuación, se detalla el análisis realizado sobre las consultas propuestas para identificar las dimensiones y medidas del modelo de datos. Después, se presenta el modelo de datos diseñado junto a los detalles de su implementación.

5.1 Análisis de los datos y las consultas enfocado al diseño del modelo de datos

Tal y como se ha mencionado en el apartado de Objetivos de trabajo, el objetivo de este proyecto es obtener una herramienta de análisis que dé respuesta a una serie de preguntas.

A continuación, se analizan las preguntas para obtener las dimensiones y medidas necesarias a tener en cuenta a la hora de diseñar el modelo de datos.

- **¿Cuál es la relación entre las actividades realizadas y los episodios de crisis graves?**

Para poder dar respuesta a esta pregunta necesitamos saber si la frecuencia de aparición de algún tipo de actividad está relacionada con el tipo de episodio SEVERE. Por tanto, la dimensión sería el tipo de actividad y el tipo de episodio y la medida sería el número de ocurrencias de cada tipo de actividad para cada tipo de episodio. Posteriormente, para este caso se filtraría por el tipo de episodio SEVERE.

Si solo se utilizan las dimensiones mencionadas, los datos se relacionarán en su totalidad. Se podría meter la fecha como dimensión para, posteriormente,

ofrecer una herramienta con más posibilidades de análisis. Por ejemplo, comprobar cómo puede variar de un año para otro la relación entre actividades y episodios. Por tanto, el año podría ser una dimensión.

- **¿Se puede establecer algún tipo de relación entre los valores de los diferentes estados de ánimo y los episodios de crisis?**

Para dar respuesta a esta pregunta se necesita saber qué estado de ánimo es el más frecuente para cada tipo de episodio. Por tanto, el tipo de estado de ánimo y el tipo de episodio son dimensiones y la frecuencia de aparición de cada estado de ánimo para cada episodio es la medida.

- **¿Estas relaciones son iguales para cualquiera de las enfermedades o, en cambio, hay relaciones más acusadas por alguna de ellas?**

Para dar respuesta a esta pregunta será necesario analizar las relaciones para cada tipo de enfermedad. Por tanto, el tipo de desorden cognitivo será una nueva dimensión que añadir. No obstante, hay que tener en cuenta que esta información está relacionada con el paciente y no con el registro diario de la información relativa al paciente.

- **¿Se puede establecer alguna relación en nivel geográfico, por ejemplo, entorno urbano rural?**

Al igual que en la anterior pregunta, si lo que se desea es tener información relacionada con el entorno, el nivel geográfico tendrá que ser también dimensión. En este caso también hay que tener en cuenta que esta dimensión se relaciona con el paciente y no con la información adquirida diariamente sobre él.

- **¿Cuál ha sido la evolución de los diferentes pacientes a lo largo del tiempo?**

Para dar respuesta a esta pregunta se podría analizar la frecuencia de aparición de cada tipo de episodio a lo largo del tiempo. Para ello, se necesitaría tener el tipo de episodio como dimensión y una dimensión de tiempo.

Puesto que en este caso se tienen registros a nivel de día y de un solo año, el mes podría ser una buena opción de granularidad para definirlo como dimensión.

Asimismo, puesto que la idea es que el sistema sea capaz de seguir aceptando más datos, tiene que ser capaz de gestionar datos de otros años, y por tanto, para que no se solapen por meses, sería necesario tener la dimensión año.

- **¿Se puede establecer alguna relación entre los episodios de crisis y el momento del día o de la semana o del año?**

Para establecer este tipo de relación se necesitará medir la frecuencia de aparición de cada tipo de episodio en una dimensión de tiempo. Como dimensión de tiempo existen diferentes posibilidades: día de la semana, número de semana del año, estación de tiempo, mes, cuatrimestre, semestre, año. No obstante, no se podrá saber en qué momento del día porque el registro se realiza diariamente y, por tanto, esto define cuál es la granularidad más pequeña.

- **¿La realización de actividades físicas mejora o empeora el estado de ánimo de los pacientes?**

Para obtener esta información se necesita tener relacionado la frecuencia de aparición de cada estado de ánimo para cada tipo de actividad. Por tanto, el tipo de actividad y el tipo de estado de ánimo son dimensión del modelo de datos.

- **¿Hay algún tipo de actividad que mejore el día a día de los pacientes?**

Para poder saber si un paciente ha mejorado se tendrá que analizar si su número de episodios se ha reducido o si al menos la gravedad ha ido disminuyendo. Por tanto, se necesitará tener como dimensión el tipo de episodio, la actividad y uno de tipo tiempo, ya que se quiere analizar una evolución en el tiempo. En este caso la medida será el número de episodios para cada tipo de actividad en una granularidad de tiempo (año, mes, semana)

5.2 Descripción del modelo de datos

Teniendo en cuenta el análisis del apartado anterior, las dimensiones principales obtenidas son: tipo actividad, tipo episodio, tipo estado de ánimo, fecha y paciente. Asimismo, la dimensión paciente tiene subdimensiones como son el tipo de enfermedad que tiene, la ciudad en la que vive y el entorno en el que vive.

En este contexto, los hechos hacen referencia al registro analítico diario del paciente. Por tanto, siguiendo los patrones estándar de diseño de modelo de datos para *data warehouse* en sistemas de B*usiness intelligence* se ha diseñado un modelo con estructura de copo de nieve, si bien se podría resumir a estructura estrella simplificando la dimensión paciente.

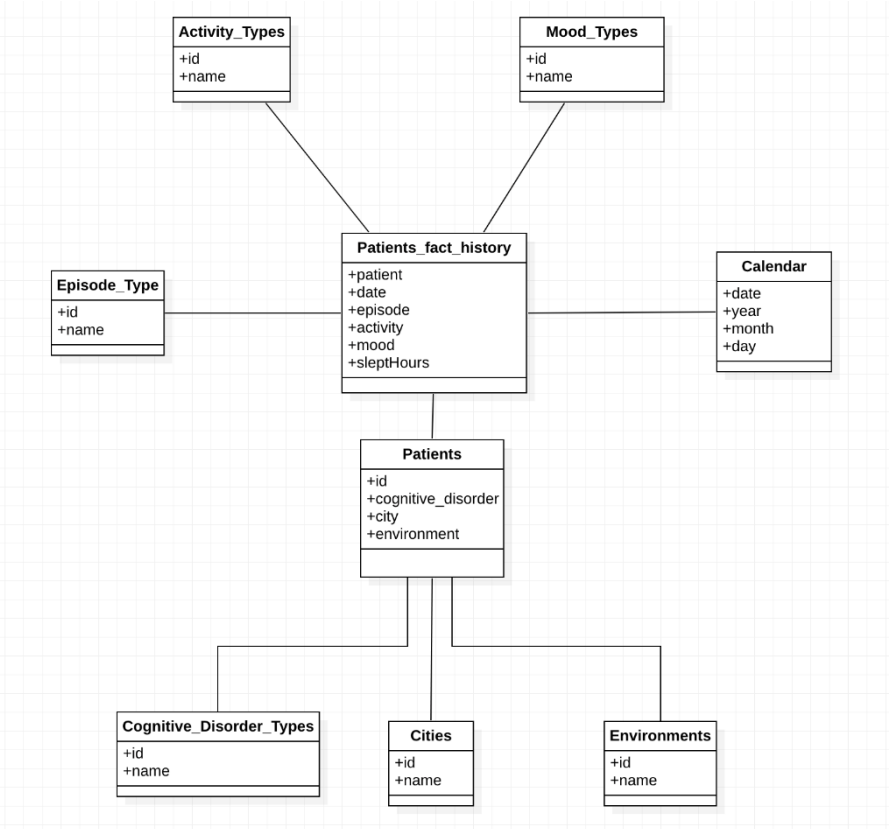


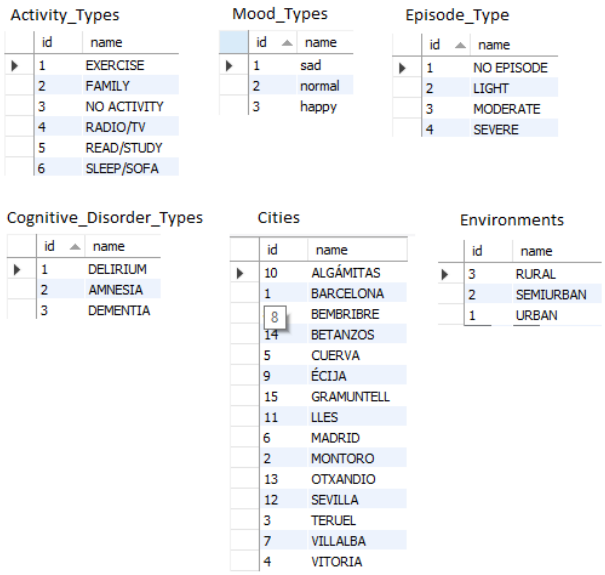
Imagen 5. Modelo de datos

En la imagen 5 se muestra el diagrama de entidades del modelo de datos diseñado. Cada entidad representa cada una de las tablas creadas en la base de datos.

