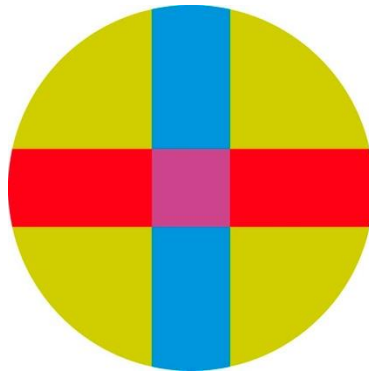


UNIVERSIDAD SAN PABLO CEU
ESCUELA POLITÉCNICA SUPERIOR
GRADO EN INGENIERÍA DE SISTEMAS DE
INFORMACIÓN



TRABAJO FÍN DE GRADO
Cuadro de Mando de Empresa Logística
Logistics company dashboard

Autor: Ignacio Martín-Peña Chinchurreta
Tutor: Pedro Garrido Gutiérrez
Co-Tutor: Luis Miguel Garay Gallastegui
Junio 2024



UNIVERSIDAD SAN PABLO-CEU

ESCUELA POLITÉCNICA SUPERIOR

División de Ingeniería

Calificación del Trabajo Fin de Grado

Datos del alumno

NOMBRE:

Datos del Trabajo

TÍTULO DEL PROYECTO:

Tribunal calificador

PRESIDENTE:

FDO.:

SECRETARIO:

FDO.:

VOCAL:

FDO.:

Reunido este tribunal el ____/____/_____, acuerda otorgar al Trabajo Fin de Grado presentado por D. Ignacio Martín-Peña Chinchurreta la calificación de _____



Resumen

Para este Trabajo Fin de Grado titulado “Cuadro de Mando para una Empresa Logística”, mi propósito principal ha sido poder optimizar la toma de decisiones de la empresa a través de una solución de Inteligencia Empresarial (Business Intelligence - BI). El proyecto se ha centrado en el diseño e implementación de un almacén de datos a través de distintos puntos de acceso (Data Warehouse) y la creación de un cuadro de mando integral e interactivo utilizando Power BI, enfocados en una empresa del sector logístico que se denomina DataCo. La herramienta permite un análisis detallado de los KPI's relacionados con las ventas, envíos, pagos y estados de pedidos. Esto le proporcionará finalmente a la empresa una visión clara y comprensible de los datos para mejorar la eficiencia operativa y la toma de decisiones estratégicas.

La solución BI ofrecerá visualizaciones interactivas y capacidades avanzadas de análisis de datos. Gracias a la herramienta Power BI permitirá a los usuarios finales filtrar, segmentar y explorar los datos de manera intuitiva, mejorando significativamente la comprensión y el uso de la información disponible. Además, se desarrollará un plan de migración y carga inicial de datos explicado más adelante para asegurar que toda la información relevante sea correctamente integrada en el sistema.

Esta solución pretende resolver las limitaciones del sistema actual en el caso de esta empresa ha sido Excel, que enfrenta dificultades en el manejo y análisis de grandes volúmenes de datos. El nuevo sistema ofrecerá consultas rápidas y flexibles, eliminando las restricciones actuales y mejorando la capacidad de la empresa para tomar decisiones informadas y oportunas en un tiempo récord.



Palabras Clave

Power BI: es una herramienta potente de Microsoft diseñada para la visualización interactiva de datos. Este programa ayuda a las empresas unir datos de diversos sitios, analizarlos y crear informes detallados que facilitan la toma de decisiones.

KPI's (Indicadores Clave de Rendimiento): ayudan a medir el éxito de una empresa en áreas específicas. Estas métricas son esenciales para evaluar el progreso hacia objetivos estratégicos y para identificar áreas que requieren mejoras.

Logística: se refiere a la gestión eficiente del flujo de bienes, servicios e información desde el punto de origen hasta el destino final. Incluye actividades como el almacenamiento, el transporte y la distribución, esenciales para asegurar que los productos lleguen a los clientes de manera oportuna y rentable.

Automatización de Reportes: herramientas tecnológicas para generar informes automáticamente, sin intervención manual constante. Esto mejora la eficiencia operativa, reduce errores humanos y garantiza que los datos estén siempre actualizados para la toma de decisiones.

Análisis de la Cadena de Suministro: significa examinar y optimizar cada fase del proceso de producción y distribución. Su objetivo es mejorar la eficiencia, reducir costos y asegurar que los productos lleguen al mercado de manera efectiva y eficiente.

Inteligencia Empresarial (BI): engloba las estrategias y tecnologías utilizadas para la recopilación, integración, análisis y presentación de información empresarial. BI permite a las organizaciones tomar decisiones basadas en datos precisos, mejorando su eficiencia y competitividad.

Cuadro de Mando: es una herramienta visual que muestra indicadores clave de rendimiento (KPI's) en un formato accesible. Facilita la supervisión de las operaciones y apoya la toma de decisiones estratégicas al proporcionar una visión clara y consolidada del desempeño empresarial.



Abstract

For this Final Degree Project titled “Dashboard for a Logistics Company,” my main goal has been to optimize the company’s decision-making process through a Business Intelligence (BI) solution. The project focused on designing and implementing a data warehouse through various access points and creating a comprehensive and interactive dashboard using Power BI, targeting a logistics company named DataCo. This tool enables detailed analysis of KPIs related to sales, shipments, payments, and order statuses, providing the company with a clear and understandable view of data to enhance operational efficiency and strategic decision-making. The BI solution will offer interactive visualizations and advanced data analysis capabilities. Thanks to Power BI, end-users will be able to filter, segment, and explore data intuitively, significantly enhancing their understanding and use of the information available. Additionally, a migration plan and initial data loading will be developed, ensuring all relevant information is correctly integrated into the system. This solution aims to overcome the limitations of the current system, previously reliant on Excel, which struggled with managing and analyzing large data volumes. The new system will provide quick and flexible queries, eliminating current restrictions and improving the company’s ability to make informed and timely decisions.



Keywords

Power BI: A powerful Microsoft tool designed for interactive data visualization, helping businesses to merge data from various sources, analyze it, and create detailed reports that facilitate decision-making.

KPIs (Key Performance Indicators): Metrics that help measure a company's success in specific areas. These metrics are crucial for assessing progress towards strategic goals and identifying areas needing improvement.

Logistics: Refers to the efficient management of the flow of goods, services, and information from the point of origin to the end destination. It includes activities like storage, transport, and distribution, essential for ensuring products reach customers timely and profitably.

Report Automation: Technological tools that automatically generate reports without constant manual intervention. This improves operational efficiency, reduces human errors, and ensures data is always up-to-date for decision-making.

Supply Chain Analysis: Involves examining and optimizing each stage of the production and distribution process. Its aim is to enhance efficiency, reduce costs, and ensure products reach the market effectively and efficiently.

Business Intelligence: Encompasses the strategies and technologies used for data collection, integration, analysis, and presentation of business information. BI enables organizations to make data-driven decisions, enhancing their efficiency and competitiveness.

Dashboard: A visual tool that displays key performance indicators (KPIs) in an accessible format. It aids in monitoring operations and supports strategic decision-making by providing a clear and consolidated view of business performance.



Índice general

Resumen	I
Palabras Clave	II
Abstract.....	III
Keywords	IV
Índice general	V
1. Introducción	1
1.1 Contexto del TFG.....	1
1.2 Objetivos	2
1.3 Organización del trabajo.....	4
2. Gestión del proyecto.....	5
2.1 Modelo de ciclo de vida	5
2.2 Planificación	7
2.3 Presupuesto	9
3. Análisis	10
3.1 Especificación de requisitos.....	10
3.2 Análisis de los Casos de Uso	12
3.3 Análisis de seguridad y RGPD	14
4. Diseño e implementación	15
4.1 Arquitectura del sistema	15
4.1.1 Arquitectura física	17
4.1.2 Arquitectura lógica	18
4.2 Diseño de datos	19
4.2.1 Migración y carga inicial de datos	22
4.3 Diseño de la interfaz de usuario	24
4.4 Diagrama de clases	36
4.5 Entorno de construcción.....	39
4.6 Referencia al repositorio de software.....	39
5. Validación del sistema	40
5.1 Plan de pruebas	40



6.	Conclusiones y líneas futuras	42
6.1	Conclusiones	42
6.2	Líneas futuras	43
7.	Bibliografía.....	44
	Anexo I – Manuales	45
	Glosario de términos.....	47



Índice de figuras

Ilustración 1: Diagrama de Gantt	8
Ilustración 2: Diagrama de Flujo de Datos.	16
Ilustración 3: Diagrama de la Tabla Principal del Dataset.....	21
Ilustración 4: Diagrama de Entidad Relación.	23
Ilustración 5: Cuadro de Mando Overview.....	25
Ilustración 6: Cuadro de Mando KPI's (versión 1).....	26
Ilustración 7: Cuadro de Mando KPI's (versión 2).	26
Ilustración 8: Cuadro de Mando Product (versión 1).	27
Ilustración 9: Cuadro de Mando Product (versión 2).....	28
Ilustración 10: Cuadro de Mando Customer.	29
Ilustración 11: Cuadro de Mando Customer (versión 2).	30
Ilustración 12: Cuadro de Mando CEO (versión 1).	31
Ilustración 13: Cuadro de Mando CEO (versión 2).....	32
Ilustración 14: Cuadro de Mando Warehouse.....	33
Ilustración 15: Cuadro de Mando Transport.	34
Ilustración 16: Cuadro de Mando Logistic (versión 1).....	35
Ilustración 17: Cuadro de Mando Logistic (versión 2).	36



Índice de tablas

<i>Tabla 1. Requisitos Funcionales.</i>	<i>12</i>
--	-----------



1. Introducción

Actualmente en la logística, la capacidad de procesar y analizar rápidamente gran cantidad de datos se ha convertido en una ventaja competitiva clave. La empresa DataCo que he elegido estudiar ha enfrentado algunas limitaciones debido a la dispersión de sus sistemas y el poco conocimiento de herramientas analíticas que permitan una visión clara de sus operaciones. Esta dispersión de datos y los informes incoherentes han complicado la toma de decisiones informadas y rápidas.

Por lo tanto, mi Trabajo de Fin de Grado (TFG) se centrará en el desarrollo de una solución de inteligencia de negocio (BI) utilizando Power BI, una herramienta líder en el mercado creada por Microsoft. El objetivo es unir diversas fuentes de datos en un único almacén diseñado para soportar consultas de alto rendimiento y proporcionar análisis en tiempo real. Los cuadros de mando interactivos desarrollados permiten a los distintos departamentos de la empresa visualizar indicadores clave de rendimiento (KPI) y métricas operacionales de manera intuitiva y accesible.

El desarrollo de este sistema BI no solo busca mejorar la eficiencia y precisión en las operaciones diarias de la empresa, sino también transformar la cultura organizacional hacia una gestión basada en datos. Al proporcionar una plataforma robusta para el análisis de datos, se facilita la identificación de tendencias, problemas y oportunidades, permitiendo así una respuesta más ágil a las dinámicas del mercado y una planificación estratégica más efectiva.

El proyecto integra conceptos de gestión de datos, análisis tecnológico y estrategia empresarial, reflejando la importancia de la inteligencia de negocio en la optimización de la cadena de suministro y la logística. La solución desarrollada aborda también un marco escalable para futuras expansiones y adaptaciones, asegurando así la sostenibilidad y el crecimiento continuo en un sector altamente competitivo.

1.1 Contexto del TFG

Durante mis prácticas en una empresa del sector, pude observar de primera mano las limitaciones causadas por la falta de herramientas analíticas avanzadas y la dispersión de sistemas. Esta experiencia fue el desencadenante para mi Trabajo Fin de Grado, donde busco mejorar significativamente la toma de decisiones a través de una solución de inteligencia de negocio (BI).

La relevancia técnica de este TFG radica en el uso de técnicas avanzadas de extracción, transformación y carga (ETL) para automatizar el procesamiento de datos. Además, el diseño y desarrollo de gráficos y medidas personalizados permite ofrecer información específica para cada área de la empresa, ayudando a directivos como el CEO, el gerente de logística y el gerente de almacén a tomar decisiones más informadas.



1.2 Objetivos

Objetivo General:

“Diseñar y construir un cuadro de mando integral (CMI) para una empresa logística (DataCo) utilizando Power BI, con la finalidad de mejorar la toma de decisiones estratégicas y operativas de la empresa”

Objetivos Específicos:

Los objetivos más específicos los he dividido en:

- Simplificación de Datos: transformar la información logística en datos más entendibles, tratando de explorar áreas de mejora y haciendo más sencilla la toma de decisiones.
 - Acciones:
 - i. Estudiar la estructura y el contenido del Dataset de la empresa logística DataCo extraída a través de Kaggle.
 - ii. Utilizar Power BI para transformar datos en visualizaciones claras y fáciles de entender.
 - iii. Generar una visión general de la situación actual de la empresa, así como de los objetivos y metas futuras, para facilitar el conocimiento del estado actual.
- Consolidación de Fuentes de Datos: unificar toda la información interesante sobre las ventas, envíos y pagos en un modelo centralizado ofreciendo una visión global.
 - Acciones:
 - i. Integrar diferentes fuentes de datos en un único modelo.
 - ii. Garantizar la coherencia y exactitud de los datos de estudio.
 - iii. Crear un modelo de datos que permita un análisis global y detallado de la información.
- Optimización en la Toma de Decisiones: proporcionar datos relevantes para cada departamento de la empresa mejorando la claridad y eficiencia operativa.
 - Acciones:
 - i. Determinar los KPI's específicos para cada departamento (CEO, Gerentes de logística, transporte y almacén)
 - ii. Creación de paneles personalizados para cada departamento en Power BI, proporcionando herramientas analíticas para tomar decisiones informadas y basadas en datos.
 - iii. Medir los datos más representativos para identificar KPI's de logística, evaluarlos frente a objetivos y crear nuevas métricas.



