

Cuadro de mando visual enfocado a la gerencia de comercios minoristas basado en un sistema de Business Intelligence y DataWarehouse

The logo of the Universitat Oberta de Catalunya (UOC) is displayed in a dark blue color, consisting of the letters 'UOC' in a stylized, bold font.

Daniel Emperador Martel

Mapexbell Business
Intelligence
Business Intelligence

Nombre Tutor/a de TF

Humberto Andrés Sanz

**Profesor/a responsable de
la asignatura**

Atanasi Daradoumis

Fecha Entrega

16/6/2024

Universitat Oberta
de Catalunya



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

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Cuadro de mando visual enfocado a la gerencia de comercios minoristas basado en un sistema de Business Intelligence y DataWarehouse</i>
Nombre del autor:	<i>Daniel Emperador Martel</i>
Nombre del consultor/a:	<i>Humberto Andrés Sanz</i>
Nombre del PRA:	<i>Atanasi Daradoumis</i>
Fecha de entrega (mm/aaaa):	<i>06/2024</i>
Titulación o programa:	Grado en Ingeniería Informática
Área del Trabajo Final:	<i>Business Intelligence</i>
Idioma del trabajo:	<i>Castellano</i>
Palabras clave	<i>Tableau Software, comercio minorista, indicadores claves de negocio</i>

Resumen del Trabajo

Este Trabajo de Fin de Grado explica la situación actual de una empresa de comercio minorista real que utiliza informes en soporte papel confeccionados en hojas de cálculo que se han vuelto impracticables y cómo se ha utilizado la fuente de datos de origen de su sistema transaccional ERP para la construcción de un nuevo sistema de información basado en *Business Intelligence* para elaborar un cuadro de mandos que haga un resumen ejecutivo con indicadores claves de negocio (KPI) enfocados a un comercio minorista.

Para ello, se definen indicadores claves de negocio, se ha construido de cero un *DataWarehouse* en una base de datos, se establecen y programan procesos de extracción, transformación y carga y se diseñan en *Tableau Software*, la herramienta de visualización de datos.

Abstract

This Bachelor's Thesis explains the current situation of a real retail company that uses paper-based reports created in spreadsheets, which have become impractical. It describes how the original data source from their transactional ERP system has been used to build a new information system based on Business Intelligence to create a dashboard that provides an executive summary with key business indicators (KPIs) focused on retail.

To achieve this, key business indicators are defined, a DataWarehouse has been built from scratch in a database, extraction, transformation, and loading processes are established as well as scheduled, and designs are created in Tableau Software, the data visualization tool.

Índice

1. Introducción.....	1
1.1 Contexto y justificación del Trabajo.....	1
1.2 Objetivos del Trabajo.....	3
1.3 Impacto en sostenibilidad, ético-social y de diversidad.....	4
1.3.1 Dimensión sostenibilidad.....	4
1.3.2 Dimensión comportamiento ético y de responsabilidad social.....	5
1.3.3 Dimensión diversidad, género y derechos humanos.....	5
1.4 Enfoque y método seguido.....	5
1.5 Planificación del Trabajo.....	6
1.5.1 Diagrama de Gantt.....	7
1.5.2 Análisis de riesgos.....	8
1.6 Breve resumen de productos obtenidos.....	9
1.7 Breve descripción de los otros capítulos de la memoria.....	10
2. Toma de requisitos.....	12
3. KPI' s para comercios minoristas abarcados en el proyecto.....	15
3.1 Clientes.....	16
3.1.1 Fidelidad de clientes.....	16
3.2 Ventas.....	17
3.2.1 Ticket medio.....	17
3.2.2 Cifra de ventas.....	18
3.2.3 Margen comercial.....	18
3.3 Producto.....	20
3.3.1 Rotación del stock.....	20
3.3.2 Cobertura.....	21
3.3.3 Diferencia de inventario.....	22
4. KPI' s de comercio minorista que quedan fuera del alcance del proyecto	23
4.1 Clientes.....	23
4.1.1 Crecimiento de tráfico.....	23
4.1.2 Tasa de conversión.....	24
4.1.3 Mystery shopper.....	25
4.2 Personal.....	25
4.2.1 Rotación de personal.....	25
4.2.2 Índice de absentismo.....	26
4.2.3 Permanencia media.....	27
4.2.4 Competitividad salarial.....	27
4.2.5 Encuesta de clima laboral.....	28
4.2.6 Retorno de la formación.....	29

4.2.7 Nivel de polivalencia.....	30
4.3 Rentabilidad.....	30
4.3.1 La cuenta de resultados.....	30
4.3.2 Umbral de rentabilidad.....	31
4.3.3 Margen de seguridad.....	32
4.3.4 Coste de personal.....	32
4.3.5 Ventas por m2.....	33
4.3.6 Tasa de esfuerzo (alquiler).....	33
4.3.7 GMROI (Gross Margin Return On Investment).....	34
4.3.8 ROI (Return On Investment).....	35
4.3.9 Productividad FTE.....	35
4.3.10 Payback.....	36
4.4 Merchandising.....	36
4.4.1 Efectividad del metro lineal.....	36
4.4.2 Nivel de ejecución.....	37
5. Análisis de métricas en la fuente de datos de origen.....	38
6. Software empleado para la solución de Business Intelligence.....	38
6.1 PostgreSQL.....	38
6.2 Pentaho Data Integration.....	39
6.3 Tableau.....	40
6.4 DBeaver.....	40
7. Diseño del Data Warehouse.....	40
7.1 Diseño lógico de la base de datos del Data Warehouse.....	40
7.2 Implementación física del diseño lógico de la base de datos del Data Warehouse.....	45
7.3 Análisis sobre la calidad de los datos.....	47
7.3.1 Dimensión factura (dim_factura).....	48
8. Diseño de procesos ETL.....	48
8.1 Cargas iniciales e incrementales.....	49
8.2 Implementación procesos ETL.....	50
8.3 Pruebas de procesos ETL.....	53
9. Cuadro de mandos.....	54
9.1 Diseño conceptual de cuadro de mandos.....	54
9.2 Implementación de los cuadros de mandos.....	55
10. Conclusiones y trabajos futuros.....	59
10.1 Conclusiones generales.....	59
10.2 Conclusiones de objetivos.....	59
10.3 Planificación y metodología.....	60
10.4 Previsiones futuras.....	60
11. Glosario.....	62
12. Bibliografía.....	65

13. Anexos.....	67
------------------------	-----------

Lista de figuras

Figura 1. Diagrama de Gantt de la planificación de tareas del proyecto	7
Figura 2. Ejemplo de informe actual con el que se analizan las ventas por producto y unidad de negocio.	13
Figura 3. Informe actual con el que se analizan las ventas por unidad de negocio y consolidado respecto al mes y año previo.	13
Figura 4. Estructura contable de Activo = Pasivo + Patrimonio Neto	21
Figura 5. Esquema estrella tabla de hechos fact_venta	44
Figura 6. Esquema estrella tabla de hechos fact_transaccion_producto	45
Figura 7. Esquema estrella tabla de hechos fact_apunte_contable	46
Figura 8. Esquema resultado de la implementación de fact_venta en la capa core	47
Figura 9. Esquema resultado de la implementación de fact_transaccion_producto en la capa core	47
Figura 10. Esquema resultado de la implementación de fact_apunte_contable en la capa core	48
Figura 11. Repositorio procesos ETL	51
Figura 12. Trabajo extracción para la tabla de hechos de líneas facturadas.	52
Figura 13. Trabajo de lectura e inserción de filas en la capa core de líneas facturadas	53
Figura 14. Transformación para eliminar filas borradas en la tabla de apuntes contables	53
Figura 15. Trabajo para el proceso de la tabla de dimensión de centros de coste	54
Figura 16. Extracto del contenido de crontab del servidor Mac Mini Server osxmapexbell	54
Figura 17. Cuadro de mandos inspiracional “Resumen Ejecutivo” de Tableau	56
Figura 18. Prototipo de cuadro de mandos para retail	57
Figura 19. Ventana emergente con detalles de la marca seleccionada.	59

1. Introducción

El área temática escogida para este Trabajo de Final de Grado en Ingeniería Informática es la de Business Intelligence enfocada a la creación de un cuadro de mandos enfocado a la gerencia de una empresa de tipo PYME con la finalidad de que pueda disponer un una rápida visualización de su situación.

Las gerencias de muchas empresas PYMES tradicionalmente toman decisiones basados en hojas de cálculo llenas de números que, si bien las personas que las emplean tienen la costumbre de usarlas, no siempre es lo más eficiente.

Los beneficios que otorga la información visual están sobradamente demostrados: el cerebro humano no es bueno en visualizar patrones o tendencias en tablas numéricas y además le es más complejo memorizar números que figuras. El estadista británico Frank Anscombe en su artículo famoso "Graphs in Statistical Analysis" (Anscombe 1973) expuso que los "cálculos numéricos son exactos, pero los gráficos son robustos". Además, la psicología define las famosas Leyes de la Gestalt que a raíz de unos estímulos (figuras en el caso que ocupa en Business Intelligence), causan percepciones en el cerebro humano y que ayudan a asimilar y memorizar información rápidamente, es decir, en tomar decisiones.

No obstante, un elevado porcentaje de PYMES siguen analizando información puramente numérica, en muchas ocasiones por costumbre o reticencia al cambio por parte de los usuarios, lo que provoca que haya una elevada probabilidad de que las decisiones tomadas no sean bien correctas o las más óptimas.

El presente proyecto pretende ser un primer paso de cambio de paradigma para convertir el proceso de toma de decisiones en una empresa PYME que ha basado sus decisiones históricamente con tablas numéricas a un nuevo modelo de análisis visual con la ayuda de una herramienta de visualización de datos.

1.1 Contexto y justificación del Trabajo

En España, según el Ministerio de Industria y Turismo en su estudio más reciente de Estadísticas y publicaciones sobre PYME (Ministerio de Industria y Turismo, n.d.), la mayor parte de las empresas son pequeñas y medianas entendiéndose como tal aquellas que tienen hasta 249 asalariados y, en su conjunto, generan empleo a 9.327.126 personas a enero de 2024. Lo que significa que estamos ante un país que se constituye principalmente por empresas de dicho tamaño.

Históricamente, siempre ha habido una elevada brecha entre acceso a tecnologías de información y tamaño de las empresas. Por fortuna, las tecnologías de información son cada vez más accesibles. Sin embargo, esto no parece ser sinónimo de empleabilidad de estas herramientas en PYMES. Según (Ingeniariak Colegio Oficial de Ingenieros Industriales de Guipúzcoa 2022) la super digitalización avanza a ritmos agigantados que acrecientan aún más las desventajas competitivas de una PYME frente a las grandes empresas y que les supone un importante lastre en la carrera tecnológica.

En la actualidad no paramos de escuchar sobre inteligencia artificial, big data, machine learning, realidad virtual, metaverso... Sin embargo, muchas de las PYMES carecen de sistemas de información sobre el estado de salud de su organización que no sea más allá de la clásica contabilidad de cuentas de obligado cumplimiento u hojas de cálculo que se van volviendo incomprensibles a medida que la empresa crece.

Las causas de que las organizaciones menores no utilicen herramientas tecnológicas son diversas: ignorancia de su existencia, desconocimiento de cómo se utilizan, dificultad de implementación, falta de personal cualificado en la organización que implemente o mantenga los sistemas, desconocimiento de indicadores claves (KPI)...

El presente trabajo pretende la implementación de un sistema de Business Intelligence para una PYME real de comercio minorista no habituada al uso de indicadores claves de negocio (KPI de aquí en adelante) y herramientas de inteligencia de negocio.

La situación de la mencionada PYME es la de una empresa familiar líder de mercado en la Comunidad Autónoma de Canarias con más de 50 años de antigüedad que comercializa productos cosméticos profesionales a través de diferentes canales de distribución: una red de 17 tiendas físicas, un comercio electrónico y una pequeña red de venta directa a salones de belleza.

En su trayectoria, la organización ha empleado principalmente diferentes sistemas transaccionales ERP para el registro de sus operaciones de compra, venta, financieros y contables... con los diversos módulos que han compuesto los ERP en cada momento. Nunca ha utilizado más de un sistema transaccional salvo en épocas de migraciones de plataformas.

Las tomas de decisiones han estado siempre fuertemente vinculadas a la figura de la gerencia que, si bien han demostrado un crecimiento en el mercado que opera por su posición líder de mercado, se han visto muy centralizadas en dicha persona basándose principalmente en su *know-how* y en impresos de listados confeccionados en hojas de cálculo puramente numéricas que han ido

creciendo debido al incremento del volumen de negocio (mayor número de puntos de venta, mayor número de productos) y que han ido mermando la síntesis de información.

La información que ha utilizado la empresa ha sido obtenida siempre directamente del sistema transaccional. Esto acarrea una serie de inconvenientes como son el hecho de que los datos no se hayan tratado previamente como es habitual en los procesos de extracción, transformación y carga (*ETL* en sus siglas del inglés *Extract, Transform, Load*) para una mayor legibilidad de la información o la sobrecarga del sistema transaccional cada vez que se desea obtener información debido a que habitualmente implica la ejecución de consultas SQL sobre la base de datos de dicho sistema en producción causando, en no pocas ocasiones, la percepción de lentitud en el sistema por parte de los usuarios en horario laboral.

La casuística de la empresa que nos ocupa no es ajena a la de muchas otras PYMES en España que no transforman los datos que registran diariamente en información.

1.2 Objetivos del Trabajo

El presente Trabajo de Fin de Grado tiene como objetivo la creación de un sistema de información de Business Intelligence para la empresa que nos ocupa con cuadros de mandos especialmente pensados para el comercio minorista, dentro sus capacidades actuales, es decir, en base a la datos que tenga registrados, de modo que los usuarios empiecen a habituarse a procesar información de manera rápida, sencilla, visual, limpia y tratada, así como evitar la sobrecarga del sistema transaccional de producción del que se abusa en ocasiones.

Para lograr este objetivo, contamos con una extensa base de datos del sistema transaccional Openbravo ERP en el que, según qué tabla, hay cientos de miles (en algunos casos se superan los millones) de registros desde el año 2015 hasta 2024 y repartidos en numerosas tablas como son pedidos de compra y venta, albaranes de entrada y salida, facturas de compra y venta, cobros y pagos, cuentas financieras, amortizaciones...

Se procederá a crear un DataWarehouse en una base de datos relacional diferente, y en custodia de la empresa, donde se almacenarán definitivamente los datos limpios y transformados. En dicho almacén de datos, se añadirá y actualizará periódicamente información de nueva generación o modificada. De esta manera, la empresa podrá seguir expandiendo el proyecto más allá del alcance de este Trabajo de Fin de Grado.

El proceso de extracción de los datos de la base de datos relacional del sistema transaccional se hará mediante la herramienta *Pentaho Data*

Integration Community Edition el cual es una herramienta ETL bajo licencia open source que permite su uso gratuito pero con ciertas limitaciones frente a la licencia de pago. No obstante, una de las importantes limitaciones de la *Community Edition* es la imposibilidad de programar con periodicidad la ejecución de trabajos a través de su interfaz gráfica de usuario *GUI*. Por tanto, cada trabajo se programará mediante el fichero especial *crontab* que emplea la herramienta *cron*, un programa de naturaleza *Daemon* presente en sistemas operativos basados en Unix.

La herramienta visual que se utilizará es Tableau, plataforma bien consolidada en la visualización de datos propiedad de la organización multinacional *Salesforce, Inc.* La empresa dispone de licencia de Tableau, pero no ha explotado su potencial y será parte del objetivo de este proyecto el sacar partido a la herramienta.

Se definirán diversos KPI enfocados a un comercio minorista que posteriormente se calcularán y mostrarán en el software Tableau.

En definitiva, diseñaremos e implementaremos (según los KPI que se desean consultar) una base de datos relacional enfocada a un *DataWarehouse* que se actualiza periódicamente con la herramienta *Pentaho Data Integration* y se construirá un cuadro de mandos que muestre los valores KPI.

1.3 Impacto en sostenibilidad, ético-social y de diversidad

1.3.1 Dimensión sostenibilidad

identificar los impactos positivos y/o negativos del TF en las tres dimensiones de la CCEG

El presente trabajo de fin de grado en lo que a la dimensión sostenibilidad se refiere puede tener un impacto negativo en lo que ha consumo energético se refiere siempre y cuando el origen de la generación de la energía consumida haya generado residuos contaminantes como los radionucleidos o los gases de efecto invernadero.

En el proyecto se prevé el uso de herramientas que requieren de hardware encendido constantemente ya sea por alojamiento de software en servidores o el uso de telecomunicaciones. Por tanto, es evidente que hará uso de consumo energético del que sería idóneo que la energía provenga de fuentes renovables.

Otra posibilidad para la reducción de la huella de carbono es que los servidores utilizados sean *hardware* energéticamente eficientes.

No obstante, también puede haber un impacto positivo en la sostenibilidad por la reducción de consumo de recursos derivados de la impresión de documentos físicos. La impresión de documentos en una PYME es frecuente en

documentos que se desfasan rápidamente o son para un mero vistazo y desechar. Por ello, el ahorro de impresión que puede suponer un cuadro de mandos con indicadores claves puede ser muy significativo.

Parte de la motivación de este trabajo, aparte de la rápida consulta de indicadores claves de negocio, es precisamente la eliminación de diferentes documentos que se pueden abreviar eficientemente en un cuadro de mandos.

Se espera que el proyecto suponga un germen en la empresa hacia un mejor análisis de la información que permita hacer compras más acertadas sobre sus necesidades y reducir el despilfarro.

1.3.2 Dimensión comportamiento ético y de responsabilidad social

El proyecto no se considera que tenga impacto alguno en esta dimensión ya que ni mejora ni empeora los objetivos de desarrollo sostenible referidos a esta dimensión que hace especial hincapié en materias como justicia social, uso fraudulento del producto obtenido, condiciones laborales...

Este trabajo se basa fundamentalmente en analizar métricas de negocio minorista que puedan ser accesibles a los usuarios cuyo trabajo se sostiene principalmente en tomar decisiones y hacer palpable la realidad de una organización.

1.3.3 Dimensión diversidad, género y derechos humanos

Considero que el presente trabajo de fin de grado no tiene impacto alguno en aspectos de género, diversidad o derechos humanos.

Sí puede haber riesgo de afectar a la legislación de privacidad debido a que el *DataWarehouse* alberga datos personales de clientes de la empresa relativos al consumo que hayan hecho y que en caso de filtración de estos datos puede suponer una grave sanción y vulneración de los derechos de privacidad. Para preservar y reducir riesgo, se utiliza un servidor *on-premise* en custodia de la empresa y se han tomado medidas como encriptación de contraseña y cambio de puertos estándar de aplicaciones.

Sí se ha procurado en la confección del cuadro de mandos tener en cuenta la accesibilidad relativa al uso de colores que personas con daltonismo puedan distinguir en gráficos.

1.4 Enfoque y método seguido

El presente proyecto se pretende implementar en una empresa real de la que, como autor de este TFG, pertenezco como línea sucesora de la gerencia de la misma. Por tanto, no solamente soy persona implicada como autor sino también como interesado en la gerencia de la organización y consciente de la

necesidad real de la implantación de un sistema de información eficaz para la toma de decisiones.

Al mismo tiempo, me permitirá verificar la calidad de la información mostrada en los cuadros de mando que se implementen.

El proyecto se guiará utilizando la metodología estándar Project Management Body of Knowledge (PMBOK) definido por el Project Management Institute.

Dicha metodología se compone de un número concreto de 5 etapas: Iniciación, planificación, ejecución, seguimiento y control, y cierre. A su vez, el estándar sugiere la gestión de hasta diez áreas de conocimiento como son la integración, el alcance, el cronograma, los costes, la calidad, los recursos, la comunicación, los riesgos, compras y contratos, y los interesados.

No todo proyecto requiere de la gestión de todas estas áreas de conocimiento. Por tanto, en este TFG se gestionará la integración, el alcance, el cronograma, la calidad por conformidad del cliente, los recursos y los interesados del proyecto.

1.5 Planificación del Trabajo

El proyecto de este TFG abarca las fechas desde el 4 de marzo de 2024 hasta el 16 de junio de 2024.