- **Patient_fact_history**. Esta es la tabla de hechos donde se almacenan los datos del estudio realizado en granularidad diaria.
- **Episode_Types**. Esta tabla almacena información de los diferentes tipos de episodio que pueden sufrir los pacientes.
- **Activity_Types**. Esta tabla almacena información de las diferentes actividades que pueden realizar los pacientes.
- **Mood_Types**. Esta tabla almacena información de los diferentes estados de ánimo en los que el paciente puede estar.
- **Calendar**. Esta tabla almacena el calendario desde la fecha de inicio configurado hasta la fecha de la última ejecución del *workflow* de procesos
- **Patients**. Esta tabla almacena información de los pacientes. Dicha tabla está compuesta por columnas que dependen de otras tablas, ya que esta es la forma en la que se representa la jerarquía de dimensiones.
- **Cognitive_Disorder_Types.** Esta tabla almacena información de las diferentes enfermedades de trastornos cognitivos que puedan existir.
- **Cities.** Esta tabla almacena información de las ciudades en las que se ha realizado el estudio y donde los pacientes viven.
- **Environments.** Esta tabla almacena información de los diferentes entornos en los que un paciente puede vivir.

En el Anexo 2 se detallan las sentencias SQL para la creación de la base de datos. No obstante, tal y como se ha diseñado la solución, las tablas de dimensiones están *a priori* pobladas, a excepción de las tablas Patients y Calendar, que se van poblando con los datos obtenidos del fichero Excel y con las iterativas ejecuciones, respectivamente.

Por lo tanto, para las siguientes dimensiones la base de datos contiene el siguiente contenido.



Activity_Types

id	name
1	EXERCISE
2	FAMILY
3	NO ACTIVITY
4	RADIO/TV
5	READ/STUDY
6	SLEEP/SOFA

Mood_Types

id	name
1	sad
2	normal
3	happy

Episode_Type

id	name
1	NO EPISODE
2	LIGHT
3	MODERATE
4	SEVERE

Cognitive_Disorder_Types

id	name
1	DELIRIUM
2	AMNESIA
3	DEMENTIA

Cities

id	name
10	ALGÁMITAS
1	BARCELONA
8	BEMBRIBRE
14	BETANZOS
5	CUERVA
9	ÉCIJA
15	GRAMUNTELL
11	LLES
6	MADRID
2	MONTORO
13	OTXANDIO
12	SEVILLA
3	TERUEL
7	VILLALBA
4	VITORIA

Environments

id	name
3	RURAL
2	SEMIURBAN
1	URBAN

Imagen 6. Inicialización de las tablas de dimensiones en la base de datos

6 Procesos ETL

Los procesos ETL son una parte importante de la integración de los datos en los sistemas BI, ya que son los encargados de adaptar los datos a un modelo de datos optimizado para su posterior análisis. La palabra ETL corresponde a las siglas de las palabras en inglés *extract* (extraer)*, transform* (transformar) *y load* (cargar). Es decir, todo proceso ETL consta de estas tres fases:

- En la fase de extracción se obtienen los datos, se analizan y se verifica, y en caso necesario, se convierten a un formato que cumpla con la estructura esperada.
- En la fase de transformación se aplican funciones a los datos extraídos para que sean cargados
- En la fase de carga se almacenan los resultados de las transformaciones en un sistema de almacenamiento

Para la solución que se plantea en este proyecto y, teniendo en cuenta el modelo de datos diseñado, se ha optado por crear un flujo de procesos, denominado *job* o *workflow*, donde cada uno de ellos tiene un objetivo en concreto y se tienen que ejecutar en un orden concreto dependiendo de la jerarquía de dimensiones. Para la implementación del flujo y los procesos se ha utilizado Pentaho Data Integration (PDI).

Imagen 7. Workflow de los process ETL

1. **Start**. En este elemento se configura cómo se quiere que se ejecute la tarea: de forma iterativa, una sola ejecución, etc.
2. **Set variable**. En este elemento se configuran las variables que se parametrizan para todo el sistema.
    a. errorFile: ruta hasta la raíz de la carpeta de log de errores.
    b. filePath: ruta del fichero de datos de donde se van a extraer los datos.

c. rootPath: ruta del directorio de las transformaciones.

3. **Calendar**. Proceso que puebla la dimensión de calendario.

4. **Patient**. Proceso ETL que carga la información de los pacientes en su correspondiente dimensión. Requiere que las dimensiones de ciudad, entorno y demencia cognitiva estén pobladas.

5. **Slept Hours**. Proceso ETL que carga la información de las horas dormidas cada día por cada paciente en la tabla de hechos. Requiere que las dimensiones dependientes paciente y fecha estén cargadas.

6. **Mood**. Proceso ETL que carga la información del estado de ánimo de cada día por cada paciente en la tabla de hechos. Requiere que las dimensiones dependientes paciente y fecha estén cargadas.

7. **Episode**. Proceso ETL que carga la información diaria de los episodios sufridos por cada paciente en la tabla de hechos. Requiere que las dimensiones dependientes paciente y fecha estén cargadas.

8. **Activity**. Proceso ETL que carga la información de la actividad diaria realizada por cada paciente en la tabla de hechos. Requiere que las dimensiones dependientes paciente y fecha estén cargadas.

6.1 Proceso Calendar

Imagen 8. Proceso de población del Calendario

Este proceso puebla la tabla Calendar en cada ejecución con las nuevas fechas hasta el día de la ejecución.

1. Obtener el nuevo rango de generación de fechas. Se obtiene la última fecha guardada en la dimensión del calendario, en caso de no tener ninguna tupla se toma como fecha de inicio la constante '01/01/2016'. Como fecha fin se estable el día de la ejecución del proceso.

2. Generar tuplas con fechas diarias en el rango establecido.

3. Obtener la fecha.

4. A partir de la fecha generar una nueva tupla con los campos de tiempo *year, month, day,* etc.

5. Añadir la fecha a la nueva tupla (*date*, *year*, *month*, *day*...).

6. Dar formato a los campos.

7. Insertar tupla en la base de datos.

6.2 Proceso Patient

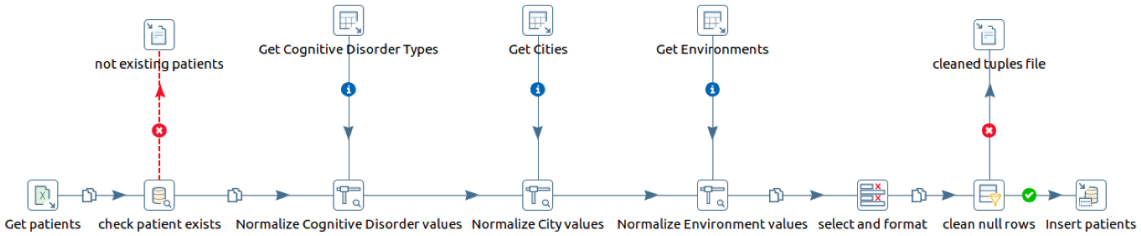


Imagen 9. Proceso ETL de información de pacientes.

Este proceso almacena información sobre los pacientes obtenida del fichero Excel. Este proceso requiere que las tablas Cities, Environments y Cognitive_Disorder_Types estén pobladas.

1. Obtener los datos de los pacientes del fichero Excel.
2. Verificar si el paciente ya existe.
3. Normalizar los nombres de los tipos de desorden cognitivo sustituyéndolos por los identificadores.
4. Normalizar los nombres de las ciudades sustituyéndolos por los identificadores.
5. Normalizar los nombres de los tipos de entorno sustituyéndolos por los identificadores.
6. Seleccionar y dar formato a los campos.
7. Verificar que, tanto el identificador del paciente, como el resto de los campos no son nulos.
8. Insertar la tupla en la tabla Patients de la base de datos.

6.3 Proceso Slept Hours

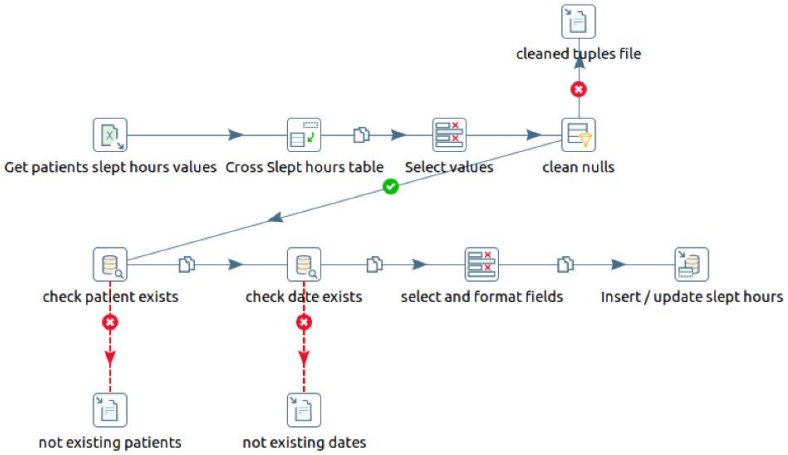


Imagen 10. Proceso ETL de información sobre horas de sueño

Este proceso almacena en la tabla de hechos los registros obtenidos y almacenados en el fichero Excel sobre las horas diarias dormidas por cada paciente.