- Automatización del Proceso ETL: implementar un flujo de extracción, transformación y carga (ETL) mediante un script de Python que transforma, limpia y prepara de forma eficiente datos de archivos CSV para su posterior análisis en Power BI.
 - Acciones:
 - i. Desarrollar un script en Python para transformar y limpiar los datos proporcionados en archivos CSV.
 - ii. Realizar conversiones de delimitadores y separadores decimales en archivos Excel.
 - iii. Cargar automáticamente los datos convertidos a Power BI.
- Creación de Puntos de Acceso a Datos: creación de dos puntos de acceso a datos para que Power BI reciba la información, facilitando la actualización y unificación de datos.
 - Acciones:
 - i. Añadir una carpeta en Sharepoint donde los empleados puedan agregar datos fiscales semanales.
 - ii. Configurar un punto de acceso en OneDrive para almacenar la mayoría de los datos históricos, donde el acceso no es compartido únicamente el propietario puede acceder y editar los datos.
 - iii. Combinar los datos de ambas fuentes en Power Query, asegurando que las tablas finales en el modelo de datos sean consistentes y completas.
- Facilitar la toma de decisiones: proporcionar una información general clara y concisa que permita conocer la empresa y su entorno.
 - Acciones:
 - i. Crear visualizaciones sencillas que demuestren el logro de los objetivos planificados y que proporcionen herramientas para eliminar inconsistencias.
 - ii. Implementar herramientas de medición para verificar el progreso y el logro de metas, identificando áreas de mejora a través del análisis de indicadores clave.
- Página Web Estática: publicar paneles creados en Power Bi en un dominio web (iglobal-store.com) para que todas las partes interesadas puedan acceder fácilmente a informes y métricas clave en tiempo real.
 - Acciones:
 - i. Configurar un entorno web estático para publicar los paneles.
 - ii. Integrar los paneles de Power BI con el sitio web.
 - iii. Proporcionar a las partes interesadas un acceso seguro a información relevante del proyecto.



1.3 Organización del trabajo

El proyecto consta de una estructura organizativa de la forma descrita a continuación:

1. **Introducción:** se proporciona el contexto general del proyecto, explicando la necesidad de implementar un sistema de análisis integral para una empresa de logística. Se describen los objetivos principales del trabajo y se justifica la elección del conjunto de datos y herramientas utilizadas.
2. **Gestión del Proyecto:** descripción del modelo de ciclo de vida aplicado para el desarrollo del sistema, así como la planificación operativa y el presupuesto estimado. Además, se identifican los posibles riesgos y se discuten las estrategias para reducirlos.
3. **Análisis:** se detallan los requisitos específicos que debe cumplir el sistema, incluyendo casos de uso que reflejan las necesidades de diferentes puestos dentro de la empresa. También se analizan cuestiones de seguridad y cumplimiento del RGPD.
4. **Diseño e Implementación:** esta sección explica la arquitectura del sistema a nivel físico y lógico incluyendo la estructura del modelo de datos en Power BI. El diseño de la interfaz de usuario se describe en detalle y se proporciona un diagrama de clases para mostrar las relaciones entre los distintos paneles. También se describe un script en Python para transformar los datos e integrarlos con el proceso ETL. Por otra parte, hay una configuración previa de puntos de acceso a datos en SharePoint y OneDrive, así como la implementación del sitio web estático para mostrar un resumen del proyecto.
5. **Validación del Sistema:** presentación del plan de pruebas realizado para verificar la precisión de las mediciones DAX y la coherencia entre los paneles. Se evalúa la disponibilidad y facilidad del uso del sistema, presentando los resultados de las pruebas de rendimientos y realizando mejoras en función de los comentarios recibidos.
6. **Conclusiones y Líneas Futuras:** por último, se realiza una aclaración de los hallazgos obtenidos durante el desarrollo e implementación del sistema mostrando los resultados alcanzados e identificando las áreas de mejora. Propone posibles mejoras en repeticiones del proyecto integrando más fuentes de datos y mejorando aún mas las métricas de los usuarios.



2. Gestión del proyecto

2.1 Modelo de ciclo de vida

Para el desarrollo de este proyecto, elegí un modelo de ciclo de vida incremental combinado con una metodología ágil. Esto nos permite que la solución evolucione a través de un proceso de iteración rápida, añadiendo constantemente los comentarios de las partes interesadas y ajustando los planes en función de las ideas que van saliendo.

- **Análisis Previo y Reuniones Iniciales:** durante la primera reunión, discutimos los objetivos del cuadro de mando e identificamos los requisitos clave. Para el análisis, elegí el conjunto de datos “DataCo Smart Supply Chain for Big Data Analysis” de Kaggle. Después de recoger los datos históricos, genere los indicadores clave de rendimiento (KPI’s) e investigue casos de estudio en la Universidad San Pablo CEU, donde hago el grado de Ingeniería de Sistemas de Información. Estos intercambios con profesores y compañeros me ayudaron a profundizar en los objetivos y estrategias clave.
- **Selección de Tecnología y Metodología:** escogí Microsoft Power BI como mi herramienta de generación de informes por su gran capacidad para la transformación de datos en informes atractivos. Para unificar de diversas fuentes, construí un almacén de datos. Implemente un enfoque iterativo y ágil, comenzando con Power BI Desktop, para conectarme a varias fuentes de datos y procesarlas mediante Power Query. Luego cree el modelo de datos con una configuración previa de los componentes y por último publique el informe en Power BI, incluyendo el acceso desde dispositivos móviles.
- **Diseño e Implementación del Modelo de Datos:** implemente relaciones entre tablas modelo y cree métricas DAX para análisis detallados como tasa de éxito de los envíos, promedio de días de retraso, tasa de cumplimiento, etc... Hice pruebas periódicas para asegurar la integridad de cada métrica, manteniendo la integración continua entre componentes. Además, he definido flujos de datos que admiten la recopilación, transformación y actualización automática.
- **Iteraciones Incrementales:** el proceso de desarrollo se ha realizado por etapas perfeccionando el sistema de BI. Desarrollé visualizaciones precisas que reflejan la situación logística y de ventas. En cada fase agregue un análisis sobre cancelaciones, incidencias y pedidos a tiempo.
- **Validación, Presentación y Despliegue:** elabore presentaciones regulares con mi tutor solicitando sus comentarios y construyendo informes basados en sus recomendaciones. Las sugerencias recibidas durante cada iteración ayudaron a mejorar aún mas el cuadro de mando integral.



- **Fase de Transferencia:** idee un plan de migración para que cualquier persona de interés que quisiera generar el entorno en cualquier otro dispositivo. Además, cree un sitio web estático en mi dominio personal iglobal-store.com, usando AWS para alojarlo. El sitio contiene una subpágina con todos los informes de Power BI creados para la empresa y todo el contenido relevante del proyecto. Esto centraliza todos los elementos de diseño, facilitando el mantenimiento o próximas actualizaciones.

Este ciclo de mejora incremental me ha permitido desarrollar una solución que no solo cumple con los requisitos iniciales sino también sienta las bases para la integración con futuros sistemas comerciales.

Para llevar a cabo este proyecto, asumí varios roles clave:

- **Programador:** las funciones eran desarrollar el modelo de datos y configurar las medidas DAX necesarias para analizar el mejor análisis logístico posible, creando informes visuales efectivos y atractivos.
- **Encargado del Seguimiento:** me aseguro de que el proyecto vaya por buen camino estableciendo métricas claras para medir el progreso y asegurarme de que cada iteración sea completada.
- **Jefe de Proyecto:** determino la estrategia general y realizo ajustes a lo largo del proceso para cumplir con los plazos objetivos con mi tutor.
- **Consultor:** recopile e implemente recomendaciones estratégicas que me ayudaron a estructurar el modelo de datos y comunicar claramente las métricas más interesantes.

Combine estos roles, con el fin de supervisar el proyecto de principio a fin, dándole más importancia siempre a las necesidades del usuario final y cumpliendo con las expectativas de calidad.



2.2 Planificación

Para garantizar el éxito del proyecto, se identificaron y planificaron cuidadosamente las tareas a realizar. Las fases del proyecto se detallan las actividades específicas, fechas estimadas de inicio y finalización.

Idea y Definición del Proyecto:

- Actividades:
 - Determinar los principales objetivos y alcance del proyecto.
 - Llevar a cabo reuniones iniciales con el tutor para establecer las principales ideas y explorar fuentes de datos para su posterior análisis.
 - Consultas periódicas con mi tutor y compañeros de clase con el fin de aclarar los objetivos finales de mi proyecto.
- Fechas estimadas: 01/11/2023 – 15/11/2023.

Análisis de Requisitos:

- Actividades:
 - Considerar los requisitos del sistema y las conclusiones de las partes interesadas.
 - Identificar métricas clave.
 - Analizar los datos ya elegidos.
- Fechas estimadas: 16/11/2023 – 30/11/2023

Diseño del Sistema:

- Actividades:
 - Desarrollar la arquitectura del sistema para empezar a añadir las múltiples fuentes de datos.
 - Creación de bocetos para los distintos informes.
 - Planificar la estructura del modelo de datos en Power BI.
- Fechas estimadas: 01/12/2023 – 31/12/2023

Construcción e Integración:

- Actividades:
 - Limpiar, transformar e integrar datos con Power BI.
 - Crear informes finales y cuadros de mando interactivos.
 - Implementar script para automatizar el proceso ETL.
- Fechas estimadas: 02/01/2024 – 15/02/2024.

Desarrollo en Iteraciones Incrementales:

- Actividades:
 - Implementar funciones por etapas.
 - Diseño de las visualizaciones de los distintos cuadros de mando.
- Fechas estimadas: 16/02/2024 – 15/04/2024.

Pruebas y Validación:

- Actividades:
 - Realizar pruebas de funcionalidad y precisión.
 - Revisar cada informe en detalle.
 - Considerar la retroalimentación y los ajustes necesarios.
- Fechas estimadas: 16/04/2024 – 30/04/2024.

**Documentación de la memoria:**

- Actividades:
 - o Redactar la memoria del TFG.
 - o Documentar las decisiones de diseño y los métodos utilizados.
- Fecha estimada: 01/05/2024 – 15/05/2024.

Conclusiones y Mejora Continua:

- Actividades:
 - o Repasar el resultado final.
 - o Identificar áreas que necesitan mejoras.
 - o Planificar posibles mejoras.
- Fecha estimada: 16/05/2024 – 31/05/2024.

Alojamiento Web en AWS:

- Actividades:
 - o Configurar Amazon Web Services para el alojamiento web.
 - o Creación de un sitio web simulando un repositorio central de todo el proyecto.
 - o Integración del informe final de Power BI.
- Fechas estimadas: 01/06/2024 – 15/06/2024.

En el siguiente diagrama de Gantt proporciona una vista detallada del calendario y orden de ejecución de cada fase del proyecto:

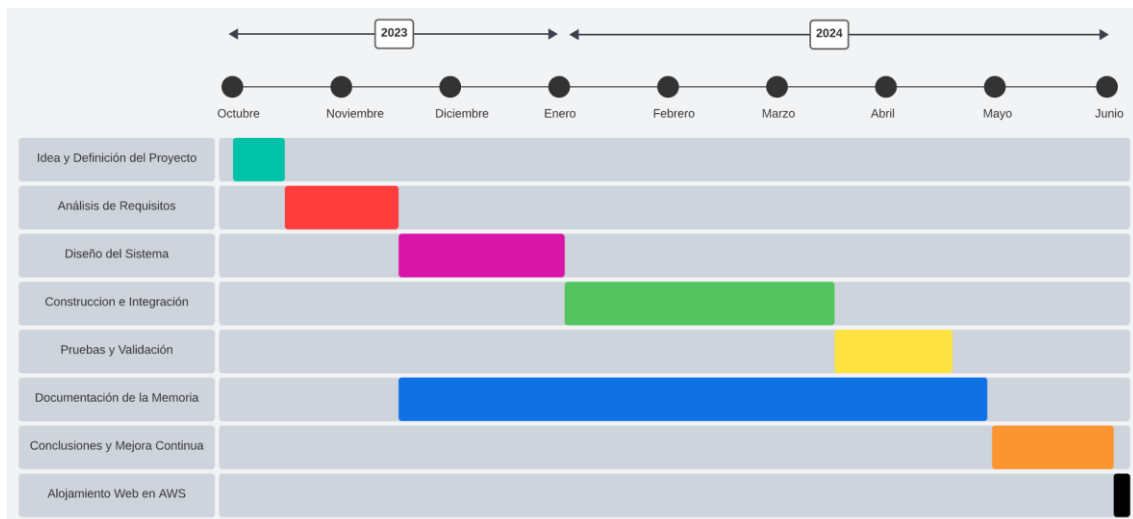


Ilustración 1: Diagrama de Gantt



2.3 Presupuesto

Para estimar los costes de desarrollo del proyecto tome en cuenta varios factores incluido el software, la infraestructura y otros gastos.