Se procede a hacer una descripción sobre las actividades previsibles que hay que acometer para alcanzar los objetivos planteados así como una estimación de tiempos.

Fase	Nombre	Fecha de inicio	Fecha de fin	Días estimados	Estado
Inicio	Planteamiento y objetivos inicio	28/02/2024	03/03/2024	5	Finalizado
PEC 1 (4/3/2024 - 14/3/2024)	Definir actividades	04/03/2024	04/03/2024	1	Finalizado
	Contextualización y justificación	04/03/2024	06/03/2024	3	Finalizado
	Definición de objetivos	04/03/2024	06/03/2024	3	Finalizado
	Enfoque y método de trabajo	06/03/2024	07/03/2024	2	Finalizado
	Planificación de actividades	06/03/2024	08/03/2024	3	Finalizado
	Análisis de riesgos y correcciones del proyecto	08/03/2024	09/03/2024	2	Finalizado
	Redacción documentación PEC 1	12/03/2024	14/03/2024	3	Finalizado
	Entrega PEC1	14/03/2024	14/03/2024	0	Finalizado
PEC 2 (15/3/2024 - 17/04/2024)	Toma de requisitos funcionales y no funcionales	15/03/2024	17/03/2024	3	Finalizado
	Definición de KPI a utilizar	17/03/2024	19/03/2024	3	Finalizado
	Análisis de métricas fuente de datos origen	19/03/2024	21/03/2024	3	Finalizado

	Instalación de base de datos PostgreSQL	18/03/2024	18/03/2024	1	Finalizado
	Instalación y configuración de Pentaho Data Integration	19/03/2024	19/03/2024	1	Finalizado
	Diseño lógico base de datos	22/03/2024	26/03/2024	5	Finalizado
	Implementación física del diseño lógico BBDD	27/03/2024	30/03/2024	4	Finalizado
	Desarrollo SQLs para obtención de datos	20/03/2024	24/03/2024	5	Finalizado
	Pruebas SQL y verificación	25/03/2024	27/03/2024	3	Finalizado
	Redacción documentación PEC 2	16/03/2024	17/04/2024	33	Finalizado
	Entrega PEC2	17/04/2024	17/04/2024	0	Finalizado
PEC 3 (18/4/2024 - 19/05/2024)	Análisis calidad de los datos	18/04/2024	22/04/2024	5	Finalizado
	Diseño de procesos ETL	21/04/2024	25/04/2024	5	Finalizado
	Implementación procesos ETL	26/04/2024	30/04/2024	5	Finalizado
	Pruebas procesos ETL	01/05/2024	02/05/2024	2	Finalizado
	Pruebas de automatización trabajos ETL	03/05/2024	04/05/2024	2	Finalizado
	Diseño conceptual de cuadro de mandos	21/04/2024	23/04/2024	3	Finalizado
	Implementación de los cuadros de mandos	24/04/2024	30/04/2024	7	Finalizado
	Pruebas iterativas con usuarios de los cuadros de mandos	01/05/2024	07/05/2024	7	No finalizado
	Verificación correcta visualización de datos en cuadros de mandos	07/05/2024	09/05/2024	3	Finalizado
	Redacción documentación PEC 3	19/04/2024	19/05/2024	31	Finalizado
	Entrega PEC 3	19/05/2024	19/05/2024	0	Finalizado
PEC 4 (20/05/2024 - 16/06/2024)	Redacción documentación PEC 4	20/05/2024	08/06/2024	20	Finalizado
	Realizar presentación visual	10/06/2024	14/06/2024	5	Finalizado
	Informe autoevaluación	15/06/2024	15/06/2024	1	Finalizado
	Entrega PEC 4	16/06/2024	16/06/2024	0	Finalizado

Tabla Planificación del proyecto

La estimación de tiempo de dedicación al proyecto es de 2 a 3 horas diarias. Hay que tener en cuenta que este proyecto será desarrollado exclusivamente por una persona que tomará diferentes roles.

1.5.1 Diagrama de Gantt

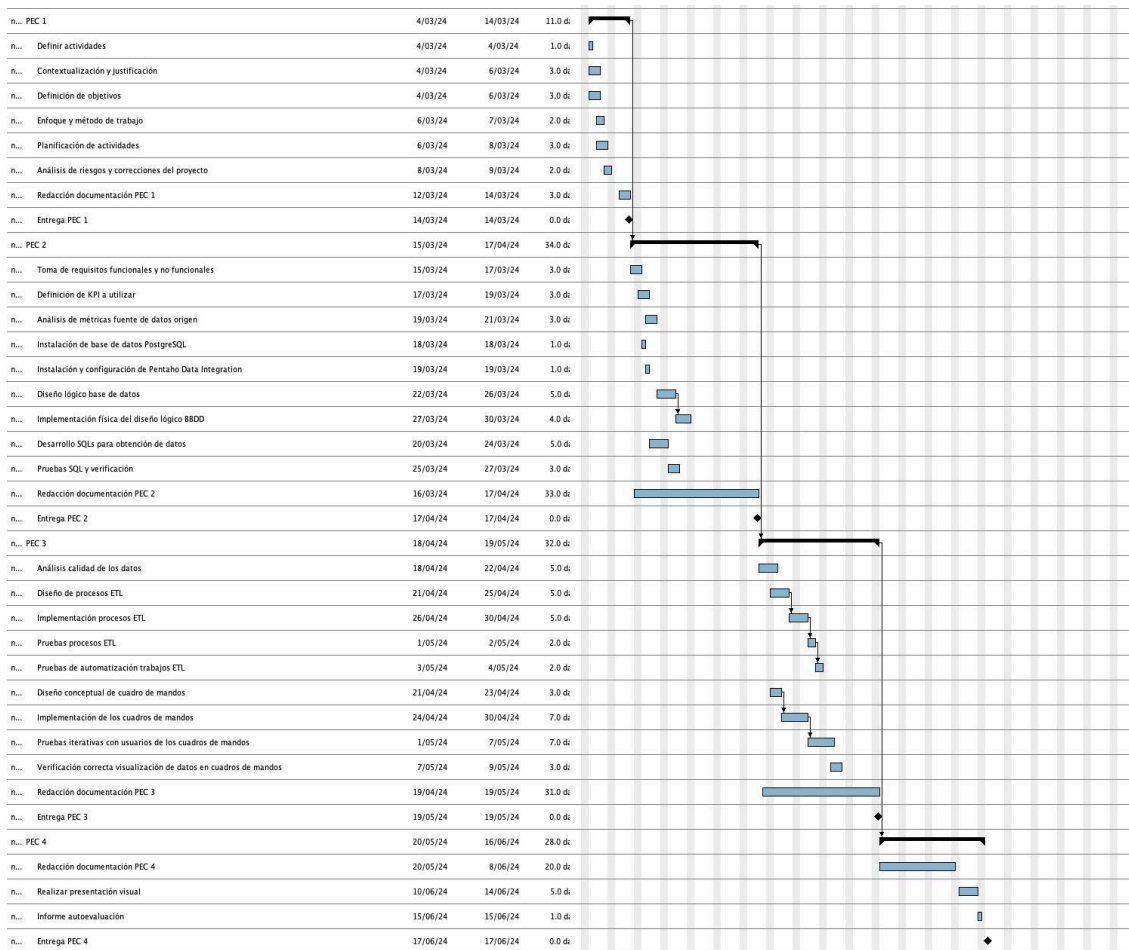


Figura 1. Diagrama de Gantt de la planificación de tareas del proyecto

Para abordar el proyecto se utilizará

1. *Mac Mini Server*: Ordenador servidor propiedad de la empresa que ejecutará tareas y contendrá las aplicaciones necesarias para el proyecto.
2. *PostgreSQL 15*: Base de datos relacional para albergar el *DataWarehouse*
3. *DBeaver*: Programa cliente gestor de base de datos
4. *Tableau*: Programa de visualización de datos
5. *Pentaho Data Integration Community Edition*: Programa de diseño y gestión de procesos ETL.

1.5.2 Análisis de riesgos

Como en todo proyecto, existe una serie de riesgos que pueden lastrar el avance del mismo. En la siguiente tabla se detallan cuáles se han detectado

ID	Riesgo	Descripción	Probabilidad	Impacto	Corrección
----	--------	-------------	--------------	---------	------------

R_001	Problemas de configuración de software	Existe la posibilidad que surjan problemas de configuración con el software que se vaya a utilizar o en el ordenador servidor que se va a utilizar	Medio	Medio	Hallar ayuda en medios telemáticos sobre la problemática concreta.
R_002	Curvatura de aprendizaje elevada	Las herramientas de software que se emplean en el proyecto pueden tener una fuerte curva de aprendizaje para cumplir los objetivos.	Alto	Alto	Hallar ayuda en medios telemáticos sobre la problemática concreta.
R_003	Baja disponibilidad horaria	Como proyecto unipersonal y encargado, existe la posibilidad de imprevistos laborales que resten la dedicación al proyecto a pesar de disponer de flexibilidad horaria.	Medio	Bajo	Priorizar lo que es importante sobre lo que es urgente. Delegar en otras personas trabajos que solapan el horario previsto para la dedicación del proyecto. Tener en cuenta fechas no laborables para dedicar al proyecto.
R_004	Fallo en estimación de dedicación	Existe el riesgo que se haya estimado (por encima o por debajo) las horas necesarias de dedicación a cada actividad	Alta	Medio	Corrección de la planificación estimada. Distribuir en más tiempo, rebajar expectativas en caso de ser una tarea muy extensa o considerar adelanto de otras tareas en caso de haber estimado por encima el tiempo de dedicación.
R_005	Fallos de Hardware	Existe la probabilidad que el hardware previsto para utilizar falle	Bajo	Alto	Sustitución inmediata del hardware dañado y restablecer copia de seguridad.

1.6 Breve resumen de productos obtenidos

Al finalizar el TFG se obtendrán los siguientes productos:

1. *DataWarehouse* que almacena los datos que genere la empresa en su sistema transaccional abierto a albergar otras fuentes de datos.
2. Modelado de datos del *DataWarehouse* para posteriores consultas.
3. Aplicación ETL *Pentaho Data Integration Community Edition* donde se diseñarán, gestionarán y almacenarán las transformaciones y trabajos que puede ejecutar.
4. Aplicación *Tableau* para la visualización de datos donde se diseñarán, gestionarán y consultarán los cuadros de mandos dentro del alcance de este TFG y más allá.

5. Documentación de instalación y configuración de las aplicaciones empleadas.

1.7 Breve descripción de los otros capítulos de la memoria

Capítulo 2 Toma de requisitos. En este capítulo se explica cuál es la situación actual de la empresa, en materia de análisis de datos, a la que se va a dedicar el proyecto. Se manifiesta brevemente qué documentos son utilizados actualmente y para qué sirven y las complejidades que conlleva para analizar la información

Capítulo 3 KPI' s para comercios minoristas. El capítulo 3 plantea un número de indicadores claves de negocio (KPIs) enfocados a los negocios de carácter minorista que entran en el proyecto y agrupados por área. Se da una breve descripción del indicador, para qué sirven, cómo se halla y qué datos son necesarios para su cálculo.

Capítulo 4 KPI' s de retail que quedan fuera del alcance del proyecto. En la misma tónica que el tercer capítulo, el cuarto sigue la explicación de más indicadores claves de negocio en comercio retail pero que por diferentes coyunturas (falta de datos, tiempo estipulado para el proyecto...) no han sido posibles incluirlas pero que se consideran relevantes y con posibilidad de futura inclusión en el almacén de datos.

Capítulo 5 Análisis de métricas en la fuente de datos de origen. Se describe muy brevemente la conclusión obtenida tras verificar el modelo de datos de la fuente de origen de los datos para la extracción de estos.

Capítulo 6 *Software* empleado para la solución de *Business Intelligence*. Se enumeran las principales herramientas que se utilizan en el proyecto para lograr el objetivo de implementación de *Business Intelligence* en la organización.

Capítulo 7 Diseño del *Data Warehouse*. El capítulo incluye detalle del diseño lógico de la capa *core* del *Data Warehouse*, los posibles esquemas comunes en un almacén de datos que se pueden utilizar, por cuál se ha optado así como el motivo, qué es un cubo *OLAP* y la granularidad de los datos utilizada. Se explica cómo se ha implementado físicamente el diseño lógico en el *Data Warehouse*. Además, se incluyen esquemas tanto del diseño lógico como el esquema resultado de la implementación. Finaliza el capítulo con una breve explicación de la calidad de los datos en la fuente de origen y los errores detectados que afectan al almacén de datos.

Capítulo 8 Diseño de procesos ETL. En este capítulo se describe en qué consiste un proceso ETL para la construcción de un *Data Warehouse*, cómo se ha organizado e implementado los procesos, es decir, la estructura del

repositorio de datos. Se finaliza con la mención de observaciones en la ejecución de pruebas: tardanza de los procesos, errores obtenidos...

Capítulo 9 Cuadro de mandos. En esta sección se explica rápidamente en qué consiste un cuadro de mandos, cuál ha sido la inspiración del cuadro de mandos a implementar y el resultado del cuadro de mandos obtenido a raíz del diseño y los datos de los que se alimenta.

Capítulo 10 Conclusiones y trabajos futuros. Este capítulo contiene las conclusiones que se han obtenido tras la elaboración del Trabajo de Fin de Grado así como cuál es la previsión de futuro del proyecto de ideas por explotar.

Capítulo 11 Glosario. Se incluye un pequeño glosario de términos utilizados en el texto de la memoria.

Capítulo 12 Bibliografía. Listado de referencias bibliográficas consultadas a lo largo del trabajo y que han sido citadas y me han servido de inspiración.

Capítulo 13 Anexos. Se incluye todo tipo de material/enlaces de interés como manuales de instalación del *software* utilizado, fichero

2. Toma de requisitos

La toma de requisitos es un proceso indispensable para toda clase de proyectos que se quiera emprender y que consiste en hallar qué se desea obtener (y qué queda fuera del alcance de un proyecto) en un plazo de tiempo estipulado.

Existen diversos tipos de técnicas para la identificación de estos requisitos, pero también es frecuente que los *stakeholders* no sean conscientes de lo que desean obtener. Esto recuerda a dos famosas citas:

“Si le hubiera preguntado a la gente, me habrían pedido caballos más rápidos” (Ford, n.d.)

“La gente no sabe lo que quiere hasta que se lo enseñas. Por eso nunca me he basado en las investigaciones de mercado. Nuestra tarea estriba en leer las páginas que todavía no se han escrito.” (Jobs, n.d.).

Con ambas citas se pretende aludir a lo expuesto en el apartado de contextualización y justificación de este proyecto. La mayoría de los *stakeholders* están habituados al análisis puramente numérico en hojas de cálculo que han crecido en tamaño junto al crecimiento del negocio de manera que se ha reducido su practicabilidad. Aunque han utilizado indicadores claves de negocio desconociendo el significado de lo que es un indicador clave de desempeño, no han llegado a plantearse su uso o se ha perseguido mostrar más datos puramente numéricos en documentos impresos.

En el mundo anglosajón existe el término *IKIWISI*, acrónimo en inglés de (*I// know it when I see it*) y que apunta a que una persona no es capaz de describir lo que desea hasta que lo vea y lo reconozca como aquello que deseaba.

La figura 2 es un ejemplo actual de la empresa para analizar las ventas, *stock* actual y *stock* mínimo. En dicha figura, cada fila corresponde a un producto y cada columna se corresponde a una organización o tienda que a la vez se descompone en otras tres columnas: ventas del producto, existencias del producto y *stock* mínimo del producto.

		Córdoba		New Lon Terna		BR		UR		GDO		BT		MA		AM		LAP		PR		TL		MR		VZ		MP		MM		LD		AM			
		EX	VE	RE	EX	VE	RE	EX	VE	RE	EX	VE	RE	EX	VE	RE	EX	VE	RE	EX	VE	RE	EX	VE	RE	EX	VE	RE	EX	VE	RE	EX	VE	RE			
9100010844	Wella blender plus toner- 35	0	14	2	6	8	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9100010845	Wella blender plus toner- 36	0	37	1	6	9	0	0	2	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9100010846	Wella blender plus toner- 35	0	19	2	6	12	0	0	2	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9100010852	Wella blender plus toner- 36	0	12	0	6	2	0	0	2	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9100010849	Wella blender plus toner- 36	0	37	0	6	9	0	0	2	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9100010848	Teñidos	0	79	5	30	37	0	0	9	1	5	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figura 2. Ejemplo de informe actual con el que se analizan las ventas por producto y unidad de negocio.

		Esencial
Como usuario quiero conocer la fidelidad de los clientes	Se desea conocer cuál es la tasa de fidelidad de los clientes nominales en cada unidad de negocio (tienda o consolidado), en una franja de tiempo determinada	Cientes
Como usuario quiero conocer cuántos clientes dejan de comprar	Se desea conocer cuál es la tasa de abandono de los clientes nominales en cada unidad de negocio (tienda o consolidado), en una franja de tiempo determinada	Cientes
Como usuario quiero conocer cuántos clientes se captan	Se desea conocer cuántos clientes nominativos se han captado en cada unidad de negocio (por tienda o consolidado), en una franja de tiempo determinada	Cientes
Como usuario quiero conocer cuál es la rotación de empleados	Se desea conocer cuál es la tasa de rotación de los empleados con la finalidad de saber si los empleados están satisfechos	Personas
Como usuario quiero conocer cuánto tiempo están ausentes los empleados	Se desea conocer cuál es la tasa de absentismo laboral de los empleados	Personas
Como usuario quiero conocer el margen de beneficio de la explotación del negocio	Como usuario quiero conocer el margen de beneficio de la explotación del negocio	Producto
Como usuario quiero conocer la rotación de los productos	Se desea conocer cuál es la rotación de cada uno de los productos en cada unidad de negocio (tienda o consolidado), en una franja de tiempo determinada	Producto
Como usuario quiero conocer cuánto tiempo estimado de stock dispone cada producto	Se desea conocer cuál es la cobertura de cada producto en cada unidad de negocio (tienda o consolidado), en una franja de tiempo determinada	Producto
Como usuario quiero conocer las diferencias de inventario	Se desea conocer cuáles son las diferencias de inventario generadas en cada unidad de negocio (tienda o consolidado) en una franja de tiempo determinada	Producto
Como usuario quiero conocer el ritmo de venta respecto a lo que se compra de un producto	Se desea conocer cuál es la tasa de ventas de un producto frente a la tasa de compra en cada unidad de negocio (tienda o consolidado) en una franja de tiempo determinada	Producto
Como usuario quiero conocer la cuenta de resultados	Se desea conocer la cuenta de resultados de cada unidad de negocio (tienda o consolidado) en una franja de tiempo determinada	Rentabilidad
Como usuario quiero conocer el punto de equilibrio/muerto	Se desea conocer el punto de equilibrio o muerto de cada unidad de negocio (tienda o consolidado) en una franja de tiempo determinada.	Rentabilidad
Como usuario quiero conocer el coste de personal	Se desea conocer cuáles son los costes de personal de cada unidad de negocio (tienda o consolidado) en una franja de tiempo determinada.	Rentabilidad
Como usuario quiero conocer la productividad de un vendedor	Se desea conocer cuál es la productividad de un vendedor en cada unidad de negocio (tienda o consolidado) en una franja de tiempo determinada	Rentabilidad
Como usuario quiero conocer las ventas por m2	Se desea conocer cuáles son las ventas por metros cuadrados de cada unidad de negocio (tienda o consolidado) en una franja de tiempo determinada	Rentabilidad
Como usuario quiero conocer la tasa de esfuerzo (alquiler)	Se desea conocer cuál es el peso relativo del alquiler de un local frente a las ventas	Rentabilidad

Como usuario quiero conocer el margen bruto de rentabilidad de la inversión	Se desea conocer cuál es el GMROI por unidad de negocio (por tienda o consolidado), en una franja de tiempo de terminada	Rentabilidad
Como usuario quiero conocer el flujo de caja	Se desea conocer cuál es el cashflow de cada unidad de negocio (por tienda o consolidado), en una franja de tiempo determinada.	Rentabilidad
Como usuario quiero conocer el margen de seguridad	Se desea conocer cuál es el margen de seguridad en el que se pueden reducir las ventas antes de entrar en pérdidas	Rentabilidad
Como usuario quiero conocer el ROI	Se desea conocer cuál es la tasa de retorno de inversión en cada unidad de negocio (tienda o consolidado), en una franja de tiempo determinada.	Rentabilidad
Como usuario quiero conocer las cifras de ventas	Se desea que la gerencia pueda consultar las ventas por unidad de negocio (por tienda o consolidado), en una franja de tiempo determinada	Ventas
Como usuario quiero conocer las unidades por transacción	Se desea conocer cuántas unidades se venden en cada transacción por unidad de negocio (por tienda o consolidado), en una franja de tiempo determinada	Ventas
Como usuario quiero consultar el ticket promedio	Se desea que conocer cuál es el ticket promedio generado por unidad de negocio (tienda o consolidado), en una franja de tiempo determinada	Ventas
Como usuario quiero consultar el precio medio de cada unidad vendida	Se desea conocer cuál es el precio medio de cada unidad vendida por unidad de negocio (tienda o consolidado), en una franja de tiempo determinada	Ventas
Como usuario quiero conocer cuáles son los 10 productos más vendidos y los 10 menos vendidos	Se desea conocer cuáles son los 10 productos más y menos vendidos en una franja de tiempo determinada.	Ventas
Como usuario quiero conocer las cifras de devoluciones	Se desea conocer cuántas devoluciones son efectuadas por unidad de negocio (tienda o consolidado), en una franja de tiempo determinada	Ventas
Como usuario quiero conocer las unidades devueltas	Se desea conocer cuántas unidades de venta son devueltas por unidad de negocio (tienda o consolidado), en una franja de tiempo determinada	Ventas
Como usuario quiero conocer el nivel de ventas cruzadas	Se desea conocer cuántas transacciones de venta incluyen otros artículos en cada unidad de negocio (tienda o consolidado), en una franja de tiempo determinada	Ventas

3. KPI' s para comercios minoristas abarcados en el proyecto

Los indicadores claves de desempeño (*Key Performance Indicators* en inglés y de ahora en adelante en sus siglas en inglés KPI) son métricas que se utilizan para medir el desempeño en diversidad de áreas como pueden ser una organización o proyectos con la finalidad de cumplir unos objetivos.