1. Obtener los datos de horas dormidas de los pacientes del fichero Excel.
2. Cruzar la información de la tabla obteniendo como campos de salida fecha, paciente y horas dormidas.
3. Seleccionar los campos.
4. Verificar que los campos no son nulos.
5. Verificar que el identificador del paciente existe en la tabla Patients
6. Verificar que la fecha existe en la tabla Calendar.
7. Dar formato a los campos.
8. Insertar o actualizar la tupla de la tabla de hechos con los valores de horas dormidas.

6.4 Proceso Mood

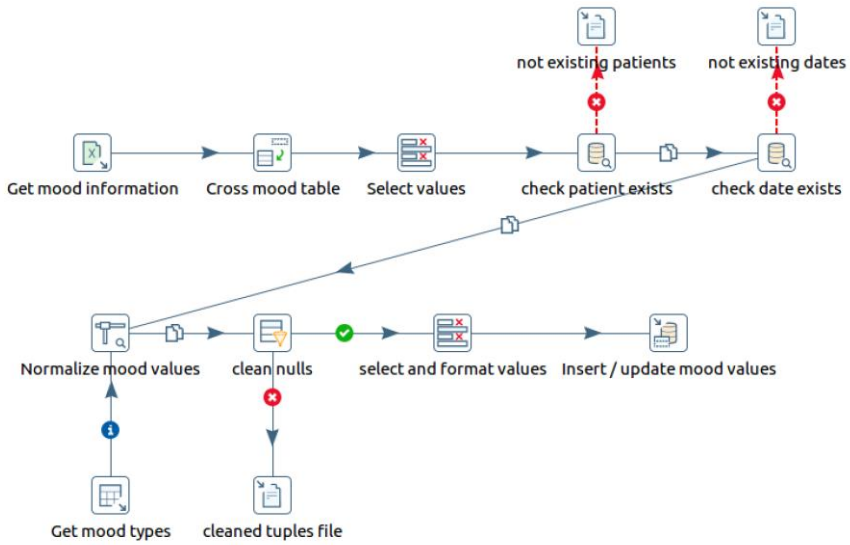


Imagen 11. Proceso ETL de información de estado de ánimo

Este proceso almacena en la tabla de hechos los registros diarios obtenidos y almacenados en el fichero Excel sobre el estado de ánimo de cada paciente.

1. Obtener los datos de estado ánimo de los pacientes del fichero Excel.
2. Cruzar la información de la tabla obteniendo como campos de salida fecha, paciente y estado de ánimo.
3. Seleccionar los campos.
4. Verificar que el identificador del paciente existe en la tabla Patients.
5. Verificar que la fecha existe en la tabla Calendar.
6. Normalizar los nombres de los estados de ánimo sustituyéndolos por los identificadores.
7. Verificar que los campos no son nulos.
8. Dar formato a los campos.
9. Insertar o actualizar la tupla de la tabla de hechos con los valores de estado de ánimo.

6.5 Proceso Episode

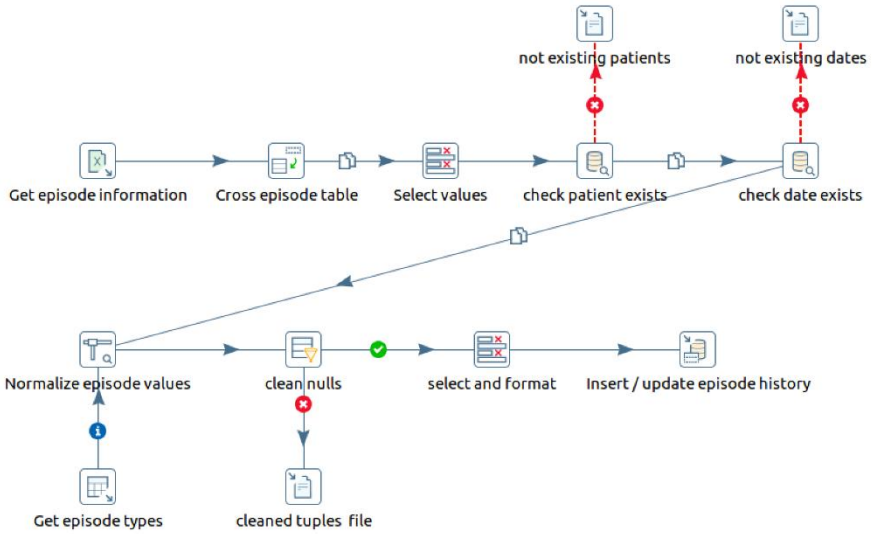


Imagen 12. Proceso ETL de información de episodios sufridos

Este proceso almacena en la tabla de hechos los registros diarios obtenidos y almacenados en el fichero Excel sobre los episodios de cada paciente.

1. Obtener los datos de los episodios de los pacientes del fichero Excel.
2. Cruzar la información de la tabla obteniendo como campos de salida fecha, paciente y episodio.
3. Seleccionar los campos.
4. Verificar que el identificador del paciente existe en la tabla Patients.
5. Verificar que la fecha existe en la tabla Calendar.
6. Normalizar los nombres de los episodios sustituyéndolos por los identificadores.
7. Verificar que los campos no son nulos.
8. Dar formato a los campos.
9. Insertar o actualizar la tupla de la tabla de hechos con los valores de episodio.

6.6 Proceso Activity

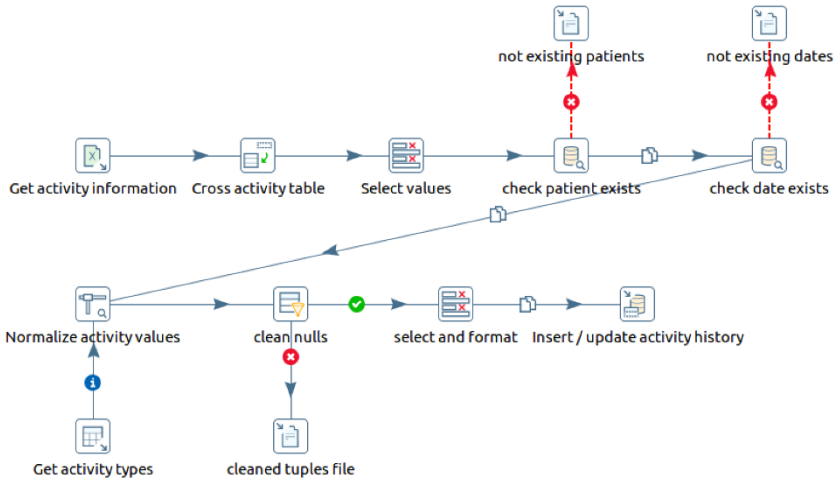


Imagen 13. Proceso ETL de actividades realizadas

Este proceso almacena en la tabla de hechos los registros diarios obtenidos y almacenados en el fichero Excel sobre la actividad realizada por cada paciente.

1. Obtener los datos de las actividades de los pacientes del fichero Excel.
2. Cruzar la información de la tabla obteniendo como campos de salida fecha, paciente y actividad.
3. Seleccionar los campos.
4. Verificar que el identificador del paciente existe en la tabla Patients.
5. Verificar que la fecha existe en la tabla Calendar.
6. Normalizar los nombres de las actividades sustituyéndolos por los identificadores.
7. Verificar que los campos no son nulos.
8. Dar formato a los campos.
9. Insertar o actualizar la tupla de la tabla de hechos con los valores de actividad.

7 Exploración de los datos. Cuadros de mandos

Imagen 14. Pantalla de cuadros de mando

Un cuadro de mando o un *dashboard* es un panel que visualiza los KPI más importantes del ámbito de análisis de los datos.

Este tipo de visualizaciones permiten ser más agiles en la toma de decisiones y en la toma de conclusiones y facilita la analitica de los datos, ya que los datos se representan de forma resumida.

Asimismo, este tipo de paneles suelen tener diferentes filtros de elementos, generalmente suelen ser los elementos dimensionales del modelo de datos, que interactúan de forma dinámica con los elementos visuales.

En la solución desarrollada, se ha implementado un menú de selectores multivalor que permiten filtrar los datos por rango de fecha, año, mes, día, paciente, ciudad, enfermedad, entorno, actividad, episodio y estado de ánimo. Por defecto, cuando la pantalla se carga se filtran los datos por el rango de fecha del último mes.

Asimismo, en el cuerpo del panel se han colocado elementos de visualización que responden a las preguntas planteadas. Estos elementos reaccionan mediante unos *listeners* que tienen para cuando los filtros cambian.

A continuación, se describen las visualizaciones implementadas con el *framework* Pentaho Dashboard Editor y las conclusiones que se han podido obtener para cada una de las preguntas.

7.1 ¿Cuál es la relación entre las actividades realizadas y los episodios de crisis graves?