Software y Herramientas:

- Microsoft 365: como estudiante, tuve acceso gratuito permitiéndome trabajar con las herramientas de SharePoint, Power BI, Microsoft Word y Excel.
- Amazon Web Services (AWS): la Universidad me proporciono un crédito de 100€ para utilizar estos servicios para el alojamiento de la página web estática en mi dominio personal.
- Dominio (GoDaddy): el dominio personal para vincular la página web tiene un costo anual de 12€.

Infraestructura:

- Equipo de desarrollo: hice uso de mi ordenador personal para la documentación de informes, análisis de datos y pruebas.

Otros gastos:

- Consumo Eléctrico: estos costes durante el desarrollo no fueron tomados en cuenta.

Dado el acceso a software gratuito y créditos de AWS, los costos directos se mantuvieron bajos. El principal gasto fue el tiempo invertido en la ejecución del proyecto, especialmente para su diseño, implementación y documentación.



3. Análisis

3.1 Especificación de requisitos

En el desarrollo del cuadro logístico en Power BI se abarca la determinación de los requisitos estándar necesarios para asegurar que el sistema sea directo y claro.

Funcionalidades del Sistema:

La visualización de datos en Power BI transforma los datos en información útil para llevar a cabo una toma de decisiones precisa. Las funcionalidades clave incluyen:

- Integración de Datos: asegurar que la información de varias fuentes este en conexión continua y actualizada.
- Visualización Dinámica: paneles interactivos que proporcionan datos en tiempo real, así como gráficos de tendencias, KPI's y tablas de rendimiento.
- Filtrado y Segmentación: ofrecer a los usuarios la opción de ordenar los datos por criterios como fechas, regiones y tipos de productos.
- Actualización Automática: mantener los tableros actualizados todos los días mediante una hora programada.

Interfaz de usuario:

El diseño aborda la usabilidad y accesibilidad según los estándares WCAB 2.1:

- Diseño Responsivo: compatible para distintos dispositivos como pueden ser los ordenadores o móviles.
- Navegación Intuitiva: funcionalidad de menús y herramientas para un uso fluido.

Rendimiento:

Debe ser posible tener una experiencia de usuario eficiente:

- Tiempos de Carga: los cuadros de mando deben cambiarse en menos de 5 segundos para obtener la información necesaria rápidamente.

***Seguridad:***

Medidas robustas para proteger los datos:

- Autenticación y Autorización: acceso al sistema permitido solo para los usuarios con los permisos correspondientes.
- Encriptación de Datos: cifrar los datos destinados al servidor y cifrar los archivos almacenados en el servidor, dificultando así el uso de los datos por parte del personal no autorizado.

Protección de datos:

El Sistema se centra en la legislación de protección de datos para obtener resultados sobresalientes en la seguridad de la información personal y empresarial.

- Cumplimiento de GDPR: formación y uso de regulaciones conformes con el GDPR no solo para la seguridad de los datos, sino también para tratar toda la información de acuerdo con las directrices recientes.

Requisitos No Funcionales:

Además de los requisitos funcionales, se han considerado varios requisitos no funcionales para garantizar lo siguiente:

- Mantenibilidad: para apoyar el Sistema, todo el proceso, incluida la implementación de nuevas funcionalidades en cualquier periodo de creación y la sustitución de problemas defectuosos, debe realizarse de manera sencilla para que el problema pueda resolverse.
- Escalabilidad: esencial para manejar la expansión del negocio y el volumen de trabajo.
- Eficiencia: en los recursos para que los servicios valgan su valor.
- Usabilidad: interfaz gráfica de usuario simple y comprensible para que el usuario se sienta bien.



Ref.	Descripción
RF01	Integración de Datos.
RF02	Visualización Dinámica
RF03	Filtrado y Segmentación.
RF04	Actualización Automática.
RF05	Diseño Responsivo.
RF06	Navegación Intuitiva.
RF07	Tiempos de Carga.
RF08	Autenticación y Autorización.
RF09	Encriptación de Datos.
RF10	Cumplimiento de GDPR.

Tabla 1. Requisitos Funcionales.

3.2 Análisis de los Casos de Uso

El análisis de casos de uso se vuelve necesario a la hora de comprender las interacciones entre los distintos actores y el sistema desarrollado. En esta parte se presentan los actores principales y los casos de uso identificados, como la identificación de las necesidades operativas y estratégicas de la empresa.

Actores:

- CEO: responsable de la organización en su totalidad, incluidas las tareas de planificación estratégica de DataCo.
- Encargado de Almacén: supervisa la gestión de los productos y la disponibilidad de estos.
- Encargado del Transporte: decide el orden, estado del envío y las rutas de entrega.
- Encargado de la Logística: encargado de la coordinación del flujo de la cadena de suministro y la optimización de procesos.

**Casos de Uso:**

- Gestión de Pedidos: permite a los operadores visualizar los pedidos, así como verificar el estado de las órdenes y la disponibilidad de productos. Esta funcionalidad es clave para asegurar el éxito de la empresa, ya que garantiza la precisión y eficiencia en las operaciones diarias.
- Visualización de Informes: permite a la gerencia y a los directores obtener informes más claros y actualizados sobre las métricas (KPI's) del negocio, que incluyen tiempos medios de entrega, niveles de eficiencia y puntualidad, porcentaje de ordenes por estado, crecimiento anual, etc... Esta función central es la base de la evaluación del rendimiento y de los cambios estratégicos, como en el caso de la estrategia empresarial.
- Análisis Predictivo: a través del análisis de datos históricos y en tiempo real a la vez, se consigue la previsión de tendencias y toma de decisiones basadas en un examen exhaustivo. Esta capacidad le da a los gerentes y directores una ventaja sobre los movimientos esperados del mercado, gracias a ello pueden reaccionar a tiempo y asegurarse de que todo funcione sin problemas.
- Integración de Datos: asegura la entrada de información sin problemas desde diferentes puntos de acceso, en mi caso desde una página de Sharepoint y OneDrive.
- Alertas en Tiempo Real: permite el lanzamiento de notificaciones en tiempo real para reaccionar inmediatamente a los cambios en los entornos operativos o nuevas demandas del mercado. Esta funcionalidad es importante ya que juega un papel en la reactividad y efectividad de la gestión logística.



3.3 Análisis de seguridad y RGPD

Realizar el análisis de seguridad para el proyecto es algo bastante importante a considerar debido a que evita brechas de seguridad de la información y de los datos. Este estudio refleja la autenticidad, confidencialidad, integridad, disponibilidad y trazabilidad, junto con sus respectivas medidas de seguridad.

Autenticidad:

Para asegurarse de que solo los usuarios permitidos puedan acceder, Power BI implementa un sistema de autenticación basado en Microsoft Entra ID, que es muy robusto, el mismo que utiliza Microsoft. Además, cuenta con autenticación de dos factores (2FA) que asegura que las credenciales se verifiquen de manera efectiva, reduciendo significativamente el riesgo de intentos de acceso no autorizados.

Confidencialidad:

La protección de los datos sensibles es clave. Es por lo que Power BI se basa en el cifrado tanto en tránsito como en reposo, asegurando que los datos estén protegidos contra acciones no autorizada durante la transmisión y el almacenamiento. En mi caso, he definido roles y permisos que permiten el acceso solo a aquellos usuarios involucrados.

Integridad:

El sistema incluye además mecanismos de validación y registro de actividades para asegurar la integridad de los datos. Cada edición de datos y acción de usuario se registra.

Disponibilidad:

El sistema utiliza la infraestructura de Azure para asegurar la disponibilidad constante y la escalabilidad de los servicios de Power BI.

Trazabilidad:

Se almacena un registro exhaustivo de todas las actividades del sistema, como el acceso y la modificación de datos.

El análisis de seguridad del sistema de Power BI para la empresa DataCo está diseñado para ser orientado a soluciones y mitigación de riesgos en todas las dimensiones, creando un entorno seguro y confiable para tratar los datos sensibles y críticos de la organización.



4. Diseño e implementación

4.1 Arquitectura del sistema

En esta sección explico cómo diseñe el sistema. La arquitectura combina diferentes componentes y servicios principales, facilitados por las tecnologías de Microsoft Azure que utiliza Power BI para analizar datos de manera efectiva y en tiempo real.

Componentes Principales del Sistema:

- **Interfaz del Usuario:** Power BI como ya he explicado en el documento es una de las mejores plataformas interactivas que te permite ver datos y darles sentido a través de paneles e informes. Se pueden acceder a las visualizaciones de varias formas:
 - Directamente en el archivo de Power BI.
 - A través de un enlace proporcionado por Power BI para quienes no tienen el programa instalado.
 - Incrustado en una página web: como comentaba, he utilizado una página web para unificar toda la información relacionada a mi TFG, y he creado una sección donde podéis ver e interactuar con los paneles. La URL es <https://iglobal-store.com/dashboard.html>.
- **Servicios Proporcionados:**
 - Recopilación de Datos: la extracción y carga de datos se produce desde carpetas de SharePoint y OneDrive. En SharePoint sólo se identifican los archivos de Excel que comienzan con "FY".
 - Transformación y Limpieza de Datos: cuando se extraen los datos de SAP, un script de Python (creado por mí y podéis verlo en el repositorio) los procesa y se encarga de limpiar y formatear las tareas antes de enviarlos a SharePoint. Power BI usa Power Query para realizar el proceso ETL (Extraer, Transformar, Cargar) después de este paso.
 - Almacenamiento de Datos: Power BI aprovecha Azure SQL Database como su almacén de datos backend para información estructurada, mientras que Azure Blob Storage se usa para administrar archivos no estructurados.



- Visualización y Reportes: en Power BI cree los paneles e informes interactivos, que permiten a los gerentes tener una representación visual de los KPI y métricas operativas que podrían revolucionar la forma en que se gestionan las empresas. En el siguiente capítulo, se analizará detalladamente la funcionalidad de cada panel: esto incluirá versiones y cualquier variación distintiva presente en cada panel.

- Componentes de Apoyo:

- Sistemas de Almacenamiento: la infraestructura de almacenamiento de datos la proporcionan Azure SQL Database y Azure Blob Storage.
- Servicios de Autenticación y Seguridad: La gestión del acceso y la integridad de los datos se garantizan mediante el uso de Azure Active Directory.
- Servicios de Comunicación: los conectores y API de Azure facilitan la comunicación entre Power BI y el almacenamiento de datos basado en la nube, así como los servicios de procesamiento.

- Capa de Acceso a Datos:

- Conectores de Datos: Power BI almacena los datos en Azure SQL Database y Azure Blob Storage.
- Procedimientos de almacenado: las consultas y las operaciones relacionadas con los datos, se encuentran en Azure SQL Database.

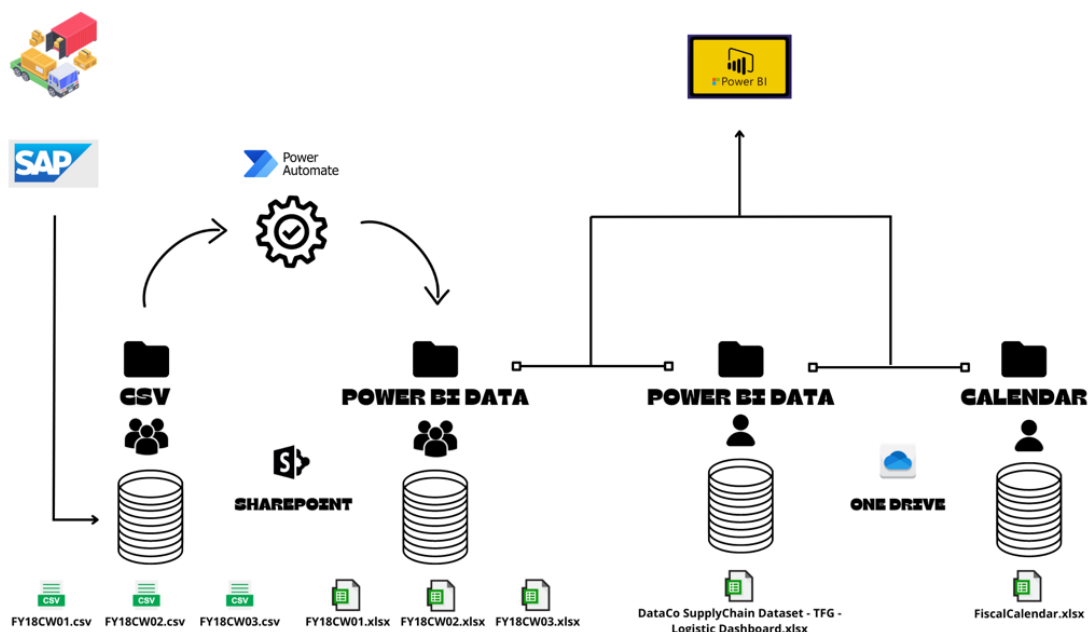


Ilustración 2: Diagrama de Flujo de Datos.



Muestra cómo los datos fluyen desde SharePoint y OneDrive al sistema de almacenamiento de Azure sin intervención directa sino a través de Power BI que los procesa después de llevar a cabo un proceso ETL y que finalmente aparecen en los paneles para su visualización.

Al proporcionar una plataforma sólida y sostenible para el análisis de datos logísticos, se desarrolla la arquitectura del sistema, lo que conduce a grandes mejoras en la toma de decisiones en virtud de una representación visual y accesible de los indicadores clave de rendimiento (KPI) junto con otros datos operativos.