Cabe mencionar que todos los KPI 's son indicadores, pero no todos los indicadores son KPI' s ya que muchos de ellos no son indispensables para la gestión de una organización. Por ello, en este proyecto nos enfocaremos en aquellos indicadores que son indispensables para monitorizar el rendimiento en una empresa de comercio minorista que no ha empleado nunca el uso de indicadores de ninguna clase.

Según (Álvarez Orozco 2015) y (“KPIs del retail: los 12 indicadores más importantes para las tiendas físicas” 2023) algunos de los KPI esenciales para todo comercio minorista son los siguientes. Se ha optado por clasificar los KPI como sugiere (Álvarez Orozco 2015, 149) por áreas esenciales: clientes, personal, producto, rentabilidad y ventas.

3.1 Clientes

3.1.1 Fidelidad de clientes

(Álvarez Orozco 2015, 31)

Existen diversidad de estudios como el de (“¿Por qué fidelizar clientes cuesta 5 veces menos que captarlos? • Nömad” 2023) en el que se hace hincapié en que captar nuevos clientes es hasta 7 veces más costoso que mantener los clientes que la empresa dispone.

Los clientes fieles son la base de la operativa de cualquier empresa. En el caso de nuestra empresa minorista, ésta dispone de un elevado número de clientes nominativos con el que se puede medir cómo de fieles son sus clientes. No obstante, al ser un comercio minorista, también existe una elevada cifra de negocio de clientes anónimos. En estos casos, se descartan estos clientes porque resulta imposible medir su fidelidad. Sin embargo, podemos saber cuál es el porcentaje de clientes nominales frente a clientes anónimos.

Existen varias formas de calcular la fidelidad de los clientes. Por una parte medir el crecimiento del número de clientes fieles

$$\text{Crecimiento fidelidad} = \frac{\text{Clientes actuales} - \text{Clientes periodo anterior}}{\text{Clientes periodo anterior}} \times 100$$

Por otra parte, medir el porcentaje de clientes fieles

$$\text{Porcentaje de fidelidad} = \frac{\text{Ventas clientes habituales}}{\text{Ventas totales}} \times 100$$

En ambos casos, necesitaremos que nuestro almacén de datos aloje una tabla de hechos de las ventas que se han realizado.

Además, podemos medir la otra cara de la moneda, es decir, calcular la tasa de abandono de clientes. En tal caso, es conveniente que este valor sea lo más bajo posible. Sólo se tendrá en cuenta los clientes que hayan hecho al menos una compra desde 2015 en adelante.

$$Tasa\ de\ abandono = \frac{Clientes\ habituales\ sin\ compra}{Total\ de\ clientes\ habituales} \times 100$$

3.2 Ventas

3.2.1 Ticket medio

(Álvarez Orozco 2015, 82)

El objetivo de este indicador es informar sobre cuál es el importe promedio en cada venta que se realiza en el comercio. Este KPI se determina por dos variables:

- Número de unidades medio vendidas en cada operación o UPT (Unidades Por Transacción). Es decir, cuántas unidades se venden en un ticket.

$$UPT = \frac{Total\ unidades\ vendidas}{Número\ de\ operaciones}$$

- Precio medio de las unidades vendidas. Esta variable tiene una relación inversa con las unidades por transacción, es decir, se presupone que a mayor precio medio, menor UPT.

$$Pme = \frac{Ventas\ totales}{Total\ unidades\ vendidas}$$

Hallando estas dos variables, podemos obtener el ticket medio:

$$Ticket\ medio = UPT \times PMe$$

El objetivo del ticket medio es maximizar el importe del ticket medio equilibrando estas dos variables.

Cabe mencionar anecdóticamente que existe la posibilidad de obtener el ticket medio mediante la siguiente fórmula:

$$Ticket\ medio = \frac{Ventas\ totales}{Número\ de\ operaciones}$$

Sin embargo, este cálculo hace que perdamos valiosa información como son las unidades por transacción y el precio medio de las unidades vendidas. Por

tanto, para este proyecto, calcularemos ambas variables mencionadas ya que si el comercio conoce sus valores, podrá afinar mejor su estrategia sobre las acciones que le conviene acometer.

Por tanto, se necesita que el almacén de datos albergue una tabla de hechos con las ventas realizadas y que cada observación contenga las unidades vendidas y el precio de cada venta.

3.2.2 Cifra de ventas

(Álvarez Orozco 2015, 89)

Las “ventas a secas” es probablemente el KPI más utilizado por todas las empresas aún sin tener conocimiento de qué es un KPI. De manera rigurosa, su cálculo resulta más complejo de lo que de primeras se pudiera esperar. Se recomienda que este indicador se calcule en función de la siguiente fórmula:

$$\textit{Ventas} = \textit{Clientes} \times \textit{Tasa de conversión} \times \textit{UPT} \times \textit{PMe}$$

$$\textit{Clientes} = \textit{Visitas de clientes}$$

$$\textit{Tasa de conversión} = \frac{\textit{Número de operaciones de venta}}{\textit{Tráfico}} \times 100$$

Sin embargo, este proyecto tiene algunas dificultades. La variable *Tasa de conversión* necesita conocer el tráfico en cada unidad de negocio y en el caso de estudio la única medida del tráfico que podemos disponer es la del número de tickets de venta realizados, lo que equivale a que el valor de la variable *Tasa de conversión* es siempre 1 y obviamente es irreal. Por tanto, es sugerible a la empresa que tome acción en implementar algún sistema que sea capaz de medir el tráfico en cada unidad de negocio y así poder hallar una tasa de conversión fiel.

Este KPI necesita que el almacén de datos albergue una tabla de hechos con las ventas realizadas y que cada observación contenga las unidades vendidas y el precio de cada venta.

En definitiva, para este KPI, tomaremos como valor de la variable *Ventas* las ventas totales en un período determinado.

3.2.3 Margen comercial

(Álvarez Orozco 2015, 94)

De poco o nada sirve conocer cuánto vende un negocio si desconoce cuánto obtiene de cada venta que realiza. El margen comercial es el KPI que mide la

diferencia entre el importe de venta y el coste de adquisición del producto vendido. Con el diferencial obtenido, la organización podrá afrontar el resto de costes directos e indirectos para finalmente obtener un beneficio económico. Por tanto, es quizás el otro indicador clave que la mayor parte de las empresas utilizan sin conocer qué es un KPI (como en el caso que nos ocupa).

El margen comercial se ve afectado por variaciones en los precios de venta (cambios de precios, descuentos, promociones) o variaciones en los precios de compra (cambios de precios, rappels de compras...).

Hay dos maneras para analizar este KPI en un negocio:

1. Por valor monetario

$$\text{Margen (valor)} = \text{Precio de venta} - \text{Coste de la venta}$$

2. Porcentualmente

$$\text{Margen porcentual} = \frac{\text{Precio de venta} - \text{Coste de la venta}}{\text{Coste de la venta}} \times 100$$

Cabe mencionar que es habitual el debate sobre si el margen comercial porcentual es sobre coste de la venta o sobre el precio de venta y puede llevar a confusiones. La fórmula previa es la comúnmente utilizada, que es sobre coste de la venta. No obstante, como anécdota, el cálculo del margen comercial porcentual sobre el precio de venta es:

$$\text{Margen porcentual (sobre precio de venta)} = \frac{\text{Precio de venta} - \text{Coste de la venta}}{\text{Precio de venta}} \times 100$$

A modo de ejemplo para realizar una comparativa, si un precio de venta es de 60 y el coste de la venta es de 40, el margen comercial es de 20, pero si se calcula el margen comercial porcentual sobre coste de la venta, entonces el margen porcentual es de 50% mientras que el margen comercial porcentual sobre el precio de venta es de 33,33%.

Para tratar de dar explicación, en el primer caso la perspectiva es que se obtiene un 50% de margen (o rendimiento) sobre el coste de una venta, es decir, que $40 + 40 * 0,5 = 60$. En el segundo caso, que la perspectiva es que el margen es $60 * 0,3333 = 19,998$ y que el coste ha supuesto $60 * 0,6666 = 39,996$.

En cualquier caso, son dos caras de una misma moneda, y en el proyecto se reflejan ambos especificando la denominación de cada margen de cara a esclarecer toda duda independientemente del usuario que consulte.

El cálculo de este KPI necesita que el almacén de datos albergue una tabla de hechos con las ventas realizadas y que cada observación contenga las unidades vendidas, el precio de cada venta y el coste de la venta.

3.3 Producto

3.3.1 Rotación del stock

(Álvarez Orozco 2015, 98)

El inventario es un activo de valor inmovilizado que se traduce en liquidez cada vez que se realiza una venta. Como se puede apreciar en la figura 4, en la clásica estructura contable $\text{Activo} = \text{Pasivo} + \text{Patrimonio Neto}$, a mayor altura del activo, menor liquidez. Las existencias de una empresa forman parte del llamado activo corriente realizable, que si bien se presuponen líquidas, tienen menor liquidez que el llamado activo corriente disponible (correspondiente al dinero en efectivo o bancario que dispone).

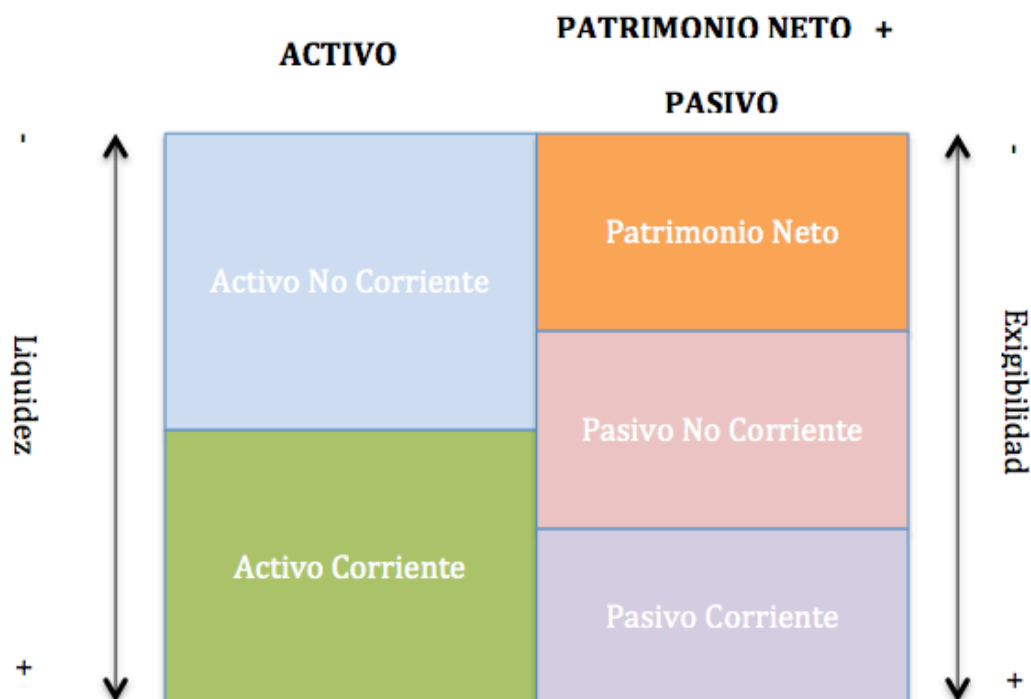


Figura 4. Estructura contable de $\text{Activo} = \text{Pasivo} + \text{Patrimonio Neto}$

En consecuencia, a mayor stock, mayor valor de activo inmovilizado que podría llegar a poner a la empresa en una delicada situación financiera en la que no pudiera afrontar a sus obligaciones de pago en el corto plazo.

Por el contrario, a menor nivel de stock, mayor liquidez tendrá la empresa a riesgo de no poder atender la demanda de sus clientes y, por consiguiente, en pérdidas de venta por roturas de stock.

Por tanto, se trata de balancear entre un nivel de stock que permita seguir generando ventas y el poder afrontar sus deudas en el corto plazo.

La rotación del stock es un KPI que mide el número de veces en un período determinado en el que la empresa ha renovado el stock en su totalidad, por ende, es un buen indicador sobre la gestión de stock e informa cuántas veces se ha sido capaz de recuperar la inversión en el tiempo analizado.

La rotación de stock se calcula con la siguiente fórmula:

$$\text{Rotación de stock} = \frac{\text{Ventas del periodo}}{\text{Stock medio del periodo}}$$

La ecuación permite emplear como valores de las variables tanto los precios de coste de los productos como el número de unidades. En este proyecto se utilizará el número de unidades.

El objetivo de este indicador es el de obtener números elevados de rotación. Además, a raíz de la rotación, se puede también obtener el llamado coeficiente de rentabilidad comercial que ciertos comerciantes minoristas utilizan para analizar la rentabilidad.

$$\text{Coeficiente de rentabilidad comercial} = \text{Margen} \times \text{Rotación de stock}$$

La valoración de este KPI necesita la tabla de hechos de los movimientos de mercancía y la tabla de hechos con las ventas.

3.3.2 Cobertura

(Álvarez Orozco 2015, 101)

El KPI de cobertura es una métrica que manifiesta por cuánto tiempo tiene el comercio disponibilidad de stock antes de entrar en rotura e íntimamente relacionada con el KPI rotación de stock. Permite obtener un conocimiento del equilibrio entre exceso y falta de existencias. Por consiguiente, posibilita adoptar una estrategia del tipo *Just In Time* y reducir los llamados costes de almacenaje derivados de obsolescencia del producto, mercancía percedera, robos, manipulación, ocupación de espacio físico, mantenimiento...

No obstante, se ha de tener especial cuidado con él porque presupone que la demanda de los productos es constante en el tiempo y existen variables que

pueden hacer variar notablemente el valor de la cobertura y guiar erróneamente al usuario. Las circunstancias que pueden afectar son, por ejemplo, el tiempo de servicio del proveedor, estacionalidad del producto, artículos sustitutos o complementarios... A pesar de ello este indicador ayuda a la organización de cara a futuras compras de mercaderías.

Se calcula con la siguiente fórmula:

$$\text{Cobertura de stock} = \frac{\text{Stock actual}}{\text{Ventas en el período}}$$

El tiempo de cobertura se expresa habitualmente en semanas o meses según la temporalidad analizada.

La evaluación de este KPI necesita la tabla de hechos de los movimientos de mercancía y la tabla de hechos con las ventas. El ERP de la organización tiene la posibilidad de conocer el stock en cualquier momento, pero con la salvedad de que lo calcula por sumatoria siempre que se desea, no almacena el stock por fechas, posiblemente el coste de espacio que supone almacenar el stock por fechas es notablemente superior al coste computacional de realizar operaciones de sumatoria.

3.3.3 Diferencia de inventario

(Álvarez Orozco 2015, 104)

El indicador clave diferencia de inventario evalúa las desviaciones en valor y en unidades entre el inventario teórico y el real.

El inventario en un comercio minorista es uno de los órganos vitales para su correcto funcionamiento además de que su registro es un requisito obligatorio de cara a la tributación obligatoria anual con la Agencia Estatal de Administración Tributaria (AEAT) en España para su declaración en las cuentas contables de existencias (300) y variación de existencias (610).

Existen dos cálculos sobre la diferencia de inventario:

1. Diferencia de inventario en valor. Se trata de la diferencia absoluta y que es la habitual que se calcula en aquellas empresas que hacen inventarios de manera regular.

$$\text{Diferencia inventario en valor} = \text{Valor stock teórico} - \text{Valor stock contado}$$

2. Porcentaje diferencia de inventario. Es la proporción de desviación de inventario que supone en una unidad de negocio debido a que cada una de ellas tendrá con toda seguridad cifras de negocio diferentes y no es lo

mismo una desviación de 1.000€ de una cifra de negocio inferior a 10.000€ que a una superior al millón de euros.

$$\text{Porcentaje diferencia de inventario} = \frac{\text{Diferencia de inventario en valor}}{\text{Ventas en el período}} \times 100$$

El cálculo de este KPI necesita una tabla con los hechos de ventas en el periodo y también tabla de hechos con los inventarios realizados, también una tabla con los movimientos de mercancía realizados para saber el stock teórico.

4. KPI' s de comercio minorista que quedan fuera del alcance del proyecto

La obra “*Cuadro de Mando Retail*” de (Álvarez Orozco 2015) sugiere hasta un total de 31 indicadores claves del desempeño para comercios minoristas en diferentes áreas esenciales de negocio. Tras revisar todas las que se sugieren en la obra mencionada y haber hecho un análisis de la fuente base de datos origen del sistema de información transaccional de la organización, se llega a la conclusión que hay ciertos KPI' s que nos son imposibles implementar para el proyecto.

El principal motivo de la exclusión de estos KPI 's se debe a que la organización carece de registros en sistemas transaccionales de los datos necesarios o los tiene registrados en soportes tradicionales como es el papel y supone una merma del plazo de cumplimiento de este TFG.

No obstante, se mencionan a modo de sugerencia para que cualquier empresa considere tomar iniciativas en registrar los datos para su posterior análisis y expandir el presente proyecto más allá del alcance de este TFG. Es recomendable que la organización contemple seriamente el modo de registrar hechos para evaluar los KPI 's excluidos.

4.1 Clientes

4.1.1 Crecimiento de tráfico

(Álvarez Orozco 2015, 26)

El crecimiento de tráfico en un comercio minorista es la variación del número de personas que acceden a un local. Este KPI se emplea para medir cuánta

gente entra en un local y cómo varía este número a lo largo del tiempo. Gracias a él, se puede conocer en mayor medida si acciones que realice la organización como las publicitarias, la competencia, acciones promocionales u otros motivos como la climatología afectan al tráfico.

La medición del tráfico supone que se implante un sistema de conteo, preferiblemente automático, de personal del que actualmente la organización carece de él. Por tanto, se descarta el uso de este KPI.

