



Sistema de Business Intelligence para Entorno Terapéutico de Enfermedades de Trastornos en la Alimentación

Natalia Medina Medina

Máster Universitario en Ingeniería Informática
Business Intelligence

David Amorós Alcaraz

María Isabel Guitart Hormigo

Junio 2016



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/)

Copyright

© Natalia Medina Medina

Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Sistema de Business Intelligence para Entorno Terapéutico de Enfermedades de Trastornos en la Alimentación</i>
Nombre del autor:	<i>Natalia Medina Medina</i>
Nombre del consultor/a:	<i>David Amorós Alcaraz</i>
Nombre del PRA:	<i>María Isabel Guitart Hormigo</i>
Fecha de entrega (mm/aaaa):	06/2106
Titulación::	<i>Máster Universitario en Ingeniería Informática</i>
Área del Trabajo Final:	<i>Business Intelligence</i>
Idioma del trabajo:	<i>Castellano</i>
Palabras clave	<i>Data Warehouse</i> <i>OLAP</i> <i>PENTAHO</i>

Resumen del Trabajo (máximo 250 palabras):

La **Inteligencia de negocio (Business Intelligence)** se define como el conjunto de estrategias, metodologías y aplicaciones cuyo objetivo es el acceso, análisis y transformación de los datos, en conocimiento que sirva de apoyo a la toma de decisiones estratégicas.

En este proyecto se diseñará y desarrollará una solución de Business Intelligence que permita realizar el análisis de la información suministrada por pacientes y terapeutas en el marco de las enfermedades de trastorno de la alimentación.

El auge de las tecnologías móviles ha posibilitado la aparición de distintas apps que, a través de un entorno atractivo, permite a los pacientes informar de sus actividades, estados de ánimo y evolución durante la terapia. Asimismo, los terapeutas pueden realizar un seguimiento más exhaustivo de los pacientes gracias a esta información.

Los objetivos de este proyecto se centran en el diseño de un Data Warehouse

que almacene los datos suministrados por los diferentes orígenes de datos, la implementación de los procesos de extracción, transformación y carga de datos en el Data Warehouse y la explotación de dichos datos.

Se pretende detectar tendencias que ayuden a determinar qué actividades producen mejor bienestar emocional al paciente, mayor reducción en la ocurrencia de episodios de vómitos, atracones o no comer.

Abstract (in English, 250 words or less):

Business intelligence (BI) can be described as a set of techniques and tools for the acquisition and transformation of raw data into meaningful and useful information for business analysis purposes

Design and develop a business intelligence solution is the main purpose of this project. It will allow the analysis of the information provided by patients and therapists in the context of disease eating disorder.

The rise of mobile technologies has enabled the emergence of different apps through an attractive environment. These apps allow patients to report their activities, moods and evolution during therapy. In addition, therapists can conduct a more thorough monitoring of patients thanks to this information.

This project is focused on the design of a Data Warehouse that stores the data supplied by the different data sources; the implementation of the processes of extraction, transformation and loading data into the Data Warehouse and exploitation of such data will also be accomplished.

It is intended to detect trends to determine what activities produce better emotional well-being to the patient and greater reduction in the occurrence of episodes of restriction, binge or not eating.

Índice

1.	Introducción	8
1.1.	Contexto y justificación del Trabajo	8
1.2.	Objetivos del Trabajo	8
1.3.	Enfoque y método seguido	9
1.4.	Planificación del Trabajo	10
1.5.	Breve resumen de productos obtenidos.....	12
1.6.	Breve descripción de los otros capítulos de la memoria	12
2.	Selección de herramientas.....	14
2.1.	Comparativa sistemas gestores de base de datos	14
2.1.1.	MySQL.....	14
2.1.2.	PostgreSQL.	15
2.2.	Selección sistema gestor de base de datos.....	15
2.3.	Comparativa plataforma BI	16
2.4.1.	Pentaho.....	16
2.4.2.	Jaspersoft	16
2.4.3.	SpagoBI	17
2.4.4.	Cuadro comparativo.....	17
2.4.	Selección plataforma BI	18
3.	Diseño del Data Warehouse.....	20
3.1.	Introducción.	20
3.2.	Indicadores clave	24
3.3.	Modelo conceptual del data warehouse.....	25
3.3.1.	Introducción	25
3.3.2.	Definición de dimensiones.	26
3.3.3.	Definición de hechos.....	27
3.3.4.	Diagrama del modelo conceptual del data warehouse.....	28
3.4.	Implementación del data warehouse.....	28
3.4.1.	Descripción tablas del data warehouse.....	28
3.4.2.	Diagrama de implementación del data warehouse	30
3.5.	Extracción, transformación y carga de datos en el data warehouse	31

3.5.1.	Introducción	31
3.5.2.	Datos origen.....	31
3.5.3.	Carga de los datos originales en la staging area	32
3.5.4.	Transformaciones de los datos en el data warehouse.....	32
4.	Análisis y diseño OLAP.....	33
4.1.	Introducción	33
4.2.	Creación de los cubos OLAP.....	33
4.2.1.	Creación de los cubos con <i>Pentaho Schema Workbench</i>	33
4.2.2.	Conexión a la base de datos.....	34
4.2.3.	Implementación de los cubos.....	34
4.2.4.	Publicación de los cubos.....	35
4.2.5.	Definición del esquema TCA_DW_SCHEMA.	36
4.3.	Navegación por los cubos OLAP.	38
4.3.1.	Navegación por los cubos con <i>Jpivot</i>	38
4.3.2.	Navegación por los cubos con <i>Saiku Analytics</i>	41
5.	Informes y Cuadros de mandos	44
5.1.	Introducción	44
5.2.	Generación de informes.....	44
5.2.1.	Informe de la relación entre actividades realizadas y episodios de crisis.....	44
5.2.2.	Informe de la relación entre los estados de ánimo y los episodios de crisis.....	49
5.2.3.	Informe de relación entre el entorno geográfico y los episodios de crisis.....	52
5.2.4.	Informe de evolución de pacientes.	57
5.2.5.	Informe de evolución de pacientes por centro médico y terapeuta.....	61
5.2.6.	Informe de relación entre los episodios de crisis y distintos momentos temporales.....	65
5.2.7.	Informe de relación entre las actividades y el estado de ánimo de los pacientes.	74
5.3.	Respuesta a las preguntas planteadas.....	78

5.3.1.	¿Cuál es la relación entre las actividades realizadas y los episodios de crisis?.....	78
5.3.2.	¿Se puede establecer algún tipo de relación entre los valores de los diferentes estados de ánimo y los episodios de crisis? 78	
5.3.3.	¿Se puede establecer alguna relación a nivel geográfico, por ejemplo entorno urbano o rural?	78
5.3.4.	¿Cómo ha sido la evolución de los diferentes pacientes a lo largo del tiempo?	79
5.3.5.	¿Hay alguna diferencia de la evolución de pacientes a nivel de centro médico? ¿Y terapeuta?	79
5.3.6.	¿Se puede establecer alguna relación entre los episodios de crisis y el momento del día o de la semana o del año?	80
5.3.7.	¿La realización de actividades físicas mejora o empeora el estado de ánimo de los pacientes?.....	82
5.4.	Cuadro de mandos.....	82
5.4.1.	Creación de un cuadro de mandos con Pentaho CDE Dashboard.	82
5.4.2.	Indicadores	85
6.	Conclusiones	86
7.	Glosario	88
8.	Bibliografía	91
9.	Anexo I. Esquema data warehouse	94
10.	Anexo II. Esquema staging area	97
11.	Anexo III. Carga de datos en el staging area	99
12.	Anexo IV. Carga de datos en el data warehouse.....	100
13.	Anexo V. Fichero XML de definición del cubo OLAP	105

Lista de figuras

Ilustración 1 - Planificación del proyecto	12
Ilustración 2 - Comparativa plataformas BI	18
Ilustración 3 - Entorno básico de un data warehouse.....	21
Ilustración 4 - Componentes de un data warehouse	24
Ilustración 5 - Diagrama modelo conceptual del data warehouse	28
Ilustración 6 - Diagrama de implementación del data warehouse	30
Ilustración 7 – Pentaho Schema Workbench. Conexión a base de datos	34
Ilustración 8 – Pentaho Schema Workbench. Conexión a base de datos	35
Ilustración 9 – Pentaho Schema Workbench. Especificación de cubos y dimensiones	35
Ilustración 10 – Pentaho Schema Workbench. Publicación del schema	35
Ilustración 11 – Pentaho Schema Workbench. Parámetros de la publicación del schema	36
Ilustración 12 – Pentaho Schema Workbench. Schema TCA_DW_SCHEMA	37
Ilustración 13 - Pantalla de acceso a Pentaho	39
Ilustración 14 - Pantalla inicial de Pentaho.....	39
Ilustración 15 - Opciones de llamada y pantalla. Jpivot.....	40
Ilustración 16 - Navegación con Jpivot.....	40
Ilustración 17 – Descarga Saiku Analytics.	41
Ilustración 18 – Versiones Saiku Analytics	42
Ilustración 19 - Confirmación de descarga Saiku Analytics	42
Ilustración 20 - Opciones de llamada y pantalla inicial. Saiku Analytics	43
Ilustración 21 - Creación de informe. Saiku Analytics	43

Ilustración 22 - Relación entre las actividades realizadas y los episodios de crisis I.....	45
Ilustración 23 - Relación entre las actividades realizadas y los episodios de crisis II.....	47
Ilustración 24 - Relación entre las actividades realizadas y los episodios de crisis III.....	49
Ilustración 25 - Relación entre los estados de ánimo y los episodios de crisis I... 50	
Ilustración 26 - Relación entre los estados de ánimo y los episodios de crisis II.. 52	
Ilustración 27 - Relación entre el entorno geográfico y los episodios de crisis I... 54	
Ilustración 28 - Relación entre el entorno geográfico y los episodios de crisis II.. 56	
Ilustración 29 - Evolución de pacientes I.....	58
Ilustración 30 - Evolución de pacientes por centro médico I	62
Ilustración 31 - Evolución de pacientes por terapeuta I.....	64
Ilustración 32 - Relación entre los episodios de crisis y la franja del día I.....	66
Ilustración 33 - Relación entre los episodios de crisis y los días de la semana I.. 70	
Ilustración 34 - Relación entre los episodios de crisis y los meses del año I	73
Ilustración 35 - Relación entre las actividades y el estado de ánimo de los pacientes I.....	75
Ilustración 36 - Relación entre las actividades y el estado de ánimo de los pacientes I.....	77
Ilustración 37 - Opciones de llamada y pantalla de CDE Dashboard	83
Ilustración 38 - Selección de plantilla CDE Dashboard	83
Ilustración 39 - Creación de gráfico OLAP CDE Dashboard	84
Ilustración 40 - Finalización gráfico OLAP CDE Dashboard.....	84
Ilustración 41 - Guardar CDE Dashboard.....	85
Ilustración 42 - Ejemplo cuadro CDE Dashboard	85

1. Introducción

1.1. Contexto y justificación del Trabajo

El objetivo del presente proyecto es la creación de una solución de **Business Intelligence** que permita analizar la información suministrada por pacientes y terapeutas, en el marco de las enfermedades de trastornos de la alimentación.

Un trastorno de la alimentación es una obsesión por los alimentos y el peso que daña el bienestar de una persona. Estas enfermedades causan graves perturbaciones en la dieta diaria; tienen su origen en diversas condiciones complejas, producto de conductas presentes por largo tiempo, factores biológicos, emocionales, psicológicos, interpersonales y sociales. La gente con trastornos alimenticios utiliza la comida y el control de la comida como un intento para controlar los sentimientos y emociones. Estos comportamientos dañan la salud física y emocional, la autoestima, la sensación de competitividad y el autocontrol de la persona

Los trastornos de la alimentación generalmente aparecen durante la adolescencia o adultez temprana, pero también se pueden dar durante la niñez o la adultez avanzada. Los trastornos de la alimentación afectan tanto a los hombres como a las mujeres. Los científicos siguen investigando y aprendiendo sobre las causas de estas enfermedades.

La información que suministren los pacientes durante su terapia resulta de extrema utilidad para aportar conocimiento que permita seguir avanzando en las investigaciones, ayudar y realizar el seguimiento a los afectados de forma cada vez más eficaz y efectiva para cada perfil.

Una solución de **Business Intelligence** aporta el análisis de los datos proporcionados por los pacientes y terapeutas, para conseguir estos propósitos.

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 proporcionados por pacientes y terapeutas, en el marco

de las enfermedades de trastorno alimentario. Por tanto, el trabajo tendrá los siguientes objetivos:

- **Diseño un Data Warehouse** que permita almacenar la información adquirida de los diferentes orígenes de datos.
- **Implementación del Data Warehouse y de los procesos ETL** (extracción, transformación y carga) que permitan alimentar el Data Warehouse a partir de los archivos base facilitados.
- **Análisis de las diferentes plataformas BI Open Source** disponibles en el mercado que nos permitan explotar la información almacenada.
- **Elección e implantación de una herramienta Open Source**, que permita el **análisis de la información** mediante una capa de software.
- **Respuesta a las siguientes preguntas:**
 - ¿Cuál es la relación entre las actividades realizadas y los episodios de crisis?
 - ¿Se puede establecer algún tipo de relación entre los valores de los diferentes estados de ánimo y los episodios de crisis?
 - ¿Se puede establecer alguna relación a nivel geográfico, por ejemplo entorno urbano o rural?
 - ¿Cómo ha sido la evolución de los diferentes pacientes a lo largo del tiempo?
 - ¿Hay alguna diferencia de la evolución de pacientes a nivel de centro médico? ¿Y terapeuta?
 - ¿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?

1.3. Enfoque y método seguido

Este TFM está enfocado en la realización de una solución de Business Intelligence. La metodología que se ha seguido ha constado de las siguientes fases:

- **Análisis de requisitos:** a partir del enunciado facilitado, se han determinado los requisitos que debe cumplir la solución de Business Intelligence, centrándonos principalmente en dar respuesta a las cuestiones planteadas en dicho enunciado.
- **Planificación:** se ha realizado una planificación de las tareas a efectuar, ajustándola al tiempo disponible para su finalización.
- **Descripción principales tareas:** las principales tareas de este proyecto son:
 - Análisis y selección de las herramientas a utilizar.
 - Diseño e implementación del Data Warehouse.
 - Implementación de los procesos ETL.
 - Definición e implementación del cuadro de mandos e informes.

1.4. Planificación del Trabajo

En base a las tareas a realizar ya descritas, se ha realizado la siguiente planificación, ajustándola a las entregas marcadas en el calendario de la asignatura.

El resumen de la planificación es el siguiente:

Nombre de tarea	Duración	Comienzo	Fin
Proyecto BI Trastornos de la alimentación	76 días	lun 29/02/16	lun 13/06/16
Preparación del Proyecto	10 días	lun 29/02/16	vie 11/03/16
Análisis del proyecto BI y definición de objetivos	3 días	lun 29/02/16	mié 02/03/16
Estudio del contexto del problema abordado	2 días	jue 03/03/16	vie 04/03/16
Elaboración del plan de trabajo	5 días	lun 07/03/16	vie 11/03/16

Nombre de tarea		Duración	Comienzo	Fin
Análisis alternativas Open Source para BI y SGBD		4 días	lun 14/03/16	jue 17/03/16
	Análisis alternativas Open Source para BI y SGBD	3 días	lun 14/03/16	mié 16/03/16
	Selección herramientas BI y SGBD	1 día	jue 17/03/16	jue 17/03/16
	Documentación	4 días	lun 14/03/16	jue 17/03/16
Diseño e implementación Data Warehouse		25 días	vie 18/03/16	jue 21/04/16
	Análisis de los datos	2 días	vie 18/03/16	lun 21/03/16
	Diseño e implementación BD	4 días	mar 22/03/16	vie 25/03/16
	Diseño e implementación DW	19 días	lun 28/03/16	jue 21/04/16
	Documentación	25 días	vie 18/03/16	jue 21/04/16
Diseño e implementación ETL		6 días	vie 22/04/16	vie 29/04/16
	Diseño e implementación procesos ETL	3 días	vie 22/04/16	mar 26/04/16
	Documentación	3 días	mié 27/04/16	vie 29/04/16
Análisis y diseño OLAP		7 días	lun 02/05/16	mar 10/05/16
	Diseño e implementación OLAP	7 días	lun 02/05/16	mar 10/05/16
	Documentación	7 días	lun 02/05/16	mar 10/05/16
Análisis y diseño de informes		7 días	mié 11/05/16	jue 19/05/16
	Diseño e implementación Informes	7 días	mié 11/05/16	jue 19/05/16
	Documentación	7 días	mié 11/05/16	jue 19/05/16

Nombre de tarea		Duración	Comienzo	Fin
Análisis y diseño de cuadro de mandos		7 días	vie 20/05/16	lun 30/05/16
	Diseño e implementación Cuadro de Mandos	7 días	vie 20/05/16	lun 30/05/16
	Documentación	7 días	vie 20/05/16	lun 30/05/16
Pruebas		26 días	lun 02/05/16	lun 06/06/16
	Documentación de pruebas	21 días	lun 02/05/16	lun 30/05/16
	Realización de pruebas	5 días	mar 31/05/16	lun 06/06/16
Cierre y Finalización del proyecto		5 días	mar 07/06/16	lun 13/06/16
	Respuesta preguntas planteadas	2 días	mar 07/06/16	mié 08/06/16
	Revisión y entrega Documentación	3 días	jue 09/06/16	lun 13/06/16

Ilustración 1 - Planificación del proyecto

1.5. Breve resumen de productos obtenidos

Los productos obtenidos con este TFM son los siguientes:

- Scripts para la generación de la estructura de tablas para la staging area y para el data warehouse.
- Scripts para la realización de los procesos ETL
- Esquema de la estructura de datos OLAP.
- Informes de datos, incluidos en esta memoria.

1.6. Breve descripción de los otros capítulos de la memoria

La presente memoria se compone de los siguientes capítulos:

- **Capítulo 2:** Análisis y selección de herramientas.
En este capítulo se realiza un estudio de las diferentes alternativas de Sistemas Gestores de Base de Datos y de plataformas BI. Todas las opciones incluidas son Open Source. En base a este análisis y a las necesidades del proyecto, se realiza la elección de las herramientas a utilizar.
- **Capítulo 3:** Diseño e implementación Data Warehouse.
En este capítulo se detalla el diseño y la implementación del data warehouse. Se incluye asimismo los procesos de extracción, transformación y carga de datos.
- **Capítulo 4:** Análisis y diseño OLAP
En este capítulo se incluye la definición del cubo OLAP, su vinculación a la plataforma Pentaho y las herramientas utilizadas para realizar la navegación por los cubos OLAP.
- **Capítulo 5:** Informes y Cuadros de mandos.
En este capítulo se detallan los informes que servirán de base para contestar a las preguntas planteadas en este proyecto.
- **Capítulo 9 – 13:** Anexos
En este capítulo se incluyen los distintos scripts que conforman el producto resultante de este proyecto.

2. Selección de herramientas

2.1. Comparativa sistemas gestores de base de datos

Hoy en día, MySQL y PostgreSQL son los dos sistemas gestores de bases de datos (SGBD, en adelante) Open Source más ampliamente utilizados. A continuación, se enumerarán brevemente las principales características de ambos SGBD.

2.1.1. MySQL.

Es un sistema de gestión de base de datos relacional, multiplataforma, multihilo y multiusuario, publicado bajo la licencia GPL. Es una de las opciones más utilizadas en aplicaciones de software libre

El desarrollo de MySQL se fundamenta en el trabajo de los desarrolladores contratados por la empresa MySQL AB quienes se encargan de dar soporte a los socios comerciales y usuarios de la comunidad MySQL y dar solución a los problemas encontrados por los usuarios.

Sus principales características son:

- **Alta velocidad de lecturas**, cuando utiliza el motor no transaccional MyISAM. Sin embargo, pueden presentarse problemas de integridad en entornos de alta concurrencia en la modificación.
- **Permite escoger entre múltiples motores de almacenamiento** para cada tabla.
- **Agrupación de transacciones**, reuniendo múltiples transacciones de varias conexiones para incrementar el número de transacciones por segundo.
- **Otras características relevantes** amplio subconjunto del lenguaje SQL, transacciones y claves foráneas, conectividad segura, replicación, búsqueda e indexación de campos de texto.

2.1.2. PostgreSQL.

Es un SGBD relacional orientada a objetos, multiplataforma y libre, publicado bajo la licencia BSD.

El desarrollo de PostgreSQL está dirigido por una comunidad de desarrolladores que trabajan de forma desinteresada, altruista, libre y/o apoyada por organizaciones comerciales. La comunidad PostgreSQL se denominada el PGDG (PostgreSQL Global Development Group).

Sus principales características son:

- **Análisis de consultas:** consta de herramientas que permiten analizar las consultas realizadas y determinar las que consumen más recursos, para poder optimizarlas y mejorar su rendimiento.
- **Tablas particionadas:** la utilización de tablas particionadas permite mejorar el rendimiento de consultas realizada a tablas con grandes cantidades de información.
- **Optimización en la creación de índices:** PostgreSQL dispone de sistemas de resolución de consultas muy eficientes, permitiendo decidir adecuadamente cuándo utilizar los índices y cuándo no.
- **Balanceo de almacenamiento:** esta característica facilita y agiliza la extracción de datos.

2.2. Selección sistema gestor de base de datos

Los requerimientos de cada proyecto serán los que permitan realizar la adecuada selección del sistema gestor de base datos. Para nuestro data warehouse, el SGBD elegido es **PostgreSQL**, ya que es un sistema robusto cuyas características principales cumplen con los requisitos que necesitamos.

PostgreSQL es una gran plataforma para alto rendimiento, muy apropiada para soluciones de Business Intelligence. PostgreSQL 9.5 (Release Date: 07-01-2016), ofrece una serie de nuevas características enfocadas a los data warehouses, así como algunas mejoras de rendimiento para manejar múltiples cargas de trabajo.

PostgreSQL ha incorporado una serie de nuevas características analíticas típicas de los data warehouses, incluyendo conjuntos de agrupación, cubos y roll-up. Aunque la funcionalidad estándar de SQL ofrece la posibilidad de ejecutarlos a través de una serie de operadores, PostgreSQL las realiza de forma mucho más fácil, acelerando los tiempos de ejecución de consultas complejas.

Otra nueva característica a destacar es un nuevo tipo de indexación llamada BRIN (Block Range Index). BRIN puede generar muy pequeños índices para describir una gran variedad de información, como los valores mínimos y máximos, que permiten realizar consultas de datos dentro de un cierto rango, saltándonos un amplio número de filas.

Otra de las nuevas características a señalar ha sido la introducción de un comando que combina INSERT y UPDATE en una sola llamada: se puede convertir automáticamente un INSERT en una actualización si los datos ya existen. PostgreSQL ofrece una cláusula especial que puede utilizarse con la inserción, que logra el mismo resultado.

2.3. Comparativa plataforma BI

Existen en el mercado actual varias alternativas de plataformas de Business Intelligence Open Source. Entre ellas, se han estudiado las opciones más populares que permitan realizar todas las fases del proyecto. Las seleccionadas son las siguientes:

2.4.1. Pentaho.

Pentaho es una plataforma de BI “orientada a la solución” y “centrada en procesos”, ya que incluye todos los principales componentes requeridos para implementar soluciones basadas en procesos. Las soluciones que Pentaho pretende ofrecer se componen fundamentalmente de una infraestructura de herramientas de análisis e informes integradas con un motor de workflow de procesos de negocio.

Incluye herramientas para la generación de informes, OLAP, ETL, cuadros de mando y minería de datos. Está enfocada principalmente en el tratamiento de los datos.

2.4.2. Jaspersoft

JasperReports es la solución Open Source más extendida para incluir la realización de informes en aplicaciones. Su solución de Business

Intelligence incluye herramientas para la generación de informes, OLAP, ETL y cuadros de mando. JasperSoft está construida en base al motor de informes, integrando proyectos ya existentes sin absorberlos. Esto ocasiona dependencias en relación a las soluciones ETL y motor OLAP.

2.4.3. SpagoBI

SpagoBI es una plataforma de integración (y no una plataforma de producto), ya que no se construye y se cierra en torno a un conjunto predefinido de herramientas. Tiene una estructura modular en la que todos los módulos se relacionan con el núcleo del sistema; esta estructura garantiza la armonía de la plataforma y la capacidad evolutiva. Integra motores de análisis específicos para cada área y permite el uso de varios motores de análisis al mismo tiempo: SpagoBI también permite la integración de soluciones propietarias para poder construir la mejor plataforma para un problema en particular.

Incluye herramientas para la generación de informes, OLAP, ETL, cuadros de mano y minería de datos.

2.4.4. Cuadro comparativo.

A continuación se incluye un cuadro comparativo de las plataformas BI estudiadas:

	Pentaho (Comunity Edition)	JasperSoft (Comunity Edition)	SpagoBI
ETL	Pentaho Data Integration (Kettle)	Talend	Talend
Servidor B.I.	Business Intelligence Server	Jasper Server	Spago BI Server
Informes	Pentaho Reporting	JasperReports Server	SpagoBI JasperReport Engine
	✓ <i>Informes Predefinidos</i>	✓ <i>Informes Predefinidos</i>	✓ <i>Informes Predefinidos</i> ✓ <i>Informes ad-hoc</i>

Cuadros de mando	Comunnity Dashboard Framework	Jaspersoft Dashboards	SpagoBI Dashboard Engine
	✓ Cuadros de mando predefinidos	✓ Cuadros de mando predefinidos	✓ Cuadros de mando predefinidos
Motor OLAP	Mondiran	Mondiran	SpagoBIJPivotEngine (embedded version of Mondrian)
Visor OLAP	Jpivot	JasperAnalysis	Jpivot

Ilustración 2 - Comparativa plataformas BI

2.4. Selección plataforma BI

La plataforma BI seleccionada para nuestro data warehouse es **Pentaho BI Suite**, ya que integra diferentes herramientas que realizar todas las fases del proyecto: procesos ETL, análisis inteligente de los datos, informes interactivos, análisis multidimensionales y minería de datos.

Las principales razones para esta elección se basan primordialmente en su alto nivel de integración entre los módulos que la componen y en su fiabilidad, avalada por ser la plataforma de referencia en los proyectos Open Source de Business Intelligence. Además, dispone de una extensa y activa comunidad y cuenta con una amplia documentación y numerosos recursos web para facilitar su aprendizaje.

Los módulos incluidos en Pentaho BI Suite son:

- **Pentaho Data Integration (PDI):** su función es la gestión de los procesos ETL, extrayendo, transformando y cargando los datos origen.
- **Pentaho Reporting (PRD):** su objetivo es la realización de informes. Pueden ser confeccionados por el usuario final o estar predefinidos para ser consultados directamente. Admite diferentes formatos de visualización.
- **Pentaho Analysis:** su cometido es permitir la consulta, exploración y análisis de la información de forma interactiva; mediante la definición de

criterios predefinidos, se podrá seleccionar una de las perspectivas disponibles.

- **Pentaho Data Mining:** el descubrimiento de patrones e indicadores a partir de la información del data warehouse es la funcionalidad realizada por este módulo.