4.1.1 Arquitectura física

Esta parte detalla todas las partes tangibles esenciales para establecer y administrar el sistema de inteligencia empresarial realizado a través de Power BI que integra las tecnologías de Microsoft Azure, creando así una plataforma potente y flexible para el análisis de datos.

- **Servidores en Azure:**

- Azure SQL Database: datos estructurados almacenados con facilidad, lo que facilita una rápida recuperación y uso analítico. Esta base de datos protege todos los detalles esenciales de la empresa, garantizando disponibilidad y escalabilidad.
- Azure Blob Storage: Utilizado para almacenar datos no estructurados, como archivos CSV que provienen de diversos orígenes, garantiza escalabilidad y durabilidad.
- Azure Virtual Machines (VMs): se utilizan para el funcionamiento de programas de procesamiento de datos, así como sus servicios de soporte relacionados. Estas máquinas ofrecen la agilidad de ajustar el tamaño de los recursos según la demanda de la carga de trabajo en las actividades de análisis y procesamiento de datos.

- **Conexiones de Red:**

- Redes Virtuales de Azure (VNet): se establecen conexiones entre otros servicios existentes dentro del ámbito de Azure. Este vínculo garantiza un flujo seguro y eficaz de canales de comunicación entre todas las diferentes partes que componen sus sistemas en Azure.
- VPN Gateway: una herramienta para crear un vínculo seguro entre la red local de la empresa y los recursos basados en Azure, garantizando que los datos se mantengan seguros y sin alteraciones cuando se transfieren.
- ExpressRoute: ofrece enlaces privados de alta velocidad que conectan la infraestructura local con los servicios de Azure, lo que reduce drásticamente los problemas de latencia que obstaculizan las aplicaciones críticas y mejora los niveles de rendimiento.



- **Dispositivos de Acceso para el Usuario:**

- PCs y Laptops: los analistas y empleados utilizan estos dispositivos para conectarse a informes y paneles de Power BI donde pueden trabajar activamente con los datos. Esto les permite tomar decisiones basadas en análisis en tiempo real que pueden realizar directamente a través de la interacción del panel.
- Tablets y Smartphones: estos dispositivos permiten a los usuarios acceder a paneles mientras están en movimiento, lo que proporciona una manera para que las personas accedan a información crítica desde cualquier ubicación. El objetivo es mejorar la flexibilidad en la toma de decisiones garantizando que las personas puedan obtener detalles incluso cuando no estén en la oficina.

El diseño de la arquitectura física del sistema pretende ser un lugar estable, seguro y eficiente para el trabajo de inteligencia de negocios. Los servicios de Azure garantizan que una organización que utiliza Power BI pueda obtener la mayor disponibilidad y seguridad que también le permitirá adaptarse a necesidades futuras. Esto asegura que el sistema sea capaz de gestionar eficazmente el análisis de cantidades significativas de datos; y además fomentar un mayor crecimiento dentro de la empresa.

4.1.2 Arquitectura lógica

La arquitectura lógica es cómo se muestran y manejan los datos. A continuación, detallare cada una de las capas.

Capa de Presentación:

- Power BI Desktop: crea informes y visualizaciones.
- Power BI Service: ayuda a publicar, compartir y administrar informes fácilmente.
- Acceso web: los informes se pueden encontrar en la página web descrita <https://iglobal-store.com/dashboard.html> lo que lo hace posible para que puedan verse sin ningún requisito de software específico.

Capa de Inserción de Datos:

- Carpeta en OneDrive: Aquí es donde se almacena el archivo principal (DataCo SupplyChain Dataset - TFG - Logistic Dashboard.xlsx) y el archivo de calendario (FiscalCalendar.xlsx). Sólo los desarrolladores pueden acceder a estos archivos.
- Carpeta en SharePoint: semanalmente, los analistas de la empresa tienen acceso al directorio de SharePoint. Sólo debe contener archivos en formato Excel que comiencen con "FY" (por ejemplo, FY18CW01.xlsx).



Capa de Procesamiento de Datos:

- Power Query: a través de un proceso conocido como Extracción y Transformación, asume la tarea de limpiar y transformar los datos extraídos de diversas fuentes. Esto incluye, entre otros, convertir formatos, eliminar errores y armonizar los nombres de las columnas con los del archivo principal. El objetivo de este proceso es tener todos los diferentes puntos de acceso representados en una tabla por Power Query; esto hace que los datos sean más uniformes y fáciles de gestionar.

Capa de Modelado de Datos:

- Modelo tabular en Power BI: el modelo tiene varias tablas de hechos (como FactSales, FactShipments, FactPayments) y tablas de dimensiones (incluidas DimCustomer, DimProduct, DimLocation, DimCalendar, DimOrderStatus). Estas tablas están organizadas de manera estructurada para facilitar el proceso de análisis de datos y generación de informes.

4.2 Diseño de datos

Aquí enfatizo primero la tabla principal del modelo de datos: "POWER BI DATA". Esta tabla ocupa una posición importante ya que todas las demás tablas de hechos y dimensiones se derivan de ella, lo que garantiza una organización estructurada y eficaz del flujo de datos. A continuación, se proporciona una descripción general de cada campo de esta tabla:

- Order Date (DATE ORDERS): Fecha en la que se realizó el pedido.
- Shipping Date (DATE ORDERS): Fecha exacta del envío.
- Benefit Per Order: Ganancias obtenidas por pedido.
- Category ID: Código de categoría del producto.
- Category Name: Descripción de la categoría del producto.
- Customer City: Ciudad del cliente.
- Customer Country: País del cliente.
- Customer First Name: Nombre del cliente.
- Customer ID: Identificador único del cliente.
- Customer Last Name: Apellido del cliente.
- Customer Segment: Tipo de cliente (Consumer, Corporate, Home Office).
- Customer State: Estado donde se encuentra la tienda que registra la compra.
- Customer Street: Calle donde se encuentra la tienda que registra la compra.
- Customer Zipcode: Código postal del cliente.
- Days for Shipment (SCHEDULED): Días programados para la entrega del producto comprado.
- Days for Shipping (REAL): Días reales de envío del producto comprado.
- Delivery Status: Estado de entrega de los pedidos (Advance Shipping, Late Delivery, Shipping Canceled, Shipping on time).
- Department ID: Código del departamento de la tienda.
- Department Name: Nombre del departamento de la tienda.
- Late Delivery Risk: Indica si el envío está atrasado (True) o no (False).
- Latitude: Latitud de la ubicación de la tienda.



- Longitude: Longitud de la ubicación de la tienda.
- Market: Región del mercado (Africa, Europe, LATAM, Pacific Asia, USCA).
- Order City: Ciudad destino del pedido.
- Order Country: País de destino del pedido.
- Order Customer ID: Código de pedido del cliente.
- Order ID: Código del pedido.
- Order Item Card Product ID: Código del producto generado a través del lector RFID.
- Order Item Discount: Valor del descuento por ítem de pedido.
- Order Item Discount Rate: Porcentaje de descuento por ítem del pedido.
- Order Item ID: Código del ítem del pedido.
- Order Item Product Price: Precio de los productos sin descuento.
- Order Item Profit Ratio: Ratio de beneficio por ítem de pedido.
- Order Item Quantity: Número de productos por pedido.
- Order Item Total: Cantidad total por pedido.
- Order Profit Per Order: Beneficio por pedido.
- Order Region: Región mundial donde se entrega el pedido.
- Order State: Estado de la región donde se entrega el pedido.
- Order Status: Estado del pedido (CANCELED, CLOSED, COMPLETE, ON_HOLD, PAYMENT_REVIEW, PENDING, PENDING_PAYMENT, PROCESSING, SUSPECTED_FRAUD).
- Order Time (DATE ORDERS): Hora en la que se realizó el pedido.
- Order Zipcode: Código postal donde se entrega el pedido.
- Product Card ID: Código del producto.
- Product Category ID: Código de categoría del producto.
- Product Description: Descripción del producto.
- Product Image: Enlace para la visita y compra del producto.
- Product Name: Nombre del producto.
- Product Price: Precio del producto.
- Product Status: Estado del stock del producto (Available, Unavailable).
- Sales: Valor de ventas.
- Sales per Customer: Ventas totales realizadas por cliente.
- Shipping Mode: Modo de envío del pedido (First Class, Same Day, Second Class, Standard Class).
- Shipping Time (DATE ORDERS): Hora exacta del envío.
- Type: Tipo de transacción realizada para el método de pago (DEBIT, CASH, PAYMENT, TRANSFER).



Aquí muestro un diagrama de la tabla “POWER BI DATA”, con el formato elegido para cada columna:

Diagrama de la Tabla Principal del Dataset	
Trabajo Fin de Grado Cuadro de Mando Para Empresa Logística DataCo SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS	
POWER BI DATA	
Order Date (DATE ORDERS)	DATE
Shipping Date (DATE ORDERS)	DATE
Benefit Per Order	DECIMAL(10, 2)
Category ID	INTEGER
Category Name	VARCHAR(100)
Customer City	VARCHAR(50)
Customer Country	VARCHAR(50)
Customer First Name	VARCHAR(50)
Customer ID	INTEGER
Customer Last Name	VARCHAR(50)
Customer Segment	ENUM('Consumer', 'Corporate', 'Home Office')
Customer State	VARCHAR(50)
Customer Street	VARCHAR(100)
Customer Zipcode	VARCHAR(10)
Days For Shipment (SCHEDULED)	INTEGER
Days For Shipping (REAL)	INTEGER
Delivery Status	ENUM('Advance Shipping', 'Late Delivery', 'Shipping Canceled', 'Shipping on Time')
Department ID	INTEGER
Department Name	VARCHAR(50)
Late Delivery Risk	BOOLEAN
Latitude	DECIMAL(10, 6)
Longitude	DECIMAL(10, 6)
Market	ENUM('Africa', 'Europe', 'LATAM', 'Pacific Asia', 'USCA')
Order City	VARCHAR(50)
Order Country	VARCHAR(50)
Order Customer ID	INTEGER
Order ID	INTEGER
Order Item Card Product ID	INTEGER
Order Item Discount	DECIMAL(10, 2)
Order Item Discount Rate	DECIMAL(5, 4)
Order Item ID	INTEGER
Order Item Product Price	DECIMAL(10, 2)
Order Item Profit Ratio	DECIMAL(5, 4)
Order Item Quantity	INTEGER
Order Item Total	INTEGER
Order Profit Per Order	DECIMAL(10, 2)
Order Region	VARCHAR(50)
Order State	VARCHAR(50)
Order Status	ENUM('CANCELED', 'CLOSED', 'COMPLETE', 'ON_HOLD', 'PAYMENT_REVIEW', 'PENDING', 'PENDING_PAYMENT', 'PROCESSING', 'SUSPECTED_FRAUD')
Order Time (DATE ORDERS)	TIME
Order Zipcode	VARCHAR(10)
Product Card ID	INTEGER
Product Category ID	INTEGER
Product Description	VARCHAR(255)
Product Image	VARCHAR(255)
Product Name	VARCHAR(100)
Product Price	DECIMAL(10, 2)
Product Status	ENUM('Available', 'Unavailable')
Sales	DECIMAL(15, 2)
Sales Per Customer	DECIMAL(15, 2)
Shipping Mode	ENUM('First Class', 'Same Day', 'Second Class', 'Standard Class')
Shipping Time (DATE ORDERS)	TIME
Type	ENUM('DEBIT', 'CASH', 'PAYMENT', 'TRANSFER')
Upload File	VARCHAR(255)

Ilustración 3: Diagrama de la Tabla Principal del Dataset.



Esta tabla se convierte en el corazón del análisis. Esto, a su vez, facilita el análisis y la visualización adicional en Power BI, lo que significa que no perdemos tiempo tratando de descubrir qué cuentan los datos de la historia, sino que me he centrado en elaborarlos creativamente en informes visuales con todos los conocimientos clave resaltados.

4.2.1 Migración y carga inicial de datos

El objetivo principal de este apartado es delinear el plan de migración y carga inicial de datos que garantice una fluida y correcta asimilación al sistema de Business Intelligence.

- **Proceso de migración y carga de datos inicial:**

La limpieza y transformación de datos fue la acción inicial que se tomó durante la migración. El conjunto de datos original en formato CSV que contenía delimitadores y formatos no admitidos por Power BI se convirtió a un formato apropiado (Excel) mediante un script de Python desarrollado en Kaggle. Esto implicó convertir los archivos al formato Excel y realizar modificaciones en los delimitadores y separadores decimales, ya que Power BI no admitía esos dos desde los archivos CSV descargados de Kaggle.

Habiendo ocurrido una metamorfosis, los datos llegan a Power BI, con un proceso ETL llevado a cabo a través de Power Query en el recinto de Power BI. En este proceso, los distintos flujos de datos encuentran su confluencia en una tabla singular: la tabla POWER BI DATA. Esta tabla adquiere una importancia fundamental ya que actúa como precursora de las tablas de hechos y de dimensiones. Por lo tanto, los datos se estructuran de manera lógica dentro de esta tabla para facilitar el análisis posterior.

- Relaciones entre tablas: el modelo de datos establece conexiones fundamentales entre las distintas tablas que permiten un escrutinio que es a la vez integral y multidimensional:
 - FactPayments y DimCustomer: se relacionan a través de la columna Order Customer ID, permitiendo así el análisis de las transacciones financieras por cliente.
 - FactSales y DimProduct: a través del ID de artículo de pedido, ambos están conectados, lo que facilita una forma sencilla de observar el rendimiento de productos individuales.
 - FactShipments y DimLocation: los detalles sobre la logística de envío están disponibles a través de su conexión a través de ID de pedido.