Imagen 15. Cuadros de mando para la representación de la relación entre actividades y episodios

Para la representación de la relación entre las actividades realizadas y los episodios se han utilizado tres tipos de elementos de visualización:

- Gráfica de barras porcentual apilada. Con este elemento de visualización se puede ver cada actividad en qué tipo de episodio tiene más impacto.

- Gráfica de barras. Con este elemento de visualización se representa la cantidad de los diferentes tipos de episodios que han ocurrido para cada tipo de actividad que se ha realizado.

- Gráfica circular. Con este elemento de visualización se representa la cantidad de episodios han ocurrido para cada tipo de actividad. A su vez se puede ver la medida de forma porcentual.

```sql
SELECT P2.episode AS episode, P2.activity AS activity, IF (P1.no_apariciones IS NULL, 0, P1.no_apariciones) AS num_days
FROM
    (
    SELECT P_F_H.episode AS idEpisode, P_F_H.activity AS idActivity, COUNT(P_F_H.date) AS no_apariciones
    FROM
    patients_fact_history P_F_H
    JOIN patients P ON P_F_H.patient = P.id
    JOIN calendar C ON P_F_H.date = C.date
    WHERE
    ((('-1' IN (${patient_param})) OR P_F_H.patient IN (${patient_param})) AND
    ((-1 IN (${episode_param})) OR P_F_H.episode IN (${episode_param})) AND
    ((-1 IN (${activity_param})) OR P_F_H.activity IN (${activity_param})) AND
    ((-1 IN (${mood_param})) OR P_F_H.mood IN (${mood_param})) AND
    ((-1 IN (${city_param})) OR P.city IN (${city_param})) AND
    ((-1 IN (${cognitive_disorder_param})) OR P.cognitive_disorder IN (${cognitive_disorder_param})) AND
    ((-1 IN (${environment_param})) OR P.environment IN (${environment_param})) AND
    ((-1 IN (${year_param})) OR C.year IN (${year_param})) AND
    ((-1 IN (${month_param})) OR C.month IN (${month_param})) AND
    ((-1 IN (${day_param})) OR C.day IN (${day_param})) AND
    P_F_H.date >= ${start_param} AND P_F_H.date <= ${end_param})
    GROUP BY P_F_H.activity, P_F_H.episode
    ) AS P1
RIGHT JOIN
    (
    SELECT idEpisode, episode, idActivity, activity
    FROM (
        (
            SELECT E_T.id as idEpisode,E_T.name as episode
            FROM episode_types E_T
            WHERE
                ((-1 IN (${episode_param})) OR E_T.id IN (${episode_param}))
        ) E_F_T
        CROSS JOIN
        (
            SELECT A_T.id as idActivity,A_T.name as activity
            FROM activity_types A_T
            WHERE
                ((-1 IN (${activity_param})) OR A_T.id IN (${activity_param}))
        ) A_F_T
    )
    ) AS P2
ON P1.idEpisode = P2.idEpisode AND P1.idActivity = P2.idActivity
ORDER BY P2.episode ASC, P2.activity ASC
```

Imagen 16. Consulta SQL para representar la relación entre actividades y episodios de forma interactiva con los filtros

Las gráficas tienen carga de datos dinámica mediante el filtrado de dimensiones. Para dar respuesta a la pregunta que se plantea hay que filtrar los datos por fecha y tipo de episodio “severe”.

Como se puede observar en la siguiente imagen los episodios críticos están relacionados con la no realización de actividades, el estímulo de la “radio/tv” y la inactividad mediante el “dormir/sofa”.

Imagen 17. Cuadros de mando de relación entre actividades y episodios con filtro “severe”

7.2 ¿Se puede establecer algún tipo de relación entre los valores de los diferentes estados de ánimo y los episodios de crisis?

Imagen 18. Cuadros de mando para la representación de la relación entre estados de ánimo y episodios

Para la representación de la relación entre los diferentes estados de ánimo y los episodios se han utilizado dos tipos de elementos de visualización:

- Gráfica de barras vertical. Con este elemento de visualización se representa la cantidad de los diferentes tipos de episodios que han ocurrido para cada tipo de estado de ánimo.

- Gráfica de barras apilada horizontal. Con este elemento de visualización se puede ver qué estado de ánimo está más relacionado con cada tipo de episodio.

Con los datos que tenemos cargados y seleccionando como rango de fecha el año 2016, en la gráfica de barras horizontal se puede observar que para todo tipo de episodios no hay un estado de ánimo que destaque y se pueda relacionar un impacto directo. Asimismo, en la gráfica de barras verticales se puede observar que independientemente del estado de ánimo, se tienden a tener más episodios *light* o directamente a no tener episodios.

7.3 ¿Estas relaciones son iguales para cualquiera de las enfermedades o, en cambio, hay relaciones más acusadas por alguna de ellas?

Imagen 19.Cuadros de mando para la representación de las relaciones entre actividades, episodios, horas de sueño y el tipo de desorden cognitivo

Para analizar si las relaciones impactan de una forma más directa en un tipo de enfermedad en concreto se ha utilizado el elemento de visualización *mapa de calor*.

Este elemento de visualización interactúa con los filtros de del menú de dimensiones. No obstante, para la visualización se utilizan las dimensiones tipo de actividad, tipo de enfermerdad y tipo de episodio y las medidas cantidad de episodios y número de horas dormidas, representadas respectivamente por un degradado de color (cuanto más oscuro más episodios) y por el tamaño del circulo.

También se ha utilizado una gráfica de barras apilada para poder analizar si el tipo de enfermedad está relacionado con los tipos de episodios.

Para dar respues a la pregunta que se plantea en este punto y, teniendo en cuenta el rango de fecha seleccionado, se puede observar que proporcionalmente la cantidad de episodios *severe* es muy parecida para todos los tipos de enfermedad. Asimismo, se puede observar que los pacientes con *dementia* y *delirium* tienden a no tener episodios, pero cuando los tienen, suelen ser light. También se observa que que los que tienen *amnesia* tienden a tener más episodios de tipo *moderate*.

7.4 ¿Se puede establecer alguna relación en nivel geográfico, por ejemplo, entorno urbano o rural?

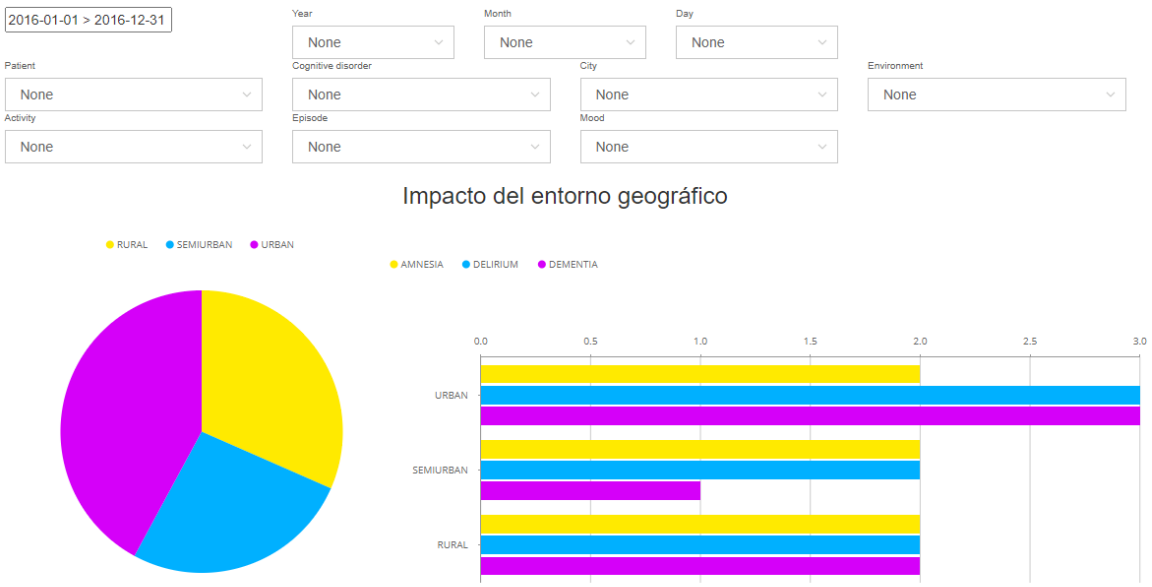


Imagen 20. Cuadros de mando para la representación el impacto del entorno geográfico

Para poder establecer algún tipo de relación a nivel geográfico se han utilizado dos elementos de visualización

- Gráfica de barras. Con este elemento de visualización se relaciona el tipo de enfermedad con el entorno geográfico mediante la cantidad de pacientes que hay.

- Gráfica circular. Con este elemento de visualización se puede observar si el tipo de entorno puede incidir en tener la enfermedad.

Si se observan los resultados obtenidos, se puede ver que en el entorno urbano hay especialmente más pacientes y que sobre todo tienen el tipo *demencia* o *delirio*.

7.5 ¿Cuál ha sido la evolución de los diferentes pacientes a lo largo del tiempo?

Para poder analizar la tendencia a lo largo del tiempo de los diferentes pacientes se ha utilizado una gráfica de lineas que interactua con los elementos de filtrado *episodio, fecha* y *paciente* entre otros y que representa la cantidad episodios que ha habido. La granularidad de tiempo escogida ha sido el mes.