3. Diseño del Data Warehouse

3.1. Introducción.

La información es uno de los activos más importantes de una organización, ya que el registro de los datos generados por los procesos operacionales constituye la base para extraer la información que dará apoyo a la toma de decisiones estratégicas. Simplificando las definiciones de sistemas operacionales (*OLTP – Online transaction processing systems*) y sistemas data warehouse, podemos decir que los primeros nos permiten registrar los datos generados por las líneas de negocio de la organización y los segundos nos permiten obtener información valiosa para la organización, a diferentes niveles.

A continuación presentamos una comparativa entre ambos sistemas¹:

	OLTP	Data Warehouse
Propósito	Soporte a los procesos operacionales	Soporte analítico para la toma de decisiones
Uso	Captura y mantenimiento de datos	Explotación de datos
Validación de los datos	Entrada de datos	A posteriori
Frecuencia de actualización	En el momento en que ocurre la transacción	Periódicamente
Datos históricos	Datos actuales	Varios años de datos
Integración de datos	Datos balanceados en el ámbito del sistema OLTP (único)	Datos integrados y balanceados desde múltiples sistemas

Ralph Kimball, destacado autor en el ámbito de *Business Intelligence*, define un data warehouse como "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis".

Una definición más amplia de data warehouse lo describe como el conjunto de datos y procesos destinados a agrupar y seleccionar los datos de los distintos

¹ *A manager's guide to Data Warehousing*, Laura L. Reeves

sistemas operacionales, para facilitar su análisis y proporcionar apoyo para la toma de decisiones. Un data warehouse ofrece una visión global, común e integrada de los datos y posee coherencia, fiabilidad e información histórica.

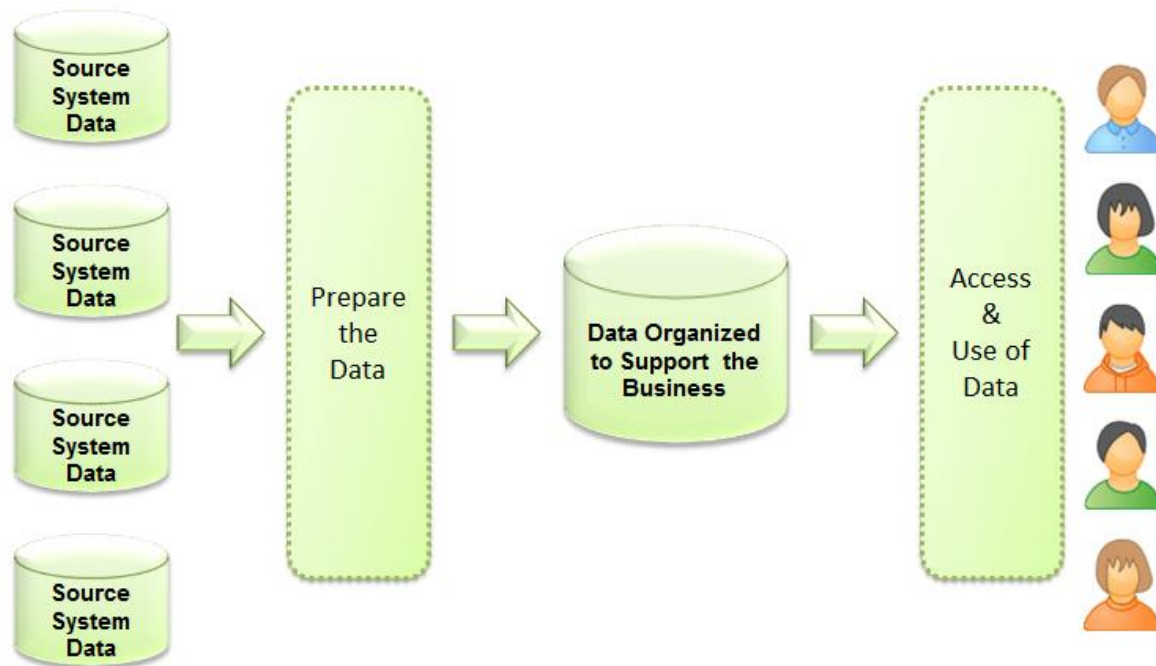


Ilustración 3 - Entorno básico de un data warehouse²

Los requisitos principales de un data warehouse son:

- **Información fácilmente accesible:** los contenidos del data warehouse deben ser fácilmente accesibles, comprensibles, intuitivos y de rápida recuperación. Es importante tener muy en cuenta que la información va dirigida a directivos y responsables de la gestión empresarial y, por tanto, la información que se les suministre debe adaptarse tanto a su vocabulario como a los procesos de la organización vinculados a sus funciones.
- **Información consistente:** la información del data warehouse debe ser coherente. La información debe ser tratada cuidadosamente, garantizando su calidad y consistencia antes de ser presentada a los usuarios. Son de especial importancia las definiciones y etiquetas que se asignen a los contenidos del data warehouse, asegurando que cada

² *A manager's guide to Data Warehousing*, Laura L. Reeves

definición queda debidamente etiquetada y diferenciada de las restantes definiciones.

- **Información adaptable al cambio:** tanto la tecnología como las necesidades empresariales están sometidas a un continuo cambio. El data warehouse debe ser capaz de adaptarse a esta circunstancia de forma transparente al usuario, respetando siempre la consistencia de los datos ya existentes.
- **Información protegida:** la información contenida en el data warehouse debe ser cuidadosamente protegida, ya que dicha información resulta muy valiosa para los competidores de la organización. Por tanto, un control de acceso eficaz para acceder a la información confidencial de la organización es de vital importancia
- **Información disponible en un tiempo apropiado:** la información que presenta el data warehouse debe presentarse a los usuarios en un tiempo razonable. La utilización intensiva del data warehouse para el soporte a la toma de decisiones, requiere que los procesos de conversión de datos en información sean diseñados con directrices específicas que permitan satisfacer la necesidad de cálculo y la disponibilidad de los datos.
- **Información para la toma de decisiones:** la información del data warehouse es una base importante para la toma de decisiones y la gestión estratégica de la organización. Los datos presentados influirán en las decisiones de negocio y su impacto en el negocio determinará el valor que se atribuya al sistema de *business intelligence*.
- **Información valiosa para la organización:** dado que la utilización de un data warehouse no facilita las tareas operativas de una organización, el valor que se le atribuya en dicha organización reside en ofrecer a los usuarios una fuente de información útil y procesable de forma rápida y sencilla. Para que un sistema de *business intelligence* sea aceptado y utilizado activamente, es importante ofrecer una solución elegante y eficaz que responda a las necesidades expuestas.

Los dos últimos requisitos detallan uno de los objetivos primordiales que debe cumplir un data warehouse: ofrecer información relevante a los responsables de la toma de decisiones para reducir la incertidumbre. Los sistemas de *business intelligence* nos servirán de ayuda para la toma de decisiones y, además, nos permitirán descubrir información de la organización que no se conocía.

La construcción de un sistema de *business intelligence* es un reto para una organización, pero también puede proporcionar grandes recompensas. Algunos de los beneficios de la utilización de un data warehouse son los siguientes:

- **Definición de indicadores clave de rendimiento:** la organización podrá conocer las **líneas** de productos de mayor aceptación o analizar la productividad de los empleados
- **Medición del éxito de las decisiones estratégicas:** la información del data warehouse permitirá comparar los rendimientos reales obtenidos con las previsiones realizadas.
- **Análisis de ventas:** detectar los cambios en los patrones de venta ayudará a la organización a **identificar** cambios en el mercado, medir la efectividad de las campañas publicitarias o valorar la confianza de los consumidores en la marca.
- **Estudio de los perfiles de clientes:** el estudio de la información del data warehouse permitirá diseñar **estrategias** para mejorar la captación de nuevos clientes, afianzar la fidelización de los actuales y ofrecer productos y servicios que respondan a las necesidades cambiantes de los clientes.
- **Desarrollo de nuevas líneas de productos o servicios:** el análisis de las líneas de **negocio** actuales permitirá detectar las que aportan mayor beneficio y posibilitará el desarrollo de nuevas líneas estratégicas.

Los componentes básicos de un sistema data warehouse (ilustración 2) son los siguientes:

- **Sistemas origen:** son todos aquellos sistemas que proporcionan datos al data warehouse. Generalmente son sistemas operacionales, cuya finalidad es la captura de datos.
- **Almacenamiento intermedio (*staging area*):** la función de este componente es recibir los datos de los diferentes sistemas origen para transformarlos y prepararlos para ser cargados en el data warehouse. El conjunto de los procesos realizados para esta transformación y preparación se denominan procesos ETL (*Extract, Transform y Load*).
- **Data Warehouse:** un data warehouse almacena los datos relevantes para realizar un análisis detallado de la información de una organización, considerando todas las perspectivas involucradas. Un data mart es un data warehouse que almacena información específica de un grupo de

trabajo dentro de la organización. La diferencia básica entre ambos es su alcance: para los primeros es la organización; para los segundos un grupo de trabajo de la organización.

- **Herramientas de acceso a datos:** estas herramientas permiten al usuario visualizar la información contenida en el data warehouse. Normalmente esta información se presenta en forma de informes, gráficas o cuadros de mando.

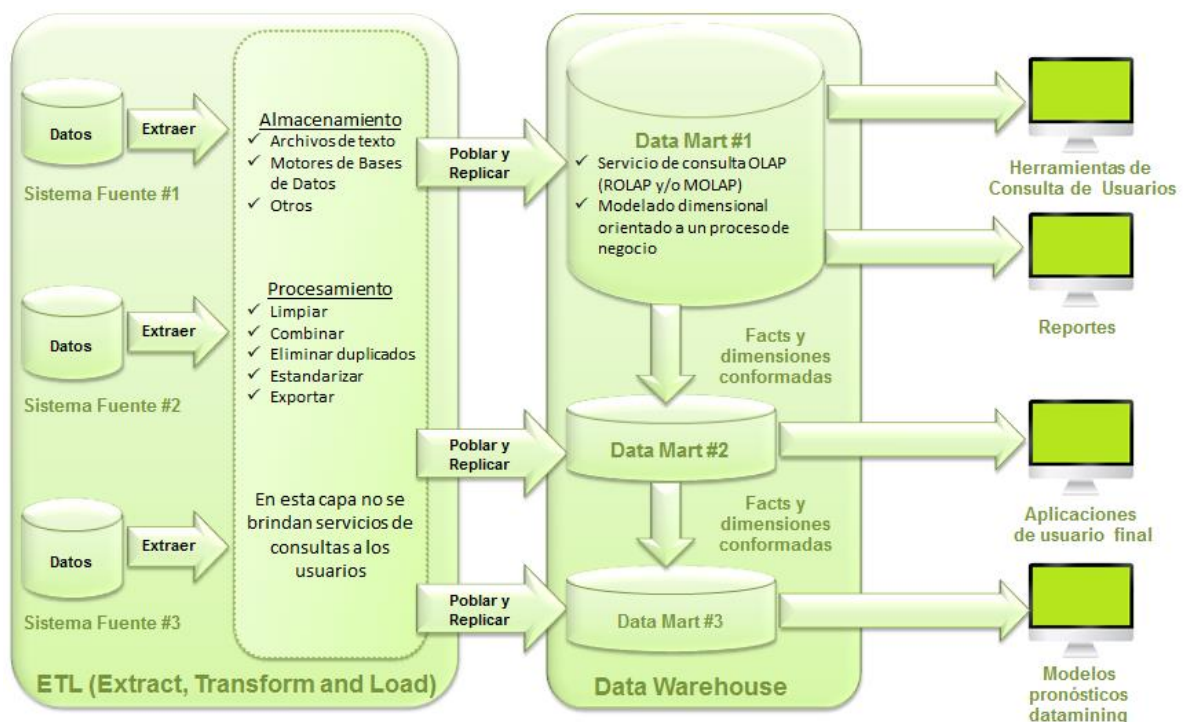


Ilustración 4 - Componentes de un data warehouse³

3.2. Indicadores clave

Los indicadores clave de un data warehouse proporcionan la información relativa a los valores sobre el rendimiento de una actividad y determinan el grado de consecución de los objetivos.

Para dar respuesta a las preguntas que se nos han planteado, se van a definir los indicadores clave a tener en cuenta en el diseño del data warehouse:

- Número de actividades por tipo de actividad.

³ Introducción al datawarehousing – Juan Zaffaroni

- Número de actividades por entorno del paciente.
 - Número de actividades por paciente.
 - Número de actividades por terapeuta.
 - Número de actividades por centro médico.
 - Número de actividades por momento del día
 - Número de actividades por semana del año.
-
- Número de episodios de crisis por tipo de episodio
 - Número de episodios de crisis por entorno del paciente
 - Número de episodios de crisis por paciente.
 - Número de episodios de crisis por terapeuta.
 - Número de episodios de crisis por centro médico.
 - Número de episodios de crisis por momento del día
 - Número de episodios de crisis por semana del año.
-
- Número de estados de ánimo por tipo de estado de ánimo.
 - Número de estados de ánimo por entorno del paciente
 - Número de estados de ánimo paciente.
 - Número de estados de ánimo terapeuta.
 - Número de estados de ánimo por centro médico.
 - Número de estados de ánimo por momento del día.
 - Número de estados de ánimo por semana del año.
-
- Número de comidas realizadas por tipo de comida.
 - Número de comidas realizadas por entorno del paciente
 - Número de comidas realizadas de ánimo paciente.
 - Número de comidas realizadas de ánimo terapeuta.
 - Número de comidas realizadas por momento del día.
 - Número de comidas realizadas por semana del año

3.3. Modelo conceptual del data warehouse

3.3.1. Introducción

El modelo multidimensional es la técnica más ampliamente utilizada para la presentación de los datos analíticos, ya que satisface dos importantes requisitos:

- **Visualización de los datos de forma comprensible** para los responsables del área de gestión de la organización.

- **Rendimiento de consultas rápido y eficaz.**

El modelo dimensional consta de los siguientes elementos:

- **Dimensiones:** representan los puntos de análisis de un hecho.
 - **Miembros:** una tabla de dimensión contiene un atributo único y un conjunto de atributos que la describen.
 - **Jerarquías:** constituyen los diferentes niveles de granularidad de una dimensión.
- **Hechos:** son representaciones de un proceso de negocio. Este tipo de tabla permite guardar dos tipos de atributos:
 - **Medidas** del proceso que se pretende modelizar.
 - **Claves ajenas** hacia registros de una tabla de dimensión.

3.3.2. Definición de dimensiones.

La definición de las dimensiones de nuestro data warehouse es la siguiente:

Dimensión	D_PACIENTE		
Descripción	Pacientes sometidos al estudio		
Atributos	✓	id_paciente	✓ nombre_paciente
	✓	ciudad	✓ sexo
	✓	id_entorno	✓ nombre_entorno
	✓	id_terapeuta	✓ nombre_terapeuta
	✓	id_centmed	✓ nombre_centmed
	✓	ciudad_centmed	
Jerarquías	Centmed	Nivel 0	Centro_médico
		Nivel 1	Terapeuta
		Nivel 2	Paciente
	Entorno	Nivel 0	Entorno
		Nivel 1	Paciente

Dimensión	D_TIEMPO	
Descripción	Atributo tiempo para las diferentes jerarquías	
Atributos	✓ id_fecha	✓ semana_anio
	✓ fecha	✓ dia_mes
	✓ anio	✓ dia_nombre
	✓ trimestre	✓ franja_dia
	✓ mes	✓ mes_nombre
Jerarquías	Nivel 0	anio
	Nivel 1	trimestre
	Nivel 2	mes
	Nivel 3	semana
	Nivel 4	dia
	Nivel 5	franja_dia

Dimensión	D_EVENTO	
Descripción	Tipos y subtipos de los eventos objetos del estudio	
Atributos	✓ id_evento	✓ nombre_evento
	✓ id_tipo_evento	✓ nombre_tipo_evento
Jerarquías	Nivel 0	Tipo_evento
	Nivel 1	Evento

3.3.3. Definición de hechos.

La definición del hecho de nuestro data warehouse es la siguiente:

Hecho	H_DIARIO
Descripción	Contiene la información relativa a los eventos y los pacientes objeto del estudio.
Dimensiones	Paciente Fecha Evento
Indicadores	<ul style="list-style-type: none"> • Número de eventos por tipo y subtipo de evento • Número de eventos por entorno del paciente • Número de eventos por paciente. • Número de eventos por terapeuta. • Número de eventos por centro médico.

- Número de eventos por momento del día
- Número de eventos por semana del año

3.3.4. Diagrama del modelo conceptual del data warehouse

El diagrama del modelo conceptual de nuestro data warehouse se muestra en la figura 3.

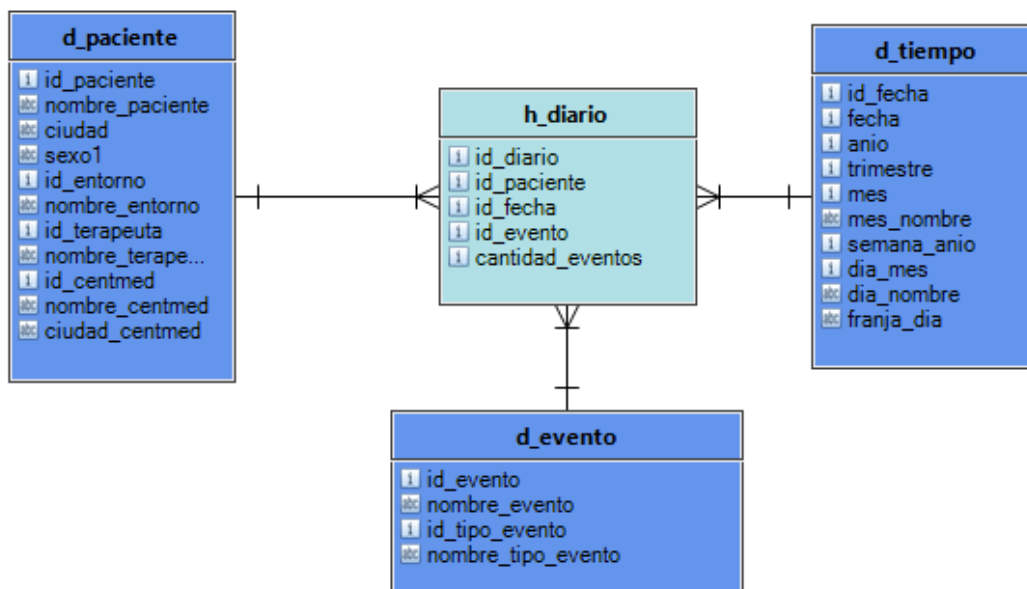


Ilustración 5 - Diagrama modelo conceptual del data warehouse

3.4. Implementación del data warehouse

3.4.1. Descripción tablas del data warehouse

Para la creación de los objetos de nuestro data warehouse se creará un nuevo usuario *TCA_DW_ADMIN*, que será el propietario de los objetos del data warehouse. El script de creación de este esquema y sus objetos puede consultarse en el **Anexo I. Esquema data warehouse**

A continuación mostramos la definición de las tablas que compondrán nuestro data warehouse:

Tabla	TCA_D_PACIENTE	
	Nombre	Tipo
Columnas	id_paciente	integer
	nombre_paciente	character
	ciudad	character
	sexo	character
	id_entorno	integer
	nombre_entorno	character
	id_terapeuta	integer
	nombre_terapeuta	character
	id_centmed	integer
	nombre_centmed	character
	ciudad_centmed	character

Tabla	TCA_D_TIEMPO	
	Nombre	Tipo
Columnas	id_fecha	integer
	fecha	date
	anio	integer
	trimestre	integer
	mes	integer
	mes_nombre	character
	semana_anio	integer
	dia_mes	integer
	dia_nombre	character
	franja_dia	character

Tabla	TCA_D_EVENTO	
	Nombre	Tipo
Columnas	Id_evento	int4
	nombre_evento	varchar
	id_tipo_evento	int4
	nombre_tipo_evento	varchar
	nombre	varchar

Tabla	TCA_H_DIARIO	
Columnas	Nombre	Tipo
	id_diario	varchar
	id_paciente	integer
	id_fecha	integer
	id_evento	integer
	cantidad_eventos	integer
Claves ajenas	Tabla referenciada	Columna referenciada
	tca_d_paciente	id_paciente
	tca_d_tiempo	id_fecha
	tca_d_evento	id_evento

3.4.2. Diagrama de implementación del data warehouse

El diagrama de implementación de nuestro data warehouse se muestra en la figura 4.

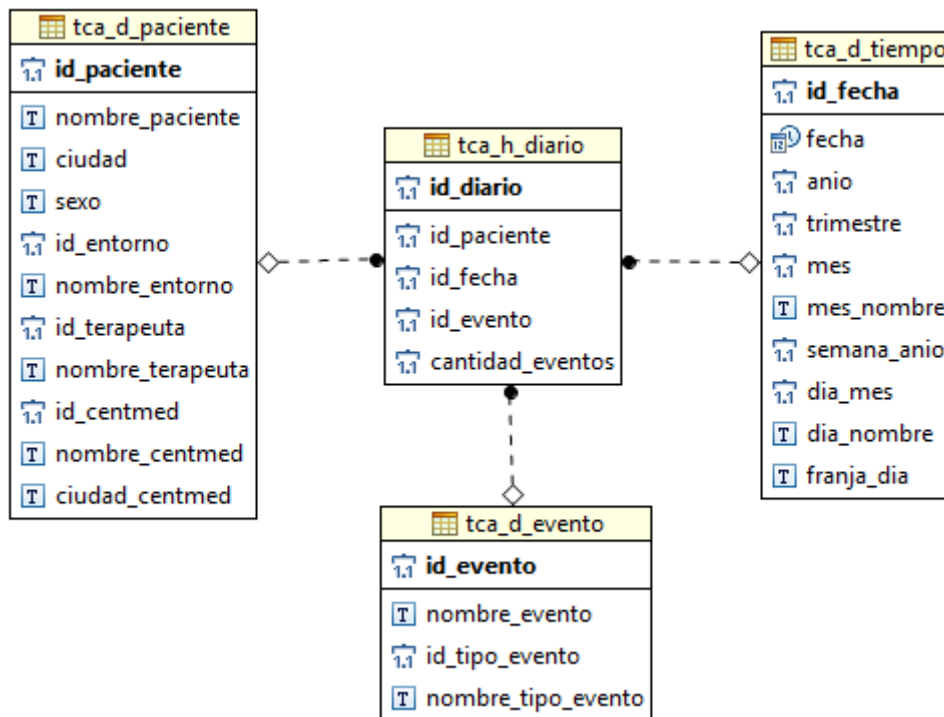


Ilustración 6 - Diagrama de implementación del data warehouse

3.5. Extracción, transformación y carga de datos en el data warehouse

3.5.1. Introducción

El proceso de extracción, transformación y carga (ETL) es la base de los datos de un data warehouse.

Específicamente, el sistema ETL:

- Elimina errores y corrige los datos que faltan.
- Proporciona medidas documentadas de la fiabilidad de los datos.
- Captura el flujo de datos transaccionales para su custodia.
- Ajusta los datos de múltiples fuentes para ser utilizados conjuntamente.
- Estructura los datos que puedan ser utilizados por las herramientas de usuario final

3.5.2. Datos origen

Los datos origen del data warehouse están contenidos en un fichero Excel con las siguientes hojas

- **Centmed:** contiene los centros médicos o terapéuticos que trabajan en el piloto.
- **Therapists:** lista de los terapeutas y el centro en el que trabajan. Un terapeuta solo trabaja para un centro médico.
- **Patient:** lista de pacientes con información sobre el sexo, su localidad, si su entorno es urbano o rural y su terapeuta.
- **Episodio:** contiene un registro de los episodios de crisis ocurridos durante un día según la franja horaria de madrugada-mañana, mediodía, tarde-noche.
- **Emotion:** registro de los estados de ánimo durante las franjas horarias.
- **Meal:** registro de ingestas.
- **Activity:** contiene el registro de actividades realizadas durante las diferentes franjas del día.

Algunos comentarios de interés sobre los datos son:

- Los registros de episodios, comida y estado de ánimo no necesariamente se hacen a la misma hora. Para simplificar se ha dividido el día en tres franjas.
- El entorno es un indicador aproximado, habrá entornos más urbanos, menos rurales, etc.
- Las actividades registradas no ocupan toda la franja, es la actividad principal realizada.

3.5.3. Carga de los datos originales en la staging area

Los datos origen del fichero Excel se cargarán en un esquema creado para esta función en la base de datos. La creación del esquema y de sus objetos puede consultarse en el **Anexo II. Esquema staging area**

Esta carga de datos se realizará mediante el script detallado en el **Anexo III. Carga de datos en el staging area.**

Mediante el comando **copy** podemos importar o exportar datos entre tablas de PostgreSQL y archivos estándar. Para cada uno de los ficheros **csv** que contienen los datos, se ha generado una instrucción con el comando **copy**, que inserta los datos en la tabla correspondiente, según el fichero origen.

3.5.4. Transformaciones de los datos en el data warehouse.

La carga de datos se realizará mediante el script detallado en el **Anexo IV. Carga de datos en el data warehouse.**

A partir de los datos de la staging area se han generado las instrucciones **insert** necesarias para cargar los datos en las diferentes tablas del data warehouse.

4. Análisis y diseño OLAP

4.1. Introducción

Un cubo OLAP (*OnLine Analytical Processing*) es una base de datos multidimensional, en la que el almacenamiento físico de los datos se realiza en un vector multidimensional. Los cubos OLAP se pueden considerar como una ampliación de las dos dimensiones de una hoja de cálculo.

A menudo se pensaba que todo lo que los usuarios pueden querer de un sistema de información se podría hacer de una base de datos relacional.

La disposición de los datos en cubos permite el análisis instantáneo de grandes cantidades de datos. Para acceder a los datos sólo es necesario indexarlos a partir de los valores de las dimensiones o ejes.

4.2. Creación de los cubos OLAP.

4.2.1. Creación de los cubos con *Pentaho Schema Workbench*.

A partir del diseño del capítulo anterior, crearemos los cubos utilizando la herramienta **Pentaho Schema Workbench**, que nos permitirá, a través de un entorno gráfico, construir los cubos y publicarlos al servidor BI, para que estén disponibles para los usuarios de la plataforma.

Cada uno de los cubos a definir debe cumplir las siguientes especificaciones:

- **Tabla de la base datos:** se debe detallar la tabla de la base de datos (tabla de hechos) asociada al cubo.
- **Dimensión:** se deben asociar al cubo las dimensiones correspondientes, siendo necesario detallar, al menos, una dimensión.
- **Medidas:** se debe definir, como mínimo una medida para cada cubo.

Cada una de las dimensiones a definir debe cumplir las siguientes especificaciones:

- **Tabla de la base datos:** se debe detallar la tabla de la base de datos (tabla de dimensión) asociada a la dimensión.
- **Jerarquía:** se debe definir, como mínimo una jerarquía para cada dimensión.