El cálculo del crecimiento de tráfico se realiza con la siguiente fórmula:

$$\text{Crecimiento de tráfico} = \frac{\text{Tráfico periodo actual} - \text{Tráfico periodo anterior}}{\text{Tráfico periodo anterior}} \times 100$$

Se sugiere a la compañía investigar sobre cómo recoger los datos de tráfico en cada ubicación ya que es crucial el conocimiento de cuántas personas visitan el establecimiento en cada período y, además, poder hallar el valor del KPI que vemos a continuación.

4.1.2 Tasa de conversión

(Álvarez Orozco 2015, 36)

Directamente relacionado con la medición de tráfico en un local es el indicador tasa de conversión. Es un KPI que se ha popularizado especialmente en los últimos años por los comercios electrónicos, pero que es igualmente necesario de aplicar en un comercio físico. Se trata de un indicador clave puesto que evalúa cómo de efectivo es un local comercial y la estrategia del comercio porque indica qué porcentaje de personas que han accedido al comercio han realizado al menos una compra. Por ende, a mayor valor de tasa de conversión, mayor efectividad tiene el comercio.

En este proyecto, como ocurre con el KPI crecimiento de tráfico, no podemos hallar la tasa de conversión por falta de registros del tráfico en el comercio. Una alternativa, como se mencionó anteriormente en el KPI cifra de ventas, es tomar como valor de tráfico el número de operaciones realizadas, pero la tasa de conversión pasará a ser una constante de valor 1 el cuál es un dato que no aporta valor. Por tanto, se descarta el uso de este KPI.

Para conocer la tasa de conversión se aplica la fórmula:

$$\text{Tasa de conversión} = \frac{\text{Número de operaciones de venta}}{\text{Tráfico}} \times 100$$

Por tanto, necesitamos de una tabla de hechos que registre el número de personas que entran en el local para medir el tráfico y también la tabla de hechos de las ventas realizadas.

4.1.3 Mystery shopper

(Álvarez Orozco 2015, 43)

Se trata de un KPI que evalúa diferentes variables de un comercio en base a la visita de un cliente especializado en realizar informes con el fin de obtener la valoración más objetiva posible del comercio.

Es importante que toda empresa minorista realice periódicamente evaluaciones objetivas de sus propios comercios, especialmente si dispone de un número elevado de locales (que también implica mayor número de empleados).

La medición se realiza mediante la confección de un informe que recoge la fecha de la visita y una serie de preguntas categorizadas a modo de cuestionario. Cada una de las preguntas tiene un grado de importancia y un valor otorgado por el *Mystery Shopper*.

Se descarta este KPI porque la empresa objeto de estudio no dispone de ningún registro formal de visitas de *Mystery Shopper*. Se sugiere que establezca de manera periódica la visita de profesionales del *mystery shopping* capaces de evaluar de manera lo más objetiva posible las tiendas de la empresa.

Se necesita una tabla de hechos con las visitas efectuadas por el *mystery shopper* con los resultados registrados.

4.2 Personal

4.2.1 Rotación de personal

(Álvarez Orozco 2015, 49)

Se conoce como rotación de personal la proporción de personal que abandona una empresa descartando motivos inevitables como jubilaciones, bajas médicas permanentes o fallecimientos en un período de tiempo concreto. Este indicador nos puede llegar a mostrar el nivel de implicación y motivación del personal de una empresa.

“Los clientes no son lo primero, lo primero son los empleados. Si cuidas de tus empleados, ellos cuidarán de tus clientes”. (Branson, n.d.)

Efectivamente, la medición de la rotación de personal es clave porque prevé el clima laboral de la organización, el compromiso e implicación de los empleados. Además, la rotación de personal conlleva altos costes que a menudo pasan desapercibidos para las empresas. Tales costes son los procesos de selección de nuevos empleados, formación, indemnizaciones, mala imagen corporativa...

Se mide la tasa de rotación de personal mediante la siguiente fórmula:

$$Tasa\ de\ rotación = \frac{Número\ de\ bajas}{Personal\ promedio} \times 100$$

Es habitual que las empresas cuenten con servicios profesionales independientes en gestorías laborales como es el caso de la organización de estudio. En dicho caso, es posible la obtención de los datos necesarios para el cálculo de la tasa de rotación, pero la transcripción supondrá un retraso en la planificación de este TFG. Por tanto, este KPI queda fuera de alcance del presente proyecto y se aconseja su aplicación en previsible expansión del proyecto.

Se necesita de una tabla de hechos que registre las fechas de contrataciones y las fechas de bajas así como la causa de la baja para lograr discernir si la baja procede o no para el cálculo.

4.2.2 Índice de absentismo

(Álvarez Orozco 2015, 53)

El índice de absentismo es un indicador que evalúa el tiempo perdido de jornadas en un periodo de tiempo. Su valor está intrínsecamente relacionado con la motivación del personal porque se presupone que un alto nivel de absentismo puede ser indicio de baja motivación. Aún más, implica un incremento de los costes ya que las horas perdidas han de ser sustituidas por más personal.

Se calcula la tasa de absentismo laboral con la siguiente fórmula:

$$Absentismo = \frac{Horas\ de\ trabajo\ perdidas\ por\ ausencia\ al\ trabajo}{Total\ horas\ contratadas\ en\ el\ periodo} \times 100$$

Es importante que toda empresa lleve un registro con las horas perdidas para conocer de primera mano cuánto tiempo se está perdiendo. Obviamente, es deseable que el valor de este índice sea lo más bajo posible.

En la empresa objeto de estudio esta información no está disponible en ningún sistema transaccional porque a fecha de este proyecto el registro de la jornada laboral se efectúa manualmente en soporte papel.

Por tanto, no se implementará en el previsible cuadro de mandos, pero se recomienda que comience a registrar estos datos electrónicamente. Se necesita la tabla de hechos que registre las horas de entrada y salida así como la tabla de hechos de contratos de los trabajadores.

4.2.3 Permanencia media

(Álvarez Orozco 2015, 54)

El índice de permanencia media nos indica la tasa de rotación del personal. Por el valor de este indicador, se puede conocer la capacidad que tiene la organización de retener talento en la empresa. Se presupone que a menor rotación, el equipo de la organización estará mejor formado, es decir, tendrá mayor capacitación y cualificación en el trabajo que desempeña, lo que transmite una imagen de estabilidad laboral en la organización y puede ser fuente de atracción de nuevos talentos.

No obstante, la baja tasa de rotación puede conllevar también la otra cara de la moneda y que es más habitual en PYMES y es que las oportunidades de desarrollo o promoción interna son más escasas y se pueda sospechar de que haya baja motivación o falta de interés por parte del empleado.

Por otra parte, una alta tasa de rotación implica que el empleo en la sociedad sea muy inestable y proyecte una imagen negativa de la calidad de los empleos.

Por tanto, es deseable una baja tasa de rotación, pero no se debe descuidar otros valores como es el clima laboral que se verá más adelante.

Se calcula la permanencia media con la siguiente fórmula:

$$\text{Permanencia Media} = \frac{\text{Suma de todas las permanencias por cada FTE}}{\text{Total FTE}}$$

El valor de FTE (*Full Time Equivalent*) se verá más adelante en el KPI Productividad FTE.

4.2.4 Competitividad salarial

(Álvarez Orozco 2015, 55)

La competitividad salarial alude al salario que los trabajadores de la compañía perciben por las labores realizadas frente al que se ofrece por las mismas labores en otras empresas. El salario se considera uno de los pilares sobre los que se sustenta un trabajador/a junto a la estabilidad laboral, el horario u otras condiciones que le puedan favorecer.

Se calcula la competitividad salarial con la siguiente fórmula:

$$\text{Competitividad salarial (\%)} = \frac{\text{Salario ofrecido} - \text{Salario competidor o de mercado}}{\text{Salario competidor o de mercado}} \times 100$$

El valor obtenido se interpreta en porcentaje y si es negativo es que la empresa ofrece un salario inferior respecto al valor referenciado (competidor o mercado) y si es positivo es que ofrece un salario superior.

Como es de esperar, su valor puede afectar a otros KPI como el de permanencia media que se ha descrito con anterioridad.

Se necesita para este cálculo por una parte la tabla de hechos de apuntes contables, concretamente los apuntes contables contenidos en la cuenta contable 465 Remuneraciones pendientes de pago (el coste de un trabajador es superior al valor contenido en dicha cuenta contable, pero lo relevante es lo que percibe la persona trabajadora), y por otra parte una tabla de hechos con los salarios ofrecidos por la competencia o por el mercado laboral.

4.2.5 Encuesta de clima laboral

(Álvarez Orozco 2015, 56)

El clima laboral es el ambiente generado en un lugar de trabajo por parte de los empleados y la organización a través de sus emociones. La razón de este KPI es para conocer cómo de satisfechos e implicados están los empleados en su puesto de trabajo. Se hace siempre de manera anónima con la finalidad de que cada persona tenga la oportunidad de poder expresarse libremente y la empresa conocer en mayor detalle cómo puede mejorar el clima dentro de sus posibilidades.

A mayor nivel de satisfacción de los empleados, se entiende que reporta mayores beneficios a la organización: mejor atención al cliente, disminución de tasa de rotación, disminución de tasas de absentismo laboral y que los propios trabajadores sean promotores de la compañía.

Las encuestas de satisfacción se ordenan por categorías para poder medir mejor los niveles de satisfacción por áreas.

Se mide la encuesta de clima laboral con la siguiente fórmula:

$$\text{Satisfacción por categoría} = \frac{\text{Puntuación categoría}}{\text{Puntuación máxima posible por categoría}}$$

$$\text{Satisfacción Total} = \frac{\text{Puntuación total}}{\text{Puntuación máxima posible total}}$$

Se sugiere a la organización objeto de estudio que considere realizar una encuesta anual de clima laboral para obtener una retroalimentación del ambiente de trabajo y acometer acciones de mejora.

Se precisa de una tabla de hechos que recoja los resultados de las encuestas formuladas.

4.2.6 Retorno de la formación

(Álvarez Orozco 2015, 66)

El retorno de la formación mide cómo de rentable son las acciones formativas que se han realizado. Si bien los resultados de una formación pueden no ser inmediatos, sí conviene tener en cuenta el plazo en el que comienzan a percibirse. Además, hay que tener en cuenta la tendencia de los resultados de ventas, es decir, el incremento de las ventas pudiera ser no debido a las acciones en formación sino por el propio mercado que tenga una tendencia alcista o bajista. Por tanto, se debe hallar el incremento de las ventas netas sobre la tendencia del mercado.

Este indicador clave se mide para conocer cómo de eficiente es la formación que reciben los trabajadores en un curso o acción formativa y por otra parte para conocer la calidad del formador/curso que reciben.

Se mide el retorno de la formación con la siguiente fórmula:

$$\text{Retorno de la formación (\%)} = \frac{\text{Incremento Ventas Neto Sobre la Tendencia} - \text{Coste de la formación}}{\text{Coste de la formación}} \times 100$$

El valor que se obtiene se interpreta en que por cada unidad monetaria invertida en formación se obtiene un porcentaje de rentabilidad en el retorno de la formación.

Una alternativa también es hallar cuánto invierte la empresa en formación. Para calcular el porcentaje de inversión en formación se puede utilizar la siguiente fórmula:

$$\text{Inversión en formación (\%)} = \frac{\text{Inversión en formación}}{\text{Ventas}} \times 100$$

El resultado será un porcentaje que la organización ha invertido en formación respecto a su cifra de ventas. Los cambios entre diferentes períodos demuestran el compromiso que tiene con la formación de los trabajadores/as.

Se requiere de la tabla de hechos de apuntes contables de donde se extraiga los valores de una cuenta contable 649 Otros Gastos Sociales que según (Omeñaca García 2008, 73), acumula los gastos de formación del personal.

4.2.7 Nivel de polivalencia

(Álvarez Orozco 2015, 70)

El nivel de polivalencia denota cómo de flexible es el personal contratado para acatar distintas tareas. En PYMES, aunque no exclusivamente, es especialmente valorada la flexibilidad de la plantilla debido a que éstas no suelen contar con un número elevado de trabajadores y que, por diferentes motivos, quizás no puedan abarcar más empleados a pesar de necesitar realizar mayor número de quehaceres.

Además, la polivalencia permite también que los trabajadores puedan rotar por distintos puestos de trabajo y evitar la dependencia de determinadas posiciones.

Se calcula el nivel de polivalencia de la siguiente manera:

$$\text{Nivel de polivalencia} = \frac{\text{Acciones realizables}}{\text{Coste de la formación}} \times 100$$

Entiéndase por acciones realizables aquellas acciones básicas que hay que desarrollar en un puesto de trabajo como es el almacén, atención al cliente, recepción de mercancías, gestión de caja...

4.3 Rentabilidad

4.3.1 La cuenta de resultados

(Álvarez Orozco 2015, 109)

La cuenta de resultados es sin lugar a dudas la métrica que toda empresa, al menos anualmente, evalúa para saber la viabilidad del negocio. Es conveniente que este KPI pueda ser desglosado por las diferentes partidas que la componen: gastos de personal, de infraestructura, marketing, gastos generales, amortizaciones e impuestos y que coincide con las cuentas contables del Plan General de Contabilidad. De esta manera, este KPI informará el peso de cada partida sobre las ventas registradas.

Además, también nos proporciona el valor de la EBITDA que es un parámetro financiero común en la contabilidad anglosajona que forma parte de la cuenta de resultados y que se ha popularizado en otros países. Este valor analiza la rentabilidad operativa de la empresa.

La cuenta de resultados requiere de una tabla de hechos contables para poder calcular cada partida.

	IMPORTE	%	FÓRMULAS
Ventas	1.500.000	100,0%	
Coste de las ventas	575.000	38,3%	
MARGEN COMERCIAL	925.000		=Ventas - Coste de las ventas
Diferencia de inventario	10.000	0,7%	
MARGEN BRUTO	915.000	61,0%	=Margen comercial - Diferencia de inventario
Personal	180.000	12,0%	
Alquileres	200.000	13,3%	
Gastos generales	130.000	8,7%	
Marketing	60.000	4,0%	
Otros varios	50.000	3,3%	
EBITDA	295.000	19,7%	= Margen Bruto - Personal - Alquileres - Gastos generales - Marketing - Otros varios
Amortizaciones	40.000	2,7%	
EBIT	255.000	17,0%	= EBITDA - Amortizaciones
Impuestos	80.000	5,3%	
BENEFICIO NETO	175.000	11,7%	=EBIT - Impuestos

4.3.2 Umbral de rentabilidad

(Álvarez Orozco 2015, 114)

También conocido como punto de equilibrio o punto muerto, el umbral de rentabilidad es un indicador clave que mide cuál es el nivel mínimo de ventas que se han de realizar para cubrir los costes fijos y en el que a partir de él se comenzará a obtener beneficios. A partir del mencionado punto, la organización comenzaría a obtener rentabilidad.

$$\text{Umbral de rentabilidad} = \frac{\text{Costes totales estimados}}{\text{Porcentaje Margen estimado}}$$

$$\text{Umbral de rentabilidad} = \frac{\text{Costes fijos del periodo}}{\text{Precio de venta} - \text{Coste de la venta}}$$

Además, se podrá obtener también cuál es el importe mínimo de compras que la organización debe afrontar para obtener ese nivel de ventas.

$$\text{Compras umbral de rentabilidad} = (1 - \% \text{ Margen estimado}) \times \text{Umbral de rentabilidad}$$

Calcular el umbral de rentabilidad requiere de la tabla de hechos contables para conocer los costes fijos y por otra parte necesitamos la tabla de hechos de venta para conocer los precios de venta y los costes de cada venta.

4.3.3 Margen de seguridad

(Grosclaude Manera 2006-2007, 8)

El margen de seguridad muestra porcentualmente cuánto pueden disminuir las ventas antes de entrar en una zona de pérdidas. Como bien es sabido, los márgenes que obtiene un negocio no son constantes en el tiempo y siempre están sujetos a variables dispares. A cada acción que una compañía haga, ha de prever cómo afectará a su cifra de negocio y, en caso que no sea fructífera, cuánto puede permitirse descender su margen comercial. A este fenómeno se le conoce como margen de seguridad.

El cálculo de este KPI se realiza de la siguiente manera:

$$\text{Margen de seguridad} = \frac{\text{Ventas del periodo} - \text{Ventas del umbral de rentabilidad}}{\text{Ventas del periodo}} \times 100$$

Necesitamos para este KPI la tabla de hechos de ventas y las tablas de hechos contables para conocer el umbral de rentabilidad.

4.3.4 Coste de personal

(Álvarez Orozco 2015, 116)

El coste del personal supone, en la inmensa mayoría de las empresas, una de las partidas de gastos más importantes por la cuantía que conlleva y más en los comercios minoristas. Por ello, es clave el control del KPI de coste de personal con una frecuencia mensual mínima.

Se entiende como coste del personal todo gasto que derivado de la empleabilidad de trabajadores: salario, indemnizaciones, seguridad social...

El cálculo de este KPI es relativamente sencillo mediante la siguiente fórmula:

$$\text{Porcentaje coste de personal} = \frac{\text{Costes de personal}}{\text{Ventas}} \times 100$$

Necesitamos para calcular este KPI la tabla de hechos contables y la tabla de hechos de ventas.

4.3.5 Ventas por m2

(Álvarez Orozco 2015, 122)

Este KPI muestra cuál es el rendimiento económico obtenido por cada metro cuadrado en un período de tiempo determinado. El espacio es un recurso sumamente importante para cada tienda dado que es la dedicación de un área en un periodo de tiempo a la generación de rentas económicas y que cada día transcurrido es irrecuperable. Los comercios disponen habitualmente de locales comerciales, en propiedad o de alquiler, en los que del total de sus metros cuadrados siempre se destina una cantidad inferior a la llamada área de tienda. Esto es que se excluye del cálculo los metros cuadrados que se destinen a office, servicios, almacenaje...

Cabe mencionar que la granularidad del cálculo pudiera ser tan fina como es el hallar las ventas por metro cuadrado de categoría de producto, marca... Esto se debe a la naturaleza diversa de los productos que ofrece un comercio y que presentan diferentes características y propiedades. Por ejemplo, la joyería presentará un ratio muy elevado frente a un comercio de alimentación.

El cálculo de este KPI se obtiene de la siguiente fórmula:

$$\text{Ventas por m2} = \frac{\text{Ventas}}{\text{Superficie neta de venta}}$$

Necesitamos para este KPI la tabla de hechos de las ventas y una tabla con la superficie neta de venta.

4.3.6 Tasa de esfuerzo (alquiler)

(Álvarez Orozco 2015, 125)

Los comercios minoristas cuentan con frecuencia con locales comerciales destinados a la actividad del negocio. Dichos espacios, bien son adquiridos o alquilados y suelen suponer, junto con la partida de gastos de personal, la segunda partida más elevada que deben confrontar mensualmente. Cuando se desarrolla una actividad económica, se trata de obtener la mayor rentabilidad posible del espacio disponible.

El KPI Tasa de esfuerzo mide la ponderación del gasto en el espacio alquilado y, en caso que el local sea de adquisición, compararlo con el precio de mercado en alquiler ya que supondría un coste de oportunidad frente a destinar el local en propiedad al alquiler.

El KPI se calcula con la siguiente fórmula:

$$Tasa\ de\ esfuerzo\ (alquiler) = \frac{Coste\ de\ alquiler\ del\ local}{Ventas} \times 100$$

Se debe tener en cuenta que se han de incluir todos los costes derivados del alquiler de un local. A pesar que un local a pie de calle suele ser exclusivamente la renta exigida por el propietario, en determinados casos como son los centros comerciales suelen exigir no solo una renta fija del alquiler sino también una renta variable por las ventas realizadas en el comercio si se supera un nivel de ventas o gastos comunitarios.

Se requiere de la tabla de hechos de apuntes contables que agregue los gastos de la cuenta contable 621 Arrendamientos y cánones que según (Omeñaca García 2008, 70) es la cuenta contable pertinente para el alquiler de inmuebles. No obstante, es conveniente utilizar una subcuenta contable específica para alquileres de inmuebles ya que la cuenta contable recoge muchos otros conceptos abstractos de alquileres y cánones. Además, se necesita también la tabla de hechos de las ventas realizadas.