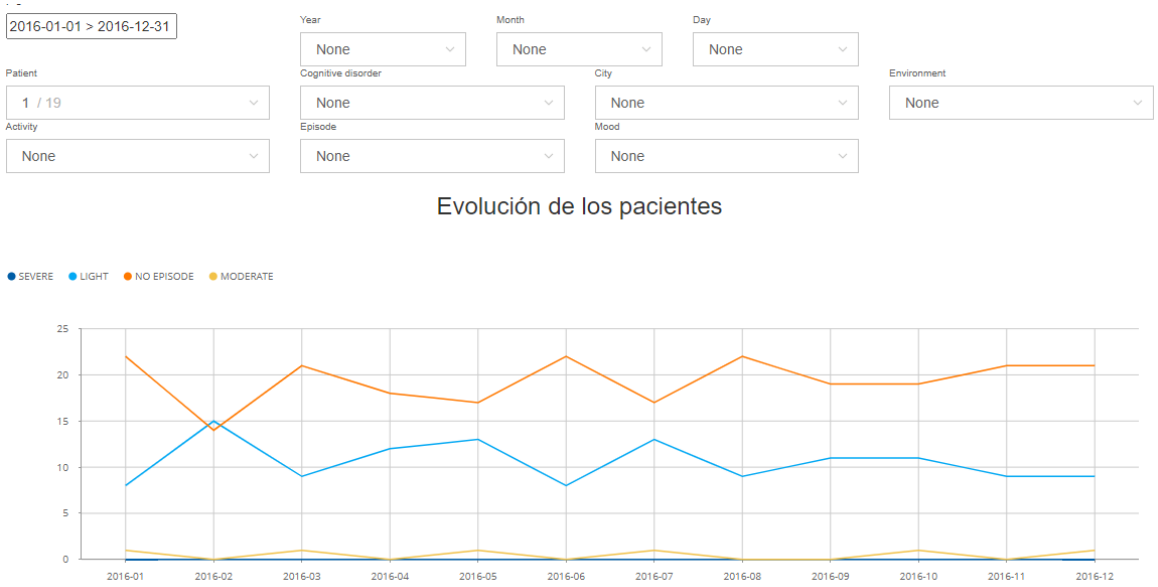


Imagen 21. Cuadro de mando para la representación de la evolución del paciente

En concreto en la imagen anterior se ha filtrado la información para el paciente P5 en el rango de fechas del año 2016. Se puede observar que sigue sin tener episodios *severos* y que aproximadamente cada dos meses sufre algunos de tipo *moderate*. Es un paciente que aproximadamente un 75% del mes suele estar sin episodios.

7.6 ¿Se puede establecer alguna relación entre los episodios de crisis y el momento del día o de la semana o del año?

Imagen 22. Cuadro de mando para la representación de la evolución de los episodios en el tiempo

Tal y como se puede observar, el mismo elemento de visualización de la consulta anterior puede ser utilizada para representar la tendencia general. Para ello, hay que seleccionar el rango de fecha y no filtrar por paciente.

En cuanto al resultado obtenido, se puede observar que en el mes de agosto del año 2016 hubo un pico de episodios *severe*, en octubre hubo un pico de episodios *light* y que a finales del mismo año la tendencia de episodios *moderate* estaba al alza.

7.7 ¿La realización de actividades físicas mejora o empeora el estado de ánimo de los pacientes?

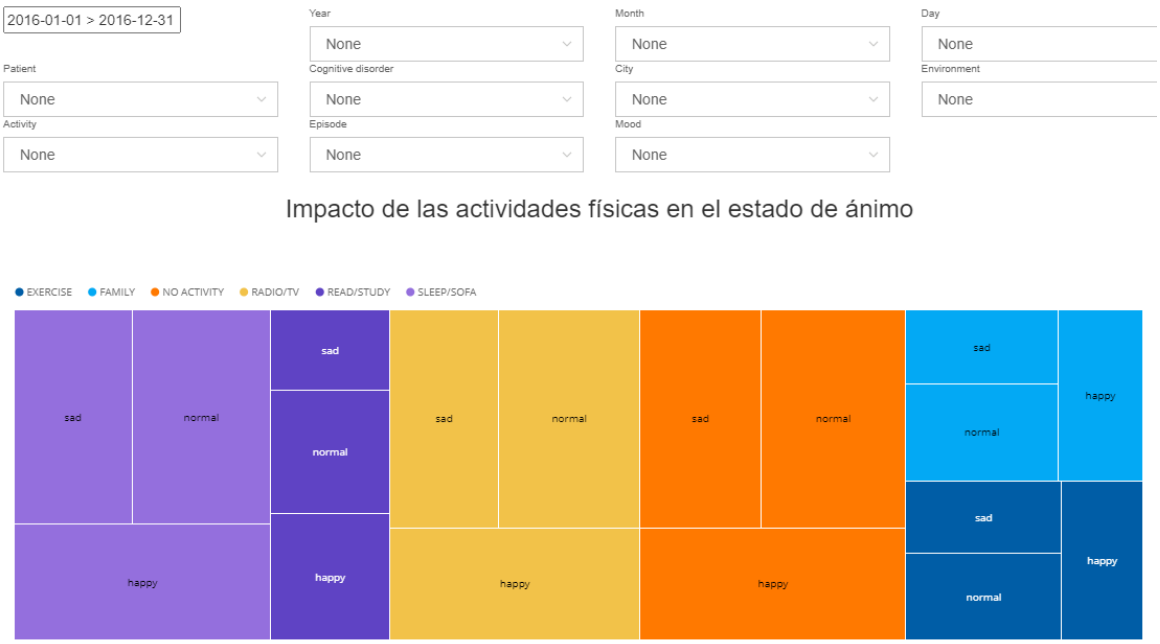


Imagen 23. Cuadro de mando para la representación de la relación entre las actividades y el estado de ánimo

Para poder analizar el impacto de las actividades físicas se ha utilizado un elemento de visualización *treemap*, donde se representa la jerarquía de dimensiones actividades > estado de ánimo. De está forma, se puede analizar qué actividad está relacionada más con los episodios (se representan las porciones con diferentes colores) y, a su vez, ver cómo afecta cada actividad en el estado de ánimo mediante los recuadros más pequeños.

Este elemento de visualización al igual que el resto está relacionado con los filtros de fecha, paciente, tipo de actividad, tipo de episodio, etc.

Tal y como se puede observar en la imagen anterior, las actividades que más hacen es la no realización de actividad, *dormir/sofá* y *radio/tv*. Asimismo, si se aplica el filtro *exercise*, tal y como se ve en la siguiente imagen, se puede observar que la realización de actividad física mejora el estado de ánimo, ya que se tiende a tener un estado *normal* o *happy*.

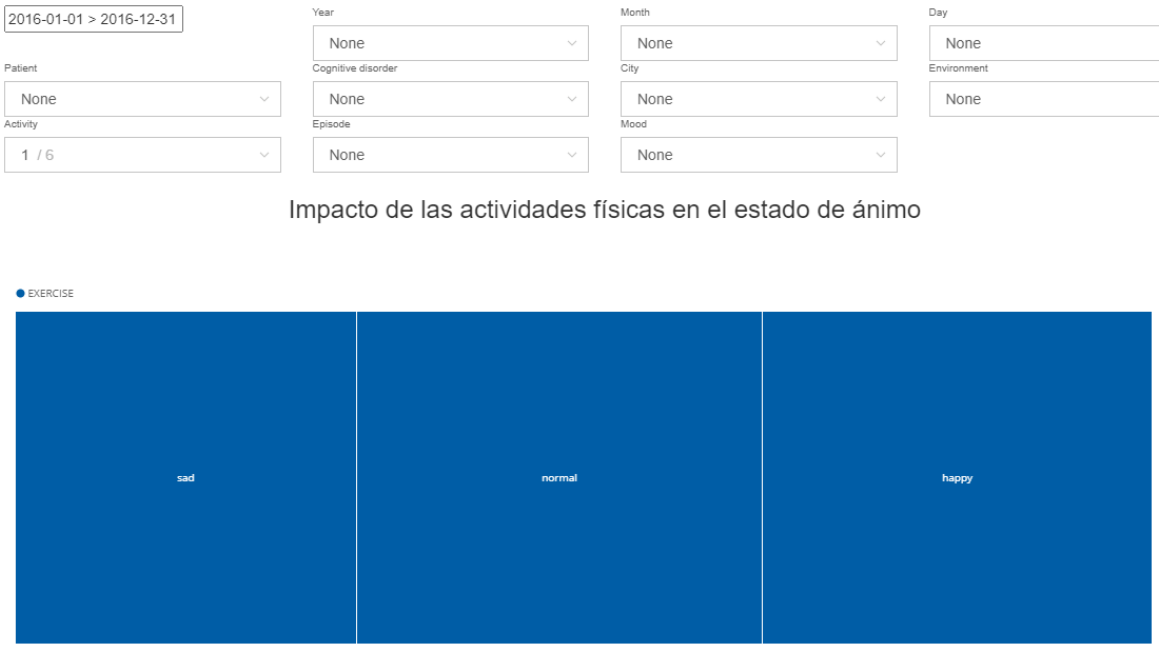


Imagen 24. Cuadro de mando para la representación de la actividad física y el estado de ánimo

7.8 ¿Hay algún tipo de actividad que mejore el día a día de los pacientes?