4.2.2. Conexión a la base de datos

El paso previo a la definición de los cubos es la creación de la conexión a la base de datos. Para ello, seguiremos las siguientes indicaciones:

- **Opciones de menú: Options → Connection**



Ilustración 7 – Pentaho Schema Workbench. Conexión a base de datos

- **Especificación de parámetros:** se deberán cumplimentar los siguientes datos:
 - Connection Name.
 - Connection Type.
 - Access.
 - Settings: host name, database name, port number, user name y password.
- **Verificación de la conexión:** desde la pantalla de especificación de parámetros, podremos comprobar su correcta configuración, mediante el botón **Test**, que comprueba si se establece la conexión.

4.2.3. Implementación de los cubos

Para la definición de los cubos, seguiremos las siguientes indicaciones:

- **Creación del esquema:** para crear el esquema, las Opciones de menú son **File → New → Schema**

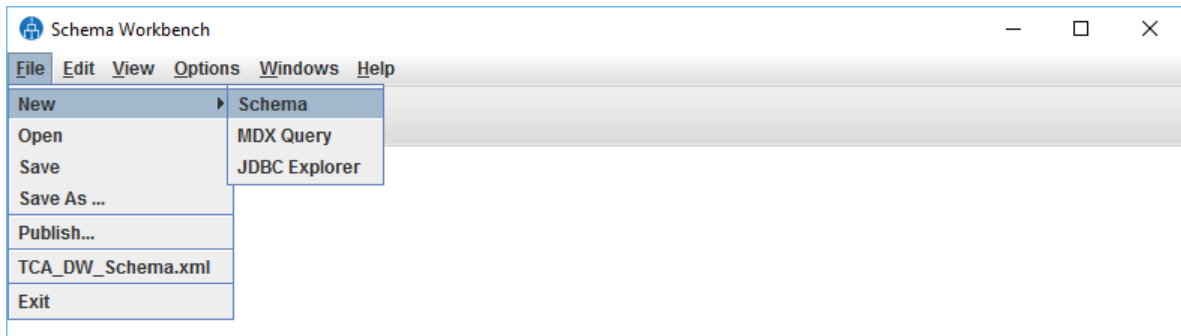


Ilustración 8 – Pentaho Schema Workbench. Conexión a base de datos

- **Especificación de cubos y dimensiones:** una vez creado el **schema**, se crearán los cubos y dimensiones en base al diseño y las especificaciones detalladas anteriormente en este documento.

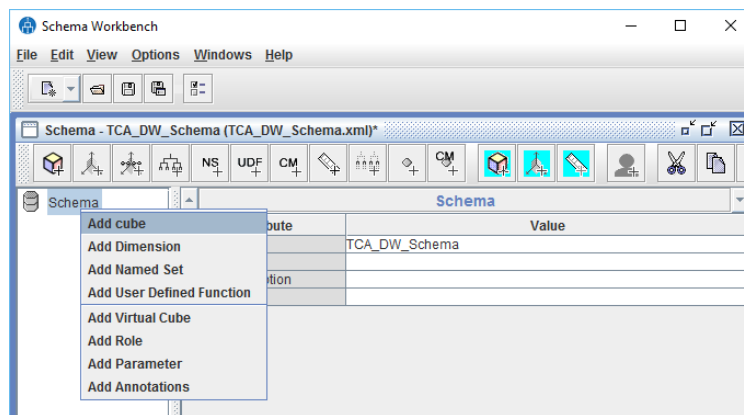


Ilustración 9 – Pentaho Schema Workbench. Especificación de cubos y dimensiones

4.2.4. Publicación de los cubos

Para la publicación de los cubos, seguiremos las siguientes indicaciones:

- **Opciones de menú:** File → Publish

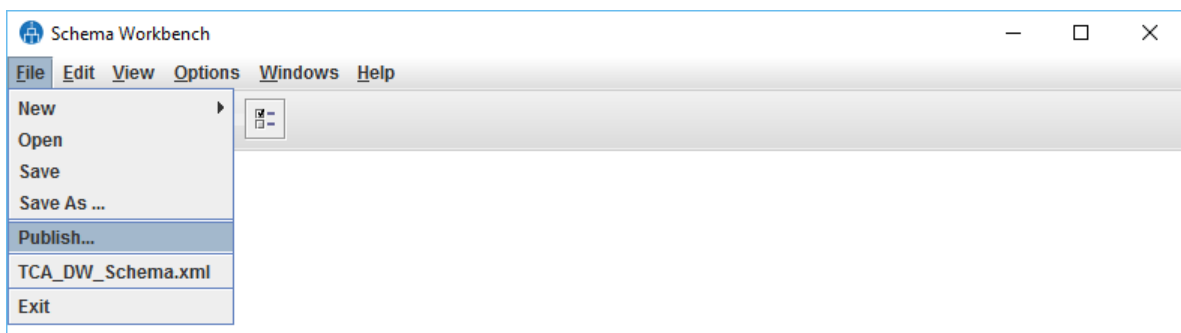


Ilustración 10 – Pentaho Schema Workbench. Publicación del schema

- **Especificación de parámetros:** se deberán cumplimentar los siguientes datos:
 - Server URL.
 - User.
 - Password.
 - Publish Settings: Pentaho or JNDI Data Source, Register XMLA Data Source.

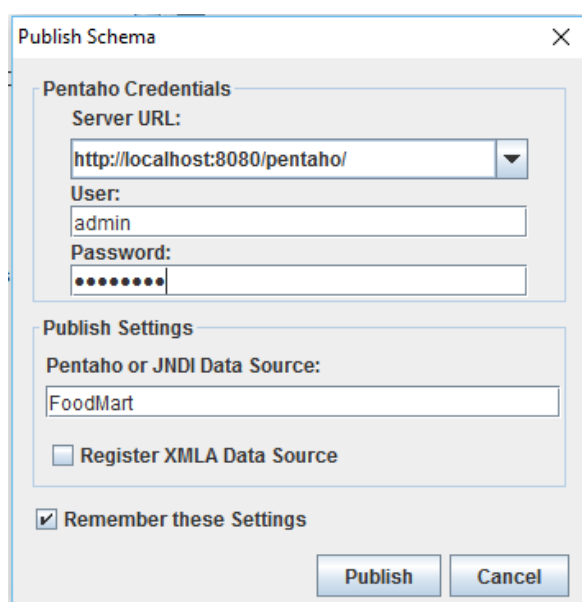


Ilustración 11 – Pentaho Schema Workbench. Parámetros de la publicación del schema

Es importante recordar que el nombre del **Pentaho or JNDI Data Source** debe coincidir con el nombre de un data source que exista en Pentaho y que sea accesible por el usuario que estamos especificando en esta misma pantalla.

4.2.5. Definición del esquema TCA_DW_SCHEMA.

Una vez realizados los pasos explicados en los apartados anteriores, hemos finalizado la definición del esquema TCA_DW_SCHEMA. En las siguientes ilustraciones, se muestra gráficamente dicho esquema:

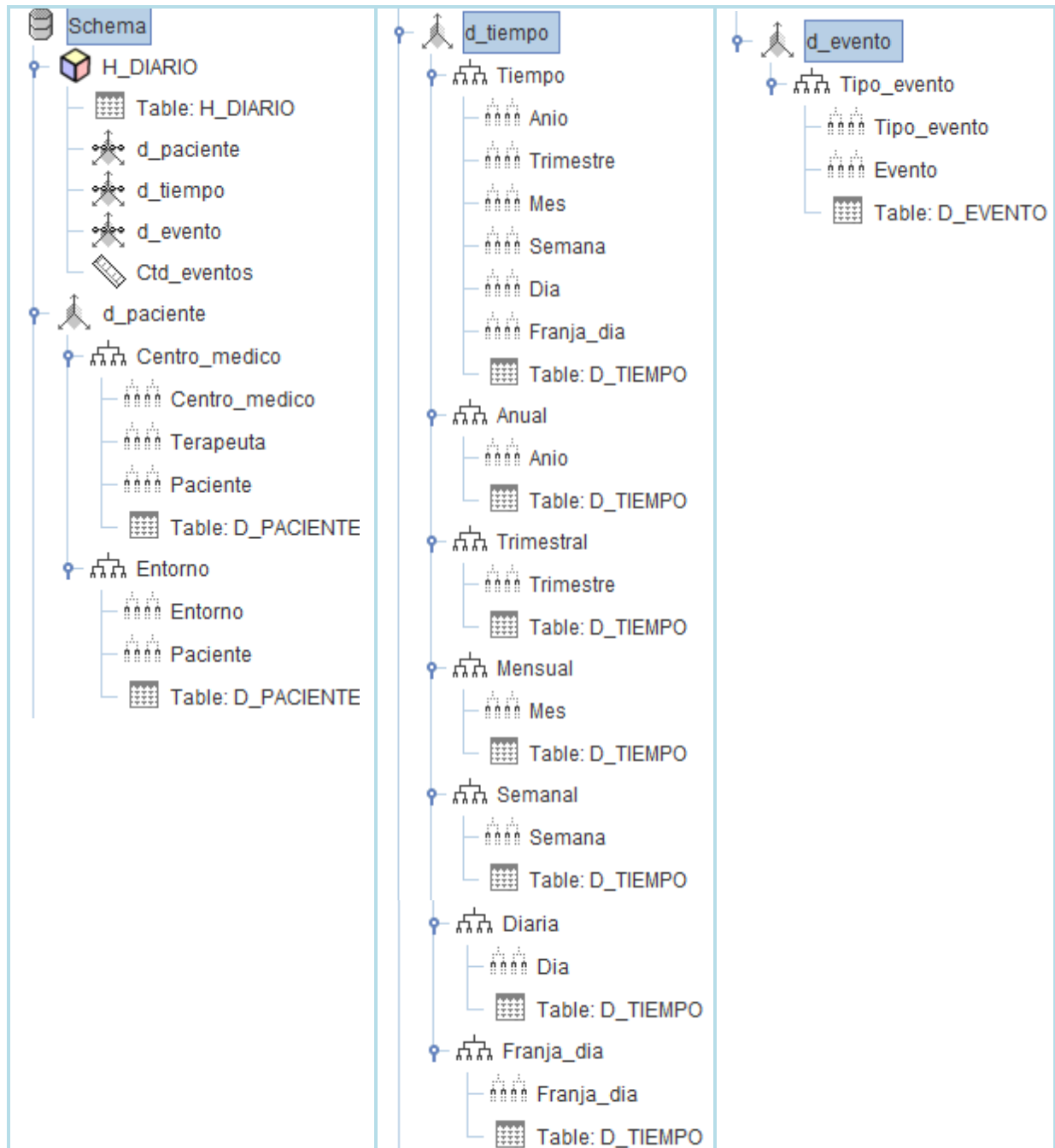


Ilustración 12 – Pentaho Schema Workbench. Schema TCA_DW_SCHEMA

En la siguiente tabla, se resume la definición del esquema TCA_DW_SCHEMA.

Schema	TCA_DW_SCHEMA		
Cubos	Nombre	Medidas	Dimensiones
	H_Diario	Número de eventos	D_Paciente D_Tiempo D_Evento

	Nombre	Jerarquía	Nivel	
Dimensiones	D_Paciente	Centro_medico	Centro_medico	
			Terapeuta	
			Paciente	
		Entorno	Entorno	
	Paciente			
	D_Tiempo	Tiempo	Anio	
			Trimestre	
			Mes	
			Semana	
			Dia	
			Franja_dia	
			Anual	Anio
			Trimestral	Trimestre
			Mensual	Mes
			Semanal	Semana
	D_Evento	Tipo_evento	Tipo_evento	
			Evento	
Evento				

El fichero XML que contiene la definición del esquema se incluye en el **Anexo V. Fichero XML de definición del cubo OLAP.**

4.3. Navegación por los cubos OLAP.

4.3.1. Navegación por los cubos con *Jpivot*.

Jpivot es una librería que permite la navegación por los cubos OLAP, y que soporta las operaciones OLAP típicas. El servidor OLAP que utiliza *Jpivot* es Mondrian.

Vamos a utilizar *Jpivot* para realizar una primera navegación por nuestro cubo. Esta librería viene incluida en la versión de Pentaho que vamos a utilizar, la 6.0.1. El primer paso es acceder a Pentaho, introduciendo en el navegador la URL de acceso. Si la instalación es local al equipo que estamos utilizando, la url de acceso es la siguiente:

<http://localhost:8080/pentaho/Login>

La pantalla principal que nos muestra Pentaho es la siguiente:



Ilustración 13 - Pantalla de acceso a Pentaho

Debemos introducir el usuario y la clave de acceso pertinentes para acceder a esta pantalla:

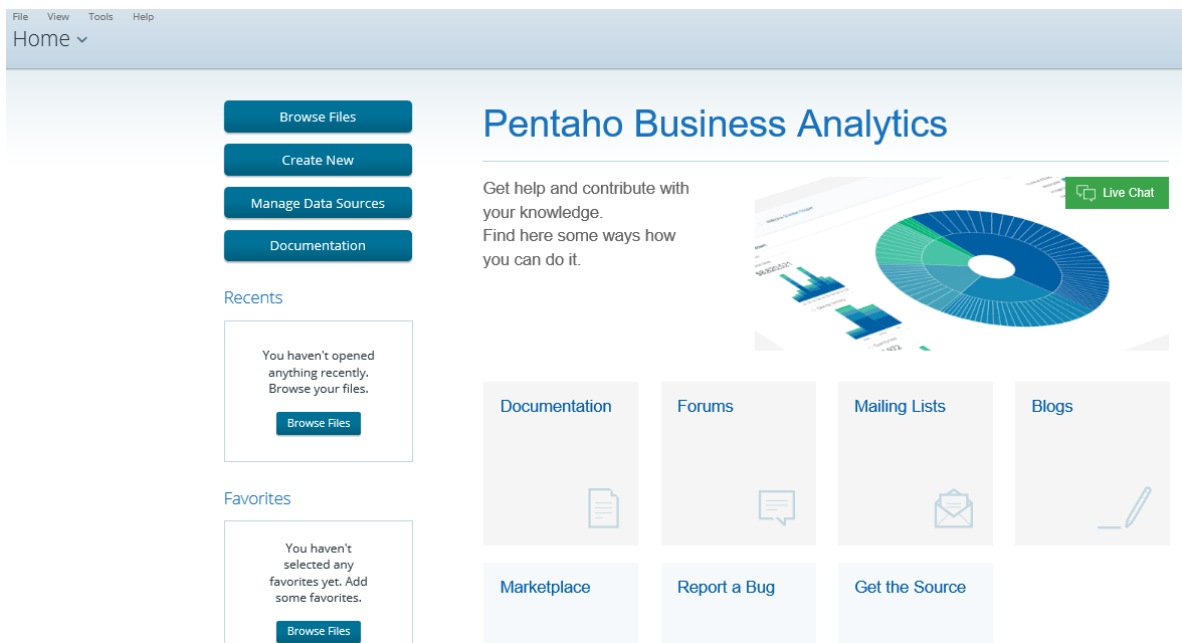


Ilustración 14 - Pantalla inicial de Pentaho

Mediante los botones **Create New** → **JPivot View** o la opción de menú **File** → **New** → **JPivot View**, accederemos a la siguiente pantalla:

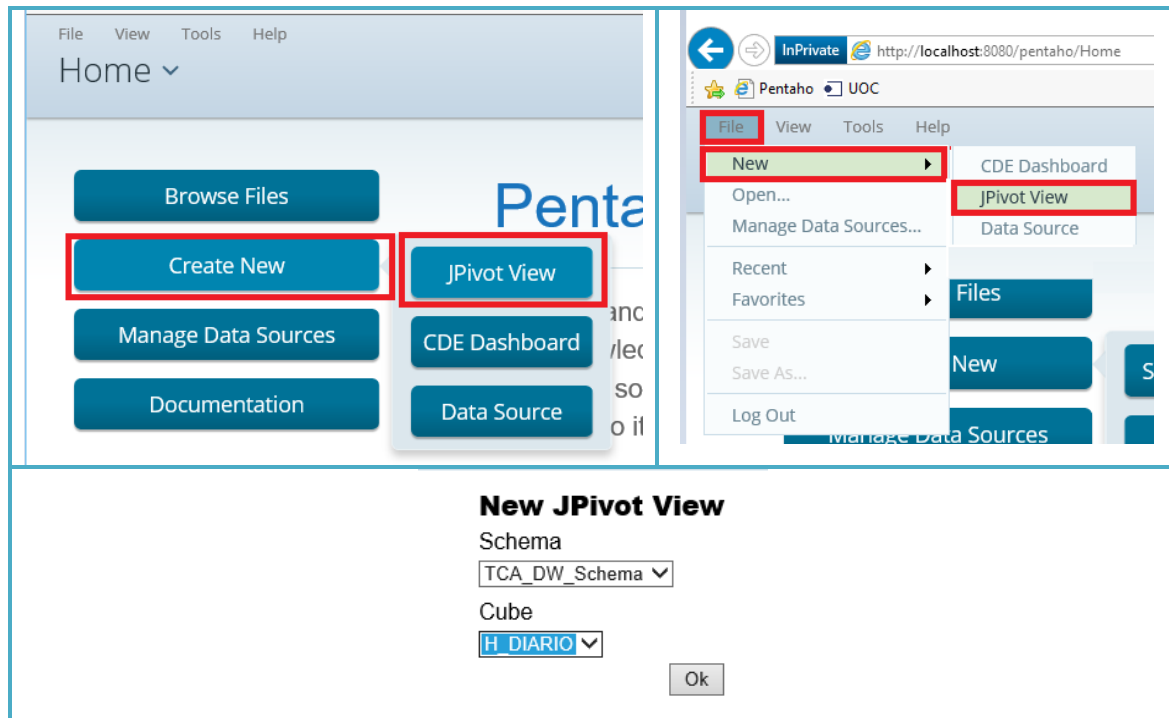


Ilustración 15 - Opciones de llamada y pantalla. Jpivot.

En los desplegables mostrados, seleccionaremos nuestro esquema y nuestro cubo y pulsaremos el botón **OK**. De esta forma, se nos mostrará la siguiente pantalla:

Centro_medico	Tiempo	Tipo_evento	Medidas
All d_paciente.Centro_medicos	All d_tiempo.Tiempos	All d_evento.Tipo_eventos	• Ctd_eventos 26.282
ARANDA TERAPEUTAS	All d_tiempo.Tiempos	All d_evento.Tipo_eventos	2.603
T13	All d_tiempo.Tiempos	All d_evento.Tipo_eventos	1.734
P19	All d_tiempo.Tiempos	All d_evento.Tipo_eventos	843
	2015	All d_evento.Tipo_eventos	556
		ACTIVITY	183
		EMOTION	183
		EPISODE	7
	MEAL	183	
2016	All d_evento.Tipo_eventos	287	
P20	All d_tiempo.Tiempos	All d_evento.Tipo_eventos	891
T14	All d_tiempo.Tiempos	All d_evento.Tipo_eventos	869

Ilustración 16 - Navegación con Jpivot

Desde esta pantalla podemos seleccionar las diferentes dimensiones, niveles y jerarquías de nuestro cubo y visualizar los datos asociados a nuestra selección. *JPivot* nos permite realizar una primera aproximación a la navegación por nuestro cubo.

Sin embargo, para dar respuesta a las preguntas planteadas en este proyecto, se ha decidido utilizar una herramienta más versátil y potente, pero igualmente sencilla e intuitiva para navegar a través de los cubos: *Saiku Analytics*.

4.3.2. Navegación por los cubos con *Saiku Analytics*.

Para poder utilizar el plugin *Saiku Analytics*, hemos de acceder al *Marketplace*, a través de la opción de menú **Home** → **Marketplace** o mediante el botón **Marketplace** de la pantalla principal. En la pantalla que se nos mostrará, introduciremos *Saiku Analytics* como término de búsqueda y visualizaremos la siguiente pantalla:

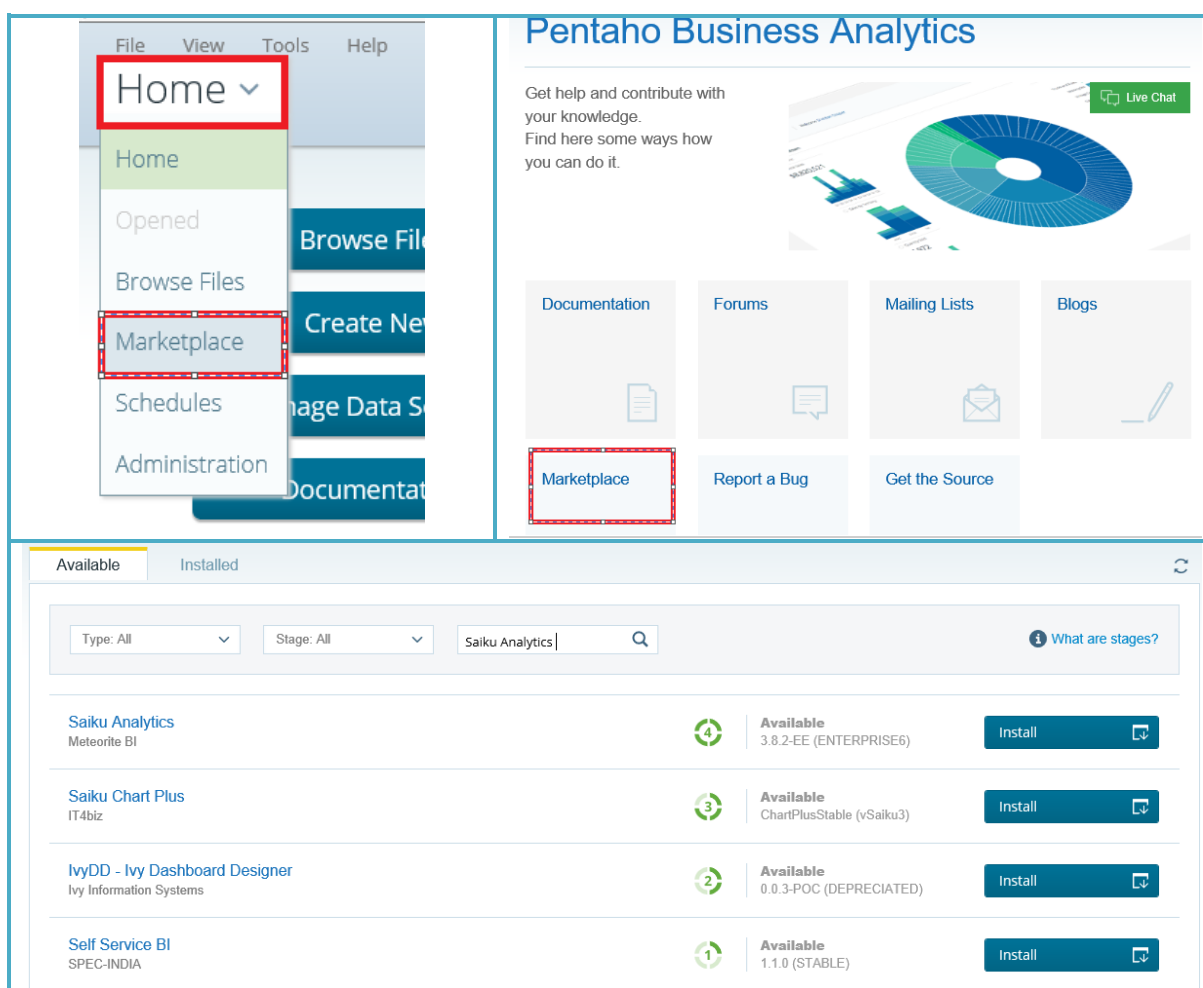


Ilustración 17 – Descarga Saiku Analytics.

Pulsando el enlace **Saiku Analytics**, nos aparecerá la siguiente pantalla, donde podremos seleccionar la versión que deseamos instalar:

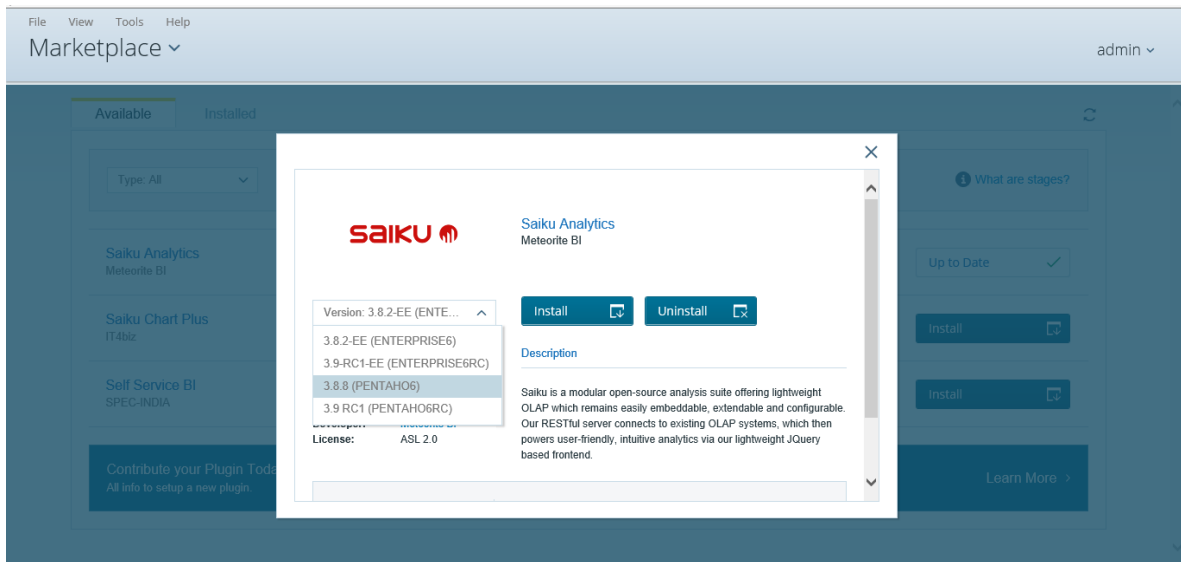


Ilustración 18 – Versiones Saiku Analytics

Sólo nos resta pulsar el botón **Install** y **OK** para terminar la instalación.

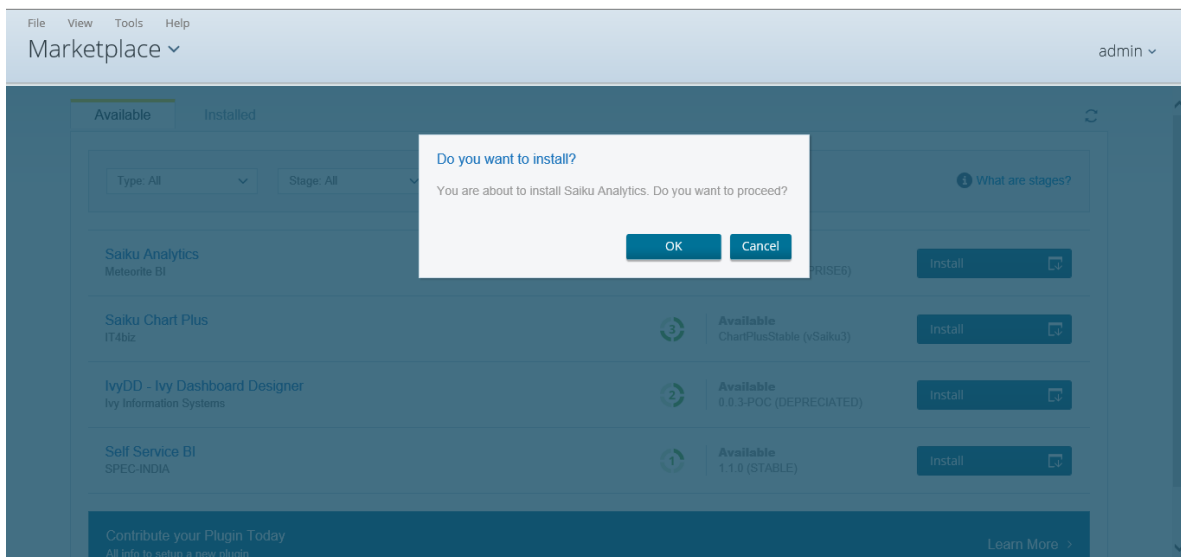


Ilustración 19 - Confirmación de descarga Saiku Analytics

Una vez instalado el plugin, debemos reiniciar el servidor BI de Pentaho para poder utilizarlo.