4.3.7 GMROI (*Gross Margin Return On Investment*)

(“KPIs del retail: los 12 indicadores más importantes para las tiendas físicas” 2023)

Este KPI evalúa la rentabilidad de la empresa midiendo el margen bruto y el coste promedio de inventario en un período de tiempo. Es decir, es el cálculo del retorno de la inversión dedicado exclusivamente al inventario. Por tanto, describe cómo de rentable es el inventario de una organización mediante la relación entre margen obtenido y el coste del inventario vendido. Así, es deseable que el valor de este indicador sea lo más alto posible ya que indica un buen equilibrio entre ventas, margen y coste de inventario.

Se calcula con la siguiente fórmula:

$$GMROI = \frac{Ventas\ del\ período}{Valor\ promedio\ del\ inventario\ del\ período} \times \%Margen\ bruto\ sobre\ precio\ de\ venta$$

Necesitamos para este KPI la tabla de hechos de ventas, una tabla con los movimientos de cada mercancía de la que se deriva el stock y una tabla con el histórico de coste en cada momento de cada producto.

4.3.8 ROI (*Return On Investment*)

(Álvarez Orozco 2015, 131)

El retorno de la inversión o *ROI* en su acrónimo del inglés, es un KPI básico en el entorno económico financiero que mide la tasa de beneficio que se obtiene por cada unidad monetaria invertida en un período de tiempo determinado. Por tanto, es un paso más frente al KPI *GMROI* porque tiene en cuenta toda la inversión en un proyecto.

Se utiliza frecuentemente para tomar la decisión de si se debe de acometer un proyecto o no según el coste de oportunidad de la organización.

Para calcular este KPI, se utiliza la siguiente fórmula:

$$ROI = \frac{\text{Beneficio}}{\text{Inversión}} \times 100$$

Necesitamos para calcular este KPI la tabla de hechos contables para hallar el beneficio neto y también necesitamos obtener cuánto se ha invertido en el proyecto a analizar que puede entenderse la inversión acometida en un período de tiempo para el funcionamiento de la empresa.

4.3.9 Productividad FTE

(Álvarez Orozco 2015, 119)

La productividad FTE (acrónimo del inglés *Full Time Equivalent*) mide cómo de productiva es cada jornada laboral de un trabajador.

Se mide en jornadas completas debido a que los comercios minoristas suelen trabajar con diversidad de jornadas: completas, parciales, fines de semana...

Se calcula mediante la siguiente fórmula:

$$\text{Productividad por FTE} = \frac{\text{Ventas}}{\text{FTE}}$$

Donde

$$FTE = \Sigma(\% \text{ Jornada} \times \text{número de personas})$$

En España, a fecha de este proyecto, una jornada completa es de 8 horas diarias y 40 horas semanales. Por tanto, en este caso el valor de (% Jornada) es del 100%.

Calcular la productividad FTE requiere disponer de una tabla que registre el número de personas empleadas así como una tabla que registre el número de horas contratadas. Actualmente la organización no dispone de ninguna tabla en base de datos con esta información y el registro de horas trabajadas se realiza en soporte de papel. Por tanto, transcribir esta información conlleva un tiempo considerable.

4.3.10 Payback

(Álvarez Orozco 2015, 129)

En finanzas, el concepto de *Payback* es muy empleado para conocer cuál es el tiempo necesario para recuperar el importe de una inversión. Se trata en conjunto a los clásicos cálculos de viabilidad de un proyecto como son el valor actual neto (VAN) y la tasa interna de retorno (TIR).

El *Payback* utiliza los flujos de caja esperados en cada año (correspondientes a los calculados con el KPI EBITDA) y se decide por aquel proyecto que tenga el *Payback* más corto.

Se calcula de la siguiente manera:

$$\textit{Payback} = \textit{Inversión inicial}(-) + \textit{Cashflow 1} + \textit{Cashflow 2}... + \textit{Cashflow n} = 0$$

Sin embargo, este KPI carece de un importante concepto y es aquel del coste de oportunidad ya que no tiene en cuenta la rentabilidad de la inversión del momento en el que quizás la empresa pudiera obtener una mayor en otro proyecto.

Necesitamos pues de una tabla de hechos con todos los flujos de caja de las diferentes cuentas financieras que posea la empresa. (Omeñaca García 2008, 119) define una cuenta financiera a las cuentas del subgrupo 57 de Tesorería: cajas de efectivo y cuentas corrientes en bancos e instituciones financieras.

4.4 Merchandising

4.4.1 Efectividad del metro lineal

(Álvarez Orozco 2015, 74)

Todo comercio dedica un espacio físico de almacenamiento para los productos que venden. El valor del espacio tiempo es crucial y es conveniente conocer la rentabilidad obtenida de cada producto en el espacio dedicado.

Por tanto, se trata de una métrica que mide si cada producto tiene adjudicado el espacio óptimo a su capacidad para ser vendido. Es evidente que cada local

físico tiene una limitación de espacio en metros lineales y común que el comercio disponga de más producto disponible que de espacio. Por ello, ha de contar con la rotación del producto, entendiéndose en este caso por la exposición, en el metro lineal para que cada referencia tenga presencia.

Para hallar la efectividad del metro lineal se emplea la fórmula:

$$\text{Efectividad exposición} = \frac{\frac{\text{Ventas del producto}}{\text{Ventas Totales}}}{\frac{\text{Exposición del producto}}{\text{Exposición Total}}}$$

El resultado que se logra se interpreta de la siguiente manera:

- Valor es superior a 1: El producto necesitaría de mayor superficie de exposición porque se logra más ventas que la superficie dedicada.
- Valor es inferior a 1: El producto no tiene bien optimizado el espacio en tienda.

Por tanto, para hallar la efectividad necesitamos una tabla de hechos con las ventas y también necesitamos conocer la exposición total así como cuánto espacio de exposición tiene un producto.

4.4.2 Nivel de ejecución

(Álvarez Orozco 2015, 77)

El nivel de ejecución del merchandising mide cómo de eficiente es la exposición de un producto en un comercio. Se trata de medir una serie de pequeñas acciones que a menudo son imperceptibles, pero que tienen una fuerte repercusión en el resultado de las ventas. Algunos ejemplos son la visibilización del precio del producto, presentación del producto en todas sus variantes, productos limpios...

Se calcula el nivel de ejecución con la siguiente fórmula:

$$\text{Nivel de ejecución} = 100 - \frac{\text{Nº de errores}}{\text{Nº de unidades} * \text{Nº de oportunidades}}$$

Para cada producto que haya en exposición, se auditan con las acciones definidas.

El significado de las variables que intervienen son:

- Número de errores corresponde a todo error o defecto detectado en la observación.
- Número de unidades son las unidades auditadas del producto afectado.
- Número de oportunidades son las acciones definidas. Así, si por ejemplo se audita el precio, el etiquetado y la presentación, el valor será 3.

El resultado se traduce en que el porcentaje obtenido es el el nivel de ejecución del producto y será deseable el valor más alto posible.

Por tanto, se requiere de una tabla de hechos que registre para cada producto las inspecciones para registrar los errores.

5. Análisis de métricas en la fuente de datos de origen

La inspección de la fuente de datos origen ha sido fundamental para poder discernir qué indicadores claves del desempeño se pueden implementar en el proyecto. Se ha concluido que muchos de ellos se pueden obtener esencialmente de tres tablas de hechos en los que se registran: tabla de ventas, tabla de transacciones de productos y tabla de apuntes contables. En cambio, muchos otros KPI de cálculos más complejos han tenido que ser descartados pero, como se ha mencionado con anterioridad, se ha considerado conveniente citarlos para que este trabajo de fin de grado sirva de futura referencia de consulta para expandir el proyecto.

Es destacable que en el análisis de la fuente de origen, el programa transaccional o su base de datos parece contemplar el registro de hechos que son necesarios para el cálculo de algunos KPI' s que se han dejado fuera como es por ejemplo el registrar la formación que tiene cada empleado o sus contratos laborales, pero que no se han registrado por desconocimiento.

En consecuencia, para el diseño lógico del *Data Warehouse* se diseñan los esquemas con los ejes centrales sobre las tres tablas de hechos mencionados anteriormente y con las tablas de dimensiones por las que se desea medir los hechos.

6. Software empleado para la solución de *Business Intelligence*

6.1 PostgreSQL

Un proyecto de *Business Intelligence* conlleva la necesidad de diseñar e implementar una base de datos que hospeda los datos generados por los diferentes sistemas de información transaccional que disponga una organización.

PostgreSQL es un sistema de base de datos objeto-relacional que utiliza el estándar SQL, con determinadas funcionalidades propias, bajo licencia *open-source* con lo que puede ser utilizado y distribuido gratuitamente.

En el presente trabajo, PostgreSQL es la solución con la que se hará la implementación física del *Data Warehouse*. Las razones de por qué se ha optado por esta solución son diversas:

1. PostgreSQL tiene una larga trayectoria superior a los 30 años como alternativa a un *RDBMS*.
2. Cumple con los requisitos *ACID* de una base de datos.
3. La licencia de uso es gratuita por lo que cualquier persona o empresa puede utilizarla sin coste alguno, por tanto, el coste del proyecto se reduce.

La organización objeto de estudio del proyecto utiliza un sistema de información ERP que también emplea PostgreSQL para registrar las transacciones.

Si bien la empresa podría contratar un servicio *cloud* especializado en *Data Warehouse* como Amazon Redshift para albergar el almacén de datos, se ha optado por la solución *on premise* por el costo y porque el proyecto es un prototipo. Más adelante, según el caso, sí es factible plantearse el servicio *cloud*.

6.2 Pentaho Data Integration

Pentaho Data Integration es un programa propiedad de la multinacional japonesa *Hitachi* bajo sus servicios de infraestructura de datos conocido por *Vantara*. El programa en cuestión es una herramienta *ETL* y se ofrece en más de una edición: la edición empresarial y la edición de comunidad.

La explicación de las diferencias entre ambas versiones es extensa y no es objeto de estudio de este proyecto. En esta ocasión por razones de coste, se opta por el uso de la edición comunitaria al ser de uso gratuito. Sin embargo, una diferencia relevante es que la edición comunitaria carece de la posibilidad de programar la ejecución de trabajos. Por tanto, para salvar esta importante característica, se opta por la programación de tareas mediante el programa *cron* presente en todo sistema basado en Unix.

No obstante, toda empresa que considere utilizar *Pentaho Data Integration* debe estimar las funcionalidades que proporciona la edición empresarial en caso que contenga características apreciables.

6.3 Tableau

Tableau Software es un *software* utilizado para la visualización de datos interactivos y es capaz de construir gráficos a raíz de la diversidad de fuentes de datos de origen. Las principales razones de por qué se ha optado por *Tableau* como herramienta para la construcción de los cuadros de mandos son que es una solución multiplataforma que funciona bajo entorno *Windows*, *Mac* o *Web*, su posición en el mercado como una de las herramientas principales de visualización de datos y que la organización cuenta ya con una licencia. Sin embargo, la organización no ha sacado partido a las posibilidades de la plataforma.

6.4 DBeaver

DBeaver es un *software* utilizado para la gestión de bases de datos de diferentes fabricantes. Se utiliza en este proyecto para la consulta en las diferentes bases de datos y crear los esquemas en el almacén de datos. Cuenta además con una edición comunitaria gratuita que cumple con los requisitos para el trabajo proyectado.

7. Diseño del *Data Warehouse*

7.1 Diseño lógico de la base de datos del *Data Warehouse*

La implementación de un KPI a un cuadro de mandos requiere reflexionar de dónde se van a obtener los valores de las variables que utilizan para su cálculo. Es habitual que las organizaciones tiendan a emplear un único sistema transaccional como un ERP y que esperen que dicho sistema proporcione toda la información que deseen.

Ahora bien, aunque esta concepción no es del todo incorrecta, tampoco es la idónea. Ciertamente es que muchos sistemas transaccionales traen de serie informes (como los clásicos informes contables de cuentas de resultados o pérdidas y ganancias) o permiten la confección de informes por parte de personal cualificado que otorgue soporte del sistema.

Sin embargo, las bases de datos de un sistema transaccional están diseñadas precisamente para el registro de transacciones, no para el propósito de analizar datos a diferencia de una base de datos diseñada como *Data Warehouse*. A modo de ejemplo, es de esperar que el sistema transaccional sea eficiente en el registro de datos como pedidos, movimientos de mercancía... y estos se registren de uno en uno y en granularidades finas. En cambio, cuando se desea analizar los datos registrados, es habitual que se consulten miles de registros a la vez porque la granularidad de la información cambia. El analista de datos desea consultar datos históricos y agregados. Lo cual significa que la base de datos del almacén de datos ha de estar optimizada para grandes consultas que permita obtener información rápidamente.

Aún más, los datos en dichos sistemas habitualmente se encuentran en estados no deseables para fines analíticos. En los sistemas transaccionales es común que sean manipulados por múltiples usuarios en una empresa e incluso por clientes si estos interactúan con algún sistema de la organización como puede ser un comercio electrónico. Por consiguiente, no es chocante que los datos registrados contengan errores de transcripción o cualquier otra índole. En cambio, se pretende que los datos destinados a ser almacenados en el *Data Warehouse* hayan pasado por variedad de operaciones como la limpieza, corrección, agrupación, renombramiento... En efecto, estas tareas consumen gran parte del esfuerzo en el procesamiento de datos.

El *Data Warehouse* de este proyecto se ha diseñado siguiendo una serie de buenas prácticas habituales en el modelado de almacenes de datos. La principal razón de realizar un modelado diferente en el almacén de datos respecto al modelado de la base de datos de origen es que la base de datos destino estará mejor organizada para la finalidad de ejecutar consultas de análisis rápidamente. Otro motivo es que el almacén de datos estará adaptado para funcionar con cualquier solución de *Business Intelligence* y podemos evitar una situación de *vendor lock-in*.

En PostgreSQL se crea una base de datos que contiene dos capas: la capa comúnmente llamada *Staging* y la capa *core* o almacén de datos.

La capa *Staging* es el esquema donde se almacenan temporalmente los datos más recientes provenientes de la fuente de datos origen. Se pretende extraer los datos de la fuente de origen cada cierto tiempo y que no hayan sido manipulados o mínimamente. El objetivo es que el proceso de extracción se efectúe lo antes posible y en horas valle de uso de la fuente de origen para evitar cualquier tipo de sobrecarga del sistema transaccional original.

La capa *core*, también conocida como capa de acceso, es el destino final de los datos y es donde se almacenarán permanentemente. Además, los datos contienen la lógica de negocio deseada y se encuentran listos para ser analizados. Esto conlleva que los datos hayan pasado previamente por tareas de transformación y carga de manera que sea amigable a los usuarios finales.

Además, se ha diseñado la capa *core* creando cubos *OLAP* siguiendo un esquema de tipo estrella.

Los cubos *OLAP* son estructuras de datos habituales en los almacenes de datos cuyas principales características son que los datos se encuentran precalculados (por lo que hay un ahorro de coste computacional cada vez que se quiera ejecutar una consulta), el rendimiento de los cubos es notablemente superior al de consultas a diferentes tablas sobre un sistema *OLTP*, están contruidos para casos de uso específicos. Cada cubo se construye en base a una tabla de hechos (*facts* en inglés) que albergan datos medibles y que se relacionan con diferentes tablas de dimensiones que definen el contexto de cada hecho ocurrido.

Cabe señalar que es posible modelar los cubos *OLAP* mediante dos esquemas: esquema *snowflake* (copo de nieve) y esquema *star* (estrella). Es frecuente que los sistemas *OLTP* se modelen con el esquema *snowflake* por el hecho de cumplir con la normalización de datos cuyo propósito son diferentes: se busca priorizar el rendimiento en tareas de lectura/escritura, evitar redundancia de datos, más sencillo de mantener. En cambio, en este proyecto se ha optado por el esquema estrella en el que el modelado contiene los datos no normalizados (hay redundancia de datos) por los principales motivos de que el esquema está diseñado para priorizar el rendimiento de obtención de datos y es más amigable al usuario para entender las relaciones de las tablas.

En base a los KPI' s que se han definido para este proyecto, se han modelado tres cubos *OLAP* cuyos hechos son los principales: líneas de productos facturados, líneas de apuntes contables y líneas de transacciones de productos. La razón de estos tres únicos cubos es que tras analizar la fuente de datos de origen, la extracción a estos cubren las necesidades previstas.

Otro asunto importante a nombrar es la granularidad del dato que se ha tenido en cuenta para el diseño lógico. Se ha optado por la granularidad más fina posible por la sencilla justificación que a menor granularidad, mejor se puede discernir la información en la herramienta de visualización de datos a gusto del usuario. Así por ejemplo, se ha escogido como venta realizada cada línea incluida en una factura y no la cabecera de la factura.

A continuación se muestra el diseño lógico realizado para el proyecto:

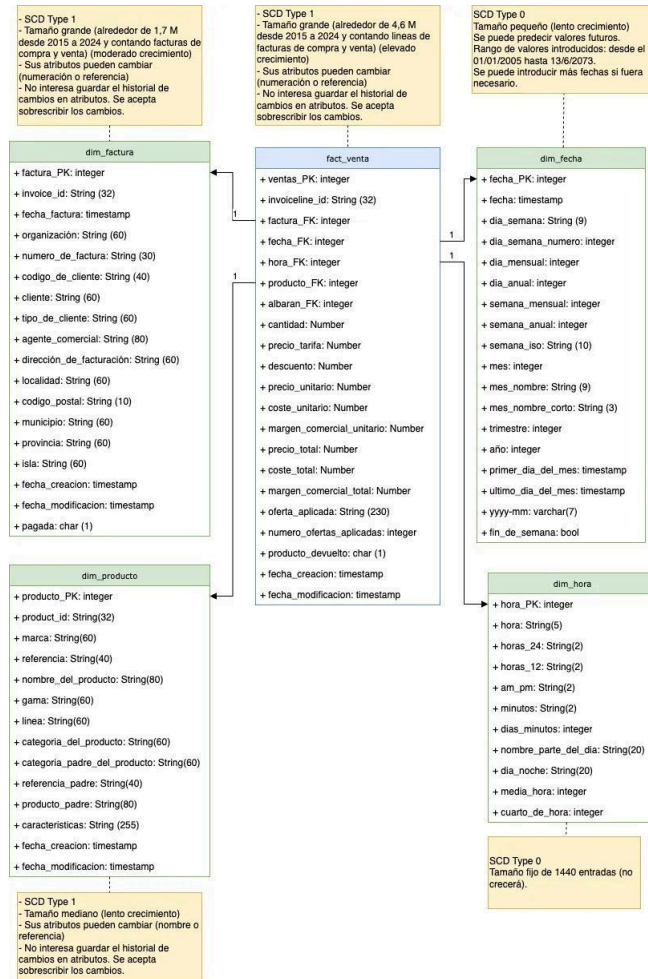


Figura 5. Esquema estrella tabla de hechos fact_venta

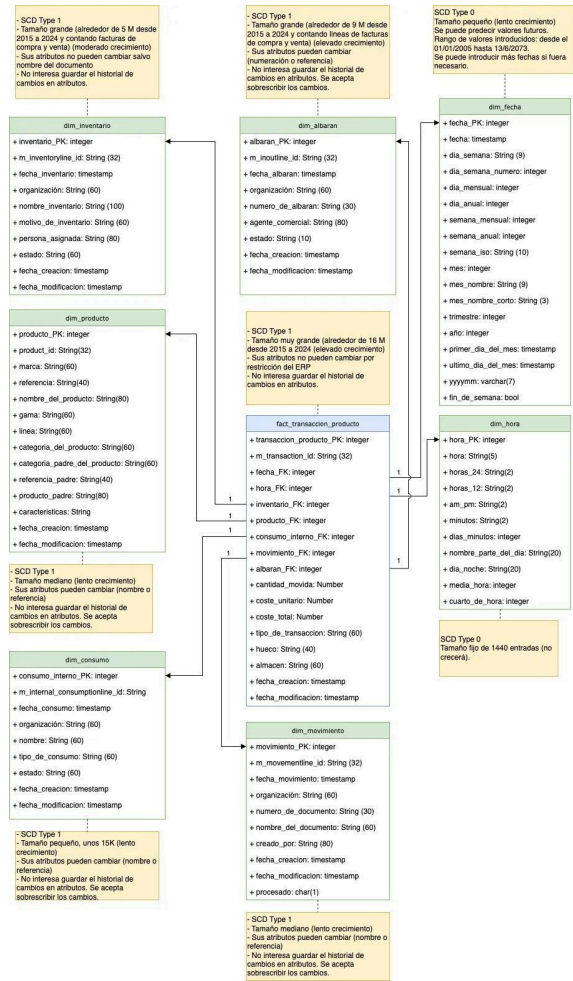


Figura 6. Esquema estrella tabla de hechos fact_transaccion_producto

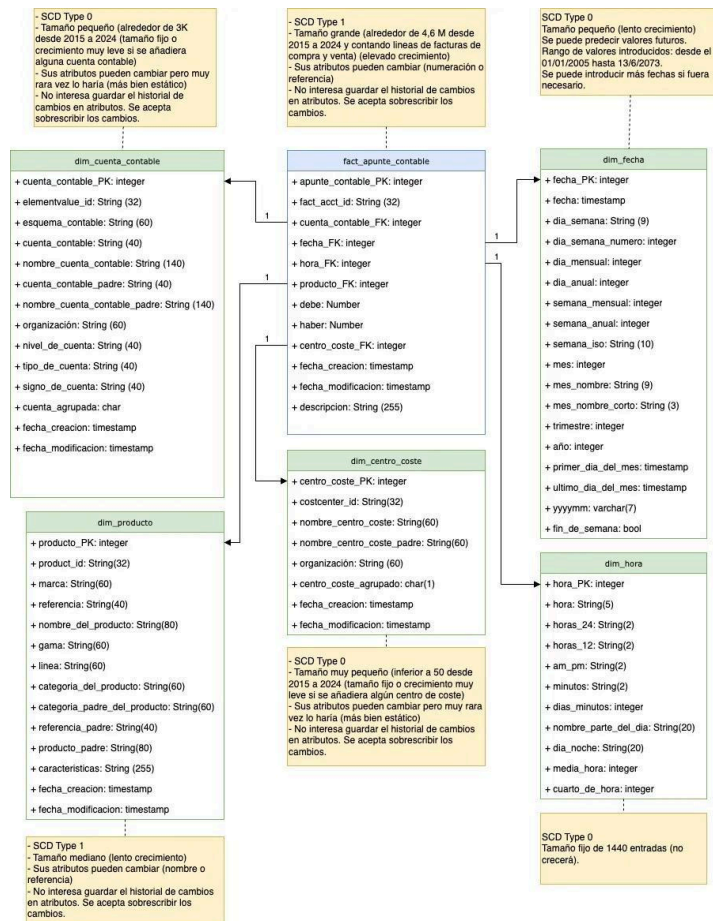


Figura 7. Esquema estrella tabla de hechos fact_apunte_contable

7.2 Implementación física del diseño lógico de la base de datos del Data Warehouse

La traslación del diseño lógico a la base de datos física en PostgreSQL se hace mediante el software *DBeaver* que se establece conexión con el servidor Mac Mini Server donde se crea una base de datos llamada *DataWarehouse_MPX* y los dos esquemas correspondientes a las capas *Staging* y *Core*.

Mediante ficheros script en SQL, se realiza la implementación física de los esquemas vistos en las figuras 5, 6 y 7 dando como resultado lo mostrado en las figuras 8, 9 y 10.

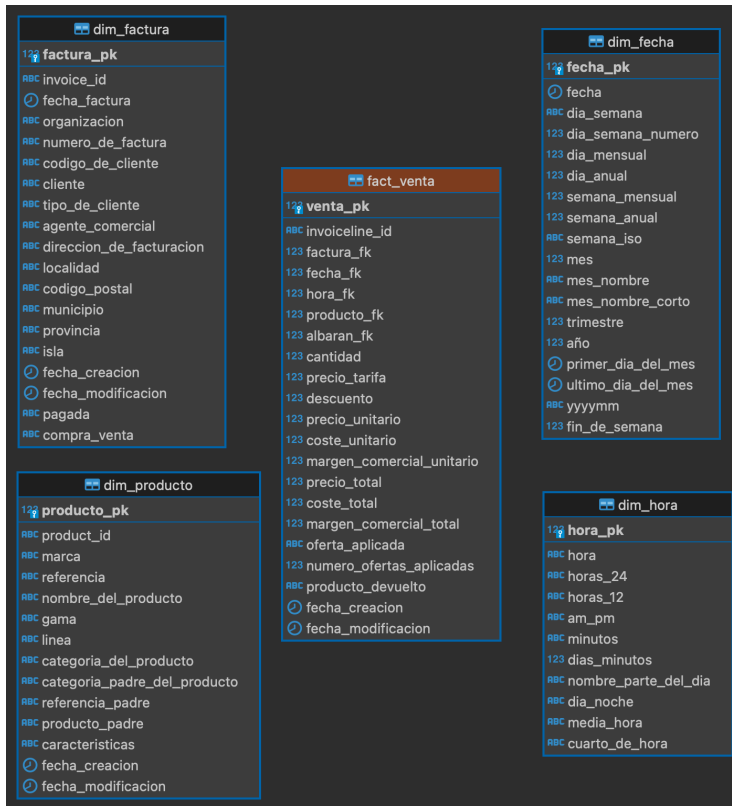


Figura 8. Esquema resultado de la implementación de fact_venta en la capa core

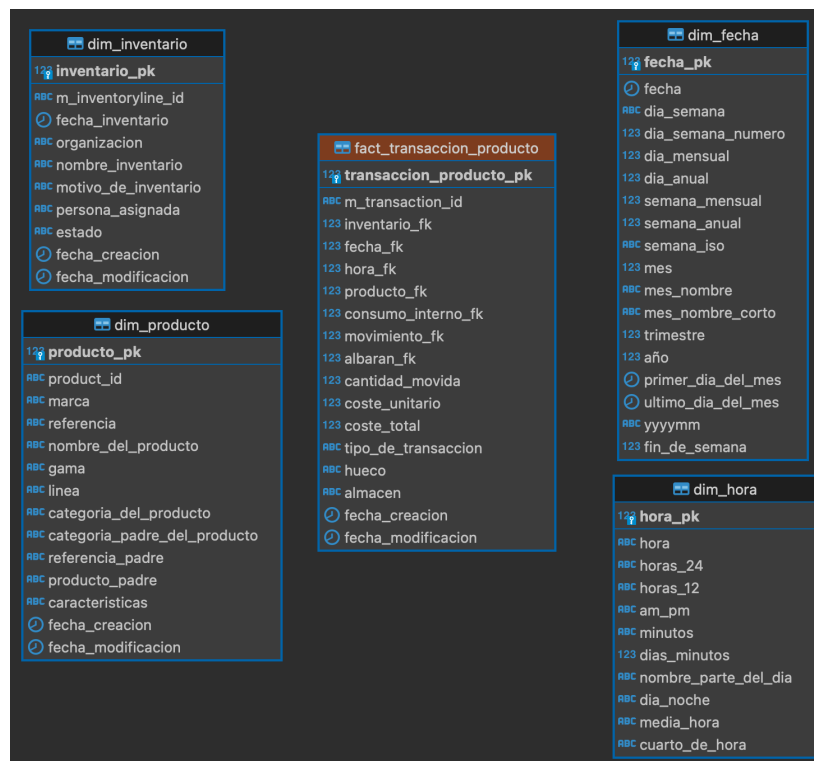


Figura 9. Esquema resultado de la implementación de fact_transaccion_producto en la capa core

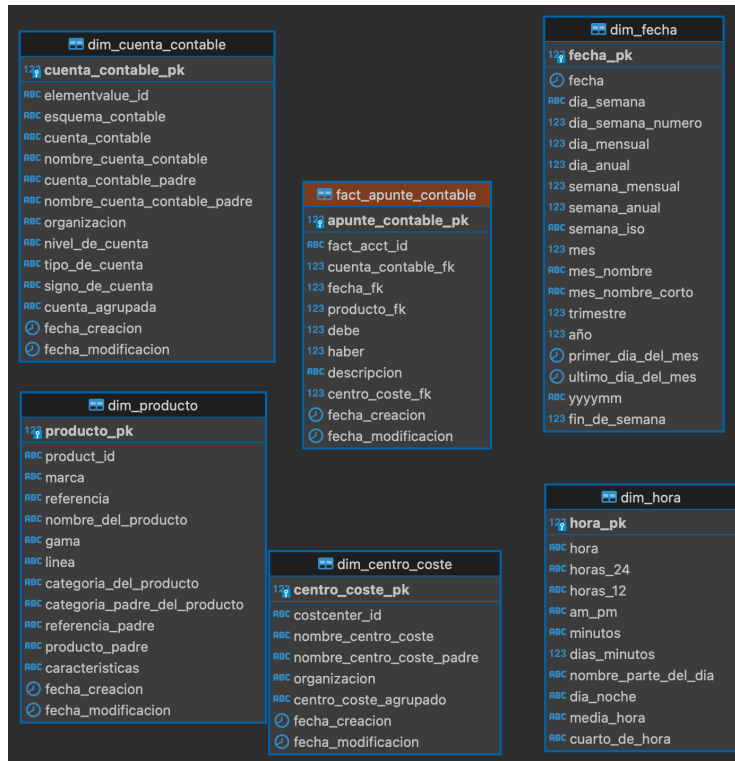


Figura 10. Esquema resultado de la implementación de fact_apunte_contable en la capa core

En este punto, el almacén de datos está preparado para alojar los datos esperados. Para cada una de las tablas de dimensiones se han desarrollado consultas SQL sobre el sistema transaccional con la previsión de insertar los resultados en la capa de *Staging*.

7.3 Análisis sobre la calidad de los datos

Con el objetivo de obtener análisis de información de calidad es indispensable que los datos sean también lo sean. Los sistemas transaccionales registran constantemente datos y en muchos casos interviene la mano humana con lo que no es inusual la aparición de errores de todo clase, especialmente de transcripción. Ello nos obliga a realizar tareas de limpieza de datos antes de almacenarlos definitivamente en un *DataWarehouse*. Cabe recordar la frase *Garbage In, Garbage out* que traducido quiere decir que si entran datos de mala calidad, se obtendrá información de mala calidad.

Según (Schuler, n.d.), el trabajo de científico de datos consume gran parte de su tiempo en efectuar tareas de limpieza. En el caso que ocupa este trabajo no es excepción alguna. A pesar que tras el análisis de métricas de la fuente de datos de origen para el alcance de este proyecto se reduce a 9 dimensiones y 3 tablas de hechos, el cual es un número inferior al inicialmente previsto.

Afortunadamente no se ha detectado un número elevado de atributos con errores. Principalmente los detectado son:

7.3.1 Dimensión factura (dim_factura)

La dimensión factura contiene los registros de las cabeceras de las facturas registradas en el sistema transaccional.

El atributo “agente_comercial” contiene errores de duplicidad de persona a la que se refiere. Es decir, al analizar la información mostrará que hay dos agentes comerciales diferentes, pero que en realidad son la misma persona. La solución ha pasado por agrupar los registros que pertenecen a la misma persona en la fuente de datos de origen. Es relevante corregir este problema puesto que permite mostrar mejor la información.

El atributo “dirección_de_facturación” es con toda probabilidad el que presenta mayores errores de transcripción debido a que los valores que puede tomar el dominio son muy extensos y que han sido fruto de malas transcripciones. Para los indicadores claves de negocio que se han propuesto, no es considerable la corrección aunque sí es recomendable.

La variable dim_factura.codigo_postal de la capa *Staging* sí presenta múltiples errores de transcripción: espacios en blanco o tabuladores, caracteres iniciales erróneos como “+ , - /” longitudes en códigos postales diferentes de 5 cifras, tratándose de canarias es que haya códigos postales que comiencen diferentes de 35 o 38, letras incluidas como códigos postales). La solución consiste en eliminar caracteres diferentes del 0-9 durante el proceso de transformación. No obstante, tampoco es una variable que se necesite para los KPI analizados.

El atributo dim_factura.municipio en la capa *Staging* contiene valores duplicados por espacios al final, textos en mayúsculas y minúsculas, comas incluidas, transcripciones erróneas... Nuevamente, no se trata de un atributo importante para los KPI propuestos.

8. Diseño de procesos ETL

Los procesos de extracción, transformación y carga consiste en diseñar los pasos que seguirá la aplicación *Pentaho Data Integration*. Se dividen principalmente en las siguientes tareas:

1. Asignar la última carga realizada para cada tabla de dimensión y de hecho. Para ello, el proceso comprobará siempre la última fecha de modificación en la capa *core* de las mencionadas tablas.
2. Extracción de los datos de la fuente de origen hacia la capa *Staging* en la que se obtienen los datos tal y cómo se encuentran en origen.
3. Posibles transformaciones sobre los datos en origen que pueden consistir en correcciones de errores cometidos en un sistema transaccional o cualquier otra índole.
4. Carga de los datos transformados hacia la capa final *core* del *DataWarehouse*.

Según qué tabla de origen, se puede dar el caso que un usuario sobre el sistema transaccional borre (consciente o inconscientemente) un registro que haya sido ya cargado en el almacén de datos o que también su modificación implique un cambio en la clave primaria del registro. Esta casuística se ha de tener en cuenta porque de lo contrario ocurrirá que se obtenga información errónea en el *DataWarehouse*. Un caso habitual es el de los registros de apuntes contables siempre y cuando el año fiscal siga en curso y no se haya procedido a su cierre.

En el caso expuesto, los procesos diseñados son muy similares entre sí y que consisten en las tareas enumeradas previamente. Sí se ha tenido en cuenta de separar los procesos de las tablas de dimensiones de las tablas de hechos, es decir, se ejecutarán siempre primero las tareas correspondientes a las tablas de dimensiones antes de las tablas de hechos. La razón de por ello es que las tablas de hechos pueden contener relaciones a otras tablas con datos que aún no han sido creados en el almacén de datos. Por tanto, se corre el riesgo de introducir líneas en las tablas de hechos que carezcan de los datos de unión. Un claro ejemplo es que la tabla de hechos de líneas facturadas contenga algún producto que no esté disponible en la tabla de dimensiones de productos. De esta forma, se evita la inserción de nuevas filas en las tablas de hechos que tengan atributos de claves foráneas con valores nulos o *dummy*.

Por ello, para los presentes y futuros procesos ETL que se implementen, será necesario que se introduzcan y actualicen previamente las dimensiones antes que los hechos.

8.1 Cargas iniciales e incrementales

En el diseño de los procesos ETL se ha de tener en cuenta la diferencia entre una carga inicial y una carga incremental. Como deja entrever sus nombres, la carga inicial supone la primera extracción y almacenamiento de datos en la capa *core* para cada tabla. Como es de esperar, con el paso del tiempo es deseable actualizar las tablas en el almacén de datos, ya sea por alteraciones en la fuente de origen como por la adición de nuevos hechos o dimensiones.

Para ello, se ha tenido en cuenta que todas las tablas incluyen un atributo habitualmente conocido como atributo delta que controla qué filas son

susceptibles de añadir o modificar. Dicho atributo suele corresponder a una variable de tipo *timestamp* o bien una clave primaria. En este proyecto se ha utilizado una variable de tipo *timestamp* que coincide con la fecha de modificación del sistema de transacción de origen. Se ha escogido la fecha de modificación porque se ha tenido en consideración el modelo de datos del sistema transaccional. De este modo, si hubiere una modificación, el ERP guarda la fecha de modificación y el proceso ETL tiene en cuenta que habrá una fecha de modificación más reciente a su fecha máxima de modificación en cualquier tabla. Si la clave primaria no está presente en el almacén de datos, se crea un nuevo registro en la tabla correspondiente, en caso de existir, actualiza los atributos del registro.

8.2 Implementación procesos ETL

La implementación de procesos ETL se ha realizado con la mencionada herramienta *Pentaho Data Integration*. Con ella se ha creado un repositorio llamado *ETL_DataWarehouse_MPX* que contiene carpetas con los archivos de las transformaciones y trabajos correspondientes a cada tabla de dimensiones y hechos. En la figura siguiente se puede apreciar cómo se ha organizado:

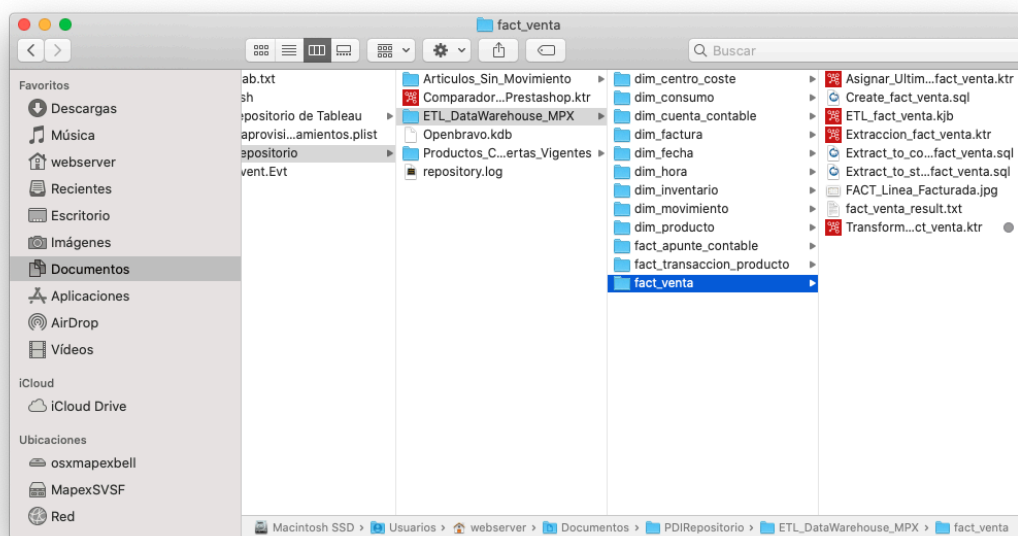


Figura 11. Repositorio procesos ETL

A modo de ejemplo, cada carpeta contiene generalmente tres transformaciones:

- Detectar la fecha de última carga,
- Extracción de los datos de la fuente de origen
- Transformaciones de la capa *Staging* hacia la capa *core*.

En el caso de la tabla de hechos de apuntes contables se incluye una transformación adicional para detectar qué apuntes contables han sido cargados previamente y que hayan sido borrados en la fuente de origen para proceder a eliminar del *Data Warehouse*. Se incluyó porque durante el proceso de pruebas, se detectó esta casuística y no se ha detectado en ningún otro caso.

Las transformaciones fueron incluidas posteriormente en trabajos, uno por tabla creada en el almacén de datos. *Pentaho Data Integration* distingue entre transformaciones (*transformations*) y trabajos (*jobs*) siendo estos últimos los que aglutinan las transformaciones y también da la posibilidad de crear trabajos que albergan trabajos de manera anidada. Además, sólo permite la programación de trabajos. Por tanto, es indispensable la creación de los trabajos para la planificación temporal.

Las siguientes figuras muestran algunos ejemplos de transformaciones y trabajos:

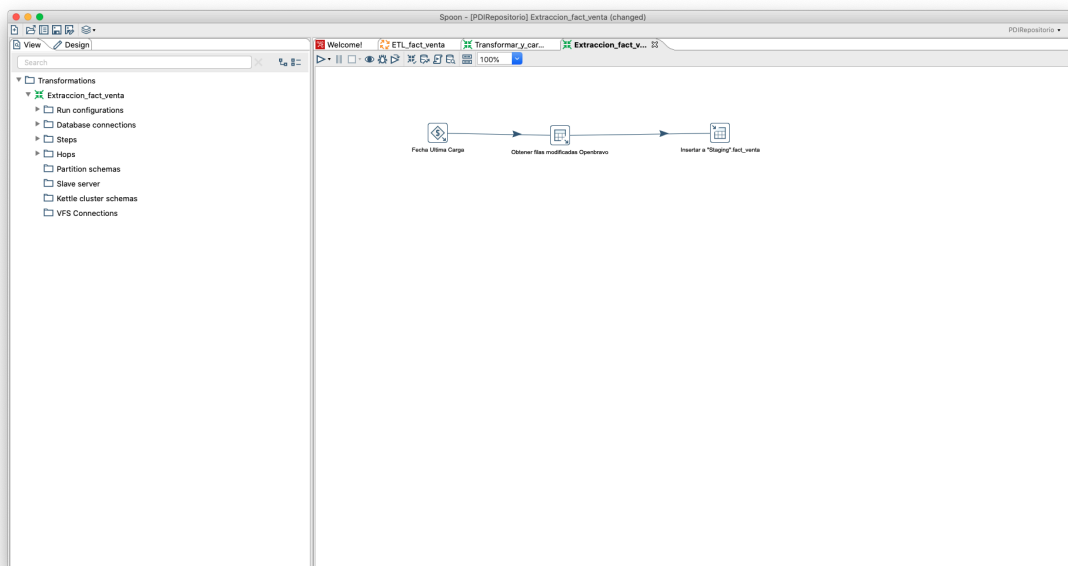


Figura 12. Trabajo extracción para la tabla de hechos de líneas facturadas.