Para dar respuesta a esta pregunta se reutilizan las visualizaciones de la consulta 7.1, ya que sin aplicar el filtro del tipo de episodio *severe*, las propias visualizaciones muestran la relación entre las actividades y los episodios.

Tal y como se puede observar en la imagen 15, cualquiera de las actividades es positiva y ayuda a que no haya episodios o a que los episodios sean *light*. No obstante, algunas actividades como *read/study*, el compartir tiempo con la familia y hacer ejercicio ayudan que el día a día de los pacientes mejore.

Asimismo, utilizan los elementos de visualización de la consulta 7.7, se puede observar que no hay una actividad que directamente haga que su estado de ánimo sea mejor.

8 Conclusiones

Finalizado este proyecto, se ha conseguido el objetivo principal del mismo: obtener una herramienta de análisis de trastornos cognitivos. La implementación se ha llevado a cabo con tecnologías en auge dentro del contexto del *software* libre, mediante una metodología tradicional de la gestión de proyectos, donde a partir de la definición de un cuestionario se han capturado los requisitos funcionales básicos e, incluso, han sido ampliados con filtros que interactúan con visualizaciones con el fin de dar más flexibilidad a la herramienta.

La primera parte del proyecto se ha dedicado al entendimiento del problema para la definición de los requisitos funcionales y el diseño y la implementación de la arquitectura y del modelo de datos. Para ello, ha sido necesario informarse sobre el contexto para poder entender mejor el cuestionario y la fuente de datos, las cuales han sido necesarias para la definición de la herramienta. Por otro lado, se ha hecho un análisis exhaustivo tanto de las cuestiones, como de la fuente de datos, que han servido de gran ayuda en el diseño del modelo de datos y las visualizaciones.

En cuanto al diseño de la arquitectura, puesto que no se tenía experiencia previa en esta tarea, se ha tenido como referencia la arquitectura general de los sistemas BI y se ha verificado que las necesidades del problema que se planteaba encajaban con esta arquitectura estándar. Tras su implementación y el desarrollo del producto se ha confirmado que ha sido una decisión acertada.

Asimismo, sin poder comparar con otras tecnologías porque no se tenía experiencia en este tipo de proyectos, se ha podido verificar que las tecnologías escogidas se han adaptado a la arquitectura y se han ajustado a las necesidades. No obstante, hay que destacar que una de las mayores dificultades que se ha tenido en el proyecto ha sido la instalación del entorno, ya que ha sido muy difícil encontrar documentación para la versión *community* 9.1 y la poca que se ha encontrado no estaba bien organizada o no era fácilmente accesible.

Además de esto, para tener un entorno de desarrollo limpio, el proyecto se comenzó sobre una máquina virtual Linux, donde Pentaho Data Integration se instaló correctamente, pero para la instalación del servidor se tuvieron problemas que no se consiguieron resolver. En concreto, el servidor de Tomcat se caía constantemente, sospecho que por falta de memoria de la máquina. Con esta problemática, opté por cambiar de entorno y tuve que continuar sobre un equipo de Windows. Posteriormente, para verificar que el problema que tenía no era una cuestión de haber hecho mal la instalación, probé sobre un equipo Linux (sin máquina virtual) y conseguí poner todo en marcha.

En cuanto a la fuente de datos, primero se verificó que los datos estaban normalizados por lo que no ha sido muy laborioso hacer la limpieza de datos. A pesar de ello, en los procesos ETL se han hecho las comprobaciones de datos y adaptaciones necesarias como medida de seguridad para garantizar el buen funcionamiento del sistema. El único cambio que se ha realizado en los datos antes de tratarlos en los procesos ETL ha sido el cambio del formato de la fecha de los valores de estado de ánimo: se ha cambiado de formato Excel a formato dd/mm/aaaa.

El diseño del modelo de datos ha sido acertado, ya que se cogió como referencia el modelo estrella, estructura comúnmente elegida para la creación de *data warehouse* que son explotados por herramientas de visualización analíticas de *Business Intelligence*, dado a la flexibilidad que permiten para el filtrado de datos. En cuanto a la tecnología escogida, MySQL ha sido perfectamente integrada con las demás tecnologías y puesto que ya se tenían conocimientos previos de su manejo, no ha supuesto ningún problema.

En la segunda parte del proyecto se ha llevado a cabo la parte de diseño e implementación de los procesos ETL y de las visualizaciones.

Para el proceso de transformación de los datos leídos de origen, se ha diseñado un workflow con diferentes procesos ETL. Todo esto ha sido desarrollado con el *framework* Pentaho Data Integration y teniendo en cuenta que la tecnología era desconocida, se ha adaptado bien a las necesidades por lo que ha sido una buena opción.

Para la parte visual, la tecnología utilizada ha sido Pentaho Dashboard Editor y con ello he creado una pantalla, donde en la parte de arriba hay un menú con los filtros y abajo los diferentes cuadros de mando que van dando respuesta a las consultas planteadas.

Tras el análisis de lo realizado en el proyecto, a continuación, se realiza una reflexión personal en cuanto a la experiencia y conocimiento adquirido, así como sobre las líneas futuras o mejoras del producto desarrollado.

8.1 Reflexión personal

Esta experiencia me ha permitido, aunque haya sido en un ámbito académico, responsabilizarme de un proyecto que planteaba una necesidad real.

La toma de decisiones en cuanto a la tecnología a utilizar, el diseño de la arquitectura y el modelo de datos, han sido unas de las tareas que más compleja me ha resultado, sobre todo por la falta de experiencia y conocimiento en el ámbito de sistemas de *Business Intelligence*.

En este sentido, creo que la decisión que tomé en cuanto la arquitectura, y el modelo de datos ha sido acertada. No obstante, creo que en la arquitectura debía haber metido una capa de modelo de datos lógico con implementación de cubos OLAP para que fuera más eficiente y para facilitar la implementación de los cuadros de mandos para que reaccionen con los filtros, ya que las consultas

SQL que he tenido que implementar han sido muy laboriosas. A pesar de que el PMBOK dice que los cambios tienen un mayor coste según avanzan en el proyecto, decidí terminar la implementación porque lo tenía muy avanzado y no podía atrasar la fecha de finalización del proyecto. En este caso, cuando se quiera mejorar incorporando el sistema de cubos OLAP habrá que desarrollar el esquema con Mondrian y rehacer las consultas a los datos.

En cuanto a conocimiento técnico-tecnológico adquirido ha sido una experiencia muy enriquecedora. Las tecnologías utilizadas eran desconocidas, por lo que ha supuesto su aprendizaje de forma autodidacta sin tener un buen soporte documental.

Pentaho Data Integration me ha parecido una herramienta muy intuitiva y que facilita mucho el desarrollo de procesos ETL. Asimismo, a pesar de que Pentaho Dashboard Editor ha dado solución a las necesidades presentadas, su usabilidad no me ha parecido muy práctica y me ha costado mucho encontrar documentación de las diferentes propiedades que ofrecen cada tipo de visualización.

La tarea de diseño del modelo de datos me ha servido para aprender sobre otras estructuras del modelo relacional. Los conceptos de modelo en estrella o en copo no los conocía y me han parecido muy interesantes para las aplicaciones de analítica de datos.

En cuanto al cumplimiento de la planificación, se puede decir que no ha habido cambios ni desviaciones significativas. El único cambio que ha habido es el adelanto de la fecha de la entrega final que al principio era el 22 de enero de 2021 y la única desviación que ha habido ha obedecido a los problemas que, como ya he comentado anteriormente, he tenido con la instalación del entorno.

Asimismo, en cuanto a la metodología tradicional de gestión de proyectos llevada a cabo, creo que ha sido adecuada, ya que la planificación estaba detallada desde el comienzo del proyecto y no era un contexto donde el cliente iba a ir definiendo los requisitos mediante el *feedback* y el avance del proyecto, donde una metodología ágil hubiera sido mejor.

Por la parte académica, en cuanto al seguimiento del proyecto mediante las entregas de las PEC creo que es una metodología de supervisión que ayuda desde un comienzo a realizar las cosas de una forma correcta, ya que mediante el *feedback* del consultor da tiempo a hacer los cambios necesarios. Lo que me ha parecido muy importante es ir escribiendo la memoria poco a poco mediante las entregas, lo que ayuda a no dejar esta tarea para el final.

En definitiva, hago un balance positivo de la realización del proyecto. Sin duda, lo recomiendo como primera toma de contacto con el mundo de los sistemas BI y como aplicación de los conocimientos que se han ido adquiriendo en los estudios.

Para finalizar, mencionar algunas de las líneas de trabajo futuro:

- Como ya se ha mencionado anteriormente, introducir una capa de modelo lógico con cubos OLAP y adaptar las visualizaciones.