La integración y el análisis de datos dependen en gran medida de estas relaciones que a su vez ayudan a producir información valiosa para ayudar en la toma de decisiones estratégicas.



Considere el siguiente diagrama entidad-relación que se muestra a continuación, entre las tablas de hechos y dimensiones. Presenta una vista que ayudará a comprender la base estructural que sustenta las diversas operaciones analíticas analizadas en el cuadro de mando.

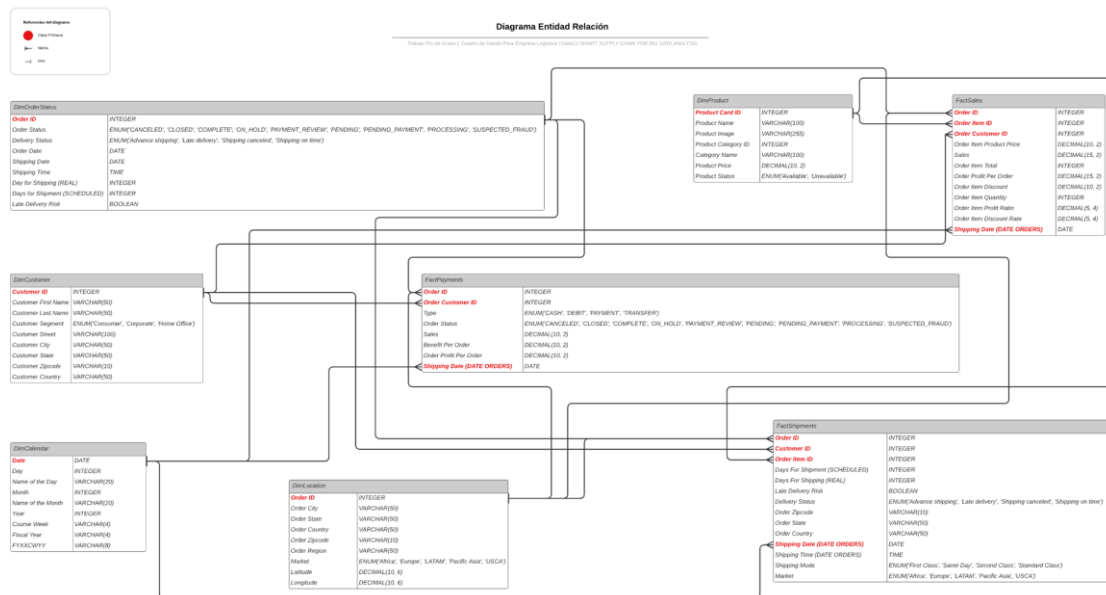


Ilustración 4: Diagrama de Entidad Relación.

La representación del modelo de datos de las relaciones de las tablas simplifica la comprensión de cómo están conectadas las tablas y la comprensión visual de la estructura de la base de datos a través de este diagrama.

Se puede acceder a los archivos CSV sin procesar obtenidos de Kaggle en el repositorio de GitHub para una fácil comprensión y revisión del procedimiento de limpieza de datos. Los dos conjuntos de archivos permiten un análisis en paralelo entre los datos no procesados y su contraparte limpia que se utilizó en el proyecto, mostrando el trabajo realizado en la normalización y limpieza.

Es un enfoque amplio hacia la migración y carga de datos que asegura al sistema de inteligencia empresarial las cualidades de ser sólido y confiable, garantizando así que sea capaz de proporcionar información confiable y oportuna que respalde la toma de decisiones comerciales efectiva.



4.3 Diseño de la interfaz de usuario

En esta sección me voy a encargar de dar la mejor perspectiva objetiva en relación a los cuadros de mando creados.

El panel "Descripción general" ofrece una instantánea de las principales métricas de desempeño de la empresa. Esto incluye:

- Total Customers: revela el recuento completo de clientes, lo cual es fundamental para evaluar el alcance y la expansión de la organización.
- Total Orders: representa la cantidad de pedidos realizados (un índice de negocio).
- Total Sales: es una representación visual de las ventas totales en euros; dando una visión fácilmente comprensible del desempeño financiero.
- New Customers per Month: muestra la cantidad de nuevos clientes obtenidos cada mes, lo que es útil para medir el crecimiento de la base de clientes.
- Revenue Grow Rate: visualización que muestra la tasa de aumento de los ingresos, y que permite una apreciación de la dirección de los mismos de manera ágil.
- Average Order Value by Year: gráfico de líneas que representa la media de los pedidos en los años, esto es útil para determinar comportamientos y transformaciones de compra.
- Product Variety by Category: un gráfico de barras que indica la diversidad de productos por categoría, esto le ayudará a ubicar los más comercializados.



El propósito de este panel es brindar una visión rápida y clara de los principales elementos del desempeño empresarial que le permitirán tomar decisiones sobre la marcha.

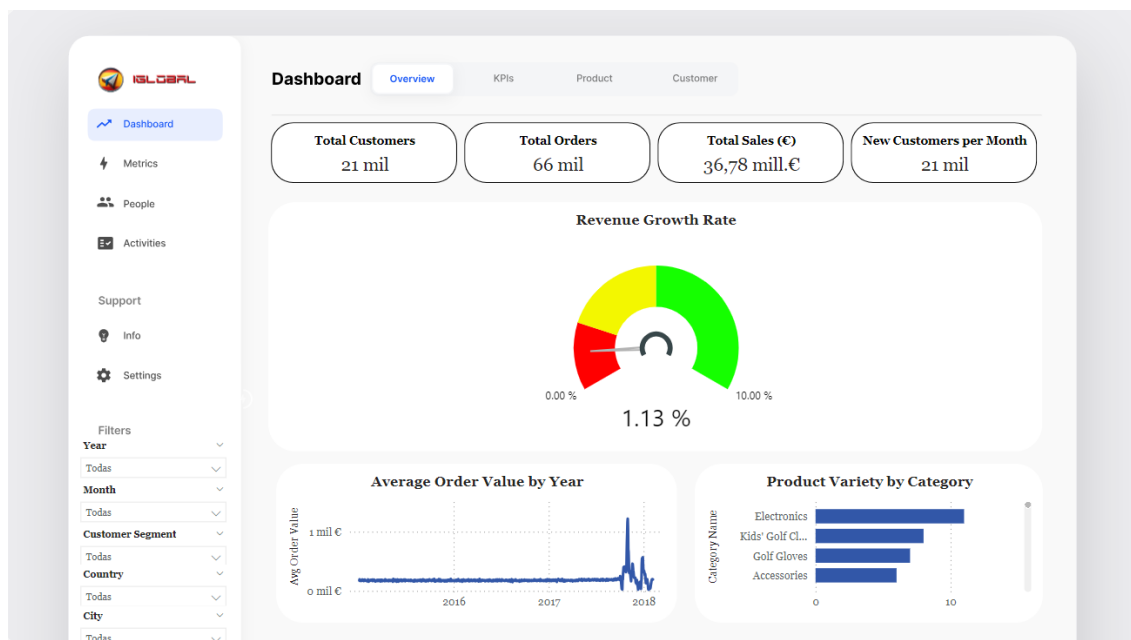


Ilustración 5: Cuadro de Mando Overview.

Una ilustración de algunas de las métricas esenciales para el desempeño de la empresa que el panel de KPI pone a la vista. Sin embargo, este panel no presenta todas las métricas, sino solo aquellas que son muy críticas, como el valor promedio del pedido, que muestra cuántos ingresos en promedio genera cada pedido y la tasa de entrega a tiempo, que representa cuántas entregas se realizan dentro del plazo estipulado. Además, se ilustra la eficiencia de cumplimiento (porcentaje de pedidos completados satisfactoriamente) y también se muestra la tasa de retención de clientes (porcentaje de clientes que realizan compras repetidas). El uso de gráficos y medidores interactivos para mostrar estos indicadores permite a los gerentes identificar rápidamente tendencias y problemas, lo que les permite tomar decisiones estratégicas informadas. Este enfoque tiene un doble impacto: mejora la capacidad de respuesta del mercado debido al fácil acceso diario a información tan vital y también fomenta la comprensión entre los diferentes departamentos sobre qué necesita más atención en función de las áreas prioritarias identificadas.



Ilustración 6: Cuadro de Mando KPI's (versión 1).

El tercer cuadro de mando se centra en mostrar las métricas clave de rendimiento. Beneficio total, orden. Tasa de cumplimiento, promedio Los descuentos son indicadores que pueden ayudar a los gerentes a ver cómo se están desempeñando las operaciones a lo largo del tiempo. Visualizar los ingresos semanales a través de gráficos de líneas ayuda a comparar las ventas de diferentes años por día, lo que permite identificar fácilmente tendencias y patrones en las ventas diarias. Esta información estratégica es vital para cualquier organización: garantiza que todas las decisiones tomadas en cualquier nivel (estratégico, táctico u operativo) estén bien informadas; facilitando así el aprovechamiento óptimo de los recursos para la gestión de la empresa.

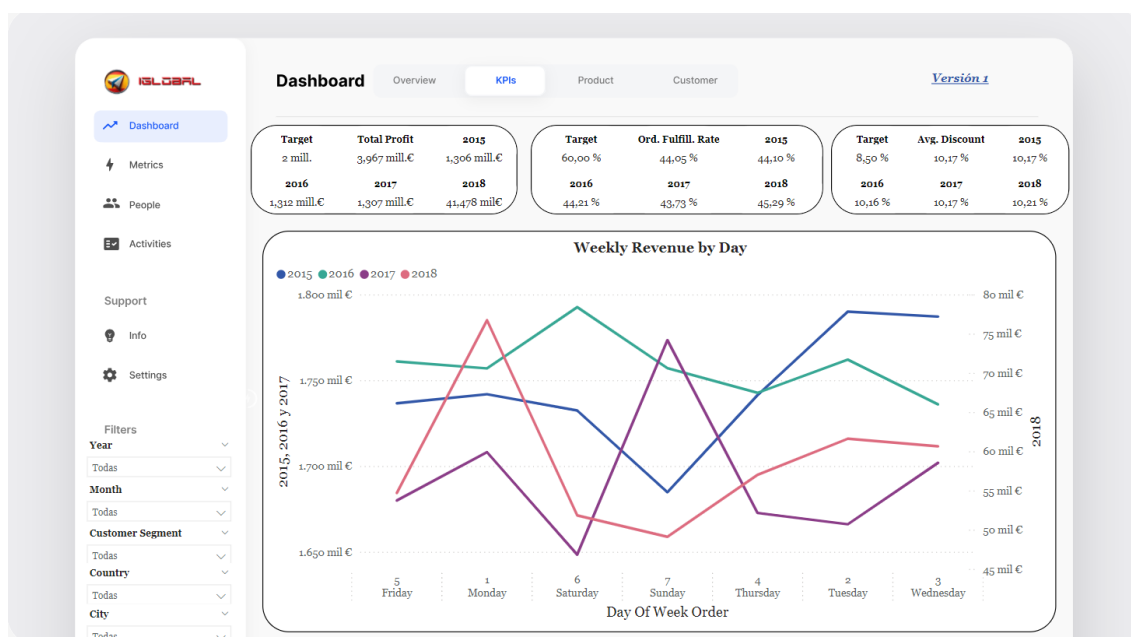


Ilustración 7: Cuadro de Mando KPI's (versión 2).



Este panel está centrado en el producto te brinda todos los detalles vitales que necesitas para un análisis de producto eficaz. También, el rendimiento de cada producto para que pueda tomar decisiones informadas basadas en datos reales. Algunas de las métricas mostradas incluyen el costo promedio por producto, el margen de beneficio y la cantidad vendida, ¡pero eso no es todo! Tomo en cuenta factores como las ventas mensuales por producto, lo que nos permite identificar los productos exitosos de aquellos que no tienen un buen desempeño. Luego se pueden establecer tendencias que conduzcan a puntos de acción apropiados, principalmente en la estrategia de ventas y en la gestión de inventario. Además de esto, no nos olvidemos de nuestros valiosos clientes: a través de agudas observaciones tanto de las tasas de devolución (la tasa a la que los clientes regresan después de realizar una compra) como del AOV (valor promedio del pedido), podemos obtener información sobre la lealtad del cliente, y qué prefieren comprar.