Tras este reinicio, a través de los botones **Create New** → **Saiku Analytics** o la opción de menú **File** → **New** → **Saiku Analytics**, accederemos a la siguiente pantalla

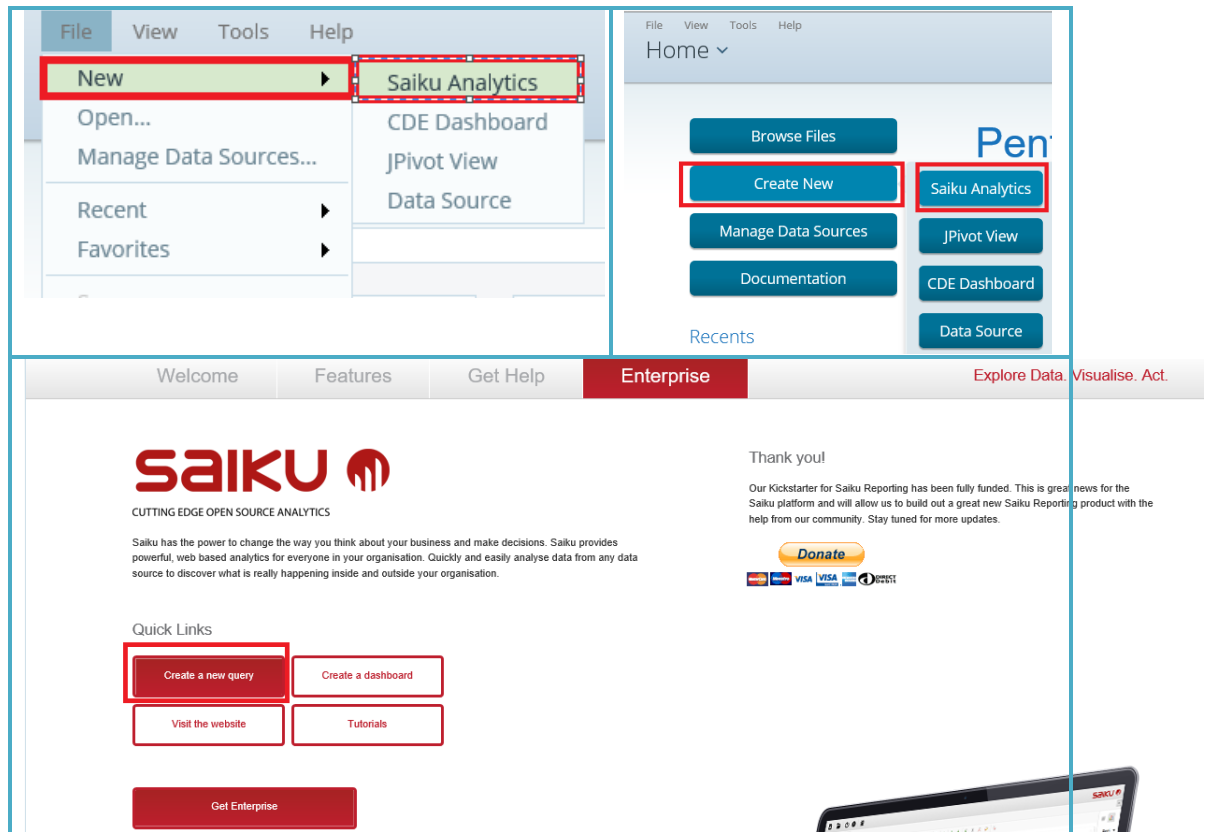


Ilustración 20 - Opciones de llamada y pantalla inicial. Saiku Analytics

La opción **Create New Query** nos llevará hasta una nueva pantalla donde podremos seleccionar el cubo, las medidas y las dimensiones que deseemos.

Como ejemplo, hemos seleccionado la medida **Ctd_eventos** y los datos mostrados son los centros médicos (filas) y los tipos y subtipos de eventos (columnas).

Centro_medico	ACTIVITY							
	ALONE	FAMILY	FRIENDS	SPORTS	STUDYING/WORKING	DISTRESSED	FUSTRATED	
ARANDA TERAPEUTAS	73	146	156	281		172	29	28
BCN TCA THERAPY	87	134	162	241		203	56	67
CSM TERRES DE L'EBRE	102	124	158	259		184	43	36
CT SEVILLA	94	128	151	261		194	42	60
CTA MERIDA	103	130	158	261		173	49	43
GAL Eating Disorder	81	136	180	246		182	33	33
GRN FOOD Disorder	94	129	153	270		181	27	45
MAD TCA Center	94	128	172	257		177	47	40
OSASUNA MUNGIA	73	141	161	278		175	19	40
TRG THERAPY CENER	71	134	166	277		177	35	26

Ilustración 21 - Creación de informe. Saiku Analytics

5. Informes y Cuadros de mandos

5.1. Introducción

Una vez se ha descrito cómo realizar la navegación por los cubos mediante *Saiku Analytics*, ya podemos comenzar a realizar la explotación de los datos que nos permitirá dar respuesta a las preguntas planteadas para este proyecto. Por cada pregunta, se generará un informe que nos permita darle respuesta.

5.2. Generación de informes

5.2.1. Informe de la relación entre actividades realizadas y episodios de crisis.

Para estudiar la relación entre las actividades realizadas y los episodios de crisis, hemos realizado la siguiente selección de parámetros:

Informe	Relación entre las actividades realizadas y los episodios de crisis	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Tipo Evento → Evento

Se ha añadido un filtro para que solamente se muestren los tipos de eventos **activity** y **episode**. Los datos obtenidos y su representación gráfica son los siguientes:



Ilustración 22 - Relación entre las actividades realizadas y los episodios de crisis I

Los datos referentes a las actividades nos permiten establecer los siguientes grupos considerando la cantidad realizada:

- Grupo 1: Sports: se aprecia que es la actividad más realizada.
- Grupo 2: Studying/Working – Friends: estas dos actividades son cercanas en cuanto a su realización.
- Grupo 3: Family: la ocurrencia de esta actividad se sitúa entre el grupo 2 y el grupo 4, el de la actividad menos realizada.
- Grupo 4: Alone: esta actividad es la menos realizada.

Con estas consideraciones iniciales, podemos comenzar a relacionar las actividades realizadas con los episodios de crisis, estableciendo las siguientes relaciones en base a la cantidad de eventos de cada categoría. Las actividades con un mayor número de ocurrencia las vincularemos con los episodios de mayor número de ocurrencias, obteniendo las siguientes conclusiones:

- Al realizar las actividades Sports, Studying/Working y Friends, lo más probable es que no suceda ningún episodio. Las ocurrencias de estas

tres actividades ascienden aproximadamente a 6000 y la ocurrencia de no episodios a 6800 aproximadamente.

- Actividad Family: dado que queda una parte por asignar de la ausencia de episodios y que la ocurrencia de la actividad Family es casi el doble que la actividad Alone, podemos decir que lo más probable es que la ausencia de episodios esté vinculada a esta actividad. Por tanto, podemos concluir que para esta actividad pueden ocurrir dos posibilidades:
 - Ausencia de episodios.
 - Episodios: dado que los tres episodios (restriction, vomited, binge) tienen ocurrencias similares, no podemos establecer cuál de estos episodios será el más probable.
- Al realizar la actividad Alone, probablemente ocurrirá uno de los tres episodios (restriction, vomited, binge). Para verificar si alguno de ellos se produce más habitualmente que los otros estudiaremos los datos con un mayor nivel de detalle.

Vamos a añadir el nivel de la franja del día, de la dimensión tiempo, para verificar estas primeras relaciones.

Informe	Relación entre las actividades realizadas y los episodios de crisis	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento
Columnas	Dimensión	D_tiempo
	Jerarquía	Franja_dia
	Nivel	Franja_dia
Columnas	Dimensión	D_tiempo
	Jerarquía	Anio
	Nivel 2	Anio
Columnas	Dimensión	D_tiempo
	Jerarquía	Mes
	Nivel 2	Mes

Los datos obtenidos y su representación gráfica son los siguientes:

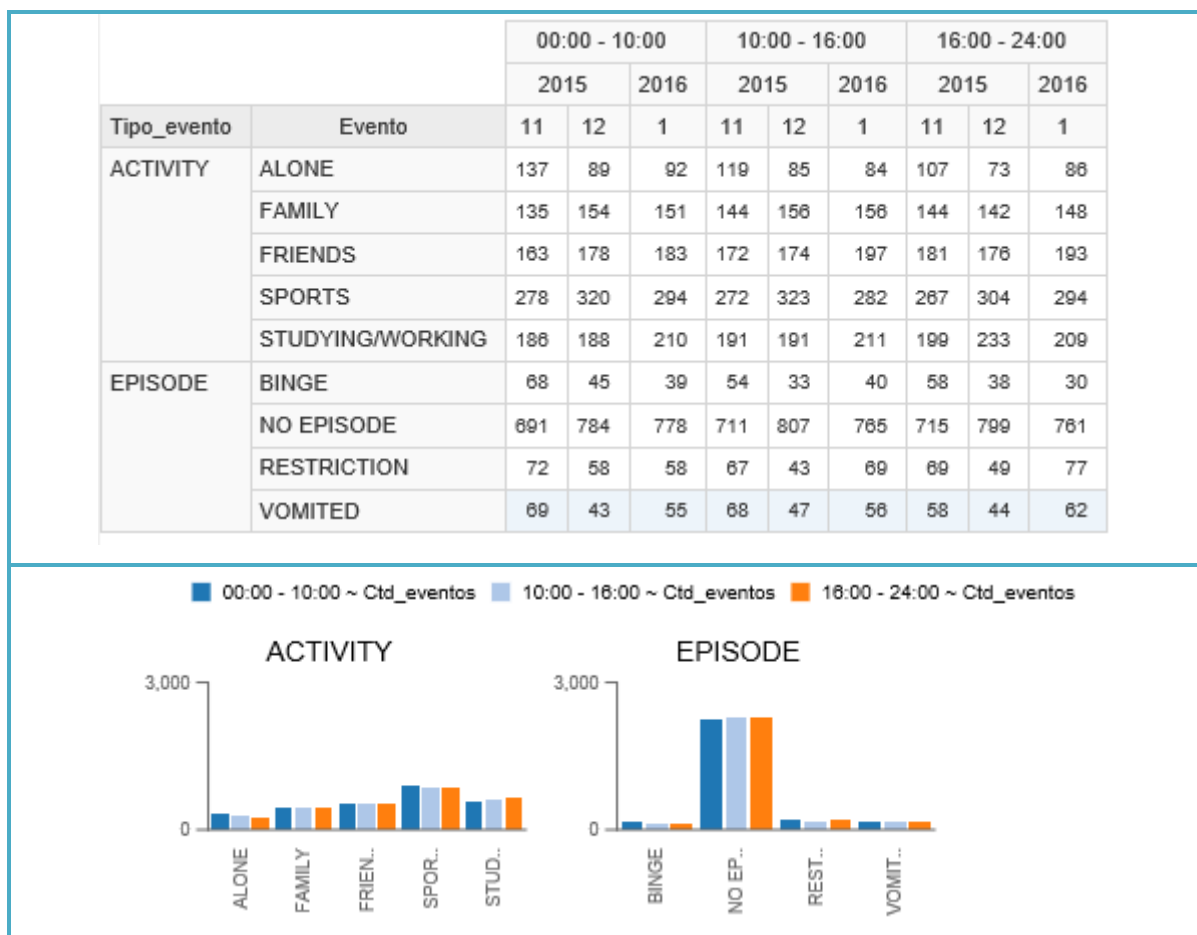


Ilustración 23 - Relación entre las actividades realizadas y los episodios de crisis II

Podemos concluir que las tendencias detectadas se confirman, ya que para cada franja del día las cantidades de cada una de las actividades y cada uno de los episodios se mantienen estables.

Para la actividad Alone, se detecta que en la última franja del día es menos habitual que durante las otras dos franjas, siendo la actividad que se reduce en mayor cantidad a lo largo del día; observando los tipos de episodios, podemos decir que el que experimenta un descenso progresivo es Binge, siendo la última franja del día la que menos episodios Binge registra. Por tanto, podemos decir que el episodio más frecuente para la actividad Alone es Binge.

La actividad Family no experimenta cambios significativos que nos permitan extraer conclusiones.

Vamos a estudiar con mayor nivel de detalle las actividades menos frecuentes (alone, family, friends) y los episodios menos frecuentes (restriction, vomited, binge), para intentar detectar alguna tendencia. Para ello, añadiremos el nivel

del año y mes, de la dimensión tiempo y un filtro para mostrar las actividades y episodios que queremos estudiar.

Informe	Relación entre las actividades realizadas y los episodios de crisis	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel 1	Evento
Columnas	Dimensión	D_tiempo
	Jerarquía	Franja_dia
	Nivel 1	Franja_dia
Columnas	Dimensión	D_tiempo
	Jerarquía	Anio
	Nivel 2	Anio
Columnas	Dimensión	D_tiempo
	Jerarquía	Mes
	Nivel 2	Mes

Los datos obtenidos y su representación gráfica son los siguientes:

		00:00 - 10:00			10:00 - 16:00			16:00 - 24:00		
		2015		2016	2015		2016	2015		2016
		11	12	1	11	12	1	11	12	1
Tipo_evento	Evento	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos
ACTIVITY	ALONE	137	89	92	119	85	84	107	73	86
	FAMILY	135	154	151	144	158	158	144	142	148
	FRIENDS	163	178	183	172	174	197	181	176	193
		435	421	426	435	415	437	432	391	427
EPISODE	BINGE	68	45	39	54	33	40	58	38	30
	RESTRICTION	72	58	58	67	43	69	69	49	77
	VOMITED	69	43	55	68	47	58	58	44	62
		209	146	152	189	123	165	185	131	169

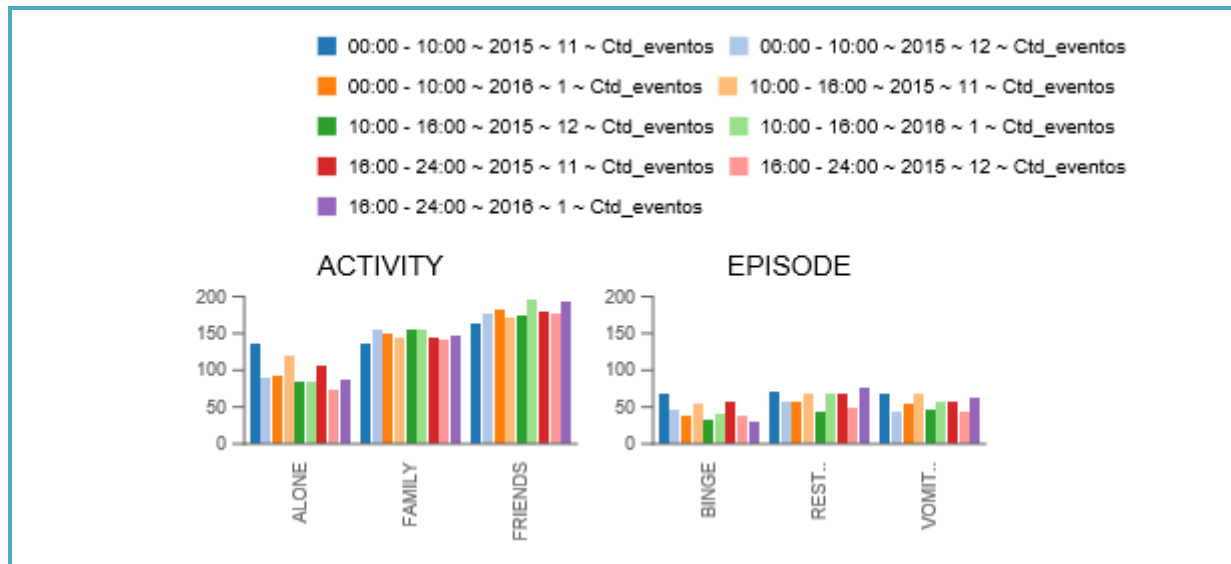


Ilustración 24 - Relación entre las actividades realizadas y los episodios de crisis III

Podemos ver nuevamente que, las fluctuaciones de la actividad Alone coinciden con las fluctuaciones del episodio Binge, por lo que se confirma que es el episodio más probable al practicar esta actividad.

Para las actividades Family y Friends observamos que sus cifras y fluctuaciones tienen un alto grado de similitud; sin embargo, no detectamos similitud con los episodios. Por tanto, las conclusiones finales son:

- Al realizar las actividades Sports, Studying/Working, lo más probable es que no suceda ningún episodio. Las ocurrencias de estas tres actividades ascienden aproximadamente a 6000 y la ocurrencia de no episodios a 6800 aproximadamente.
- Al realizar las actividades Family y Friends, pueden ocurrir dos posibilidades:
 - Ausencia de episodios.
 - Episodios: no se aprecia una tendencia para la ocurrencia de un episodio concreto.
- Al realizar la actividad Alone, probablemente ocurrirá uno de los tres episodios (restriction, vomited, binge). El episodio con mayor probabilidad de ocurrencia será Binge.

5.2.2. Informe de la relación entre los estados de ánimo y los episodios de crisis.

Para estudiar la relación entre los estados de ánimo y los episodios de crisis, hemos realizado la siguiente selección de parámetros:

Informe	Relación entre los estados de ánimo y los episodios de crisis	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento
Columnas	Dimensión	D_tiempo
	Jerarquía	Franja_dia
	Nivel	Franja_dia

Se ha añadido un filtro para que solamente se muestren los tipos de eventos **emotion** y **episode**. Los datos obtenidos y su representación gráfica son los siguientes:

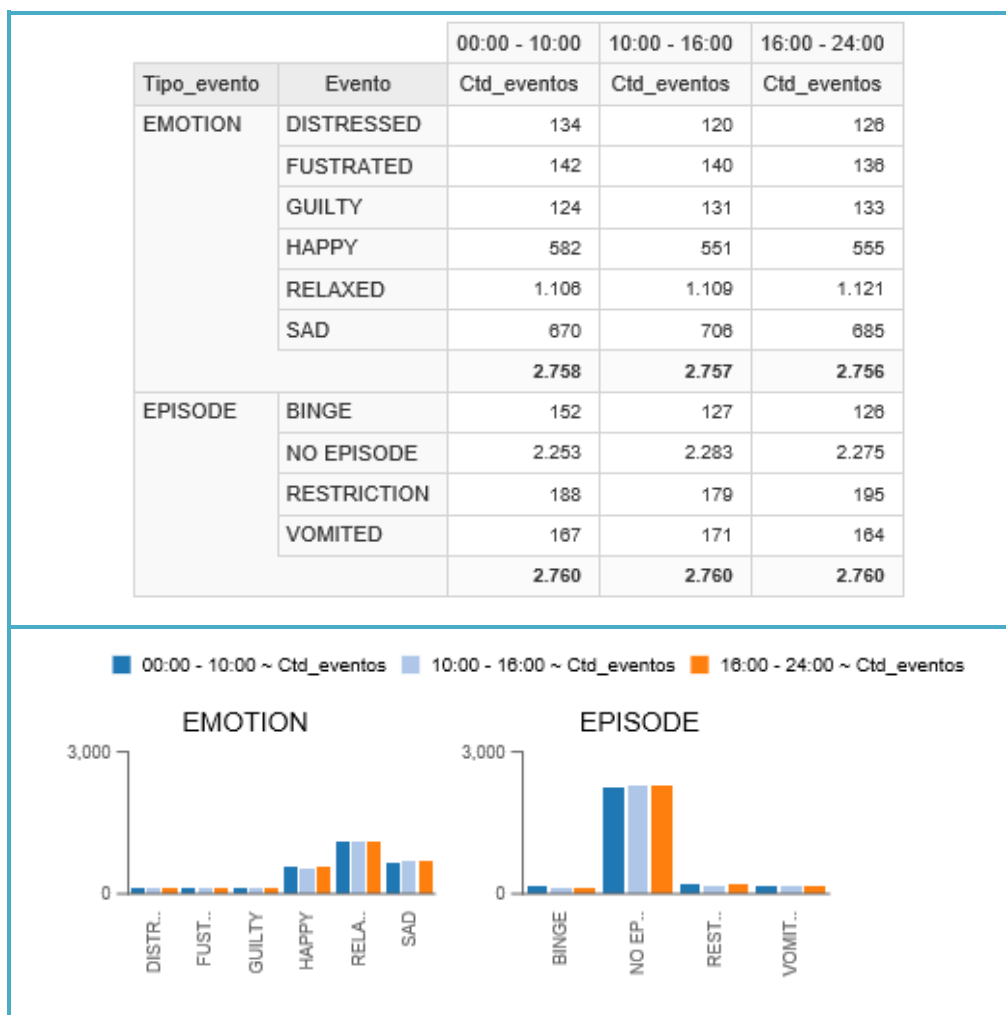


Ilustración 25 - Relación entre los estados de ánimo y los episodios de crisis I

Los datos referentes a los estados de ánimo nos permiten establecer los siguientes grupos considerando la cantidad realizada:

- Grupo 1: Relaxed, Sad, Happy: estos tres estados de ánimo son los más habituales. Si los relacionamos con el episodio más habitual, obtenemos que para estos tres estados de ánimo lo más probable es que no suceda ningún episodio.
- Grupo 2: Distressed, Frustrated, Guilty: estos tres estados de ánimo, además de ser los menos habituales, tienen tasas de ocurrencias muy similares. Por tanto, podemos deducir que lo más probable es que estos tres estados de ánimo produzcan un episodio Binge, Restriction, Vomited. Para poder determinar qué episodio es el más habitual en función del estado de ánimos vamos a seguir estudiando los datos.

Para ello, añadiremos el nivel del año y mes, de la dimensión tiempo y un filtro para mostrar los estados de ánimo y episodios que queremos estudiar.

Informe	Relación entre los estados de ánimo y los episodios de crisis	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento
Columnas	Dimensión	D_tiempo
	Jerarquía	Franja_dia
	Nivel 1	Franja_dia
Columnas	Dimensión	D_tiempo
	Jerarquía	Anio
	Nivel 2	Anio
Columnas	Dimensión	D_tiempo
	Jerarquía	Mes
	Nivel 3	Mes

Los datos obtenidos y su representación gráfica son los siguientes:



Ilustración 26 - Relación entre los estados de ánimo y los episodios de crisis II

Estos datos tienen un frecuencia de ocurrencia muy similar, tanto para los estados de ánimo como para los episodios; por tanto, no nos permiten extraer conclusiones para determinar qué episodio es el más probable según el estado de ánimo.

Por tanto, las conclusiones finales son:

- Al experimentar los estados de ánimo Relaxed, Sad, Happy, lo más probable es que no suceda ningún episodio.
- Al experimentar los estados de ánimo Distressed, Frustrated, Guilty, lo más probable es que ocurra un episodio Binge, Restriction, Vomited. No es posible determinar qué episodio es el probable en función del estado de ánimo.

5.2.3. Informe de relación entre el entorno geográfico y los episodios de crisis.

Para estudiar la relación entre el entorno geográfico y los episodios de crisis, hemos realizado la siguiente selección de parámetros:

Informe	Relación entre el entorno geográfico y los episodios de crisis	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento
Columnas	Dimensión	D_tiempo
	Jerarquía	Franja_dia
	Nivel 1	Franja_dia
Columnas	Dimensión	D_paciente
	Jerarquía	Entorno
	Nivel 2	Entorno

Se ha añadido un filtro para que solamente se muestren los tipos de eventos **episode**. Los datos obtenidos y su representación gráfica son los siguientes:

		00:00 - 10:00		10:00 - 16:00		16:00 - 24:00	
		RURAL	URBAN	RURAL	URBAN	RURAL	URBAN
Tipo_evento	Evento	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos
EPISODE	BINGE	81	71	64	63	62	64
	NO EPISODE	1.117	1.138	1.143	1.140	1.153	1.122
	RESTRICTION	137	51	113	66	121	74
	VOMITED	45	122	60	111	44	120
		1.380	1.380	1.380	1.380	1.380	1.380

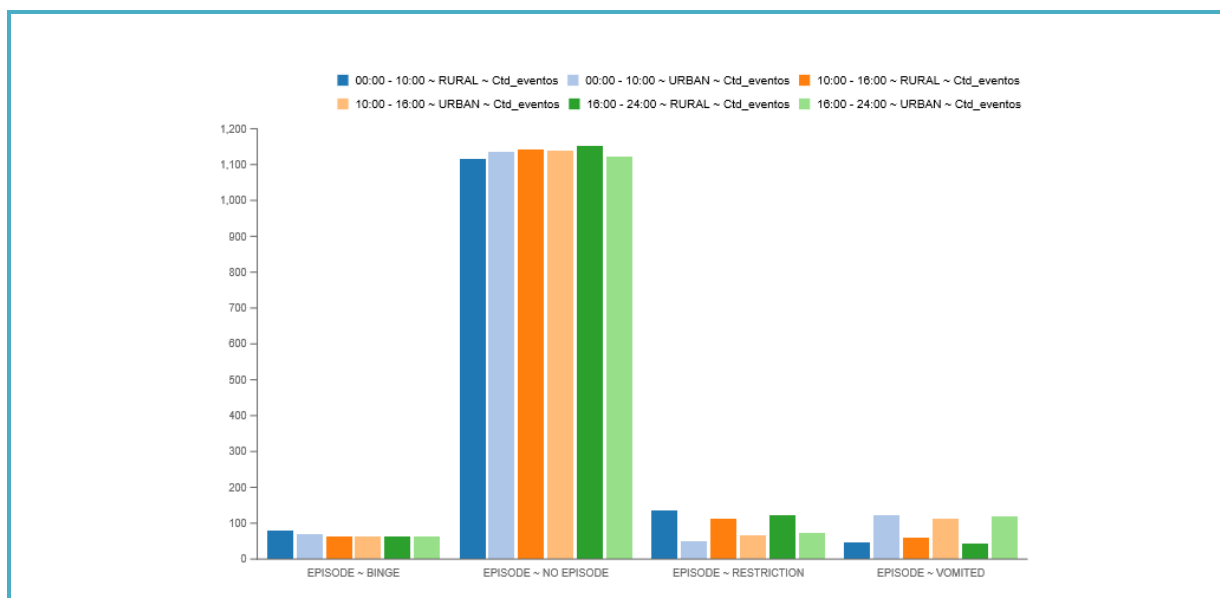


Ilustración 27 - Relación entre el entorno geográfico y los episodios de crisis I

Los datos referentes a los episodios de crisis nos permiten establecer los siguientes grupos considerando la cantidad realizada:

- Grupo 1: No episode: la ausencia de episodios de crisis es el evento más numeroso. Con estos datos, se aprecia que no existe diferencia significativa de este evento en función del entorno del paciente.
- Grupo 2: Binge: este episodio de crisis registra una ocurrencia parecida en ambos entornos, el rural y el urbano.
- Grupo 3: Restriction: este episodio sí presenta diferencias significativas entre ambos entornos; mientras que el entorno rural registra un número alto de ocurrencias en las tres franjas de día, en el entorno urbano este episodio se produce en un número mucho más bajo, siendo inferior a la mitad de los episodios rurales de la primera franja de la mañana. Por tanto, podemos afirmar que en el entorno rural los episodios Restriction son más frecuentes que en el urbano, alcanzando la mayor diferencia en la primera franja de la mañana.
- Grupo 4: Vomited: este episodio presenta una mayor ocurrencia en entornos urbanos que en rurales. En la primera y la tercera franja del día, el número de episodio urbano triplica el número de episodios rurales.