- Mejorar el aspecto y la usabilidad cambiando los colores por defecto y asignando un color a cada elemento, añadiendo títulos a los ejes, mejorando los textos de los elementos de visualización, así como utilizando otro tipo de calendario que permite directamente ir a la fecha deseada.

- Introducir nuevos datos para verificar que el sistema sigue siendo válido.

- Trabajar con el experto en el dominio para capturar nuevos datos y poder crear nuevas relaciones y enriquecer el modelo de datos. Por ejemplo, en lugar de categorizar los tipos de trastornos cognitivos en categorías genéricas, bajar a niveles más bajos y utilizar subcategorías.

9 Glosario

B

BI.

Véase Business Intelligence

Business Intelligence

Inteligencia de negocio. Es el conjunto de procesos requeridos para ofrecer una solución informática que nos permita analizar cierto ámbito del negocio.

C

Commodity hardware

Hardware básico. Dispositivos asequibles sin gran capacidad de procesamiento

Cubos OLAP

Base de datos multidimensionales

D

Data marts

Son los subconjuntos de datos de los almacenes de datos (data warehouse)

Data warehouse

Almacén de datos. Conjunto de datos orientada a un determinado ámbito

K

KPI

Acrónimo de key performance indicator. Conocido también como indicador clave o medidor de desempeño o indicador clave de rendimiento, es una medida del nivel del rendimiento de un proceso

10 Bibliografía

[1] ISES, «Trastornos cognitivos: Qué son y cómo nos afectan,» Instituto Superior de Estudios Sociales y Sociosanitarios, [En línea]. Available: https://www.isesinstituto.com/noticia/trastornos-cognitivos-que-son-y-como-nos-afectan. [Último acceso: 11 Enero 2021].

[2] Airbnb developers, «Superset,» Airbnb.io, [En línea]. Available: https://airbnb.io/projects/superset/. [Último acceso: 9 Enero 2021].

[3] Wikipedia, «Data Warehouse,» Wikipedia, [En línea]. Available: https://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos. [Último acceso: 17 Enero 2021].

[4] Hitachi, «CCC Playground,» [En línea]. Available: https://webdetails.github.io/ccc/. [Último acceso: 9 Enero 2021].

[5] Wrike, «Guía de Gestión de proyectos,» [En línea]. Available: https://www.wrike.com/es/project-management-guide/. [Último acceso: 9 Enero 2021].

[6] Project Management Institute, Guía de los fundamentos para la dirección de proyectos (Guía del PMBOK) (6ª edición), Newtown Square, Pennsylvania: Project Management Institute, Inc., 2017.

[7] Postgresql Tutorial, «PostgreSQL vs. MySQL,» Postgresql Tutorial, [En línea]. Available: https://www.postgresqltutorial.com/postgresql-vs-mysql/. [Último acceso: 9 Enero 2021].

[8] Wikipedia, «Esquema en copo de nieve,» Wikipedia, [En línea]. Available: https://es.wikipedia.org/wiki/Esquema_en_copo_de_nieve. [Último acceso: 17 Enero 2021].

[9] Wikipedia, «Esquema en estrella,» Wikipedia, [En línea]. Available: https://es.wikipedia.org/wiki/Esquema_en_estrella. [Último acceso: 17 Enero 2021].

[10] Escuela de negocios Feda, «Gestión ágil vs gestión tradicional de proyectos ¿cómo elegir?,» Escuela de negocios Feda, 20 Mayo 2019. [En línea]. Available: https://www.escueladenegociosfeda.com/blog/50-la-huella-de-nuestros-docentes/471-gestion-agil-vs-gestion-tradicional-de-proyectos-como-elegire. [Último acceso: 9 Enero 2021].

[11] Wikipedia, «Pentaho,» Wikipedia, [En línea]. Available: https://en.wikipedia.org/wiki/Pentaho. [Último acceso: 9 Enero 2021].

[12] Pentaho Community Forums, «Pentaho Community Forums,» Pentaho Community, [En línea]. Available: https://forums.pentaho.com/threads/161089-CCC-FAQ-Frequently-Asked-Questions-About-CCC/. [Último acceso: 9 Enero 2021].

[13] Wikipedia, «SpagoBI,» Wikipedia, [En línea]. Available: https://es.wikipedia.org/wiki/SpagoBI. [Último acceso: 9 Enero 2021].

[14] K. Hristozov, «MySQL vs PostgreSQL -- Choose the Right Database for Your Project,» Okta, [En línea]. Available: https://developer.okta.com/blog/2019/07/19/mysql-vs-postgres. [Último acceso: 9 Enero 2021].

Anexo 1. Manual de instalación Pentaho Community Edition 9.1

Preparar entorno

Crear usuario Pentaho (Sólo en Linux)

1. Crear un usuario administrador llamado “pentaho”

```
adduser pentaho
usermod -aG sudo username
```

2. Verificar que el nuevo usuario tiene permisos de lectura, escritura y ejecución en el *home directory*.

3. Verificar que el nuevo usuario tiene permisos de escritura en el directorio de instalación de Pentaho Server

Crear estructura de directorios

1. Entrar en la máquina

2. Crear los siguientes directorios

```
<your home directory>/pentaho/server
<your home directory>/pentaho/client-tools
<your home directory>/.pentaho
```

3. Verificar que los directorios tienen permiso de lectura, escritura y ejecución

Instalación de Java

Para esta versión de Pentaho Server necesitaremos Java 8.Instalar java y configurar la variable de entorno.

- Comandos de Linux

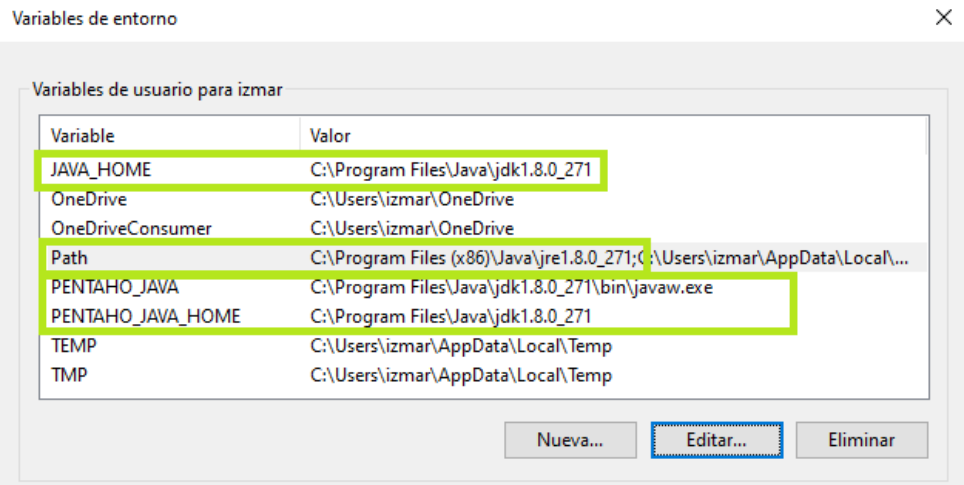
```
sudo apt-get install openjdk-8-jdk
export JAVA_HOME=/usr/lib/jvm/jdk8.x.
export PENTAHO_JAVA_HOME=/usr/lib/jvm/jdk8.x.
export PENTAHO_JAVA=/usr/lib/jvm/jdk8.x./
echo $JAVA_HOME
```

- Pasos en Windows.

    1. Descargar el ejecutable en el siguiente enlace: https://www.java.com/es/download/

    2. Ejecutar el instalador.

    3. Configurar las variables de entorno PENTAHO_JAVA, JAVA_HOME, PENTAHO_JAVA_HOME y Path



Instalación del Repositorio de información

En este caso se ha escogido MySQL, pero Pentaho soporta otros tipos de SGBD como PostgreSQL, Oracle, etc.

- Comandos de instalación de MySQL y Workbench, un cliente GUI de MySQL en Linux

```
sudo apt-get install mysql-server
sudo apt-get install mysql-workbenchecho
```

- Pasos de instalación para Windows.

    1. Descargar ejecutable en el siguiente enlace: https://dev.mysql.com/downloads/windows/installer/8.0.html

    2. Ejecutar el instalador

    3. Elegir la instalación de MySQL Server y Workbench

Descargar y descomprimir ficheros

1. Descarga de ficheros en el siguiente enlace: https://sourceforge.net/projects/pentaho/files/Pentaho%209.1/

2. Dentro de la web se encuentran los diferentes componentes descargaremos el servidor

    a. Descargar Pentaho Data Integration (PDI) de la carpeta client_tools

    b. Descargar el Pentaho Server de la carpeta server, la versión que no pone “manual”

3. Descomprimir

    a. Descomprimir Pentaho Data Integration en <your home directory>/pentaho/client-tools

    b. Descomprimir Pentaho Server en <your home directory>/pentaho/server

Arranque y configuración de Pentaho Server

1. Arranque del servidor:

    - Comandos para Linux

```
<your home directory>/pentaho/server/pentaho-server
sudo ./start_pentaho
```

    - Windows. Doble click en el ejecutable start_pentaho.bat de <your home directory>/pentaho/server/pentaho-server

2. Entrar por el navegador en http://localhost:8080/pentaho y se cargará Pentaho User Console, donde Pentaho viene incorporado y se puede comenzar a crear visualizaciones.