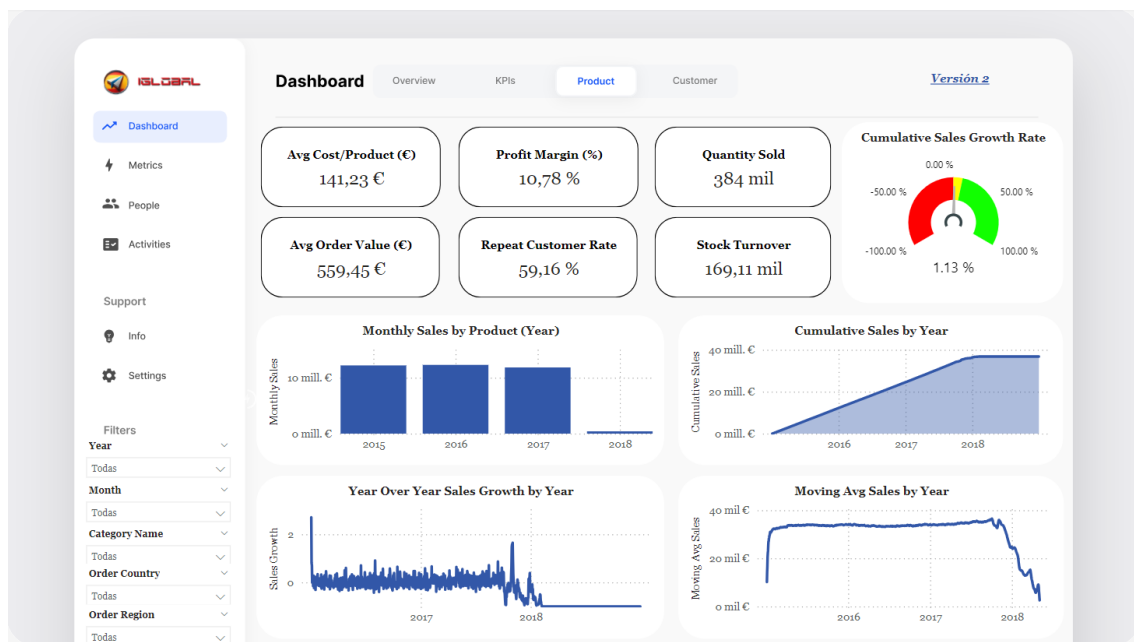


Ilustración 8: Cuadro de Mando Product (versión 1).

La sección "Producto" del panel ofrece una vista gráfica que proporciona información detallada sobre el rendimiento de los productos. Es muy útil para estar atento a las ventas por sus categorías y regiones, así como conocer las cifras actuales más el histórico. Los gráficos proporcionan información como el desglose de las ventas por categoría durante varios años y lo que constituye estas cifras: detalles como los productos más vendidos. Aparte de esto, también hay una representación geográfica que muestra dónde son populares ciertos productos, lo que puede ayudar a tomar decisiones sobre estrategias de marketing o distribución. La sección tiene como objetivo reconocer productos exitosos (a través de su desempeño) para que se puedan realizar los ajustes apropiados para maximizar la rentabilidad a partir de aquellas estrategias identificadas con indicadores de éxito.

Estos datos son esenciales en la toma de decisiones que ayuda a los gerentes a detectar patrones de ventas, evaluar el desempeño por región o producto e incluso realizar ajustes estratégicos rápidamente efectivos en base a informes en tiempo real para mejorar la eficacia operativa y aumentar los ingresos.

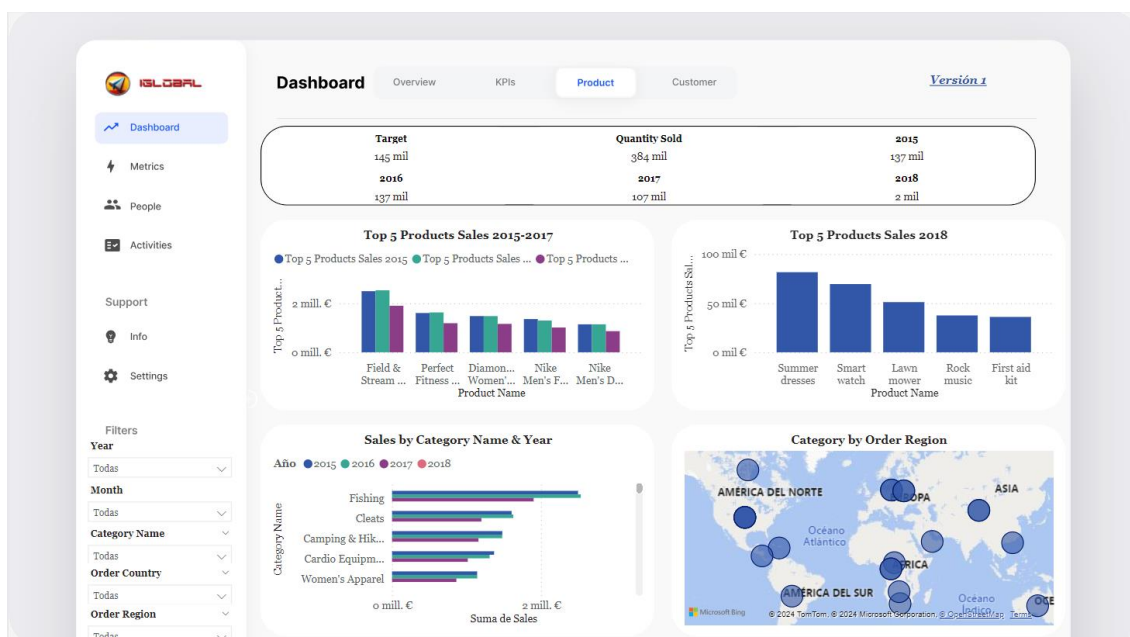


Ilustración 9: Cuadro de Mando Product (versión 2).

Este panel tiene como objetivo comprender a nuestros clientes. Muestra cifras importantes en la parte superior, como la frecuencia media de compra, que se sitúa impresionantemente en 8,74, el considerable número de clientes que regresan, fijado en 21.000, y sin olvidar el llamativo índice de fidelización del cliente valorado en la friolera de 4,21 millones de euros. Estas cifras dan una idea tanto de la lealtad de los clientes como de la frecuencia con la que realizan compras. Y no pasemos por alto la importancia del número revelado de nuevos clientes por año fiscal: esta métrica en particular nos ayuda a medir el crecimiento de nuestra base de clientes año tras año.

La parte central detalla el recuento de clientes de alto valor, que representa aquellos clientes que han realizado compras importantes. Por otro lado, las Ventas Promedio por Cliente dan una idea de cuánto vale en promedio cada cliente. La proporción de clientes que repiten compras, como lo muestra la Tasa de Fidelización del Cliente (59,16%), es otro aspecto importante para considerar al evaluar la satisfacción y la lealtad. Este valor se deriva de dividir una medida por otra, y aunque no siempre es perfecto o exacto a la hora de representar la realidad, puede ser muy revelador dependiendo de cómo se calcula: dos empresas diferentes pueden tener valores idénticos para estas dos medidas, pero llegan a ellos usando completamente diferentes fuentes de datos o metodologías.

La sección inferior revela las ventas totales por ID de cliente, una función que revela a sus principales clientes y su valor. También muestra la contribución a los ingresos por segmento de clientes (consumidores, empresas y oficinas centrales), que desempeña un papel fundamental a la hora de determinar los principales segmentos generadores de ingresos. Esto ayuda a idear estrategias de marketing enfocado. Recuerda esto cuando diseñes estrategias de marketing: apunta siempre a los segmentos de clientes más rentables.

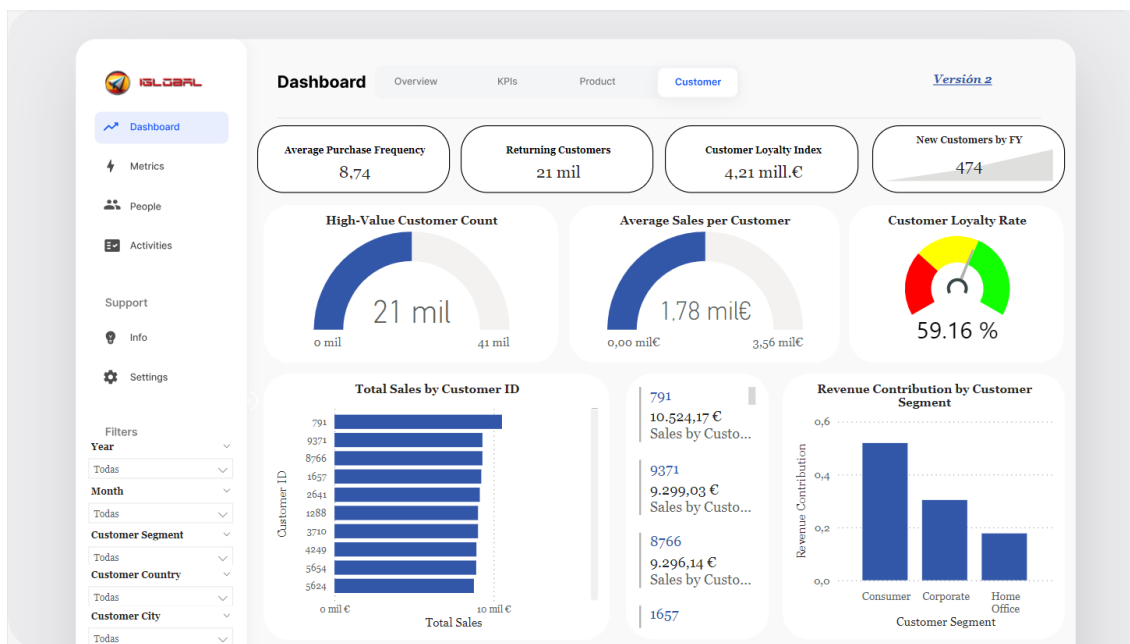


Ilustración 10: Cuadro de Mando Customer.

Una perspectiva detallada es lo que el panel de Clientes pretende ofrecer sobre el comportamiento y la contribución del cliente. Muestra dos puntos principales; la frecuencia promedio de compra, que permite evaluar la frecuencia con la que los clientes realizan compras, y el número de clientes recurrentes, que es un indicador clave para medir la lealtad y la efectividad de cualquier estrategia de retención implementada. El valor aportado por los clientes leales se puede entender a través del índice de fidelización de clientes, mientras que el análisis del gráfico que representa los nuevos clientes por año fiscal ayuda a medir el crecimiento de la base de clientes. Además, los clientes de alto valor se destacan junto con las ventas promedio por cliente: esto no solo da una idea sobre el gasto promedio, sino que también revela oportunidades sobre dónde se debe centrar la atención para aumentar los ingresos. Los clientes más valiosos se pueden identificar según la tasa de fidelidad del cliente y las ventas totales por ID de cliente; De manera similar, conocer la contribución a los ingresos de los diferentes segmentos de clientes proporciona información sobre qué segmento aporta más ingresos, lo que ayuda al marketing a priorizar los esfuerzos de manera más efectiva hacia ciertos segmentos.

Los datos que ofrecen estas métricas son esenciales para informar decisiones estratégicas sobre cómo mejorar la retención de clientes, fomentando la lealtad y optimizando las estrategias de marketing y ventas. Ser inteligente en la toma de decisiones implica confiar en la información correcta en el momento adecuado.

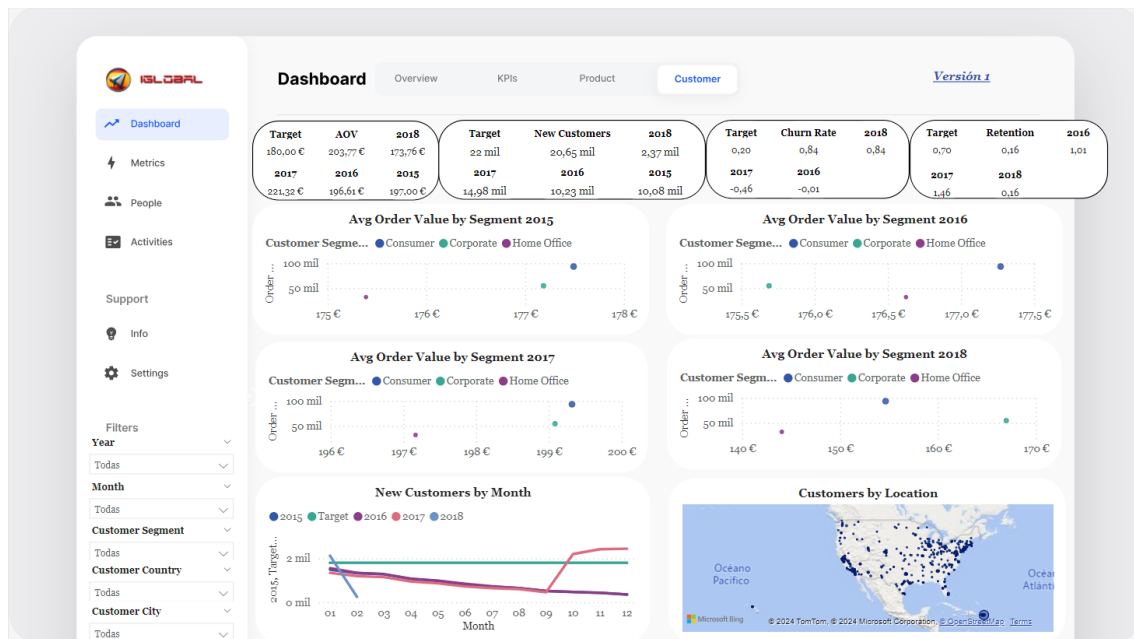


Ilustración 11: Cuadro de Mando Customer (versión 2).

Esta parte del panel de control ejecutivo que diseñamos para el director general muestra un factor importante para medir la efectividad operativa y la satisfacción del cliente: la Tasa de Cancelaciones. Además, el margen de beneficio bruto y los ingresos totales le muestran al director ejecutivo una imagen de la rentabilidad general de la empresa. También es posible ver otros indicadores clave, como las ventas totales por ID de cliente y el margen de beneficio promedio por producto, que ayudan a identificar los clientes más valiosos y los productos rentables. La presentación no termina aquí; Con el crecimiento de ventas anual junto con el CLV por ID de cliente y las cifras de crecimiento de ventas mensuales mostradas, resulta más fácil entender los indicadores de desempeño financiero y las tendencias de crecimiento a lo largo del tiempo.