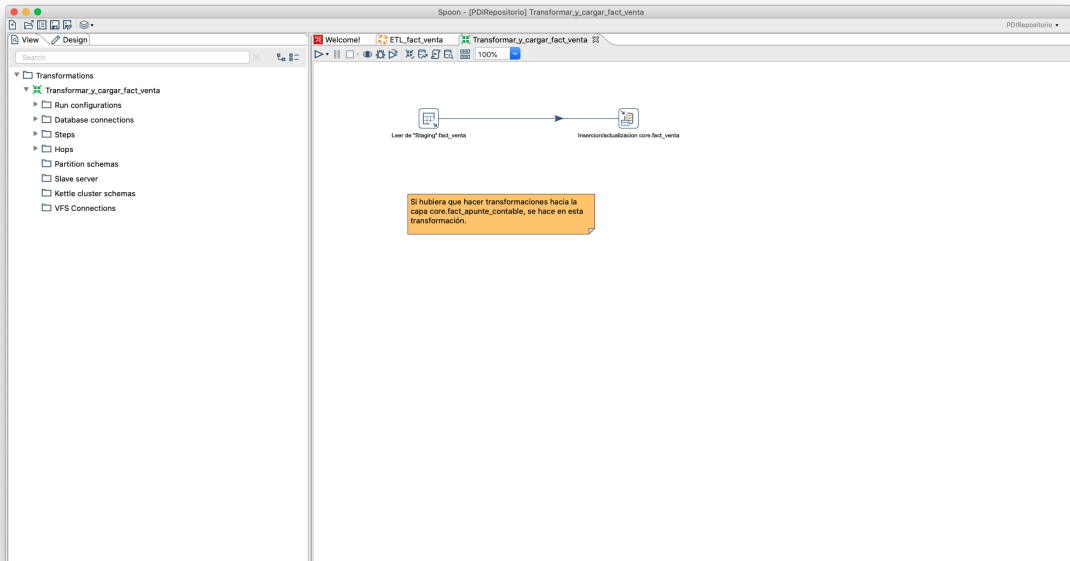


Figura 13. Trabajo de lectura e inserción de filas en la capa core de líneas facturadas

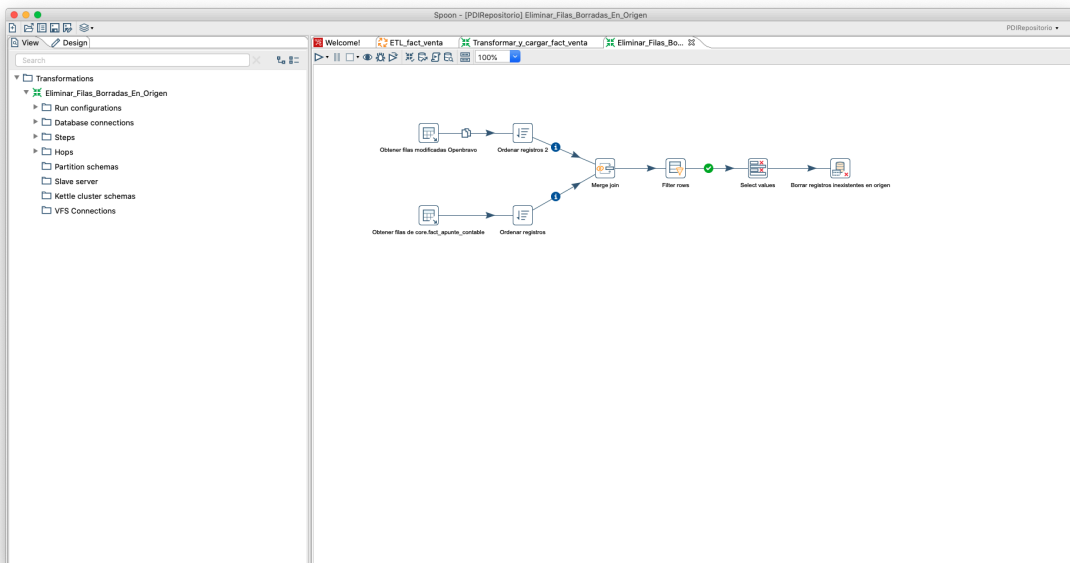


Figura 14. Transformación para eliminar filas borradas en la tabla de apuntes contables

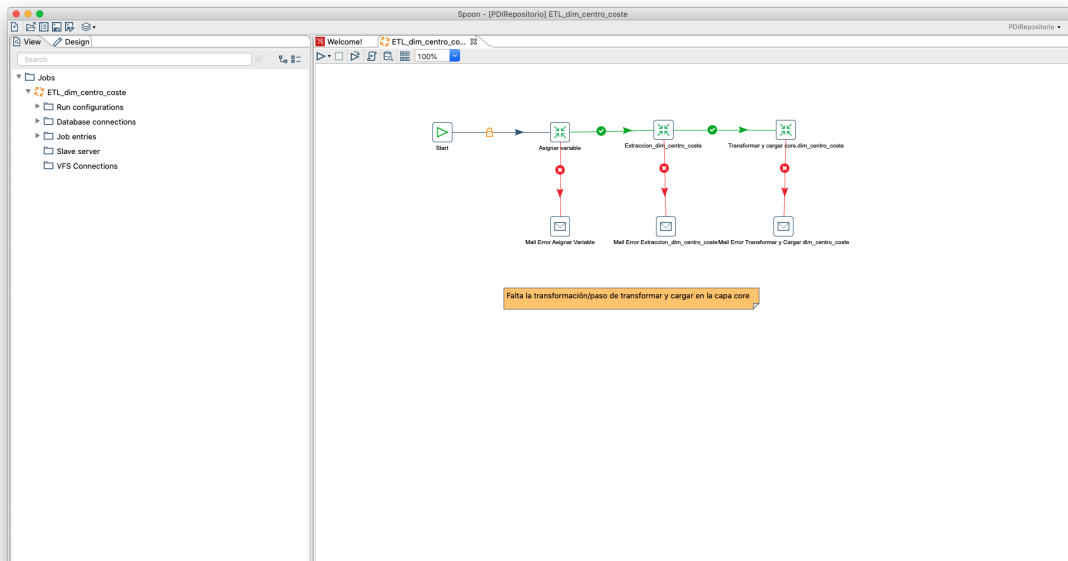


Figura 15. Trabajo para el proceso de la tabla de dimensión de centros de coste

Cabe recordar que en el momento del análisis de la herramienta *Pentaho Data Integration* se mencionó la imposibilidad de realizar la programación de ejecución de trabajos a través de su interfaz gráfica en su edición comunitaria. La alternativa es utilizar el administrador regular de procesos presente en todo sistema basado en Unix llamado *cron*. En la figura siguiente, se aprecia la planificación en el fichero especial *crontab* realizada para este proyecto:

```

#Trabajos ETL de dimensiones de DataWarehouse_MPX
30 22 * * * /Applications/data-integration/kitchen.sh -rep:PDIREpositorio -dir:/ETL_DataWarehouse_MPX/dim_centro_coste
-job:ETL_dim_centro_coste &> /Users/webserver/Documents/PDIREpositorio/ETL_DataWarehouse_MPX/dim_centro_coste/
dim_centro_coste_result.txt
31 22 * * * /Applications/data-integration/kitchen.sh -rep:PDIREpositorio -dir:/ETL_DataWarehouse_MPX/dim_consumo -job:ETL_dim_consumo
&> /Users/webserver/Documents/PDIREpositorio/ETL_DataWarehouse_MPX/dim_consumo/dim_consumo_result.txt
32 22 * * * /Applications/data-integration/kitchen.sh -rep:PDIREpositorio -dir:/ETL_DataWarehouse_MPX/dim_cuenta_contable
-job:ETL_dim_cuenta_contable &> /Users/webserver/Documents/PDIREpositorio/ETL_DataWarehouse_MPX/dim_cuenta_contable/
dim_cuenta_contable_result.txt
35 22 * * * /Applications/data-integration/kitchen.sh -rep:PDIREpositorio -dir:/ETL_DataWarehouse_MPX/dim_factura -job:ETL_dim_factura
&> /Users/webserver/Documents/PDIREpositorio/ETL_DataWarehouse_MPX/dim_factura/dim_factura_result.txt
40 22 * * * /Applications/data-integration/kitchen.sh -rep:PDIREpositorio -dir:/ETL_DataWarehouse_MPX/dim_inventario
-job:ETL_dim_inventario &> /Users/webserver/Documents/PDIREpositorio/ETL_DataWarehouse_MPX/dim_inventario/dim_inventario_result.txt
45 22 * * * /Applications/data-integration/kitchen.sh -rep:PDIREpositorio -dir:/ETL_DataWarehouse_MPX/dim_movimiento
-job:ETL_dim_movimiento &> /Users/webserver/Documents/PDIREpositorio/ETL_DataWarehouse_MPX/dim_movimiento/dim_movimiento_result.txt
50 22 * * * /Applications/data-integration/kitchen.sh -rep:PDIREpositorio -dir:/ETL_DataWarehouse_MPX/dim_producto
-job:ETL_dim_producto &> /Users/webserver/Documents/PDIREpositorio/ETL_DataWarehouse_MPX/dim_producto/dim_producto_result.txt

#Trabajos ETL de hechos de DataWarehouse_MPX
53 22 * * * /Applications/data-integration/kitchen.sh -rep:PDIREpositorio -dir:/ETL_DataWarehouse_MPX/fact_apunte_contable
-job:ETL_fact_apunte_contable &> /Users/webserver/Documents/PDIREpositorio/ETL_DataWarehouse_MPX/fact_apunte_contable/
fact_apunte_contable_result.txt
55 22 * * * /Applications/data-integration/kitchen.sh -rep:PDIREpositorio -dir:/ETL_DataWarehouse_MPX/fact_transaccion_producto
-job:ETL_fact_transaccion_producto &> /Users/webserver/Documents/PDIREpositorio/ETL_DataWarehouse_MPX/fact_transaccion_producto/
fact_transaccion_producto_result.txt
0 23 * * * /Applications/data-integration/kitchen.sh -rep:PDIREpositorio -dir:/ETL_DataWarehouse_MPX/fact_venta -job:ETL_fact_venta
&> /Users/webserver/Documents/PDIREpositorio/ETL_DataWarehouse_MPX/fact_venta/fact_venta_result.txt

```

Figura 16. Extracto del contenido de crontab del servidor Mac Mini Server osxmapexbell

8.3 Pruebas de procesos ETL

Una vez implementadas las transformaciones y los trabajos, *Pentaho Data Integration* sí permite la ejecución individual de cada transformación en su

interfaz gráfica. También es posible la ejecución de los trabajos a través de la ejecución del comando correspondiente en la terminal del sistema.

En el proyecto se realizaron pruebas por pasos de cada transformación para la verificación de su correcta ejecución. Más allá de pequeños problemas de configuración de los pasos de la herramienta *Pentaho Data Integration*, las pruebas fueron mayormente exitosas. Los errores en las pruebas provenían principalmente de errores en sentencias SQL o ajustes en pasos dentro de una transformación.

La mayor parte de las transformaciones no supuso un coste elevado de tiempo. No obstante, las tablas de hechos de apuntes contables y de transacciones de productos sí llegaron a consumir tiempos elevados en sus cargas iniciales ya que sendas tablas contienen alrededor de los 15 millones de registros. Esto se tradujo en unas 5 horas para cada tabla.

9. Cuadro de mandos

9.1 Diseño conceptual de cuadro de mandos

Un cuadro de mandos supone una visualización general de la que se desea obtener cierta información rápidamente. Diseñar cuadros de mandos eficaces supone en cierto modo un reto en arte visual en el que entran las Leyes de la Gestalt, la composición e incluso la imaginación para transmitir información y que puede extenderse más allá de la finalidad de este Trabajo de Fin de Grado. Se ha de tener en cuenta una arquitectura de información en el que se agrupen indicadores relacionados, ordenados de mayor relevancia a menor, si es intuitivo para los usuarios, si éstos presentan problemas de daltonismo... Para que sea eficiente, el usuario ha de ser capaz de interpretar los datos en cuestión de segundos.

Por ello, en cuestión de diseño de cuadro de mandos, se ha optado por la inspiración dada por un ejemplo de muestra de la aplicación *Tableau Software* basado en un cuadro de mandos de retail, pero con la excepción de haber sustituido la información del ejemplo por los indicadores clave de negocio propuestos en este proyecto que el margen de tiempo del proyecto me ha permitido.

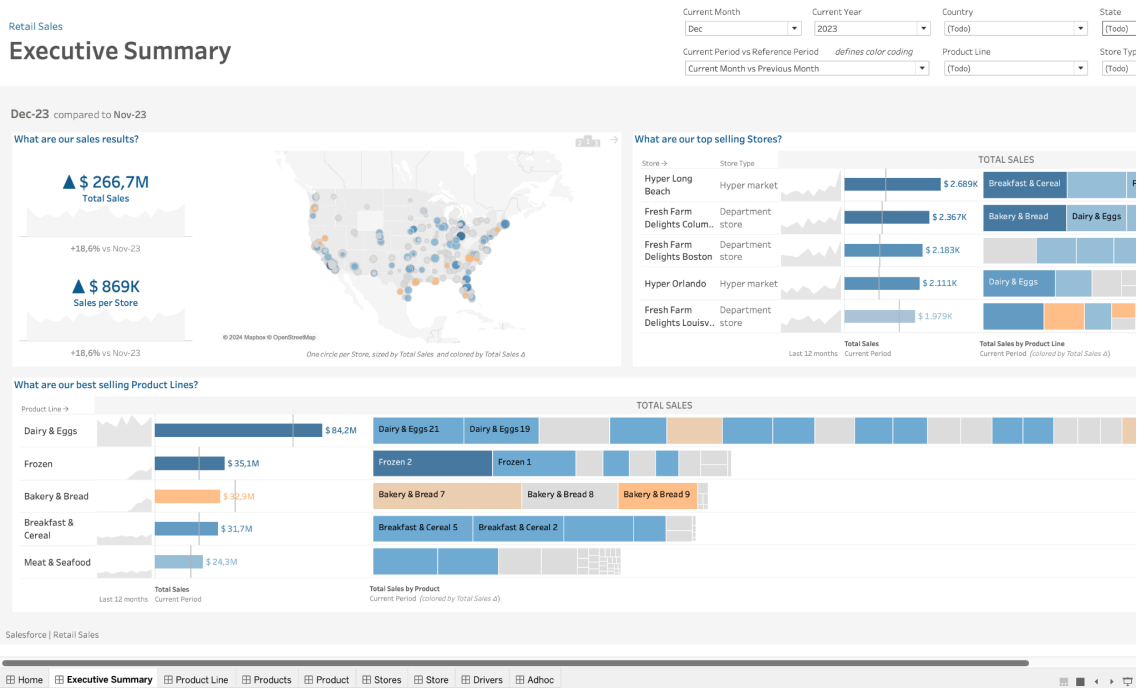


Figura 17. Cuadro de mandos inspiracional "Resumen Ejecutivo" de Tableau

9.2 Implementación de los cuadros de mandos de mandos

La implementación del prototipo del cuadro de mandos se presenta a continuación. Se trata de un cuadro de mandos que contiene un resumen ejecutivo de algunos de los KPI's que se propusieron en el proyecto.

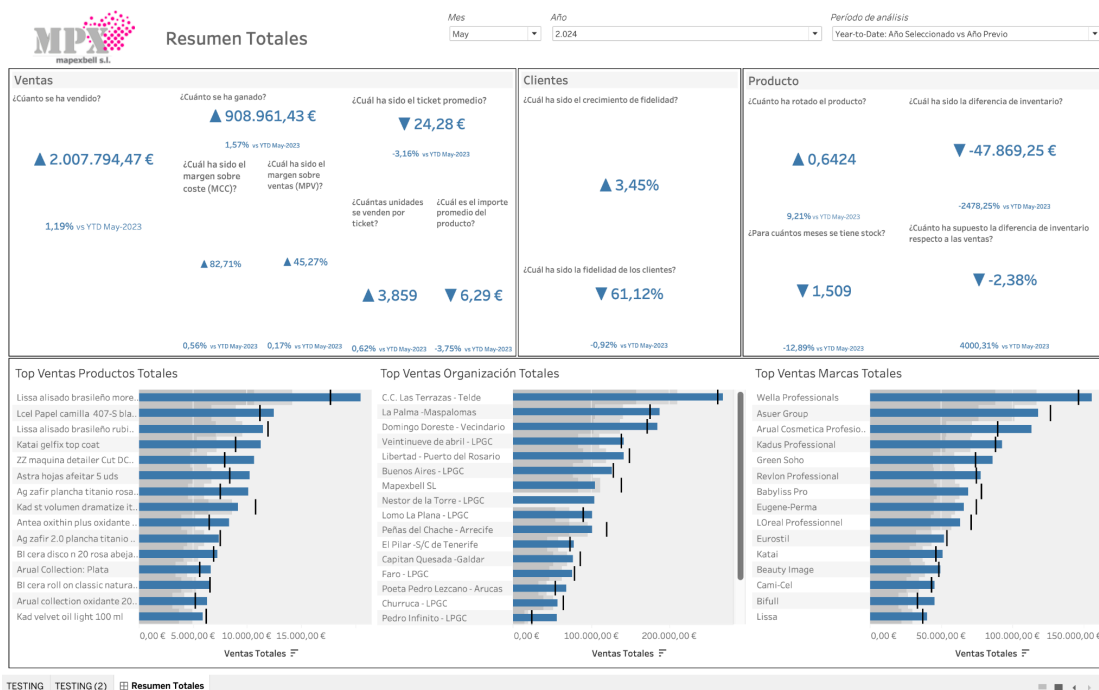


Figura 18. Prototipo de cuadro de mandos para retail

La figura 18 representa el cuadro de mandos implementado, si nos fijamos en la fila superior del cuadro el usuario tiene la potestad de seleccionar varios parámetros:

- Mes de análisis
- Año de análisis
- Periodo de análisis contra el periodo de referencia: Mes seleccionado vs mes previo, mes seleccionado vs mismo mes del año previo, total acumulado hasta mes seleccionado vs total acumulado del mismo mes del año previo.

Se han agrupado los KPI' s en base al área funcional: ventas, clientes y producto. Para ello, se ha sectorizado en recuadros en cuyo título se ha nombrado el área y la disposición se ha ordenado de mayor importancia a menor siguiendo el orden natural de la lectura: de arriba a abajo y de izquierda a derecha. Así, se considera que el recuadro de ventas tiene mayor relevancia que el de clientes y que el de producto. La fila inferior se ha dedicado a mostrar un top 15 de ventas por dimensiones: producto, organización/tienda y marca del producto. Cada ranking se ha considerado por volumen en ventas generadas en el período seleccionado.

Al entrar en detalle en cada recuadro, se incluyen diferentes indicadores claves de negocio. Con la finalidad de facilitar la lectura al usuario, cada KPI tiene un pequeño título formulado a modo de pregunta y cuyo valor la responde. Se ha seguido esta inspiración del cuadro de mandos original porque se comparte la opinión de que para el usuario final del cuadro de mandos le será más sencillo de interpretar el valor si lee la pregunta que si se pusiera como título el nombre del indicador clave, es decir, si un usuario no está familiarizado con los nombre de los KPI entonces el valor UPT puede que no lo asimile tan fácilmente como si leyera la pregunta “¿cuántas unidades se venden por ticket?” o “¿Cuánto se ha vendido?” en lugar de Cifra de negocio.