Vamos a analizar los datos con un mayor nivel de detalle para validar las conclusiones anteriores. Para ello, añadiremos el nivel del año y mes, de la dimensión.

Informe	Relación entre el entorno geográfico y los episodios de crisis	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_tiempo
	Jerarquía	Anio
	Nivel 1	Anio
Filas	Dimensión	D_tiempo
	Jerarquía	Mes
	Nivel 2	Mes
Filas	Dimensión	D_tiempo
	Jerarquía	Franja_dia
	Nivel 3	Franja_dia
Filas	Dimensión	D_paciente
	Jerarquía	Entorno
	Nivel 4	Entorno
Columnas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento

Los datos obtenidos y su representación gráfica son los siguientes:

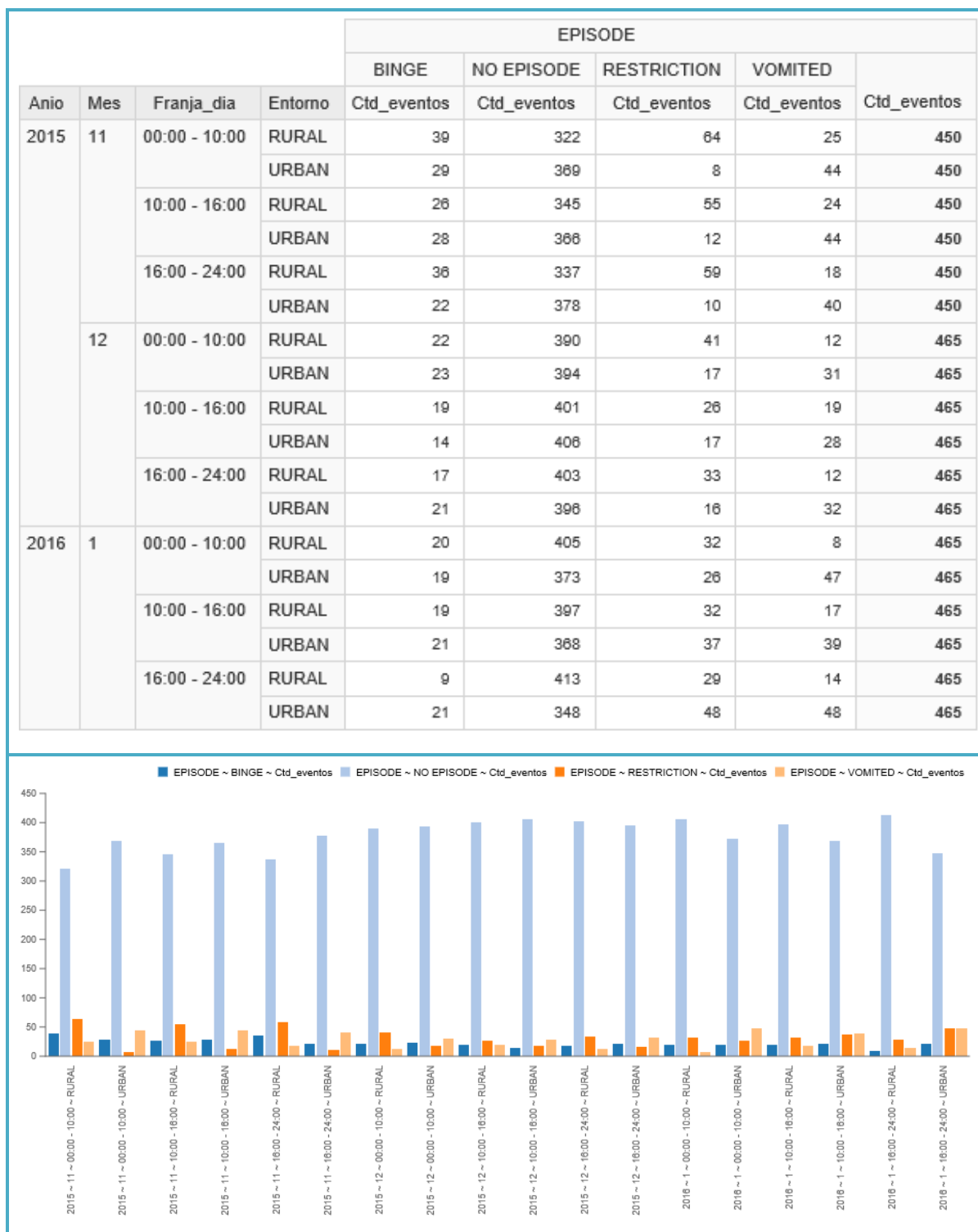


Ilustración 28 - Relación entre el entorno geográfico y los episodios de crisis II

Como se puede observar, se ratifica que el evento más numeroso en todos los casos en la ausencia de episodios. Asimismo, para el episodio Binge, también se confirma que el número de ocurrencias es similar en ambos entornos.

Para los episodios Restriction se registran más casos en el entorno rural que en el urbano, excepto en el mes de enero de 2016: para este mes, en la segunda y tercera franja del día, el número de episodios es superior en el

entorno rural. Esto puede ser debido a la evolución de los paciente, pero sería necesario contar con datos de meses subsiguientes para apreciar hacia que entorno se decanta el mayor número de episodios.

Para el episodio Vomited, vemos que se sigue cumpliendo la tendencia detectada: el número de episodio urbano es bastante superior al número de episodios rurales.

5.2.4. Informe de evolución de pacientes.

Para estudiar la evolución de los pacientes, más fácilmente estos datos, agrupado a los pacientes por su entorno y hemos realizado la siguiente selección de parámetros:

Informe	Evolución de pacientes	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento
Columnas	Dimensión	D_paciente
	Jerarquía	Entorno
	Nivel 1	Entorno
Columnas	Dimensión	D_tiempo
	Jerarquía	Anio
	Nivel 2	Anio
Columnas	Dimensión	D_tiempo
	Jerarquía	Mes
	Nivel 3	Mes

Los datos obtenidos y su representación gráfica son los siguientes:

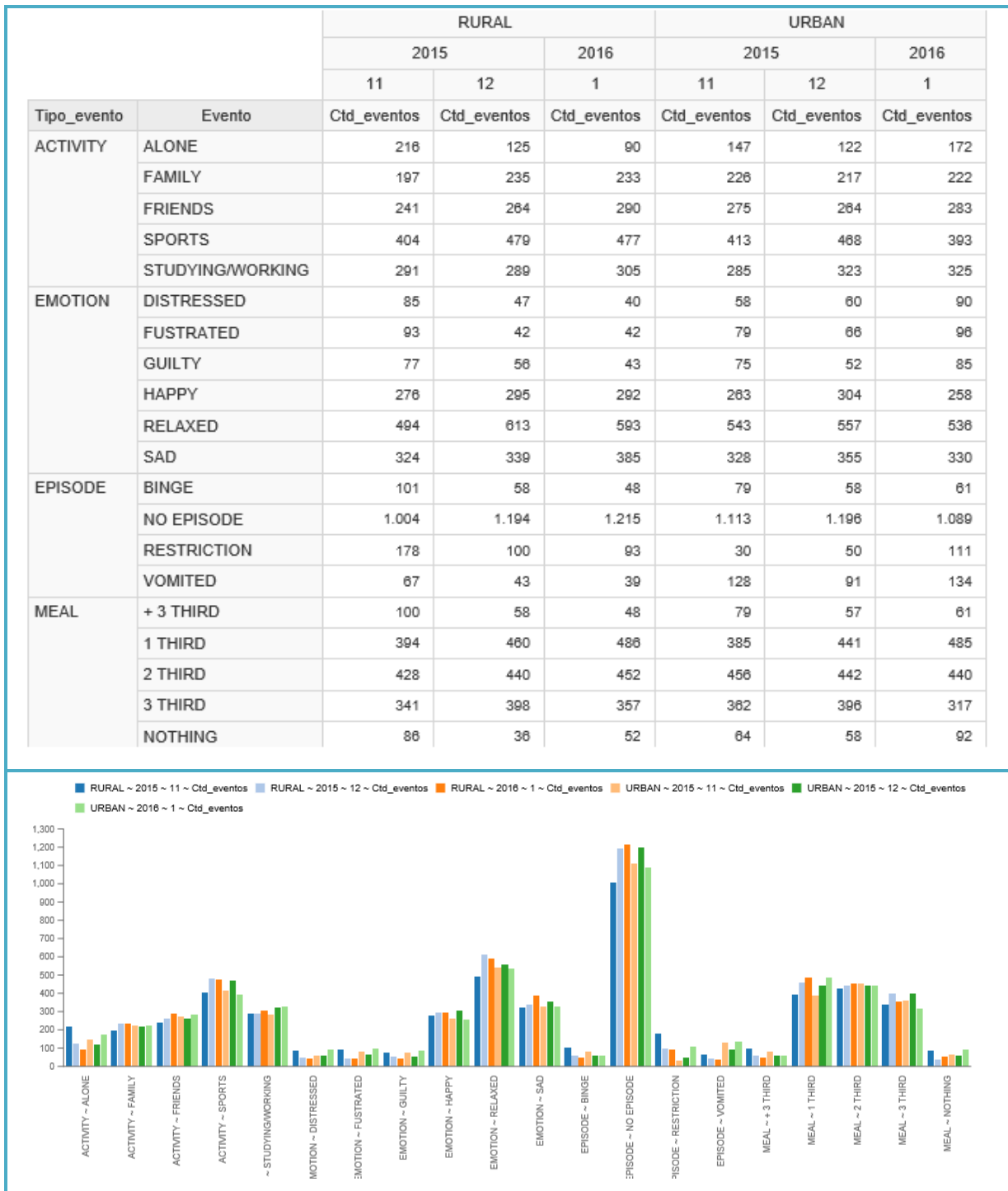


Ilustración 29 - Evolución de pacientes I

Estos datos nos permiten realizar las siguientes observaciones:

- Entorno rural:
 - Actividades: la única actividad que se ha ido reduciendo a lo largo del tiempo es Alone. El resto, Family, Friends, Sports y Studying/Working, han ido aumentando de forma continuada.

En el último mes, la actividad con mayor número de ocurrencias es Sports. La de menor número de ocurrencias es Alone y después Family.

- Emotion: los estados de ánimo Distressed, Frustrated y Guilty se han ido reduciendo a lo largo del tiempo. El resto de estados de ánimo Happy, Relaxed y Sad han ido aumentando de forma continuada.

En el último mes, el estado de ánimo con mayor número de ocurrencias es Relaxed. Los de menor número de ocurrencias son Distressed, Frustrated y Guilty, que registran cifras muy parecidas.

- Episode: los episodios Binge, Restriction y Vomited han ido disminuyendo a lo largo del tiempo.

En el último mes, el episodio con mayor número de ocurrencias es Restriction. El de menor número de ocurrencias Vomited.

- Meal: las ingestas Nothing y +3 Third se han ido reduciendo a lo largo del tiempo. Las ingestas, 1 Third, 2 Third han ido aumentando de forma continuada. La ingesta 3 Third ha sufrido fluctuaciones, pero la cantidad final es superior a la inicial, por lo que se engloba en la categoría de ingestas que han aumentado.

En el último mes, la ingesta con mayor número de ocurrencias es 1 Third. La de menor número de ocurrencias es 3 Third.

De estas observaciones para el entorno rural, podemos concluir:

- Realizar actividades que permitan al paciente interactuar con otras personas, reduce los sentimientos negativos (Distressed, Frustrated y Guilty) y potencia los positivos (Happy, Relaxed).
- A pesar de que el sentimiento Sad también ha aumentado, se aprecia que se ha reducido el número de episodios y que la no ingesta de comida también se ha reducido.
- En términos generales, la evolución de los pacientes ha sido satisfactoria, ya que se han incrementado y potenciado los sentimientos positivos y las ingestas, y se han reducido los episodios.
- Entorno urbano:
 - Actividades: la única actividad que se ha ido reduciendo significativamente a lo largo del tiempo es Sports. Las actividades

Family y Friends han sufrido reducciones tan bajas, que hemos considerado estable su número. Las actividades Alone y Studying/Working, han ido aumentando, aunque no siempre de forma continuada.

En el último mes, la actividad con mayor número de ocurrencias es Sports. La de menor número de ocurrencias es Alone.

- Emotion: los estados de ánimo Happy, Relaxed y Sad se han reducido en una cantidad muy pequeña, por lo que consideramos que se mantienen estables a lo largo del tiempo. Los estados de ánimo Distressed, Frustrated y Guilty han aumentando de forma significativa.

En el último mes, la actividad con mayor número de ocurrencias es Relaxed. La de menor número de ocurrencias es Guilty.

- Episode: el episodio Binge disminuye ligeramente, por lo que consideramos que se mantiene estable. Restriction aumenta significativamente y Vomited se mantiene estable.

En el último mes, el episodio con mayor número de ocurrencias es Vomited. El de menor número de ocurrencias Binge.

- Meal: las ingestas +3 Third, 3 Third y 2 Third se han ido reduciendo, con alguna fluctuación puntual. Las ingestas 1 Third y Nothing, han aumentado.

En el último mes, la ingesta con mayor número de ocurrencias es 1 Third. La de menor número de ocurrencias es 3 Third.

De estas observaciones para el entorno urbano, podemos concluir:

- A pesar de que la actividad Sports sea la más realizada, ha sufrido un decremento a lo largo de la evolución de los pacientes. Sin embargo, la actividad Alone se ha incrementado. Analizando estas circunstancias junto con el aumento de los episodios Restriction y la estabilidad de Binge y Vomited, podemos concluir que los pacientes no están evolucionando satisfactoriamente. Asimismo, las actividades realizadas por los pacientes no les están reportando sentimientos positivos, ya que han sido los negativos los que han registrado un incremento. También se ve reflejado en las ingestas, aumentando la ingesta Nothing y la 1 Third.

- En términos generales, la evolución de los pacientes ha sido poco significativa. Los sentimientos negativos no se han visto superados a través de la actividades realizadas. Podemos decir que dejar de practicar la actividad Sport y practicar Alone incide en un aumento de los episodios y una ingesta de comida pobre.

5.2.5. Informe de evolución de pacientes por centro médico y terapeuta.

Para estudiar la evolución de los pacientes según el centro médico hemos realizado la siguiente selección de parámetros:

Informe	Evolución de pacientes por centro médico	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento
Columnas	Dimensión	D_paciente
	Jerarquía	Centro_medico
	Nivel 1	Centro médico

Los datos obtenidos y su representación gráfica son los siguientes:

		ARANDA TERAPEUTAS	BCN TCA THERAPY	CSW TERRES DE L'EBRE	CT SEVILLA	CTA MERIDA	GAL Eating Disorder	GRN FOOD Disorder	MAD TCA Center	OSASUNA MUNGIA	TRG THERAPY CENTER
Tipo_evento	Evento	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos	Ctd_eventos
ACTIVITY	ALONE	73	87	102	94	103	81	94	94	73	71
	FAMILY	146	134	124	128	130	136	126	128	141	134
	FRIENDS	156	162	158	151	158	180	153	172	161	166
	SPORTS	281	241	259	261	261	249	270	257	278	277
	STUDYING/WORKING	172	203	184	194	173	182	181	177	175	177
EMOTION	DISTRESSED	29	56	43	42	49	33	27	47	19	35
	FUSTRATED	28	67	36	60	43	33	45	40	40	26
	GUILTY	22	53	40	51	45	37	40	41	27	32
	HAPPY	181	137	166	142	157	182	176	190	176	178
	RELAXED	346	305	339	344	322	331	315	307	366	361
	SAD	222	209	203	189	209	212	221	203	200	193
EPISODE	BINGE	40	31	51	42	41	41	41	44	34	40
	NO EPISODE	709	699	653	640	649	664	674	675	707	711
	RESTRICTION	29	81	93	106	95	33	26	42	19	35
	VOMITED	50	17	31	40	43	60	84	67	68	42
MEAL	+ 3 THIRD	40	30	51	42	40	41	41	44	34	40
	1 THIRD	281	243	260	282	249	271	275	264	257	269
	2 THIRD	244	276	271	249	274	273	270	258	264	259
	3 THIRD	234	212	212	213	206	216	201	232	216	229
	NOTHING	29	66	33	42	56	27	40	30	37	28

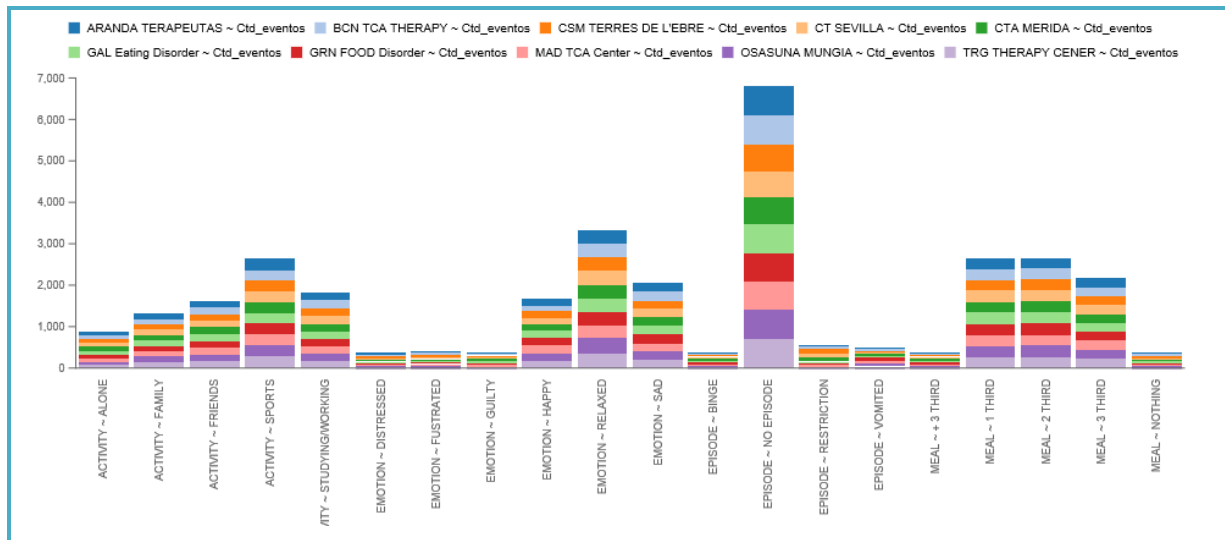


Ilustración 30 - Evolución de pacientes por centro médico I

Vamos a revisar los episodios registrados para valorar la evolución de los pacientes por centro médico. Hemos elegido el episodio Restriction por ser el más numeroso; vamos a seleccionar con los dos centros que tengan más episodios y con los dos centros que tengan menos episodios:

- Mayor número de episodios: estos centros son los que tienen una mala evolución de pacientes. Son los siguientes:
 - **CT SEVILLA:** 106 episodios de Restriction.
 - **CTA MERIDA:** 95 episodios de Restriction.

Si observamos las actividades de estos centros, vemos que tienen los valores más altos para la actividad ALONE y valores medios para la actividad SPORTS.

En relación a los estados de ánimos, para los sentimientos negativos (Distressed, Frustrated y Guilty) tienes valores medios – altas.

Las ingestas con mayor número de ocurrencias son 1 Third y 2 Third.

Con estos datos, vemos que la evolución de los pacientes no es muy satisfactoria. El alto número de la actividad Alone y las cifras de los estados de ánimo negativos inciden en el alto número de episodios.

- Menor número de episodios: estos centros son los que tienen una buena evolución de pacientes. Son los siguientes:
 - **ARANDA TERAPEUTAS:** 29 episodios de Restriction.
 - **OSASUNA MUNGIA:** 19 episodios de Restriction.

Si estudiamos las actividades de estos centros, vemos que tienen el segundo valor más bajo para la actividad ALONE y los valores más altos para la actividad SPORTS.

En relación a los estados de ánimos, para los sentimientos negativos (Distressed, Frustrated y Guilty) tienen valores bajos y tienen los valores más altos para el estado de ánimo Relaxed.

Las ingestas con mayor número de ocurrencias son 1 Third y 2 Third.

Con estos datos, vemos que la evolución de los pacientes es más satisfactoria. El bajo número de la actividad Alone, el alto de la actividad Sports y los estados de ánimo más cercanos al bienestar del paciente, nos dan la pauta de una evolución adecuada.

Para estudiar la evolución de los pacientes según el centro médico hemos realizado la siguiente selección de parámetros:

Informe	Evolución de pacientes por centro médico y terapeuta	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento
Columnas	Dimensión	D_paciente
	Jerarquía	Centro_medico → Terapeuta
	Nivel 1	Centro médico

Los datos obtenidos y su representación gráfica son los siguientes:

		NDA TERAPELÍN TCA THERAPY		TERRES DE L'		CT SEVILLA		CTA MERIDA		Eating Disor		N FOOD Disor		IAD TCA Cent		ASUNA MUNÇ THERAPY CE					
Tipo_evento	Evento	T13	T14	T1	T2	T15	T16	T11	T12	T17	T18	T7	T8	T10	T9	T5	T6	T19	T20	T3	T4
ACTIVITY	ALONE	50	23	66	21	71	31	66	28	66	37	49	32	28	66	54	40	57	16	46	25
	FAMILY	97	49	94	40	83	41	86	42	89	41	92	44	42	87	95	33	85	56	92	42
	FRIENDS	90	66	104	58	108	50	106	45	104	54	117	63	49	104	115	57	112	49	113	53
	SPORTS	197	84	156	85	165	94	165	96	185	76	166	83	95	175	169	88	178	100	183	94
	STUDYING/WORKING	118	54	132	71	124	60	129	65	108	65	128	54	61	120	119	58	120	55	115	62
EMOTION	DISTRESSED	18	11	46	10	28	15	34	8	35	14	20	13	7	20	28	19	14	5	25	10
	FUSTRATED	16	12	50	17	28	8	43	17	28	15	23	10	12	33	24	16	30	10	14	12
	GUILTY	15	7	39	14	26	14	31	20	27	18	26	11	13	27	14	23	4	23	9	
	HAPPY	120	61	89	48	111	55	96	46	102	55	122	60	61	118	124	66	108	68	125	53
	RELAXED	230	116	202	103	220	119	227	117	226	96	219	112	106	209	208	99	240	126	236	125
	SAD	153	69	126	83	138	65	121	68	134	75	142	70	76	145	141	62	137	63	126	67
EPISODE	BINGE	28	12	16	15	36	15	28	14	25	16	28	13	11	30	27	17	26	8	30	10
	NO EPISODE	474	235	472	227	428	225	419	221	440	209	464	230	233	441	458	217	458	249	477	234
	RESTRICTION	1	28	47	34	57	36	65	41	44	51		33		29		42		19	3	32
MEAL	VOMITED	49	1	17	31		40			43		60		32	52	67		68		42	
	+ 3 THIRD	28	12	15	15	36	15	28	14	25	15	28	13	11	30	27	17	26	8	30	10
	1 THIRD	181	100	146	97	174	86	194	88	170	79	159	112	89	166	170	94	165	92	175	94
	2 THIRD	165	79	185	91	181	90	159	90	179	95	188	85	89	181	181	77	191	93	176	83
	3 THIRD	160	74	153	59	137	75	141	72	145	81	157	59	74	127	157	75	139	77	152	77
	NOTHING	18	11	53	13	23	10	30	12	33	23	20	7	12	28	17	13	31	6	16	12

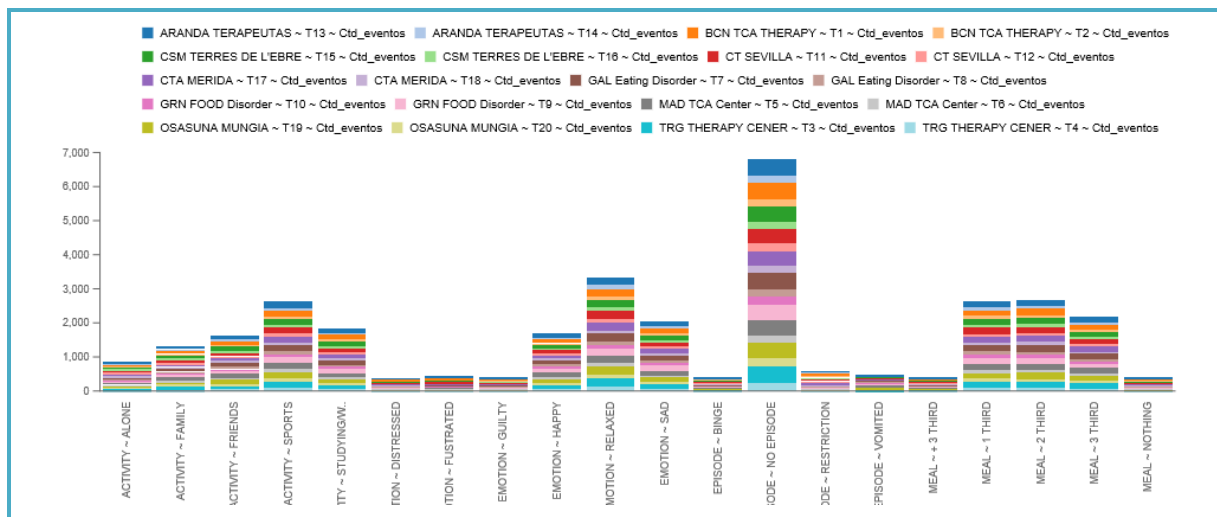


Ilustración 31 - Evolución de pacientes por terapeuta I

Aplicando la misma metodología que para el estudio de la evolución por centro médico, seleccionamos los siguientes terapeutas:

- Mayor número de episodios: estos terapeutas son los que tienen una mala evolución de pacientes. Son los siguientes:
 - **T15**, centro **CSM TERRES DE L'EBRE**: 57 episodios de Restriction.
 - **T11**, centro **CT SEVILLA**: 65 episodios de Restriction.