Arranque y configuración de Pentaho Data Integration

1. (Sólo en Linux) En el fichero README indica que es necesaria la instalación de la librería libwebkitgtk-1.0.0.

```
sudo apt-get installlibwebkitgtk-1.0.0
```

2. Configuración del controlador MySQL:

    ➢ Descargar JAR desde https://dev.mysql.com/downloads/connector/j/5.1.html

    ➢ Insertar el JAR en <your home directory>/pentaho/client-tools/data-integration/drivers

3. Arrancar PDI

    ➢ Comandos en Linux

```
<your home directory>/pentaho/client-tools/data-integration
sudo ./spoon
```

- Windows. Doble *click* en el ejecutable Spoon.bat de <your home directory>/pentaho/client-tools/data-integration

4. Añadir el Plugin desde la interfaz. Tool→marketplace→PDI MySQL Plugin

Anexo 2. SQL SCRIPT

Las siguientes instrucciones SQL son del script generado para la instalación de la base de datos de la herramienta implementada.

```sql
--
-- Table structure for table `Activity_Types`
--

CREATE TABLE `Activity_Types` (
  `id` int(11) NOT NULL,
  `name` varchar(45) NOT NULL,
  PRIMARY KEY (`id`),
  UNIQUE KEY `nombre_UNIQUE` (`name`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

--
-- Dumping data for table `Activity_Types`
--

LOCK TABLES `Activity_Types` WRITE;
/*!40000 ALTER TABLE `Activity_Types` DISABLE KEYS */;
INSERT INTO `Activity_Types` VALUES (1,'EXERCISE'),(2,'FAMILY'),(3,'NO ACTIVITY'),
(4,'RADIO/TV'),(5,'READ/STUDY'),(6,'SLEEP/SOFA');
/*!40000 ALTER TABLE `Activity_Types` ENABLE KEYS */;
UNLOCK TABLES;

--
-- Table structure for table `Calendar`
--

CREATE TABLE `Calendar` (
  `date` date NOT NULL,
  `year` int(11) DEFAULT NULL,
  `month` int(11) DEFAULT NULL,
  `day` int(11) DEFAULT NULL,
  PRIMARY KEY (`date`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

--
-- Table structure for table `Cities`
--

CREATE TABLE `Cities` (
  `id` int(11) NOT NULL,
  `name` varchar(45) NOT NULL,
  PRIMARY KEY (`id`),
  UNIQUE KEY `nombre_UNIQUE` (`name`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

--
-- Dumping data for table `Cities`
--

LOCK TABLES `Cities` WRITE;
/*!40000 ALTER TABLE `Cities` DISABLE KEYS */;
INSERT INTO `Cities` VALUES (10,'ALGÁMITAS'),(1,'BARCELONA'),(8,'BEMBRIBRE'),(14,'BETANZOS'),
(5,'CUERVA'),(9,'ÉCIJA'),(15,'GRAMUNTELL'),(11,'LLES'),(6,'MADRID'),(2,'MONTORO'),
(13,'OTXANDIO'),(12,'SEVILLA'),(3,'TERUEL'),(7,'VILLALBA'),(4,'VITORIA');
/*!40000 ALTER TABLE `Cities` ENABLE KEYS */;
UNLOCK TABLES;
```

```sql
--
-- Table structure for table `Cognitive_Disorder_Types`
--

CREATE TABLE `Cognitive_Disorder_Types` (
  `id` int(11) NOT NULL,
  `name` varchar(45) NOT NULL,
  PRIMARY KEY (`id`),
  UNIQUE KEY `nombre_UNIQUE` (`name`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;

--
-- Dumping data for table `Cognitive_Disorder_Types`
--

LOCK TABLES `Cognitive_Disorder_Types` WRITE;
/*!40000 ALTER TABLE `Cognitive_Disorder_Types` DISABLE KEYS */;
INSERT INTO `Cognitive_Disorder_Types` VALUES (2,'AMNESIA'),(1,'DELIRIUM'),(3,'DEMENTIA');
/*!40000 ALTER TABLE `Cognitive_Disorder_Types` ENABLE KEYS */;
UNLOCK TABLES;

--
-- Table structure for table `Environments`
--


CREATE TABLE `Environments` (
  `id` int(11) NOT NULL,
  `name` varchar(45) NOT NULL,
  PRIMARY KEY (`id`),
  UNIQUE KEY `nombre_UNIQUE` (`name`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;

--
-- Dumping data for table `Environments`
--

LOCK TABLES `Environments` WRITE;
/*!40000 ALTER TABLE `Environments` DISABLE KEYS */;
INSERT INTO `Environments` VALUES (3,'RURAL'),(2,'SEMIURBAN'),(1,'URBAN');
/*!40000 ALTER TABLE `Environments` ENABLE KEYS */;
UNLOCK TABLES;

--
-- Table structure for table `Episode_Types`
--

CREATE TABLE `Episode_Types` (
  `id` int(11) NOT NULL,
  `name` varchar(45) NOT NULL,
  PRIMARY KEY (`id`),
  UNIQUE KEY `nombre_UNIQUE` (`name`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;

```

```sql
--
-- Dumping data for table `Episode_Types`
--

LOCK TABLES `Episode_Types` WRITE;
/*!40000 ALTER TABLE `Episode_Types` DISABLE KEYS */;
INSERT INTO `Episode_Types` VALUES (2,'LIGHT'),(3,'MODERATE'),(1,'NO EPISODE'),(4,'SEVERE');
/*!40000 ALTER TABLE `Episode_Types` ENABLE KEYS */;
UNLOCK TABLES;

--
-- Table structure for table `Mood_Types`
--

CREATE TABLE `Mood_Types` (
  `id` int(11) NOT NULL,
  `name` varchar(45) NOT NULL,
  PRIMARY KEY (`id`),
  UNIQUE KEY `nombre_UNIQUE` (`name`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
/*!40101 SET character_set_client = @saved_cs_client */;

--
-- Dumping data for table `Mood_Types`
--

LOCK TABLES `Mood_Types` WRITE;
/*!40000 ALTER TABLE `Mood_Types` DISABLE KEYS */;
INSERT INTO `Mood_Types` VALUES (3,'happy'),(2,'normal'),(1,'sad');
/*!40000 ALTER TABLE `Mood_Types` ENABLE KEYS */;
UNLOCK TABLES;

--
-- Table structure for table `Patients`
--

DROP TABLE IF EXISTS `Patients`;
/*!40101 SET @saved_cs_client     = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `Patients` (
  `id` varchar(45) NOT NULL,
  `city` int(11) DEFAULT NULL,
  `cognitive_disorder` int(11) DEFAULT NULL,
  `environment` int(11) DEFAULT NULL,
  PRIMARY KEY (`id`),
  KEY `fk_city_idx` (`city`),
  KEY `fk_cognitive_disorder_idx` (`cognitive_disorder`),
  KEY `fk_environment_idx` (`environment`),
  CONSTRAINT `fk_city` FOREIGN KEY (`city`)
    REFERENCES `Cities` (`id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_cognitive_disorder` FOREIGN KEY (`cognitive_disorder`)
    REFERENCES `Cognitive_Disorder_Types` (`id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_environment` FOREIGN KEY (`environment`)
    REFERENCES `Environments` (`id`) ON DELETE NO ACTION ON UPDATE NO ACTION
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
--
-- Table structure for table `Patients_Fact_History`
--

DROP TABLE IF EXISTS `Patients_Fact_History`;
/*!40101 SET @saved_cs_client     = @@character_set_client */;
/*!40101 SET character_set_client = utf8 */;
CREATE TABLE `Patients_Fact_History` (
  `patient` varchar(45) NOT NULL,
  `date` date NOT NULL,
  `mood` int(11) DEFAULT NULL,
  `activity` int(11) DEFAULT NULL,
  `sleptHours` int(11) DEFAULT NULL,
  `episode` int(11) DEFAULT NULL,
  PRIMARY KEY (`patient`,`date`),
  KEY `fk_calendar_idx` (`date`),
  KEY `fl_activity_idx` (`activity`),
  KEY `fk_mood_idx` (`mood`),
  KEY `fk_episode_idx` (`episode`),
  CONSTRAINT `fk_calendar` FOREIGN KEY (`date`) REFERENCES `Calendar` (`date`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_episode` FOREIGN KEY (`episode`) REFERENCES `Episode_Types` (`id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_mood` FOREIGN KEY (`mood`) REFERENCES `Mood_Types` (`id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fk_patient` FOREIGN KEY (`patient`) REFERENCES `Patients` (`id`) ON DELETE NO ACTION ON UPDATE NO ACTION,
  CONSTRAINT `fl_activity` FOREIGN KEY (`activity`) REFERENCES `Activity_Types` (`id`) ON DELETE NO ACTION ON UPDATE NO ACTION
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```