Si seguimos examinando un recuadro, se aprecia que todos tienen un triángulo hacia arriba o hacia abajo a mano izquierda del valor del KPI. Se utiliza el concepto de figura porque se cree que de esta manera, el usuario final verá de inmediato si el valor obtenido ha sido superior o inferior (cabe la rara posibilidad de que fuese igual, en tal caso el símbolo sería un triángulo apuntando hacia la derecha) respecto al periodo previo. De lo contrario, es altamente probable que el usuario, al ver el valor obtenido, su siguiente cuestión inmediata sea si el valor ha sido superior o inferior.

El empleo de una fuente superior para el valor de cada KPI no es casual. Se utiliza para que destaque el valor por encima de cualquier otra información que

contenga el recuadro del KPI. En cambio, en la parte inferior de cada recuadro se ha incluido en una fuente menor cuál ha sido la variación del valor respecto al periodo de referencia debido a que sería la siguiente pregunta natural que se preguntaría un usuario después de preguntarse: ¿Cuál ha sido el valor del indicador? ¿Ha sido superior o inferior respecto al periodo de referencia? Por tanto, en un solo recuadro el usuario sabrá el valor, la variación y cuánto ha variado respecto al periodo de referencia.

Al entrar en detalle con la fila inferior en el que se encuentran los ranking, en el que se ha decidido emplear gráficas de barras que miden el valor del periodo analizado y una marca que indica el valor del periodo de referencia. Las barras en este caso particular resultan más visuales en los siguientes aspectos: el usuario es capaz de hacer una comparación rápida en cuál ha vendido más, cuánto más ha vendido una posición respecto a la posición inferior, y cuánto ha variado respecto al periodo de referencia. Una opción que es posible que un usuario tendiera a querer es la de mostrar el valor de cada marca, el cuál convertiría cada gráfica ilegible debido a la saturación de información. Para remediar esta situación, *Tableau Software* ofrece la posibilidad de incluir un menú emergente con el que el usuario pueda entrar en detalle de una marca si realiza una acción como seleccionar el valor o pasar el ratón por encima. Evidentemente esta posibilidad sólo existe mientras no se emplee el soporte impreso.

En la figura 19 se puede ver que el menú emergente incluye el periodo de análisis y su valor, el periodo de referencia y su valor, y la variación porcentual de ambos periodos. Así, si el usuario quiere entrar en más detalle, puede consultar sin aglomerar el cuadro de mandos ya que esta información es secundaria.

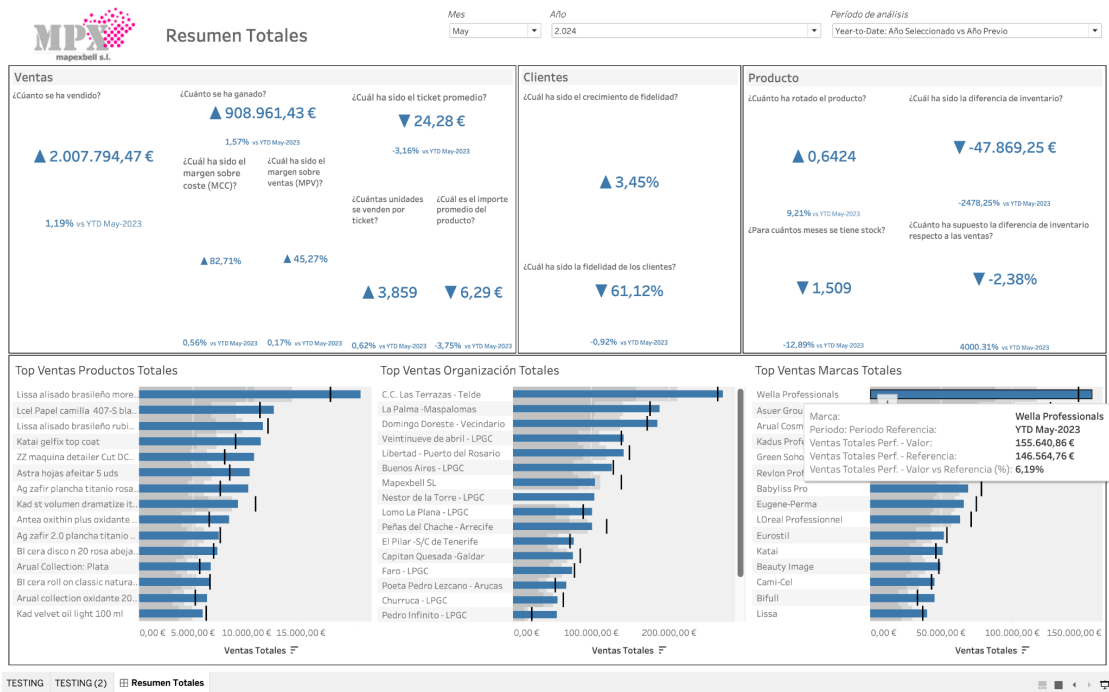


Figura 19. Ventana emergente con detalles de la marca seleccionada.

El cuadro de mandos obtenido pretende ser un primer acercamiento a la empresa objeto de estudio que pretenda valorar toda la información que transmite rápidamente. Los KPI que se muestran se han considerado como los primordiales que más pueden interesar a la gerencia de la empresa objeto de estudio además de ser los más plausibles a obtener tras el análisis del modelo de datos de la fuente de origen.

10. Conclusiones y trabajos futuros

10.1 Conclusiones generales

La selección de la temática para el Trabajo de Fin de Grado fue una decisión personal importante ya que la otra inquietud que tenía era escoger el área temática de desarrollo de videojuegos. Sin embargo, la selección por la temática de *Business Intelligence* primó a causa de las ganas de aprender a confeccionar y sintetizar información en la situación laboral personal, ampliar la trayectoria profesional y las ganas de superación.

En líneas generales, el resultado creo que ha sido satisfactorio porque en un pequeño cuadro de mandos se logra sintetizar información que en otros casos habría supuesto la confección de diferentes hojas de cálculo y que además permite a un usuario final interactuar con el cuadro con lo que adquiere la información deseada reduciendo la dependencia de un tercero para obtener la misma información pero con diferentes valores de parámetros.

Según el prisma en que se contemple, por una parte los resultados alcanzados han sido superiores a los esperados gracias a la inspiración de un cuadro de mandos ejemplo de *Tableau Software*. Por otro lado, el resultado obtenido ha sido inferior porque se esperaba haber elaborado más indicadores claves de negocio en el mismo cuadro de mandos. Sin embargo, se llegó a la conclusión que quizás se pretendió englobar demasiados en un espacio reducido y por la dificultad que entrañó valorar ciertos indicadores.

10.2 Conclusiones de objetivos

Respecto a la pregunta de si se han alcanzado todos los objetivos propuestos, por la parte de obtención de productos es que sí ya que se ha creado el sistema de información de *Business Intelligence* con los componentes que lo conforman: el almacén de datos ha sido definido y creado, se ha creado unos procesos ETL en un repositorio de *Pentaho Data Integration* y se ha creado un primer cuadro de mandos con la herramienta de visualización de datos *Tableau Software*.

Se planteó también el objetivo de que el *Data Warehouse* se actualice periódicamente mediante la herramienta *Pentaho Data Integration* y se ha logrado. Acto seguido, se previó el uso de *Tableau Software* y también se ha construido el cuadro de mandos.

En dónde sólo se ha logrado formalizar parcialmente el objetivo es en el número de indicadores claves de negocio a implementar. Se definieron los KPI relevantes y se quería haber añadido más indicadores claves de negocio que por razones de dificultad me he visto obligado a renunciar.

En el proyecto se ha definido un total de 29 KPI de los cuales algunos de ellos requieren cálculos de mayor complejidad para hallar valores de variables. Finalmente, se han mostrado en el cuadro de mandos un total de siete KPI, variables que se han considerado relevantes mostrar cómo son las unidades por transacción y el precio medio de producto, y tres rankings que también se valoraron de importancia para la gerencia de la empresa.

Las mayores dificultades halladas en el proyecto han sido el diseño lógico del almacén de datos, la curva de aprendizaje con *Tableau Software* y tratar de optimizar la base de datos PostgreSQL

10.3 Planificación y metodología

El uso de la metodología del *PMBOK* y la confección del diagrama de Gantt ha simplificado enormemente la consecución de los objetivos planteados. Ciertamente es que la inexperiencia en utilizar esta metodología supuso un reto al principio, pero según fue avanzando el proyecto, me facilitó marcar día a día lo que había que tener finalizado, qué se tenía que estar ejecutando y qué vendría después.

El principal reto fue prever el tiempo de dedicación de cada tarea/actividad y de todas las tareas definidas únicamente se ha echado en falta la ejecución de una de ellas que fue la de realizar pruebas iterativas con usuarios del cuadro de mandos para que recibieran *feedback* de estos y ver posibilidades de mejora. Estas pruebas se harán más allá de la finalización del proyecto.

En líneas generales, no hubo desviaciones relevantes en el cumplimiento de lo planificado a excepción de la salvedad mencionada anteriormente y que se tomó la decisión de realizar las pruebas iterativas después de finalizar el Trabajo de Fin de Grado.

10.4 Previsiones futuras

La previsión futura del proyecto es amplia porque hay numerosas ideas que no se han llegado a desarrollar. Existe una serie de aspectos que se pretende cumplir para que el proyecto se expanda.

En primer lugar, se prevé calcular los KPI que han quedado fuera del alcance. Estos son numerosos y algunos de ellos se pretendió ser añadidos dentro del alcance del proyecto como son aquellos relacionados con la tabla de hechos de apuntes contables.

En segundo lugar, se quiere mejorar el rendimiento del *Data Warehouse* porque el cálculo de los indicadores claves de negocio que se han implementado para ser mostrados en el cuadro de mandos conlleva una demora entre 2 y 3 minutos, lo cual puede suponer un problema de mala experiencia de usuario.

En tercer lugar, se desea que el cuadro de mandos pueda ser filtrado por dimensiones como la organización/tienda porque el implementado sólo muestra los indicadores a nivel global.

En cuarto lugar, se desea también perforar sobre el cuadro de mandos elaborado para que el usuario pueda saber con más detalle sobre el valor de algunos KPI. Por ejemplo, si el usuario desea más información sobre el KPI ticket medio, tenga la posibilidad de perforar sobre este KPI y mostrar este valor de cada agente comercial.

En último lugar, se desea incluir para los KPI ya desarrollados (y futuros), las tendencias. Actualmente, sólo se muestra el valor del periodo de análisis y cuánto ha variado respecto el periodo de referencia, pero también es interesante ver gráficamente cuál ha sido la evolución del valor.

Para la consecución de estos logros, se prevé realizar una nueva planificación con nuevos plazos.

Estas son algunas de las previsiones futuras para este proyecto. Desde su concepción, se planteó el Trabajo de Fin de Grado que fuera el comienzo de la implementación de un sistema de información de *Business Intelligence* en una organización real de comercio minorista y se pueda sacar provecho desde la finalización del mismo.

11. Glosario

ACID: Siglas del inglés *Atomic, Consistency, Isolation and Durability*. Propiedades de los sistemas de gestión de bases de datos.

Atributo delta: Columna en una tabla de datos que se emplea para controlar cargas incrementales o modificaciones.

Business Intelligence: También conocido por sus siglas en inglés BI e inteligencia de negocio en castellano. Es un sistema de información empleado para el análisis de datos y la toma de decisiones. Habitualmente se nutre de un *Data Warehouse* donde los datos han sido tratados previamente.

Core layer: Capa en un almacén de datos transformados y destinado a almacenarlos definitivamente

Coste de oportunidad: Segundo mejor valor disponible al que se renuncia por tomar la decisión a otra opción.

Coste directo: Costes que se pueden relacionar directamente con el objeto del coste final porque se puede conocer una relación cualitativa, unidad de medida exacta entre coste y unidad objeto del coste

Coste indirecto: Costes que no guardan una relación directa con el objeto del coste final y del que se desconoce unidad de medida exacta entre coste y unidad objeto del coste.

Cron: Daemon que despierta periódicamente para ejecutar peticiones requeridas.

Crontab: Fichero que contiene entradas para el programa cron para ser ejecutadas en los horarios/tiempos especificados.

Daemon: Programa informático que se ejecuta en segundo plano en lugar de ser controlado directamente por un usuario.

Data Mart: Subconjunto de tablas o registros de un almacén de datos enfocados a resolver dudas de unidades de negocio como por ejemplo un departamento de almacén, financiero-contable...

Data Warehouse: Base de datos utilizada como almacén que ha sido optimizada para propósitos de análisis.

EBITDA: Parámetro de análisis financiero que son los beneficios previos a los intereses, impuestos, depreciación y amortización. Las siglas provienen del inglés *Earnings Before Interest, Taxes, Depreciation and Amortization*.

ELT: Siglas del inglés *Extract, Transform and Load* que es un proceso similar a ETL, pero enfocado a otras necesidades como *Machine Learning, Big Data* o requerimientos de información a tiempo real.

ETL: Siglas del inglés *Extract, Transform and Load* que es un proceso por el cual se extraen datos de una o más bases de datos origen, se realizan tareas de transformación como son la limpieza de datos, cálculo de nuevos atributos, renombrar... y posterior carga en una base de datos.

GUI: Del inglés *Graphical User Interface*, es la interfaz gráfica del usuario.

KPI: Del inglés *Key Performance Indicator*. Métrica clave para evaluar el desempeño de un proceso u objeto.

Leyes de la Gestalt: Normas que explican el origen de las percepciones a partir de los estímulos por el que las personas asimilan mejor la información y que se rigen en dos reglas generales: la ley de buena forma y la ley de figura y fondo

OLAP: Siglas del inglés *On-Line Analytical Purposes*. Cubos multidimensionales con el objetivo de analizar eficientemente grandes cantidades de datos.

OLTP: Siglas del inglés *On-Line Transaction Processing*. Tipo de procesamiento de datos diseñado para facilitar el registro, recuperación y procesamiento sobre sistemas de información transaccionales.

On premise: *Software* o *hardware* instalado en una ubicación física propiedad de una persona o empresa.

Open source: Modelo de desarrollo de *software* de manera que el código fuente sea accesible para todos y fomente la colaboración/participación

Payback: Número de años que se necesitan para recuperar, a través del flujo de caja, la inversión inicial.

PYME: Acrónimo de Pequeña y Mediana Empresa

RDBMS: Siglas del inglés *Relational Database Management System*. Programa informático de gestión de bases de datos de tipo relacional.

RDMS: Ver RDBMS

Retail: Comercio minorista en inglés.

Rotura de stock: Situación en la que una empresa no puede atender la demanda de sus clientes de uno o varios productos.

SQL: Siglas del inglés *Structured Query Language*. Es un lenguaje declarativo estándar para la administración y obtención de datos en bases de datos relacionales.

Staging layer: Capa en un almacén de datos destinado a albergar temporalmente los datos más recientes generados por un sistema transaccional con la finalidad de no sobrecargar el sistema transaccional y someterlos a un proceso de transformación.

Stakeholder: Parte interesada o involucrada en alguna empresa o proyecto.

TIR: Acrónimo de tasa interna de retorno, es la tasa mínima de rentabilidad que se exige a una inversión/proyecto que permite que un inversionista se decante frente a otra inversión/proyecto.

VAN: Acrónimo de valor actual neto, es el valor presente de los flujos de caja netos restados de la inversión inicial de un proyecto.

Vendor lock-in: Situación en la que el cambio de solución informática de un proveedor a otro es muy elevado y evita que finalmente el cliente pueda migrar a otra solución.

12. Bibliografía

Álvarez Orozco, Marcos. 2015. *Cuadro de mando retail: Los indicadores clave (KPI) de los comercios altamente efectivos*. Barcelona, España: Profit Editorial.

Anscombe, Frank J. 1973. "Graphs in Statistical Analysis." *The American Statistician* 27, no. 1 (febrero): 17-21.
<https://www.sjsu.edu/faculty/gerstman/StatPrimer/anscombe1973.pdf>.

Branson, Richard C. n.d. "Los clientes no son lo primero, lo primero son los empleados. Si cuidas de tus empleados, ellos cuidarán de tus clientes".

Ford, Henry. n.d. "Si le hubiera preguntado a la gente qué querían, habrían dicho 'caballos más rápidos'".

"GMROI, Margen Bruto del Retorno de la Inversión en Inventario." 2009.
Gravitar.
<https://gravitar.biz/tecnologia-negocios/gmroi-margen-bruto-retorno-inversion-inventario/>.

Grosclaude Manera, Luis. 2006-2007. "Tema 2 Clasificación y conceptos de costes," Apuntes. In *Contabilidad de Gestión*. 2º Cuatrimestre ed. N.p.: Universidad Carlos III de Madrid.

Ingeniariak Colegio Oficial de Ingenieros Industriales de Guipúzcoa. 2022.
"¿Existe una brecha digital entre grandes y pequeñas empresas?"
¿Existe una brecha digital entre grandes y pequeñas empresas? –
Ingeniariak – OAP.
<https://oap.ingeniariak.eus/2022/08/01/existe-una-brecha-digital-entre-grandes-y-pequenas-empresas/>.

Jobs, Steve P. n.d. *“La gente no sabe lo que quiere hasta que se lo enseñas. Por eso nunca me he basado en las investigaciones de mercado. Nuestra tarea estriba en leer las páginas que todavía no se han escrito.”*

Ministerio de Industria y Turismo. n.d. “Ministerio de Industria y Turismo - Estadísticas y publicaciones sobre PYME.” Industria y PYME. Accessed marzo 11, 2024.

<https://industria.gob.es/es-es/estadisticas/paginas/estadisticas-y-publicaciones-sobre-pyme.aspx>.

Omeñaca García, Jesús. 2008. *CONTABILIDAD GENERAL. 11a*. N.p.: Deusto.

Romero, Rocío. 2023. “¿Por qué fidelizar clientes cuesta 5 veces menos que captarlos? • Nömad.” Nömad.

<https://nomad.ooo/blog/por-que-fidelizar-clientes-cuesta-5-veces-menos-que-captarlos/>.

Schuler, Nikolai. n.d. “Data Warehouse - The Ultimate Guide.” Udemy.

Accessed marzo 15, 2024.

<https://www.udemy.com/course/data-warehouse-the-ultimate-guide/>.

ShippyPro. 2023. “KPIs del retail: los 12 indicadores más importantes para las tiendas físicas.” ShippyPro Blog.

<https://www.blog.shippypro.com/es/kpi-retail>.

Woodie, Alex. 2020. “Data Prep Still Dominates Data Scientists' Time, Survey Finds.” Datanami.

<https://www.datanami.com/2020/07/06/data-prep-still-dominates-data-scientists-time-survey-finds/>.

13. Anexos

Instrucciones de instalación del gestor de paquetes Brew para Mac o Linux

<https://brew.sh/>

Instrucciones de instalación de PostgreSQL 15 para Mac a través de Brew

<https://formulae.brew.sh/formula/postgresql@15#default>

Instrucciones de instalación de Pentaho Data Integration

<https://www.hitachivantara.com/en-us/pdf/implementation-guide/three-steps-to-install-pentaho-data-integration-ce.pdf>

Enlace de descarga de la aplicación Pentaho Data Integration Community Edition (requiere registro)

<https://pentaho.com/pentaho-community-edition/>

Instrucciones de instalación de DBeaver

<https://dbeaver.com/docs/dbeaver/Installation/>

Enlace de descarga de la aplicación DBeaver Community Edition

<https://dbeaver.io/download/>

Instrucciones de instalación de Tableau

https://help.tableau.com/current/desktopdeploy/en-us/desktop_deploy_download_and_install.htm

Enlace de descarga de la aplicación Tableau (requiere registro)

https://www.tableau.com/support/releases/desktop/2024.1.3?utm_source=discover_pane&utm_medium=product#esdalt