Si observamos las actividades de estos terapeutas, vemos que tienen los valores más altos para la actividad ALONE y valores medios altos para la actividad SPORTS.

En relación a los estados de ánimos, para los sentimientos negativos (Distressed, Frustrated y Guilty) tienes valores medios – altos. Aunque el estado de ánimo Relaxed también tiene una alta tasa de ocurrencia, vemos ue el estado Sad es mayor que el estado Happy.

Las ingestas con mayor número de ocurrencias son 1 Third y 2 Third.

Con estos datos, vemos que la evaluación de la evolución de los pacientes no es muy satisfactoria.

- Menor número de episodios: estos terapeutas son los que tienen una buena evolución de pacientes. Son los siguientes:
 - **T13**, centro **ARANDA TERAPEUTAS**: 1 episodio de Restriction.
 - **T3**, centro **TRG THERAPY CENER**: 3 episodios de Restriction.

Si estudiamos las actividades de estos centros, vemos que tienen valores medios-bajos para la actividad ALONE y los valores más altos para la actividad SPORTS.

En relación a los estados de ánimos, para los sentimientos negativos (Distressed, Frustrated y Guilty) tienen valores bajos y tienen los valores más altos para el estado de ánimo Relaxed.

Las ingestas con mayor número de ocurrencias son 1 Third y 2 Third.

Con estos datos, confirmamos las tendencias detectadas y se observa que la evolución de los pacientes es más satisfactoria. Una alta tasa de la actividad Sports, incide en unos estados de ánimo más cercanos al bienestar del paciente; todo ello, permite evaluar la evolución como satisfactoria.

5.2.6. Informe de relación entre los episodios de crisis y distintos momentos temporales.

Para estudiar entre los episodios de crisis en base a criterios temporales, hemos realizado la siguiente selección de parámetros:

Informe	Relación entre los episodios de crisis y la franja del día	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento
Columnas	Dimensión	D_tiempo
	Jerarquía	Franja_día
	Nivel	Franja_día
Columnas	Dimensión	D_tiempo
	Jerarquía	Anio
	Nivel 2	Anio
Columnas	Dimensión	D_tiempo
	Jerarquía	Mes
	Nivel 3	Mes

Se ha añadido un filtro para que solamente se muestren los tipos de eventos **episode**. Los datos obtenidos y su representación gráfica son los siguientes:

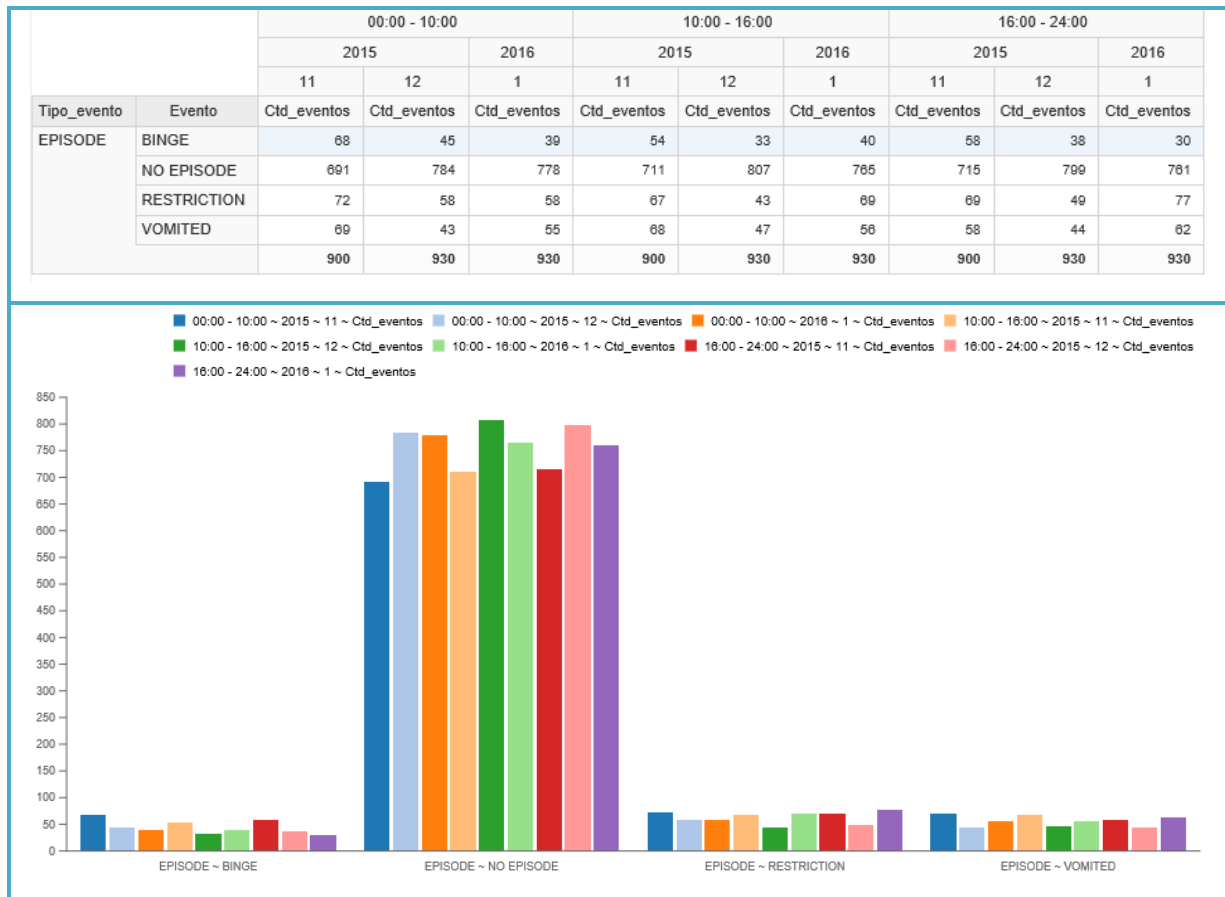


Ilustración 32 - Relación entre los episodios de crisis y la franja del día I

Los datos referentes a las actividades nos permiten realizar las siguientes observaciones considerando la cantidad realizada:

- Binge: el orden de las franjas diarias, de mayor a menor número de episodios, es el siguiente
 - Noviembre 2015:
 - Primera franja → Tercera franja → Segunda franja .
 - Diciembre 2015:
 - Primera franja → Tercera franja → Segunda franja .
 - Enero 2016:
 - Segunda franja → Primera franja → Tercera franja.

Según estos datos, a medida que evoluciona el paciente, la franja en que menos episodios Binge se producen es la tercera franja; las otras dos tienen un número similar de ocurrencias para este episodio.

Se observa también una tendencia a la reducción de estos episodios en todas las franjas a medida que transcurre el tiempo.

- Restriction: el orden de las franjas diarias, de mayor a menor número de episodios, es el siguiente
 - Noviembre 2015:
 - Primera franja → Tercera franja → Segunda franja .
 - Diciembre 2015:
 - Primera franja → Tercera franja → Segunda franja .
 - Enero 2016:
 - Tercera franja → Segunda franja → Primera franja.

Según estos datos, a medida que evoluciona el paciente, la franja en que menos episodios Restriction se producen es la primera franja; a continuación, la segunda y finalmente la tercera.

Se observa también una tendencia a la reducción de estos episodios en las dos primeras franjas a medida que transcurre el tiempo. Para la tercera franja se observa un incremento leve pero paulatino.

- Vomited: el orden de las franjas diarias, de mayor a menor número de episodios, es el siguiente
 - Noviembre 2015:
 - Primera franja → Segunda franja → Tercera franja.
 - Diciembre 2015:
 - Primera franja → Tercera franja → Segunda franja .
 - Enero 2016:
 - Tercera franja → Segunda franja → Primera franja

Según estos datos, a medida que evoluciona el paciente, las franjas en que menos episodios Vomited se producen son las dos primeras franjas, que tienen un número de ocurrencias muy similar; la tercera franja sí tiene un número de ocurrencias ligeramente mayor.

Se observa también una tendencia a la reducción de estos episodios en las primeras franjas a medida que transcurre el tiempo. Para la segunda y la tercera franja se observa un ligero incremento de los números de episodios.

- No episode: el orden de las franjas diarias, de mayor a menor número de episodios, es el siguiente
 - Noviembre 2015:
 - Tercera franja → Segunda franja → Primera franja.
 - Diciembre 2015:

- Segunda franja → Tercera franja → Primera franja .
- Enero 2016:
 - Primera franja → Segunda franja → Tercera franja.

Según estos datos, a medida que evoluciona el paciente, las franjas en que se producen menos ausencia de episodios son la tercera y la segunda franja. Es decir, la mejor franja del día es la primera franja, ya que es donde los pacientes registran menos cantidad de episodios.

Para estudiar entre los episodios de crisis según los días de la semana, hemos realizado la siguiente selección de parámetros:

Informe	Relación entre los episodios de crisis y los días de la semana	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento
Filas	Dimensión	D_tiempo
	Jerarquía	Dia_semana
	Nivel	Dia_semana
Columnas	Dimensión	D_tiempo
	Jerarquía	Anio
	Nivel 2	Anio
Columnas	Dimensión	D_tiempo
	Jerarquía	Mes
	Nivel 3	Mes

Los datos obtenidos y su representación gráfica son los siguientes:

Tipo_evento	Evento	Dia_semana	2015		2016
			11	12	1
			Ctd_eventos	Ctd_eventos	Ctd_eventos
EPISODE	BINGE	Domingo	42	18	20
		Jueves	27	28	15
		Lunes	21	12	11
		Martes	24	12	17
		Miercoles	15	19	13
		Sabado	30	18	18
		Viernes	21	11	15
	NO EPISODE	Domingo	333	310	357
		Jueves	278	387	293
		Lunes	358	308	302
		Martes	291	395	301
		Miercoles	291	381	305
		Sabado	283	315	388
		Viernes	283	314	378
	RESTRICTION	Domingo	41	19	40
		Jueves	30	23	27
		Lunes	34	24	24
		Martes	19	25	20
		Miercoles	30	27	22
		Sabado	22	15	37
		Viernes	32	17	34
	VOMITED	Domingo	34	13	33
		Jueves	25	34	25
		Lunes	37	16	23
		Martes	26	18	22
		Miercoles	24	23	20
		Sabado	25	12	27
		Viernes	24	18	23

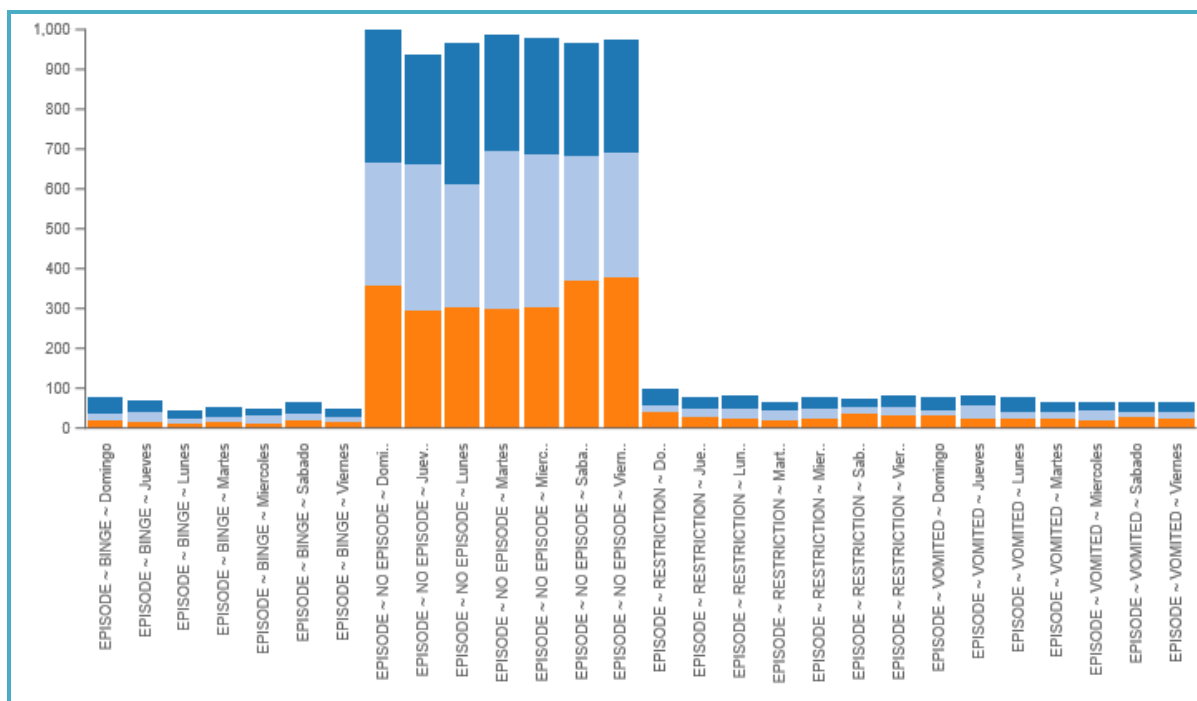


Ilustración 33 - Relación entre los episodios de crisis y los días de la semana I

Los datos referentes a las actividades nos permiten realizar las siguientes observaciones considerando la cantidad realizada:

- Binge: el día de la semana con mayor número de episodios, según el mes, es:
 - Noviembre 2015: domingo
 - Diciembre 2015: jueves
 - Enero 2016: domingo

El día de la semana con menor número de episodios, según el mes, es:

- Noviembre 2015: miércoles
- Diciembre 2015: viernes
- Enero 2016: lunes

Se observa que el mejor día de la semana en diciembre de 2015 y enero de 2016, coincide con el día siguiente al peor día.

No se observa un incremento a decremento continuo de lunes a domingo. Únicamente el primer mes se detecta que el fin de semana han sido los dos días con mayor número de episodios, pero no se ha podido registrar esa tendencia en los meses posteriores.

Se detecta, que al evolucionar el paciente, existe una tendencia de reducción de episodios en todos los días de la semana. No se advierte que exista algún día con mayor probabilidad de ocurrencia del episodio.

- Restriction: el día de la semana con mayor número de episodios, según el mes, es:
 - Noviembre 2015: domingo
 - Diciembre 2015: miércoles
 - Enero 2016: domingo

El día de la semana con menor número de episodios, según el mes, es:

- Noviembre 2015: martes
- Diciembre 2015: sábado
- Enero 2016: martes

No se observa relación aparente entre los mejores días y los peores días de la semana. Tampoco se detecta un incremento a decremento continuo de lunes a domingo.

Al evolucionar el paciente, no existe una tendencia de reducción de episodios en todos los días de la semana. Se advierten incrementos y decrementos en varios días de la semana. No se ha detectado una tendencia entre estos incrementos y decrementos.

- Vomited: el día de la semana con mayor número de episodios, según el mes, es:
 - Noviembre 2015: lunes
 - Diciembre 2015: jueves
 - Enero 2016: domingo

El día de la semana con menor número de episodios, según el mes, es:

- Noviembre 2015: miércoles, jueves y viernes
- Diciembre 2015: sábado y domingo
- Enero 2016: miércoles

No se observa relación aparente entre los mejores días y los peores días de la semana.

En el primer mes, se observa que tras el domingo y el lunes, existe una tendencia a la reducción de episodios a medida que transcurre la semana. En los otros dos meses no se detecta, por lo que no podemos extraer ninguna conclusión válida.

Al evolucionar el paciente, no existe una tendencia de reducción de episodios en todos los días de la semana.

- No episode: el día de la semana con mayor número de ausencia de episodios, según el mes, es:
 - Noviembre 2015: domingo
 - Diciembre 2015: martes
 - Enero 2016: viernes

El día de la semana con menor número de ausencia episodios, según el mes, es:

- Noviembre 2015: viernes y sábado
- Diciembre 2015: lunes
- Enero 2016: jueves

Como en los otros casos, no se han podido establecer tendencias de la ausencia de episodios según el día de la semana.

Para estudiar entre los episodios de crisis según el momento del año, hemos realizado la siguiente selección de parámetros:

Informe	Relación entre los episodios de crisis y los meses del año	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_tiempo
	Jerarquía	Anio
	Nivel 2	Anio
Filas	Dimensión	D_tiempo
	Jerarquía	Mes
	Nivel 3	Mes
Columnas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento

Los datos obtenidos y su representación gráfica son los siguientes:

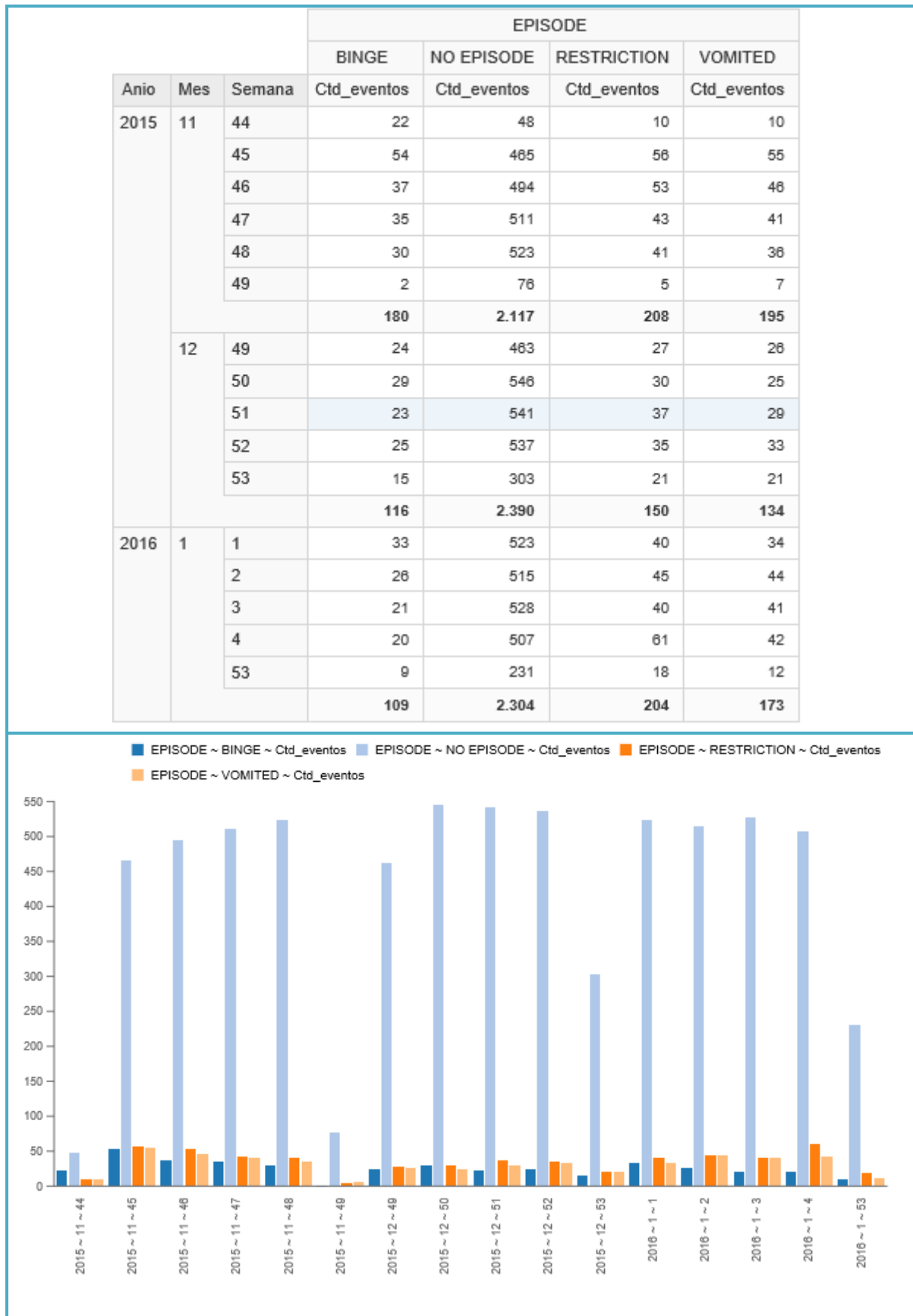


Ilustración 34 - Relación entre los episodios de crisis y los meses del año I

Los datos, tal como se ha señalado anteriormente, han sido recogidos en noviembre y diciembre de 2015 y enero de 2016. Es la época de Navidad y se han observado los datos teniendo en cuenta esta circunstancia para determinar su posible impacto.

Antes del análisis de los datos, hay que hacer constar que los días 1, 2 y 3 de enero se han asociado a la semana 53.

Se observa que en enero de 2016 se produce un incremento en los episodios Restriction y Vomited. Sin embargo, la semana que registra mayor número de episodios Restriction es la última semana de enero y la semana que registra el mayor número de episodios Vomited es la segunda semana de enero. Al estar alejados los episodios de la semana 53 y de la semana 1 (semanas fiestas navideñas), no se puede establecer relación directa con las fiestas navideñas.

Para el episodio Binge se observa una disminución continua en el número de episodios por mes. Al observar las semanas, se registran pequeñas fluctuaciones sin llegar a producirse un aumento significativo en las semanas navideñas.

Por tanto, no se puede concluir que las fechas navideñas hayan influido en la aparición de episodios.

5.2.7. Informe de relación entre las actividades y el estado de ánimo de los pacientes.

Para estudiar la relación entre las actividades y los episodios de crisis, hemos realizado la siguiente selección de parámetros:

Informe	Relación entre las actividades y el estado de ánimo de los pacientes	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento
Columnas	Dimensión	D_tiempo
	Jerarquía	Franja_dia
	Nivel	Franja_dia

Se ha añadido un filtro para que solamente se muestren los tipos de eventos **activity** y **emotion**. Los datos obtenidos y su representación gráfica son los siguientes:

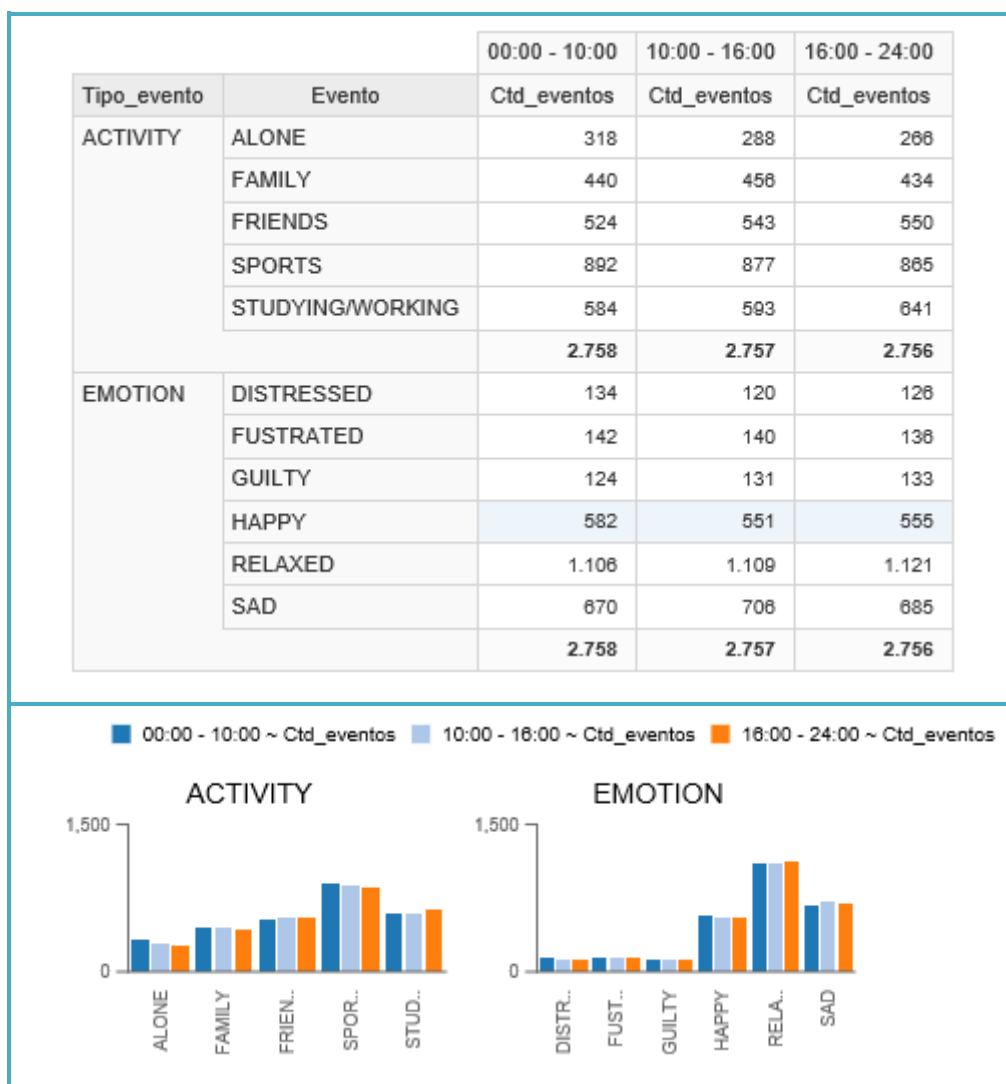


Ilustración 35 - Relación entre las actividades y el estado de ánimo de los pacientes I

Los datos referentes a las actividades nos permiten establecer los siguientes grupos considerando la cantidad realizada:

- Grupo 1: Alone: este estado de ánimo es el menos frecuente. Para relacionarlo con los estados de ánimo, vemos que la cantidad de eventos de esta actividad coincide aproximadamente con la suma de los eventos de los estados de ánimo Distressed, Frustrated, Guilty.
- Grupo 2: Sports – Studying/Working – Family - Friends: en este grupo la actividad Sports supera a las otras tres. Sin embargo, en términos generales podemos agruparlas ya que la diferencia, aunque apreciable, no es destacadamente superior. Por tanto, podemos asociar este grupo con los sentimientos Relaxed, Happy y Sad. En este grupo de estados

de ánimo, vemos que se produce una situación análoga a la de las actividades: el estado de ánimo Relaxed es el más numeroso, pudiendo apreciarse una diferencia con los otros dos que, aun siendo destacable, no nos impide agruparlos en un mismo conjunto.

Por tanto, relacionando este grupo de actividades con este grupo de estado de ánimo, vemos que practicar cualquiera de estas actividades puede ocasionar uno de estos estados de ánimo. De los tres estados de ánimo el más probable será Relaxed, seguido de Happy y Sad.

Para estudiar con un mayor nivel de detalle estos datos, añadiremos el nivel del año y mes, de la dimensión tiempo y el entorno del paciente.