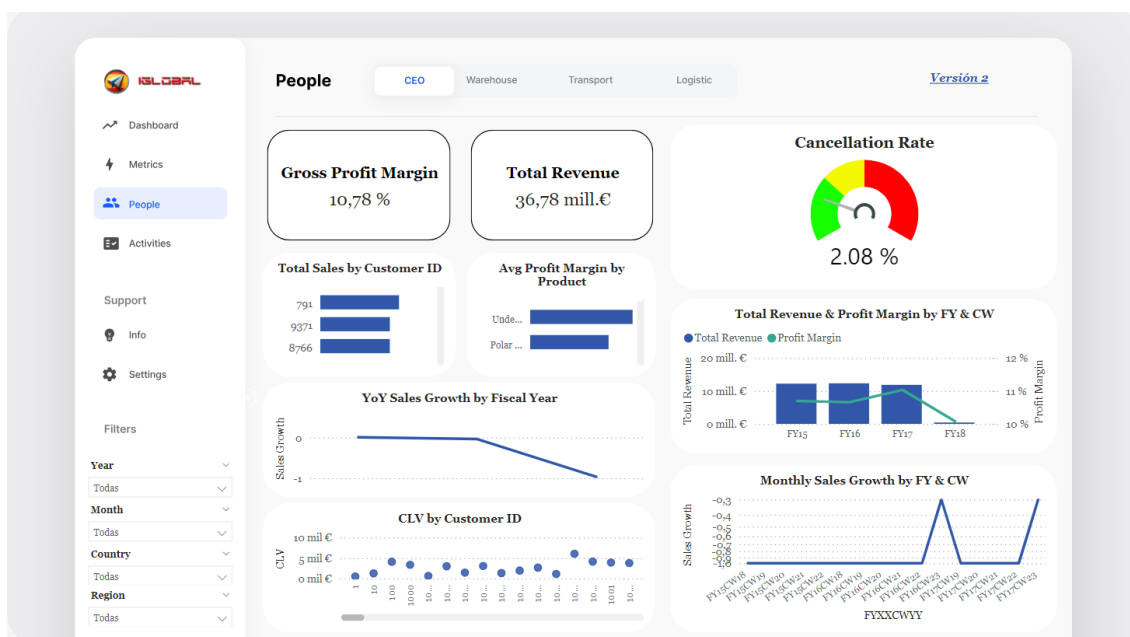


Ilustración 12: Cuadro de Mando CEO (versión 1).

Se ha construido un panel denominado "CEO" con un propósito específico: mostrar todas las áreas de productividad financiera y operativa de un vistazo. Significa que las métricas más importantes (como el margen de beneficio bruto y los ingresos totales) se encuentran en la parte superior. Esto permite al director de la empresa ver fácilmente si su negocio es rentable o no. En otras palabras, el indicador de tasa de cancelación proporciona información sobre cuánto tiempo se dedica al cumplimiento del pedido, si es efectivo o no.

Una representación visual de los datos que muestran las ventas totales según la identificación del cliente y el margen de beneficio promedio basado en los productos ayuda a identificar los clientes y productos más rentables. Por otro lado, el crecimiento de las ventas año tras año y el valor de vida del cliente (CLV) brindan una visión de la perspectiva a largo plazo, que es esencial para reconocer a los clientes valiosos junto con la tendencia de ventas continua.

En el ámbito de la toma de decisiones estratégicas, esta junta desempeña un papel indispensable: señalando qué es necesario mejorar y dónde se encuentran las oportunidades de crecimiento. La presentación de estadísticas específicas junto con gráficos visualmente atractivos anima a los usuarios a observar más de cerca el análisis de datos que sustenta la actividad operativa.



Ilustración 13: Cuadro de Mando CEO (versión 2).

El Warehouse Scorecard no se centra en todas las métricas, solo en las métricas clave del almacén. Brinda una perspectiva integral del desempeño de las operaciones logísticas. El tiempo promedio de envío y el promedio de días de retraso más la tasa de cumplimiento de pedidos y la tasa de éxito del envío son algunos de los números en este panel. Con estos indicadores, los gerentes de almacén pueden utilizar datos para tomar decisiones que mejoren la eficiencia operativa. Además, los detalles sobre la distribución estatal y la dispersión regional de los pedidos se muestran junto con los diferentes modos de envío utilizados, lo que ayuda a identificar áreas donde se puede mejorar el control de inventario y la logística. Poder ver los pedidos pendientes ayuda a priorizar las acciones y la asignación de recursos, lo que a su vez garantiza que la entrega de productos a los clientes sea eficiente y oportuna.

Las dos métricas desempeñan un papel importante en la comprensión y el ajuste del rendimiento del almacén para identificar áreas que pueden reducir los retrasos que conducen a una mayor precisión en la entrega de los pedidos. Al utilizar estos instrumentos, los responsables de logística pueden desarrollar decisiones basadas en análisis que conduzcan a una mejora tanto de la eficacia como de la satisfacción del cliente.

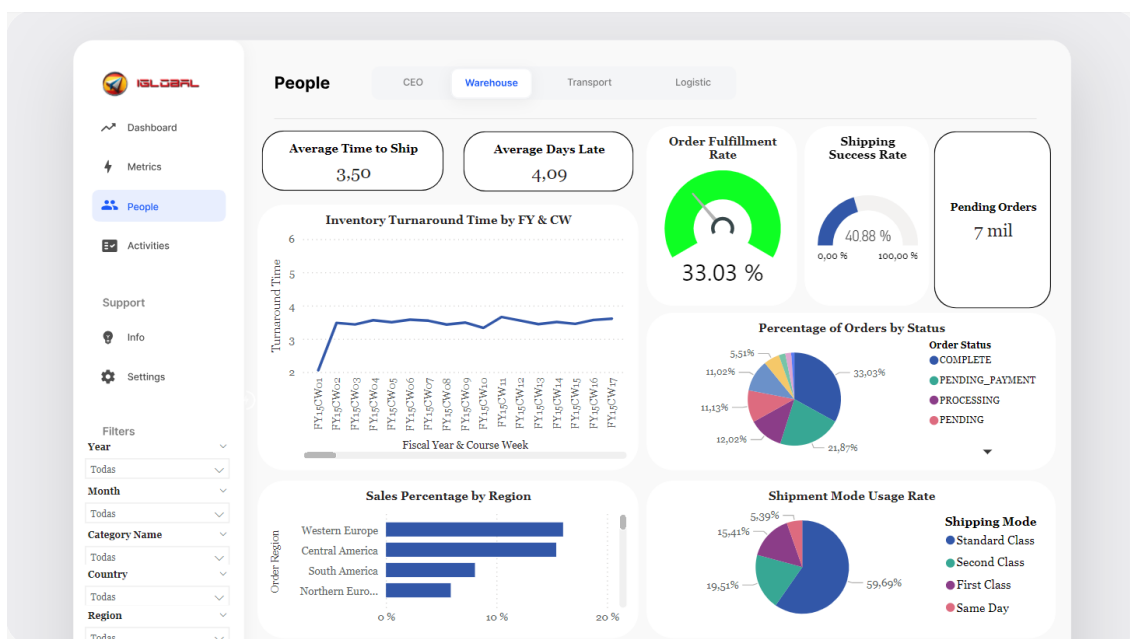


Ilustración 14: Cuadro de Mando Warehouse.

Este tablero está destinado a la supervisión y control de las operaciones de transporte de la empresa. Los indicadores clave de rendimiento, como el tiempo promedio de envío y el número de pedidos completados y cancelados, aparecen en la parte superior para facilitar una evaluación de la efectividad operativa. El gráfico de retrasos en las entregas por semana fiscal permite reconocer cualquier patrón cíclico o períodos comunes de retrasos; De manera similar, la segmentación geográfica representada a través del gráfico de distribución del mercado puede ayudar con las estrategias de optimización de rutas basadas en la asignación de recursos de ubicación. Además, ver el desempeño logístico a través de las tasas de retraso junto con la eficiencia por modo ofrece una perspectiva distinta que permite la acción gerencial para mejorar la puntualidad y la utilización de los recursos de transporte. A pesar de ser enfoques algo poco ortodoxos, estas métricas desempeñan un papel fundamental como señales que señalan áreas que merecen más atención y, en última instancia, apuntan a la satisfacción del cliente no solo porque son capaces de identificar un sistema de entrega mejorado en puntualidad sino también en eficiencia, sino a través de ellos.

Estas dos métricas son cruciales en el análisis y optimización de las operaciones de almacén para detectar áreas propensas a retrasos, lo que en última instancia fomenta una mayor precisión en la entrega de pedidos. Mediante el uso de dichas herramientas, los líderes de logística pueden diseñar estrategias basadas en datos: una práctica que aumenta significativamente la efectividad operativa y, a su vez, satisface las demandas de los clientes con facilidad.

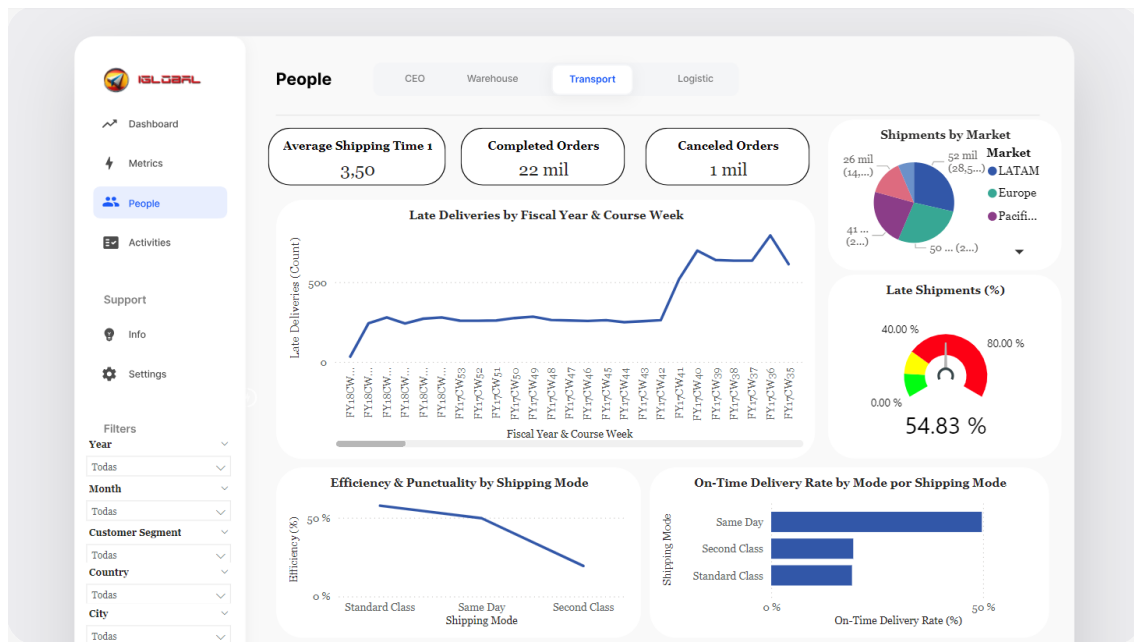


Ilustración 15: Cuadro de Mando Transport.

El “Dashboard Logistic” nace con un objetivo concreto: optimizar la eficiencia logística de la empresa. Destaca la tasa de cumplimiento y el costo proyectado por envíos tardíos junto con los pedidos pendientes de revisión de pago, una revelación sobre fallas operativas. Pero la cosa no termina ahí; la métrica Ventas totales por categoría de producto delinea las ventas según la contribución de cada categoría de producto, brindando una visión integral del desempeño. Además, la visualización de la eficiencia logística por producto asegura que las áreas que demandan mejoras en los procesos de envío estén debidamente anotadas. A través de la representación gráfica que representa los envíos críticos por región y el análisis de los envíos a lo largo del tiempo por año fiscal y semana del curso, podemos identificar fácilmente puntos críticos y reconocer patrones, lo que respalda la toma de decisiones estratégicas para mejorar nuestra eficacia logística.

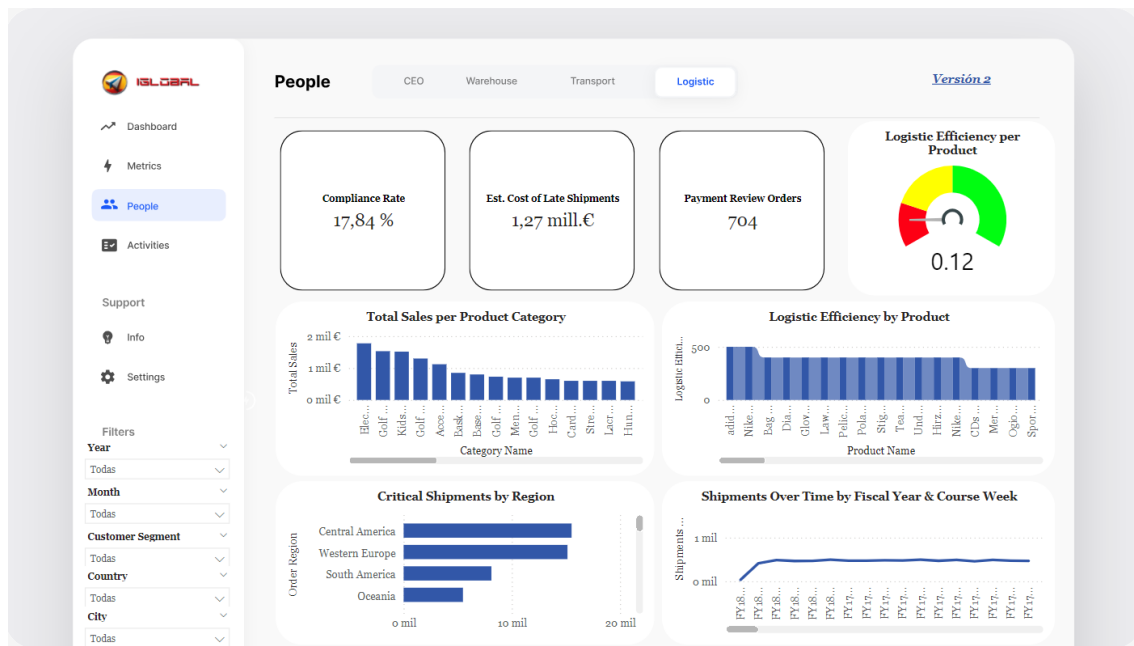


Ilustración 16: Cuadro de Mando Logistic (versión 1).