Informe	Relación entre las actividades y el estado de ánimo de los pacientes	
Cubo	H_DIARIO	
Medida	Ctd_eventos	
Filas	Dimensión	D_paciente
	Jerarquía	Entorno
	Nivel	Entorno
Filas	Dimensión	D_evento
	Jerarquía	Tipo_evento
	Nivel	Evento
Columnas	Dimensión	D_tiempo
	Jerarquía	Franja_día
	Nivel 1	Franja_día
Columnas	Dimensión	D_tiempo
	Jerarquía	Anio
	Nivel 2	Anio
Columnas	Dimensión	D_tiempo
	Jerarquía	Mes
	Nivel 3	Mes

Los datos obtenidos y su representación gráfica son los siguientes:

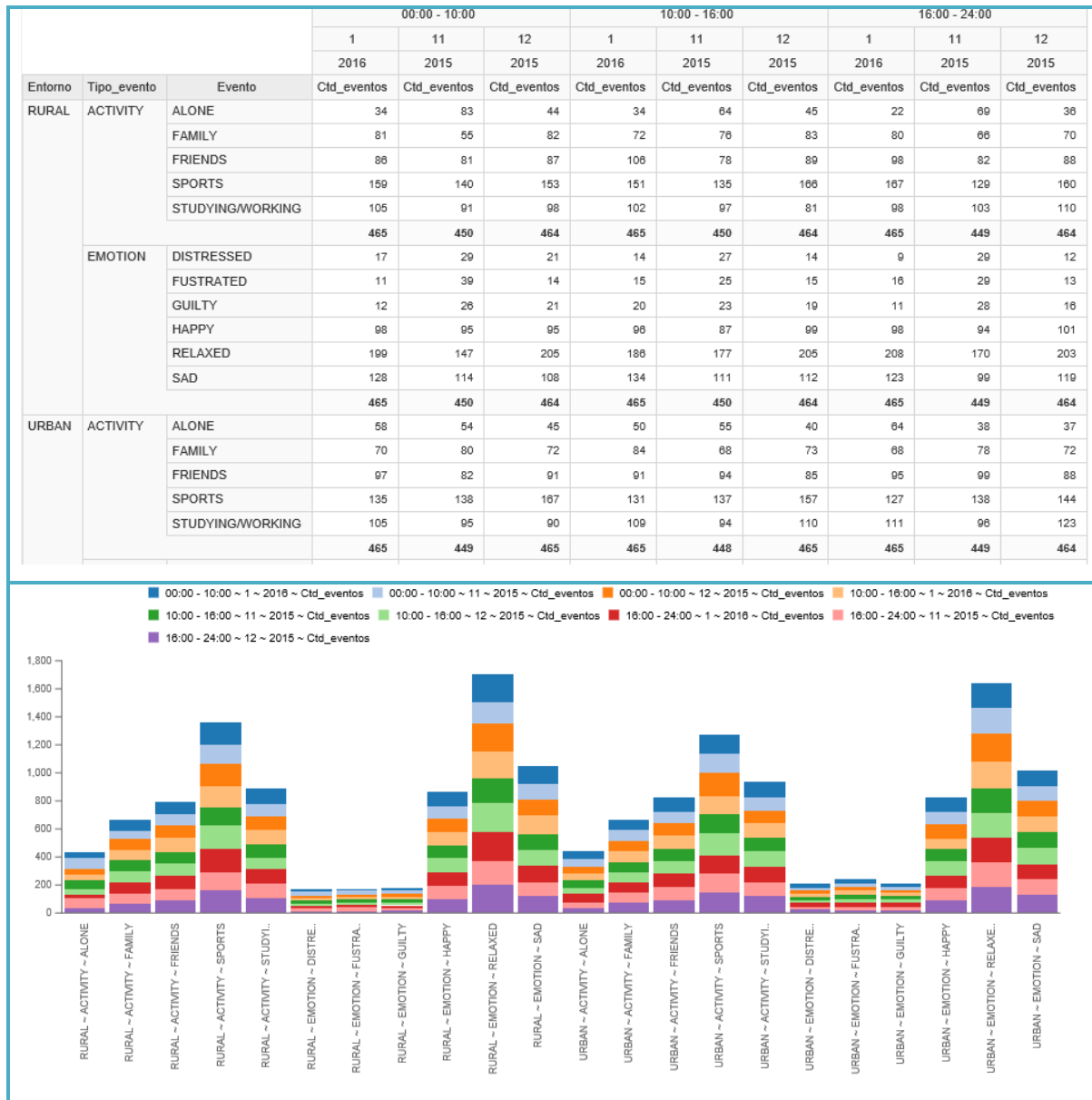


Ilustración 36 - Relación entre las actividades y el estado de ánimo de los pacientes I

Estos datos confirman nuestras conclusiones, ya que se observa que para ambos entornos, la actividad menos realizada es Alone, con tasas de realización similares. Asimismo, las tasas de ocurrencia de los estados de ánimo Distressed, Frustrated, Guilty son igualmente parecidas en ambos entornos, con un ligero incremento para el entorno Urbano.

En el grupo formado por las actividades Sports – Studying/Working – Family – Friends, la actividad Sports es la más realizada en ambos entornos. Los sentimientos Relaxed, Happy and Sad confirman las tendencias descritas anteriormente.

Por tanto, podemos concluir lo siguiente:

- Al practicar la actividad Alone, probablemente se producirá un episodio de Distressed, Frustrated, Guilty.
- Al practicar las actividades Sports – Studying/Working – Family – Friends, el paciente probablemente se sentirá Relaxed, Happy o Sad. Estos estados de ánimo están ordenados por orden de probabilidad.

5.3. Respuesta a las preguntas planteadas.

5.3.1. ¿Cuál es la relación entre las actividades realizadas y los episodios de crisis?

- Al realizar las actividades Sports, Studying/Working, lo más probable es que no suceda ningún episodio.
- Al realizar las actividades Family y Friends, pueden ocurrir dos posibilidades:
 - Ausencia de episodios.
 - Episodios: no se aprecia una tendencia para la ocurrencia de un episodio concreto.
- Al realizar la actividad Alone, probablemente ocurrirá uno de los tres episodios (restriction, vomited, binge). El episodio con mayor probabilidad de ocurrencia será Binge.

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

- Al experimentar los estados de ánimo Relaxed, Sad, Happy, lo más probable es que no suceda ningún episodio.
- Al experimentar los estados de ánimo Distressed, Frustrated, Guilty, lo más probable es que ocurra un episodio Binge, Restriction, Vomited. No es posible determinar qué episodio es el probable en función del estado de ánimo.

5.3.3. ¿Se puede establecer alguna relación a nivel geográfico, por ejemplo entorno urbano o rural?

La relación entre los episodios y los entornos de los pacientes es la siguiente:

- Binge: este episodio de crisis registra una ocurrencia parecida en ambos entornos, el rural y el urbano.
- Restriction: este episodio de crisis es más frecuentes en el entorno rural que en el urbano, alcanzando la mayor diferencia en la primera franja de la mañana.
- Vomited: este episodio de crisis es más frecuente en entornos urbanos que en rurales.

5.3.4. ¿Cómo ha sido la evolución de los diferentes pacientes a lo largo del tiempo?

Se ha analizado la evolución de los pacientes en base a su entorno. Para el entorno rural, podemos concluir:

- Se realizan en gran número las actividades que permiten al paciente interactuar con otras personas; esto reduce los sentimientos negativos (Distressed, Frustrated y Guilty) y potencia los positivos (Happy, Relaxed).
- En términos generales, la evolución de los pacientes ha sido satisfactoria, ya que se han incrementado y potenciado los sentimientos positivos y las ingestas, y se han reducido los episodios.

Para el entorno urbano, podemos concluir:

- En términos generales, la evolución de los pacientes ha sido poco significativa. Los sentimientos negativos no se han visto superados a través de la actividades realizadas. Podemos decir que dejar de practicar la actividad Sport y practicar Alone incide en un aumento de los episodios y una ingesta de comida pobre.

5.3.5. ¿Hay alguna diferencia de la evolución de pacientes a nivel de centro médico? ¿Y terapeuta?

En relación a la evolución de los pacientes por centro médico, se han localizado los centros con peor evolución y con mejor evolución:

- Centros con peor evolución. Son los siguientes: **CT SEVILLA** y **CTA MERIDA**. Los pacientes de estos centro destacan en la actividad Alone y no lo hacen en la actividad Sports. Por tanto, también se aprecia que hay muchos pacientes con estados de ánimo Distressed, Frustrated y Guilty.

- Centros con mejor evolución. Son los siguientes: **ARANDA TERAPEUTAS** y **OSASUNA MUNGIA**. La evolución de los pacientes es satisfactoria ya que hay un número bajo de pacientes que practica la actividad Alone, un número alto de pacientes que practica la actividad Sports, lo que ocasiona que los estados de ánimo sean los positivos, los relacionados con el bienestar del paciente.

En relación a la evolución de los pacientes por terapeuta, se han localizado los terapeutas con peor evolución y con mejor evolución:

- Terapeutas con peor evolución. Son los siguientes: **T15**, centro **CSM TERRES DE L'EBRE** y **T11**, centro **CT SEVILLA**. Los pacientes de estos terapeutas tienen los valores más altos para la actividad ALONE y valores medios altos para la actividad SPORTS. Por ello,
- Terapeutas con mejor evolución. Son los siguientes: **T13**, centro **ARANDA TERAPEUTAS** y **T3**, centro **TRG THERAPY CENER**. Los pacientes de estos terapeutas experimentan mayoritariamente el estado de ánimo Relaxed, debido a la práctica numerosa de la actividad Sports. Todo ello, permite evaluar la evolución como satisfactoria.

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

Las conclusiones para la relación entre los episodios de crisis y la franja del día son los siguientes:

- Episodio Binge: la franja en que menos episodios Binge se producen es la tercera franja; las otras dos tienen un número similar de ocurrencias para este episodio.
A medida que evoluciona el paciente, estos episodios se reducen en todas las franjas.
- Episodio Restriction: la franja en que menos episodios se producen es la primera franja; a continuación, la segunda y finalmente la tercera.
A medida que evoluciona el paciente, estos episodios se reducen en las dos primeras franjas. Para la tercera franja se observa un incremento leve.

- Episodio Vomited: las franjas en que menos episodios Vomited se producen son la dos primeras franjas, que tienen un número de ocurrencias muy similar; la tercera franja sí tiene un número de ocurrencias ligeramente mayor.

A medida que evoluciona el paciente, estos episodios se reducen en la primera franja. Para la segunda y la tercera franja se observa un ligero incremento.

La mejor franja del día para la ausencia de episodios es la primera franja.

Las conclusiones para la relación entre los episodios de crisis y la franja del día son los siguientes:

- Episodio Binge: al evolucionar el paciente, existe una tendencia de reducción de episodios en todos los días de la semana. No se advierte que exista algún día con mayor probabilidad de ocurrencia del episodio.
- Episodio Restriction: no se observa relación aparente entre los mejores días y los peores días de la semana. Al evolucionar el paciente, no existe una tendencia de reducción de episodios en todos los días de la semana. Se advierten incrementos y decrementos en varios días de la semana. No se ha detectado una tendencia entre estos incrementos y decrementos.
- Episodio Vomited: no se observa relación aparente entre los mejores días y los peores días de la semana. Al evolucionar el paciente, no existe una tendencia de reducción de episodios en todos los días de la semana.

Las conclusiones para la relación entre los episodios de crisis y los meses del año son las siguientes:

- Episodios Restriction y Vomited: Se observa que en enero de 2016 se produce un incremento en estos episodios. Sin embargo, la semana que registra mayor número de episodios Restriction es la última semana de enero y la semana que registra el mayor número de episodios Vomited es la segunda semana de enero. Al estar alejados los episodios de la semana 53 y de la semana 1 (semanas

fiestas navideñas), no se puede establecer relación directa con las fiestas navideñas.

- Episodio Binge: se observa una disminución continua en el número de episodios por mes. Al observar las semanas, se registran pequeñas fluctuaciones sin llegar a producirse un aumento significativo en las semanas navideñas. Por tanto, no se puede concluir que las fechas navideñas hayan influido en la aparición de episodios

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

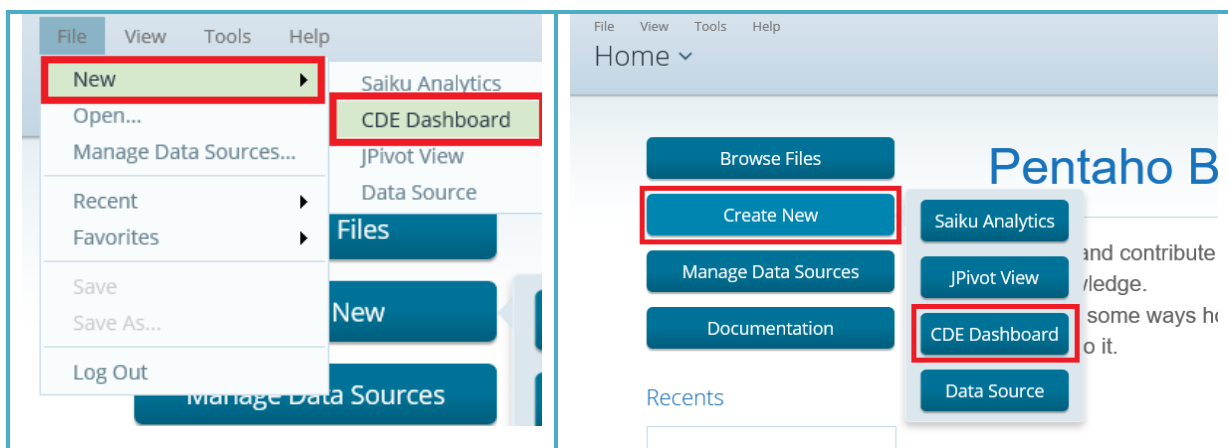
Como ya se ha señalado anteriormente, practicar la actividad Sports, ocasiona una mejora en el estado de ánimo de los pacientes. Los estados de ánimo más probables, ordenados por orden de probabilidad, son: Relaxed, Happy o Sad.

5.4. Cuadro de mandos.

5.4.1. Creación de un cuadro de mandos con Pentaho CDE Dashboard.

Un cuadro de mando puede definirse como un conjunto de indicadores, cuyo objetivo es facilitar la toma de decisiones y el control. Se requiere un sistema completo de análisis para determinar los indicadores a incluir.

Para crear un cuadro de mando en Pentaho, debemos seleccionar la opción **Create New → CDE Dashboard** o la opción de menú



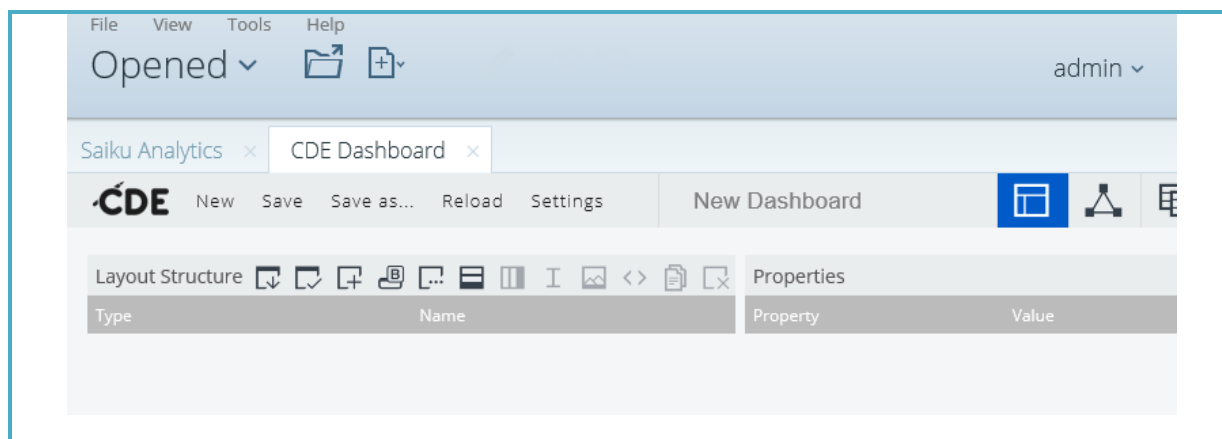


Ilustración 37 - Opciones de llamada y pantalla de CDE Dashboard

Para facilitar la creación del cuadro de mandos, utilizaremos una de las plantillas predefinidas que ofrece Pentaho. Para ello pulsamos el botón **Apply Template**:

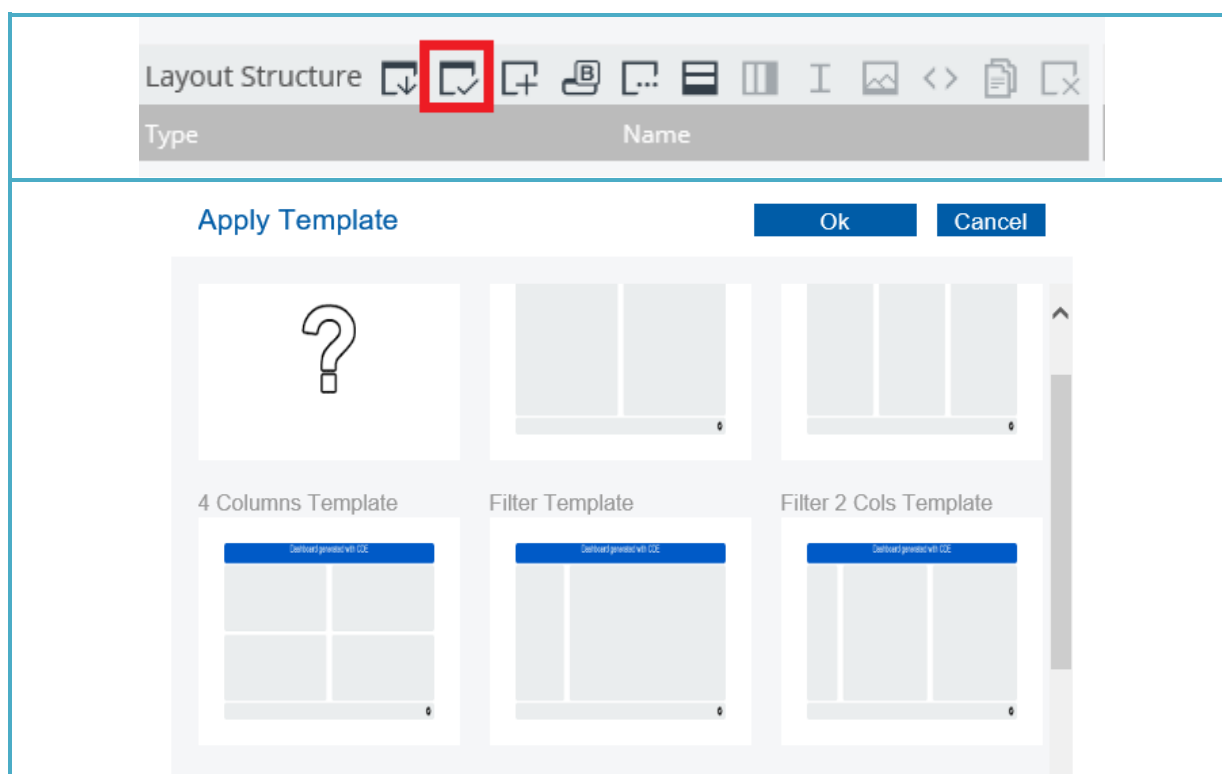


Ilustración 38 - Selección de plantilla CDE Dashboard

La pantalla que aparece nos muestra las plantillas disponibles. En nuestro caso seleccionaremos la opción **3 Columns Template**.

A continuación, crearemos los gráficos que queremos mostrar en el cuadro de mandos. Para ello, en la parte derecha de la barra de iconos, tenemos la opción **Datasources Panel**. Seleccionamos, en el menú de la izquierda, la

opción **Wizards** → **OLAP Chart Wizard**. De esta forma, se nos mostrará la pantalla de diseño de nuestro gráfico.

The screenshot shows the CDE Dashboard interface. At the top, there is a 'Datasources Panel' and a menu with icons. The 'Wizards' menu is open, showing options like 'OLAP Selector Wizard', 'OLAP Chart Wizard' (highlighted with a red dashed box), and 'Saiku OLAP Wizard'. Below this, the 'OLAP Chart Wizard' configuration screen is visible. It includes a 'Preview Area' with a bar chart showing 'Value' on the y-axis (0 to 1,000) and categories on the x-axis: ALONE, FUSTRATED, NO EPISODE, and 3 THIRD. The 'Dimensions' panel on the left lists categories like Centro_medico, Entorno, Tiempo, Anual, Trimestral, Mensual, Semanal, Diaria, Franja_dia, and Tipo_evento. The 'Measures' and 'Filters' panels are also visible. The 'Columns' section shows 'Entorno', 'Anio', and 'Mes' selected.

Ilustración 39 - Creación de gráfico OLAP CDE Dashboard

Mediante esta pantalla, crearemos todos los gráficos que deseemos incluir en el cuadro de mandos. Debemos seleccionar del menú de la izquierda los elementos que configurarán cada gráfico. Una vez finalizada la creación de un gráfico, pulsaremos el botón **OK** de la parte superior de la pantalla:

The screenshot shows the 'OLAP Wizard' dialog box. At the top, there are 'Ok' and 'Cancel' buttons, with 'Ok' highlighted by a red box. Below the buttons, the dialog displays the configuration for the OLAP Chart Wizard. It includes a 'Catalog' dropdown set to 'TCA_DW_Schema', a 'Cube' dropdown set to 'H_DIARIO', a 'Name' field containing 'Eventos y Entornos', and an 'Html Object' dropdown set to 'Select...'. The main area of the dialog shows the same bar chart as in the previous screenshot, with the 'Dimensions' panel on the left and the 'Columns' section at the bottom.

Ilustración 40 - Finalización gráfico OLAP CDE Dashboard

Por último, guardaremos nuestro cuadro de mando con la opción **Save as...**

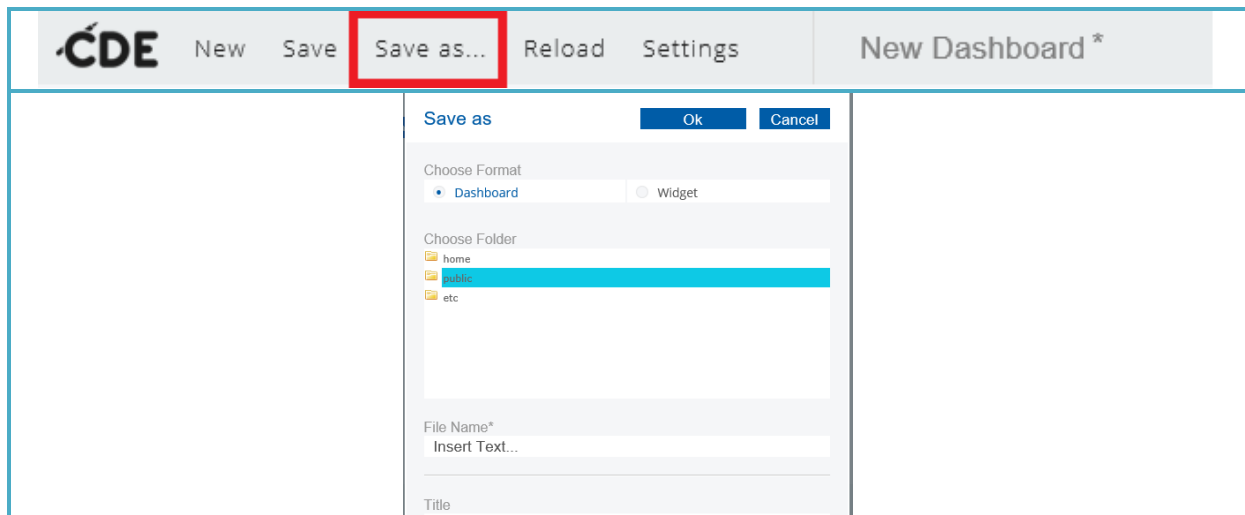


Ilustración 41 - Guardar CDE Dashboard

5.4.2. Indicadores

Los indicadores de nuestro cuadro de mando son:

- Número de eventos por tipo de actividad.
- Número de eventos por tipo de estado de ánimo
- Número de eventos por ingesta.
- Número de eventos por episodio.

El cuadro de mando una vez realizado se muestra en la siguiente figura:

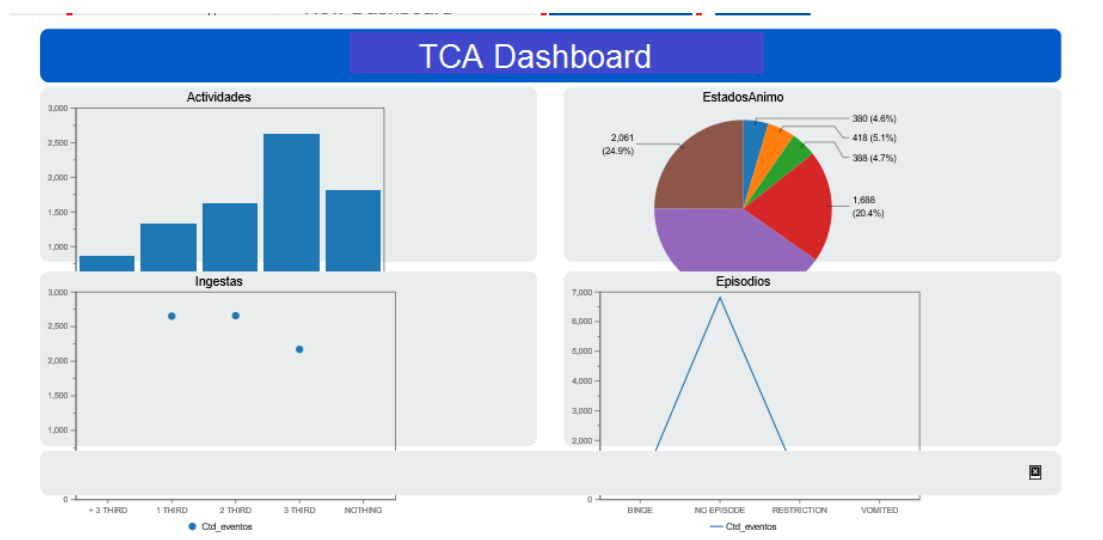


Ilustración 42 - Ejemplo cuadro CDE Dashboard

6. Conclusiones

Business Intelligence es, hoy en día, uno de los ámbitos con mayor crecimiento. El desarrollo de un sistema de Business Intelligence requiere de una planificación y metodología que permitan finalizar con éxito este tipo de proyectos y que garanticen su utilización y su utilidad para los usuarios a los que esté dirigido.

Durante la realización de este TFM, he podido aprender y profundizar en los siguientes aspectos básicos del Business Intelligence:

- Disponer de un sistema BI adecuado es, actualmente, un requisito extremadamente importante para el apoyo a la toma de decisión y para la gestión estratégica de las organizaciones.
- Un diseño apropiado del sistema BI hará que se convierta rápidamente en una fuente de información útil, eficaz y consistente. La aceptación del sistema BI estará directamente relacionada con la respuesta eficaz y elegante a las necesidades planteadas.

Durante la realización de este TFM, he intentado mantener estas ideas principales como objetivo al realizar cada una de las tareas. Asimismo, me ha permitido realizar un estudio de herramientas Open Source y la elección de la más adecuada: existen numerosas herramientas Open Source que nos permiten abordar los distintos proyectos de Business Intelligence. Es importante conocer las principales características de cada una para poder realizar una elección adecuada que permita maximizar las ventajas de la alternativa escogida.

En relación a la planificación del proyecto, ha resultado un poco ajustada. Por motivos laborales, el tiempo de dedicación al proyecto se ha visto reducido, ocasionando con ello retrasos en las entregas planificadas. Sin embargo, haciendo un esfuerzo en la recta final, se ha conseguido cumplir con los objetivos fijados inicialmente.

Las líneas futuras de este proyecto se centran principalmente en lo siguiente:

- Profundizar y mejorar el cuadro de mando: se ha realizado un cuadro de mando muy básico que no permite aprovechar todo el potencial de los datos registrados.

- Estudio de herramientas alternativas para la generación de listados: las capacidades de Saiku Analytics en ocasiones han resultado algo limitadas. Existen alternativas que nos aportan mayor flexibilidad y permite dotar al producto de valor añadido adicional.

En el plano personal, ha sido una experiencia enriquecedora y que me ha proporcionado la oportunidad de adquirir experiencia y conocimientos vitales en un ámbito que desconocía, pero que había despertado mi interés desde hacía tiempo. Ha sido un proyecto muy interesante y provechoso, ya que he conseguido familiarizarme con las herramientas y técnicas básicas para la realización de proyectos de Business Intelligence.

7. Glosario

Base de datos relacional. La base de datos relacional (BDR) es un tipo de base de datos (BD) que cumple con el modelo relacional (el modelo más utilizado actualmente para implementar las BD ya planificadas).

Permite establecer interconexiones o relaciones entre los datos (que están guardados en tablas), y a través de dichas conexiones relacionar los datos de ambas tablas, de ahí proviene su nombre: "modelo relacional".

Business Intelligence. Se denomina inteligencia empresarial, inteligencia de negocio o BI, al conjunto de estrategias y aspectos relevantes enfocados a la administración y creación de conocimiento sobre el medio, a través del análisis de los datos existentes en una organización o empresa.

Cuadro de mando. La gestión de las empresas requiere un sistema de indicadores (en inglés KPIs o Key Performance Indicators) que nos faciliten la toma de decisiones y el control. Se requiere un sistema completo de análisis.

El sistema de indicadores debe organizarse en un cuadro de mando. El cuadro de mando recoge los principales indicadores y los presenta de un modo claro y útil. El cuadro de mando es un sistema que nos informa de la evolución de los parámetros fundamentales del negocio. Los cuadros de mando han de presentar sólo aquella información que sean imprescindible, de una forma sencilla y por supuesto, sinóptica y resumida.

Data Mart. Un Data mart es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

Data Warehouse. En el contexto de la informática, un almacén de datos (del inglés data warehouse) es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Se trata, sobre todo, de un expediente completo de una organización, más allá de la información transaccional y operacional, almacenado en una base de datos diseñada para favorecer el análisis y la divulgación eficiente de datos (especialmente OLAP, procesamiento analítico en línea). El almacenamiento

de los datos no debe usarse con datos de uso actual. Los almacenes de datos contienen a menudo grandes cantidades de información que se subdividen a veces en unidades lógicas más pequeñas dependiendo del subsistema de la entidad del que procedan o para el que sean necesario.

ETL. Extract, Transform and Load («extraer, transformar y cargar», frecuentemente abreviado ETL) es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio.

OLAP. Es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo de la llamada Inteligencia de negocios (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales (o Cubos OLAP) que contienen datos resumidos de grandes Bases de datos o Sistemas Transaccionales (OLTP). Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares.

OLTP. Es el acrónimo en inglés de Procesamiento de Transacciones En Línea (OnLine Transaction Processing) es un tipo de procesamiento que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional). Los paquetes de software para OLTP se basan en la arquitectura cliente-servidor ya que suelen ser utilizados por empresas con una red informática distribuida.

Open Source. Es el software distribuido y desarrollado libremente. Se focaliza más en los beneficios prácticos (acceso al código fuente) que en cuestiones éticas o de libertad que tanto se destacan en el software libre. Para muchos el término «libre» hace referencia al hecho de adquirir un software de manera gratuita, pero más que eso, la libertad se refiere al poder modificar la fuente del programa sin restricciones de licencia, ya que muchas empresas de software encierran su código, ocultándolo y restringiéndose los derechos a sí misma..

Pentaho BI Suite. Es un conjunto de programas libres para generar inteligencia empresarial (Business Intelligence). Incluye herramientas integradas para generar informes, minería de datos, ETL, etc. 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 basados en procesos y ha sido concebido desde el principio para estar basada en procesos

PostgreSQL. Es un Sistema de gestión de bases de datos relacional orientado a objetos y libre, publicado bajo la licencia PostgreSQL1 , similar a la BSD o la MIT.

Saiku Analytics. Es una herramienta que permite a los usuarios de negocios explorar fuentes de datos complejas, utilizando un interfaz drag and drop familiar, todo dentro de un navegador.

Script. En informática, un script, archivo de órdenes, archivo de procesamiento por lotes o, cada vez más aceptado en círculos profesionales1 y académicos, script2 3 4 es un programa usualmente simple, que por lo regular se almacena en un archivo de texto plano.

SGBD. Un sistema gestor de base de datos (SGBD) es un conjunto de programas que permiten el almacenamiento, modificación y extracción de la información en una base de datos, además de proporcionar herramientas para añadir, borrar, modificar y analizar los datos. Los usuarios pueden acceder a la información usando herramientas específicas de interrogación y de generación de informes, o bien mediante aplicaciones al efecto.

Staging Area. es un área intermedia de almacenamiento de datos utilizada para el procesamiento de los mismos durante procesos de extracción, transformación y carga (ETL). Esta área se encuentra entre la fuente de los datos y su destino, que a menudo son almacenes de datos, data marts u otros repositorios de datos. Las áreas de stage de datos son a menudo de naturaleza transitoria, su contenido se borrará antes de ejecutar un proceso de ETL o inmediatamente después de haberlo finalizado con éxito.

8. Bibliografía

Libros

- **Introducción al Business Intelligence**

Autores: Jordi Conesa Caralt. Josep Curto Díaz

Edición: Primera edición en lengua castellana

Segunda reimpresión

Editorial: Editorial UOC

- **The Data Warehouse Toolkit**

Autores: Ralph Kimball. Margy Ross

Edición: Third edition

Editorial: Editorial Wiley

- **A manager's guide to Data Warehousing**

Autores: Laura L. Reeves

Edición: First edition

Editorial: Editorial Wiley

- **The Data Warehouse ETL toolkit**

Autores: Ralph Kimball. Joe Caserta.

Edición: First edition

Editorial: Editorial Wiley

Web

- Inteligencia empresarial.

https://es.wikipedia.org/wiki/Inteligencia_empresarial

Fecha consulta: Abril 2016

- Inteligencia empresarial.

https://es.wikipedia.org/wiki/Inteligencia_empresarial

Fecha consulta: Abril 2016

- Almacén de datos

https://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos

Fecha consulta: Abril 2016

- PostgreSQL

<https://es.wikipedia.org/wiki/PostgreSQL>

Fecha consulta: Abril 2016

- MySQL

<https://es.wikipedia.org/wiki/MySQL>

Fecha consulta: Mayo 2016

- Pentaho

<https://es.wikipedia.org/wiki/Pentaho>

Fecha consulta: Mayo 2016

- JasperReports

<https://es.wikipedia.org/wiki/JasperReports>

Fecha consulta: Mayo 2016

- SpagoBI

<https://es.wikipedia.org/wiki/SpagoBI>

Fecha consulta: Mayo 2016

- Cubo OLAP

https://es.wikipedia.org/wiki/Cubo_OLAP

Fecha consulta: Mayo 2016

- Cuadro de mando

https://es.wikipedia.org/wiki/Cuadro_de_mando

Fecha consulta: Mayo 2016

- PostgreSQL 9.5 Geared to Liberate Enterprises from the Data Warehouse

<http://thenewstack.io/postgresql-9-5-geared-liberate-enterprises-data-warehouse/>

Fecha consulta: Abril 2016

- Porque deberías pensar en PostgreSQL para tu Data Warehouse
https://www.linkedin.com/pulse/porque-deber%C3%ADas-pensar-en-postgresql-para-tu-data-qui%C3%B1ones-azc%C3%A1rate?trk=pulse-det-nav_art

Fecha consulta: Abril 2016

- ¿Por qué utilizar Pentaho?
<http://www.stratebi.com/pentaho>

Fecha consulta: Abril 2016

9. Anexo I. Esquema data warehouse

```
CREATE DATABASE "TCA_DB"
```

```
create schema tca_dw_admin;
```

```
GRANT USAGE ON SCHEMA tca_dw_admin TO tca_dw_admin ;
```

```
ALTER DEFAULT PRIVILEGES IN SCHEMA tca_dw_admin GRANT ALL  
PRIVILEGES ON TABLES TO tca_dw_admin;
```

```
ALTER DEFAULT PRIVILEGES IN SCHEMA tca_dw_admin GRANT ALL  
PRIVILEGES ON SEQUENCES TO tca_dw_admin;
```

```
SET search_path TO tca_dw_admin;
```

```
create sequence seq_id_entorno start with 1 increment by 1;
```

```
create sequence seq_id_terapeuta start with 1 increment by  
1;
```

```
create sequence seq_id_tipo_evento start with 1 increment  
by 1;
```

```
CREATE TABLE tca_d_paciente          (id_paciente          serial  
primary key,
```

```
                                     nombre_paciente  varchar(255)  not  
null unique,
```

```
                                     ciudad varchar(255),
```

```
                                     sexo varchar(1),
```

```
                                     id_entorno integer,
```



```
nombre_entorno varchar(255),
id_terapeuta integer,
nombre_terapeuta varchar(255),
id_centmed integer,
nombre_centmed varchar(255),
ciudad_centmed varchar(255)
);
```

```
CREATE TABLE tca_d_tiempo (id_fecha serial primary
key,
fecha date not null ,
anio integer,
trimestre integer,
mes integer,
mes_nombre varchar(25),
semana_anio integer,
dia_mes integer,
dia_nombre varchar(25),
franja_dia varchar(50),
unique (fecha,franja_dia)
);
```

```
CREATE TABLE tca_d_evento (id_evento serial
primary key,
nombre_evento varchar(255) not
null unique,
```

```
        id_tipo_evento integer,  
        nombre_tipo_evento varchar(255)  
not null);
```

```
CREATE TABLE tca_h_diario (id_diario serial  
primary key,  
        id_paciente integer references  
tca_d_paciente,  
        id_fecha integer references  
tca_d_tiempo,  
        id_evento integer references  
tca_d_evento,  
        cantidad_eventos numeric  
);
```

10. Anexo II. Esquema staging area

```
create schema tca_admin;
```

```
GRANT USAGE ON SCHEMA tca_admin TO tca_admin ;
```

```
ALTER DEFAULT PRIVILEGES IN SCHEMA tca_admin GRANT ALL  
PRIVILEGES ON TABLES TO tca_admin;
```

```
ALTER DEFAULT PRIVILEGES IN SCHEMA tca_admin GRANT ALL  
PRIVILEGES ON SEQUENCES TO tca_admin;
```

```
SET search_path TO tca_admin;
```

```
CREATE TABLE tca_centmed (code integer,  
                           name varchar(255) not null,  
                           cp varchar(5),  
                           city varchar(250));
```

```
CREATE TABLE tca_therapist (name varchar(255) not null,  
                             centmed integer );
```

```
CREATE TABLE tca_patient (name varchar(255) not null,  
                            city varchar(250),  
                            gender varchar(1) ,
```

```
environment varchar(250),  
therapist varchar(250));
```

```
CREATE TABLE tca_episode (patient varchar(250),  
day date,  
hour varchar(250),  
episode varchar(250));
```

```
CREATE TABLE tca_emotion (patient varchar(250),  
day date,  
hour varchar(250),  
emotion varchar(250));
```

```
CREATE TABLE tca_meal (patient varchar(250),  
day date,  
hour varchar(250),  
meal varchar(250));
```

```
CREATE TABLE tca_activity (patient varchar(250),  
day date,  
hour varchar(250),  
activity varchar(250));
```

11. Anexo III. Carga de datos en el staging area

```
SET search_path TO tca_admin;
```

```
copy          tca_centmed(code,name,cp,city)          from
E'c:\\UOC\\TFM\\Data\\01 centmed.csv' USING delimiters ';'
CSV HEADER encoding 'LATIN1' ;
```

```
copy          tca_therapist(name,centmed)            from
E'c:\\UOC\\TFM\\Data\\02 therapist.csv' USING delimiters
';' CSV HEADER ;
```

```
copy    tca_patient(name,city,gender,environment,therapist)
from E'c:\\UOC\\TFM\\Data\\03 patient.csv' USING delimiters
';' CSV HEADER encoding 'LATIN1';
```

```
copy          tca_episode(patient,day,hour,episode)   from
E'c:\\UOC\\TFM\\Data\\04 episodes.csv' USING delimiters ';'
CSV HEADER ;
```

```
copy          tca_emotion(patient,day,hour,emotion)   from
E'c:\\UOC\\TFM\\Data\\05 emotion.csv' USING delimiters ';'
CSV HEADER ;
```

```
copy          tca_meal(patient,day,hour,meal)         from
E'c:\\UOC\\TFM\\Data\\06 meal.csv' USING delimiters ';' CSV
HEADER ;
```

```
copy          tca_activity(patient,day,hour,activity) from
E'c:\\UOC\\TFM\\Data\\07 activity.csv' USING delimiters ';'
CSV HEADER ;
```

12. Anexo IV. Carga de datos en el data warehouse

```
begin;

insert      into      tca_dw_admin.tca_d_evento(nombre_evento,
nombre_tipo_evento)
select      distinct      activity,      'ACTIVITY'from
tca_admin.tca_activity where activity is not null;

insert      into      tca_dw_admin.tca_d_evento(nombre_evento,
nombre_tipo_evento)
select      distinct      meal, 'MEAL' from tca_admin.tca_meal where
meal is not null;

insert      into      tca_dw_admin.tca_d_evento(nombre_evento,
nombre_tipo_evento)
select      distinct      emotion,      'EMOTION'      from
tca_admin.tca_emotion where emotion is not null;

insert      into      tca_dw_admin.tca_d_evento(nombre_evento,
nombre_tipo_evento)
select      distinct      coalesce(episode,'NO      EPISODE'),
'EPISODE'from tca_admin.tca_episode;
-----
-----
insert      into
tca_dw_admin.tca_d_tiempo(fecha,franja_dia,anio,mes,dia_mes
, trimestre, semana_anio,mes_nombre,dia_nombre)
select      distinct      day as fecha, hour as franja_dia,
extract(year from day) as anio ,extract(month from day) as
mes,extract(day from day) as dia_mes ,extract(quarter from
day)as trimestre,
extract(week from day) as semana_anio,
case extract(month from day)
when 1 then 'Enero'
when 2 then 'Febrero'
when 3 then 'Marzo'
when 4 then 'Abril'
when 5 then 'Mayo'
when 6 then 'Junio'
when 7 then 'Julio'
when 8 then 'Agosto'
when 9 then 'Septiembre'
when 10 then 'Octubre'
when 11 then 'Noviembre'
```

```

else 'Diciembre' end as mes_nombre,

case extract(dow from day)
when 1 then 'Lunes'
when 2 then 'Martes'
when 3 then 'Miercoles'
when 4 then 'Jueves'
when 5 then 'Viernes'
when 6 then 'Sabado'
else 'Domingo' end as dia_nombre
from (      select      distinct      day,      hour      from
tca_admin.tca_activity where day is not null
      union
      select distinct day, hour from tca_admin.tca_meal
where day is not null
      union
      select distinct day, hour from tca_admin.tca_emotion
where day is not null
      union
      select distinct day, hour from tca_admin.tca_episode
where day is not null
      order by 1,2 )as Fechas
;

```

```

-----
insert into tca_dw_admin.tca_d_paciente(nombre_paciente,
ciudad, nombre_entorno,sexo, nombre_terapeuta, id_centmed,
nombre_centmed, ciudad_centmed)
select pat.name, pat.city, pat.environment, pat.gender,
the.name, cent.code,cent.name, cent.city
from tca_admin.tca_patient pat, tca_admin.tca_therapist
the, tca_admin.tca_centmed cent
where pat.therapist = the.name
and the.centmed=cent.code
;
-----

```

```

insert
into
tca_dw_admin.tca_h_diario(id_paciente,id_fecha,id_evento,ca
ntidad_eventos)
select
pat.id_paciente,
tpo.id_fecha,
eve.id_evento,count(1)
from
tca_dw_admin.tca_d_paciente
pat,
tca_dw_admin.tca_d_tiempo tpo, tca_dw_admin.tca_d_evento
eve, tca_admin.tca_activity act
where pat.nombre_paciente = act.patient
and tpo.fecha = act.day
and tpo.franja_dia = act.hour
and eve.nombre_evento=act.activity
and eve.nombre_tipo_evento='ACTIVITY'

```

```
group by pat.id_paciente, tpo.id_fecha, eve.id_evento;

insert                                     into
tca_dw_admin.tca_h_diario(id_paciente,id_fecha,id_evento,ca
ntidad_eventos)
select                                     pat.id_paciente,          tpo.id_fecha,
eve.id_evento,count(1)
from                                     tca_dw_admin.tca_d_paciente      pat,
tca_dw_admin.tca_d_tiempo  tpo,   tca_dw_admin.tca_d_evento
eve, tca_admin.tca_meal meal
where pat.nombre_paciente = meal.patient
and   tpo.fecha = meal.day
and   tpo.franja_dia = meal.hour
and   eve.nombre_evento=meal.meal
and   eve.nombre_tipo_evento='MEAL'
group by pat.id_paciente, pat.id_terapeuta,pat.id_centmed,
tpo.id_fecha, eve.id_evento;

insert                                     into
tca_dw_admin.tca_h_diario(id_paciente,id_fecha,id_evento,ca
ntidad_eventos)
select                                     pat.id_paciente,          tpo.id_fecha,
eve.id_evento,count(1)
from                                     tca_dw_admin.tca_d_paciente      pat,
tca_dw_admin.tca_d_tiempo  tpo,   tca_dw_admin.tca_d_evento
eve, tca_admin.tca_episode epi
where pat.nombre_paciente = epi.patient
and   tpo.fecha = epi.day
and   tpo.franja_dia = epi.hour
and   eve.nombre_evento=coalesce(episode,'NO EPISODE')
and   eve.nombre_tipo_evento='EPISODE'
group by pat.id_paciente, pat.id_terapeuta,pat.id_centmed,
tpo.id_fecha, eve.id_evento;

insert                                     into
tca_dw_admin.tca_h_diario(id_paciente,id_fecha,id_evento,ca
ntidad_eventos)
select                                     pat.id_paciente,          tpo.id_fecha,
eve.id_evento,count(1)
from                                     tca_dw_admin.tca_d_paciente      pat,
tca_dw_admin.tca_d_tiempo  tpo,   tca_dw_admin.tca_d_evento
eve, tca_admin.tca_emotion emo
where pat.nombre_paciente = emo.patient
and   tpo.fecha = emo.day
and   tpo.franja_dia = emo.hour
and   eve.nombre_evento=emo.emotion
and   eve.nombre_tipo_evento='EMOTION'
group by pat.id_paciente, pat.id_terapeuta,pat.id_centmed,
tpo.id_fecha, eve.id_evento;
```

```
commit;

DO $$
DECLARE
    c_terapeutas CURSOR FOR select p1.nombre_terapeuta ,
coalesce (min(p1.id_terapeuta),
nextval('tca_dw_admin.seq_id_terapeuta')) id_terapeuta
FROM tca_dw_admin.tca_d_paciente
p1,tca_dw_admin.tca_d_paciente p2
where p1.nombre_terapeuta =
p2.nombre_terapeuta
group by p1.nombre_terapeuta;

    c_entornos CURSOR FOR select p1.nombre_entorno ,
coalesce (min(p1.id_entorno),
nextval('tca_dw_admin.seq_id_entorno')) id_entorno
FROM tca_dw_admin.tca_d_paciente
p1,tca_dw_admin.tca_d_paciente p2
where p1.nombre_entorno =
p2.nombre_entorno
group by p1.nombre_entorno;

    c_tipos_eventos CURSOR FOR select
e1.nombre_tipo_evento , coalesce (min(e1.id_tipo_evento),
nextval('tca_dw_admin.seq_id_tipo_evento')) id_tipo_evento
FROM tca_dw_admin.tca_d_evento
e1,tca_dw_admin.tca_d_evento e2
where e1.nombre_tipo_evento =
e2.nombre_tipo_evento
group by e1.nombre_tipo_evento;

BEGIN
    FOR terapeutas_record IN C_terapeutas LOOP

        update tca_dw_admin.tca_d_paciente
        set id_terapeuta = terapeutas_record.id_terapeuta
        where id_terapeuta is null
        and
        nombre_terapeuta=terapeutas_record.nombre_terapeuta;
    END LOOP;
    FOR entornos_record IN C_entornos LOOP

        update tca_dw_admin.tca_d_paciente
        set id_entorno = entornos_record.id_entorno
        where id_entorno is null
        and nombre_entorno=entornos_record.nombre_entorno;

    END LOOP;
    FOR tipos_eventos_record IN C_tipos_eventos LOOP
```

```
        update tca_dw_admin.tca_d_evento
           set          id_tipo_evento          =
tipos_eventos_record.id_tipo_evento
           where id_tipo_evento is null
           and
nombre_tipo_evento=tipos_eventos_record.nombre_tipo_evento;

        END LOOP;
END$$;
```

13. Anexo V. Fichero XML de definición del cubo OLAP

```
<Schema name="TCA_DW_Schema">

  <Dimension type="StandardDimension" usagePrefix="d_pac" visible="true"
highCardinality="false" name="d_paciente">

    <Hierarchy name="Centro_medico" visible="true" hasAll="true"
primaryKey="id_paciente">

      <Table name="tca_d_paciente" schema="tca_dw_admin"
alias="D_PACIENTE">

        </Table>

        <Level name="Centro_medico" visible="true" column="nombre_centmed"
nameColumn="nombre_centmed" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

          </Level>

          <Level name="Terapeuta" visible="true" column="nombre_terapeuta"
nameColumn="nombre_terapeuta" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

            </Level>

            <Level name="Paciente" visible="true" column="nombre_paciente"
nameColumn="nombre_paciente" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

              </Level>

            </Hierarchy>

          <Hierarchy name="Entorno" visible="true" hasAll="true"
primaryKey="id_paciente">

            <Table name="tca_d_paciente" schema="tca_dw_admin"
alias="D_PACIENTE">
```

```
</Table>

<Level name="Entorno" visible="true" column="nombre_entorno"
nameColumn="nombre_entorno" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

</Level>

<Level name="Paciente" visible="true" column="nombre_paciente"
nameColumn="nombre_paciente" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

</Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" usagePrefix="d_tpo" visible="true"
highCardinality="false" name="d_tiempo">

<Hierarchy name="Tiempo" visible="true" hasAll="true"
primaryKey="id_fecha">

<Table name="tca_d_tiempo" schema="tca_dw_admin"
alias="D_TIEMPO">

</Table>

<Level name="Anio" visible="true" column="anio" nameColumn="anio"
type="Numeric" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">

</Level>

<Level name="Trimestre" visible="true" column="trimestre"
nameColumn="trimestre" type="String" internalType="int"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

</Level>

<Level name="Mes" visible="true" column="mes" nameColumn="mes"
type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">

</Level>
```

```
<Level name="Semana" visible="true" column="semana_anio"
nameColumn="semana_anio" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
<Level name="Dia" visible="true" column="dia_mes"
nameColumn="dia_mes" type="String" uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
<Level name="Franja_dia" visible="true" column="franja_dia" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
<Hierarchy name="Anual" visible="true" hasAll="true"
primaryKey="id_fecha">
```

```
<Table name="tca_d_tiempo" schema="tca_dw_admin"
alias="D_TIEMPO">
```

```
</Table>
```

```
<Level name="Anio" visible="true" column="anio" nameColumn="anio"
type="String" uniqueMembers="false" levelType="Regular"
hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>
```

```
<Hierarchy name="Trimestral" visible="true" hasAll="true"
primaryKey="id_fecha">
```

```
<Table name="tca_d_tiempo" schema="tca_dw_admin"
alias="D_TIEMPO">
```

```
</Table>
```

```
<Level name="Trimestre" visible="true" column="trimestre" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
```

```
</Level>
```

```
</Hierarchy>

<Hierarchy      name="Mensual"      visible="true"      hasAll="true"
primaryKey="id_fecha">

  <Table        name="tca_d_tiempo"      schema="tca_dw_admin"
alias="D_TIEMPO">

    </Table>

    <Level      name="Mes"      visible="true"      column="mes"      type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

    </Level>

</Hierarchy>

<Hierarchy      name="Semanal"      visible="true"      hasAll="true"
primaryKey="id_fecha">

  <Table        name="tca_d_tiempo"      schema="tca_dw_admin"
alias="D_TIEMPO">

    </Table>

    <Level      name="Semana"      visible="true"      column="semana_anio"
nameColumn="semana_anio"      type="String"      uniqueMembers="false"
levelType="Regular" hideMemberIf="Never">

    </Level>

</Hierarchy>

<Hierarchy      name="Diaria"      visible="true"      hasAll="true"
primaryKey="id_fecha">

  <Table        name="tca_d_tiempo"      schema="tca_dw_admin"
alias="D_TIEMPO">

    </Table>

    <Level      name="Dia_semana"      visible="true"      column="dia_nombre"
type="String"      uniqueMembers="false"      levelType="Regular"
hideMemberIf="Never">

    </Level>

</Hierarchy>
```

```
<Hierarchy      name="Franja_dia"      visible="true"      hasAll="true"
primaryKey="id_fecha">

  <Table          name="tca_d_tiempo"          schema="tca_dw_admin"
alias="D_TIEMPO">

  </Table>

  <Level name="Franja_dia" visible="true" column="franja_dia" type="String"
uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

  </Level>

</Hierarchy>

</Dimension>

<Dimension type="StandardDimension" visible="true" highCardinality="false"
name="d_evento">

  <Hierarchy      name="Tipo_evento"      visible="true"      hasAll="true"
primaryKey="id_evento">

  <Table          name="tca_d_evento"          schema="tca_dw_admin"
alias="D_EVENTO">

  </Table>

  <Level name="Tipo_evento" visible="true" column="nombre_tipo_evento"
type="String"      uniqueMembers="false"      levelType="Regular"
hideMemberIf="Never">

  </Level>

  <Level name="Evento" visible="true" column="nombre_evento"
type="String"      uniqueMembers="false"      levelType="Regular"
hideMemberIf="Never">

  </Level>

</Hierarchy>

</Dimension>

<Cube name="H_DIARIO" visible="true" cache="true" enabled="true">

  <Table name="tca_h_diario" schema="tca_dw_admin" alias="H_DIARIO">
```

```
</Table>

<DimensionUsage source="d_paciente" name="d_paciente" visible="true"
foreignKey="id_paciente" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="d_tiempo" name="d_tiempo" visible="true"
foreignKey="id_fecha" highCardinality="false">

</DimensionUsage>

<DimensionUsage source="d_evento" name="d_evento" visible="true"
foreignKey="id_evento" highCardinality="false">

</DimensionUsage>

<Measure name="Ctd_eventos" column="cantidad_eventos"
aggregator="count" visible="true">

</Measure>

</Cube>

</Schema>
```