Proporciona información importante para la gestión logística. Entre ellos, se encuentran la Tasa de Cumplimiento, el Costo Estimado de Envíos Tardíos y los Pedidos en Revisión de Pago. Además de las Ventas totales por categoría de producto que puede ver, también eche un vistazo a la Eficiencia logística por producto: rendimiento y costos vinculados a los envíos. El gráfico "Envíos críticos por región" señala áreas problemáticas según el historial, mientras que "Envíos a lo largo del tiempo" rastrea las tendencias temporales; ambos útiles para la toma de decisiones sobre optimización de inventario o mejora de procesos y reducción de costos operativos. Información: útil para la gestión logística.

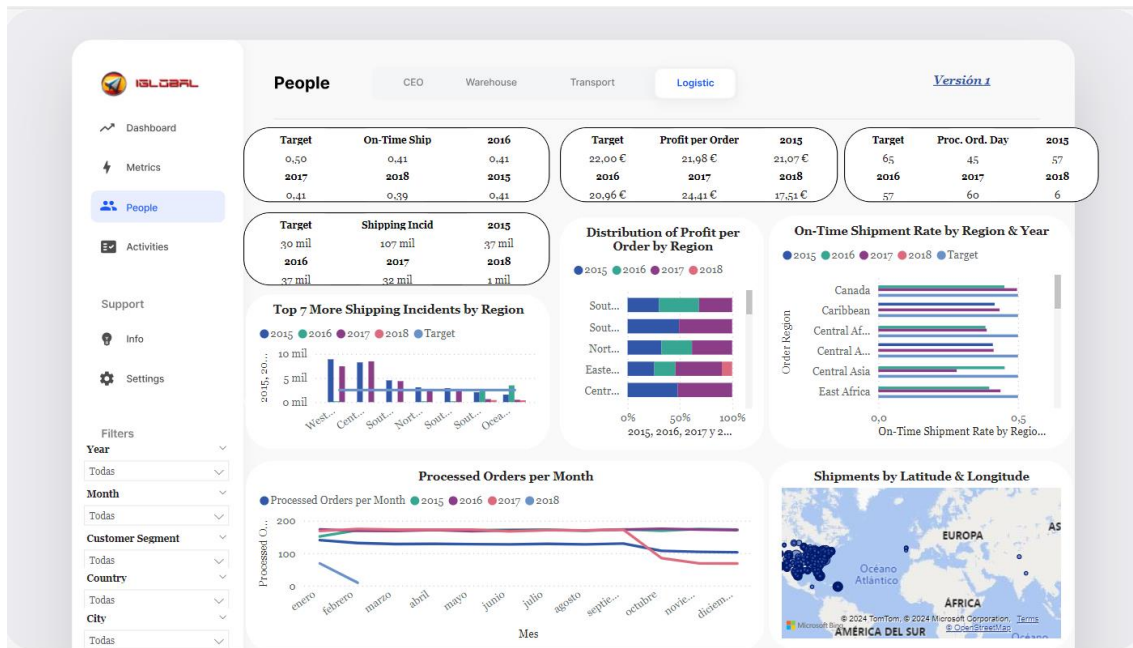


Ilustración 17: Cuadro de Mando Logistic (versión 2).

4.4 Diagrama de clases

En este segmento, voy a presentar el diagrama de clases del sistema e ilustrare el marco orientado a objetos que se adoptó para el desarrollo del proyecto. Aunque el proceso de desarrollo se ha centrado principalmente en un enfoque basado en la manipulación y el análisis de datos, el diagrama de clases aún ofrece una idea de cómo se estructuran varias entidades y sus interrelaciones dentro del sistema sin comprometer la calidad debido a compromisos con otros métodos o estrategias.

Esta representación visual ayuda enormemente a comprender la lógica interna y cómo los distintos componentes interactúan entre sí.

Clases principales del Sistema:

- **Clase PowerBiData:** descrita al comienzo del apartado de diseño de datos.
- **Clase FactSales:**
 - OrderID
 - OrderItemID
 - OrderCustomerID
 - OrderItemProductPrice
 - Sales
 - OrderItemTotal
 - OrderProfitPerOrder
 - OrderItemDiscount
 - OrderItemQuantity
 - OrderItemProfitRatio



- OrderItemDiscountRate
- ShippingDate

- **Clase FactShipments:**

- OrderID
- CustomerID
- OrderItemID
- DaysForShipmentScheduled
- DaysForShippingReal
- LateDeliveryRisk
- DeliveryStatus
- OrderZipcode
- OrderState
- OrderCountry
- ShippingDate
- ShippingTime
- ShippingMode
- Market

- **Clase FactPayments:**

- OrderID
- OrderCustomerID
- Type
- OrderStatus
- Sales
- BenefitPerOrder
- OrderProfitPerOrder
- ShippingDate

- **Clase DimCustomer:**

- CustomerID
- CustomerFirstName
- CustomerLastName
- CustomerSegment
- CustomerStreet
- CustomerCity
- CustomerState
- CustomerZipcode
- CustomerCountry

- **Clase DimProduct:**

- ProductCardID
- ProductName
- ProductImage
- ProductCategoryID
- CategoryName
- ProductPrice
- ProductStatus



- **Clase DimLocation:**

- OrderID
- OrderCity
- OrderState
- OrderCountry
- OrderZipcode
- OrderRegion
- Market
- Latitude
- Longitude

- **Clase DimCalendar:**

- Date
- Day
- NameOfTheDay
- Month
- NameOfTheMonth
- Year
- CourseWeek
- FiscalYear
- FYXXCWYY

- **Clase DimOrderStatus:**

- OrderID
- OrderStatus
- DeliveryStatus
- OrderDate
- ShippingDate
- ShippingTime
- DaysForShippingReal
- DaysForShipmentScheduled
- LateDeliveryRisk



4.5 Entorno de construcción

En esta parte ilustro las herramientas y tecnologías para poder llevar a cabo el proyecto, lo he separado en distintas secciones.

IDE y herramientas de desarrollo:

- Script de Python: creados con precisión utilizando Visual Studio Code para el aspecto de codificación y al mismo tiempo editando archivos de datos.
- Power BI Desktop: su aplicación preferida para crear paneles visualmente atractivos e informes interactivos que cautiven a su audiencia.

Marcos y bibliotecas:

- Pandas: para tareas de limpieza y manipulación de datos dentro de scripts de Python.
- NumPy: profundice en el ámbito de las operaciones numéricas avanzadas.
- OpenPyXL: una biblioteca de Python que permite la manipulación de archivos de Excel.

Servicios y Plataformas:

- OneDrive y SharePoint: Puntos de acceso de carga y almacenamiento de datos.
- Kaggle: fuente principal de datos del proyecto.

Gestión de repositorios y versiones:

- GitHub: para el repositorio con todos lo relevante a la documentación de mi TFG.

Estos componentes y herramientas fueron clave para el desarrollo y la implementación del sistema de inteligencia empresarial.

4.6 Referencia al repositorio de software

<https://github.com/imartinpena/TFG-LogisticsDashboard-PowerBI.git>



5. Validación del sistema

5.1 Plan de pruebas

Para garantizar la operatividad y confiabilidad del sistema creado para este Trabajo Fin de Grado, he llevado a cabo un exhaustivo plan de pruebas que abarcó pruebas unitarias, pruebas de integración y pruebas de usuarios del mundo real.

Durante las pruebas unitarias, cada módulo fue evaluado individualmente. Los scripts de Python y las transformaciones de Power Query se sometieron a pruebas para validar el funcionamiento de cada componente de forma aislada.

La correcta transformación de archivos CSV a Excel se ha verificado utilizando varias configuraciones de datos en Python Script. De manera similar, se han probado la limpieza de datos y la división de columnas a través de Power Query.

En relación con las pruebas de integración, ver que todos los componentes del sistema funcionan juntos como se esperaba. Por ejemplo, en las pruebas de integración de Power BI me asegure de que los datos que se introducían en él salían después de que se lleven a cabo los procesos de extracción, transformación y carga adecuados, incluida la verificación de si se producen correctamente desde diferentes fuentes como OneDrive o SharePoint.

Pruebas de usuario: se realizaron en colaboración con personas con las que solía trabajar durante mis prácticas universitarias, que son analistas de datos. He dividido estas pruebas en dos fases: una primera fase que involucra a usuarios elegidos y una segunda fase que involucra a un grupo más grande de usuarios.

Lo que se hizo fue detectar problemas tempranos y obtener retroalimentación inicial.

Verifique la estabilidad y rendimiento del sistema en condiciones realistas. Seguidamente implemente pruebas de carga y estrés en el sistema simulando un gran volumen de actividad del usuario. Finalmente, encontré algunos cuellos de botella que de otro modo no serían evidentes durante la operación normal.

El sistema fue sometido a pruebas de carga y estrés. Esto fue para garantizar que se pudieran procesar grandes cantidades de datos junto con muchos usuarios al mismo tiempo sin ningún problema en el rendimiento.



Simulación de varios usuarios: se simuló la generación de tráfico para varios usuarios que accedían a datos en Power BI.

Se hizo un seguimiento de métricas de rendimiento que incluyen métricas de rendimiento de monitoreo, como tiempos de respuesta y uso de recursos durante la prueba.

El plan de pruebas que se llevó a cabo tenía como objetivo garantizar que el sistema cumpla con todos los requisitos funcionales y no funcionales. También garantizo una experiencia de usuario sólida y eficiente. La combinación de pruebas unitarias, de integración y de usuario (además de las pruebas de carga y estrés) ha permitido que cualquier problema descubierto durante la fase de implementación del problema antes de llegar al final. Es así como he podido asegurar la calidad del sistema desarrollado con total confianza.



6. Conclusiones y líneas futuras

6.1 Conclusiones

Durante la elaboración de este apasionante TFG ha sido posible la implementación de una solución sólida y eficiente para la gestión de datos logísticos mediante Power BI. Este trabajo me ha permitido poner en práctica todos los conocimientos que he ido adquiriendo a lo largo de la carrera, y poder conocer otras tecnologías y metodologías.

El objetivo principal era mejorar la toma de decisiones dentro de una empresa de logística, lograda mediante la utilización de herramientas de inteligencia empresarial (BI). Se hizo posible fusionar diversas fuentes de datos, limpiar y transmutar los detalles, y luego exhibirlos de una manera inequívoca y obtenible con paneles interactivos, parte integral de lo cual es la automatización del proceso de extracción en Power BI: un factor considerable para reducir la cantidad, del tiempo y la energía necesarios para los esfuerzos de presentación de informes.

Durante el proceso de curación de los datos de Kaggle surgió un obstáculo de considerable magnitud. El desarrollo y la integración de un script de Python diseñado para moldear estos datos sin procesar en un formato funcional ha desempeñado un papel fundamental a la hora de dirigir el proyecto hacia el éxito. Además, la delimitación en la composición estructural de los nodos de acceso a datos dentro de OneDrive y SharePoint ha facilitado una administración eficaz e inexpugnable de la información. Esto garantiza la accesibilidad, así como la actualidad y exactitud de los datos de forma perpetua. Los deportes, al igual que los juegos o las artes, muchas veces nos ayudan a escapar de la realidad; aunque en ocasiones utilizamos el deporte para mejorar nuestras habilidades o destrezas para otros ámbitos de la vida.

La adopción de Power BI es un poderoso instrumento probado para visualizar y analizar datos, una tarea que comúnmente realizan los gerentes.



6.2 Líneas futuras

Este proyecto me ha expuesto a varias áreas donde podemos desarrollar y ampliar los sistemas futuros. A continuación, se muestran algunas posibles líneas de trabajo futuras:

- **Automatización Total del Proceso ETL:** incluir más herramientas en el sistema existente para una automatización completa de las tareas de Extracción, Transformación y Carga (ETL). Esto reducirá significativamente los esfuerzos de intervención manual y eliminará errores humanos importantes.
- **Integración de inteligencia artificial y aprendizaje automático:** desarrollar modelos basados en la predicción a través del aprendizaje automático con el fin de realizar predicciones futuras: esto incluye, entre otros, la logística (por ejemplo, previsión de la demanda y determinación de la ruta de entrega óptima).
- **Experiencia de usuario mejorada:** revolucionar el nivel de diseño y personalización de la interfaz de usuario de Power BI; Tal revolución facilitaría que los usuarios tengan una mejor interacción con los datos y adquieran detalles más granulares.
- **Unidad con los Sistemas ERP:** mejorar aún más la interconectividad del sistema mediante la fusión con otras plataformas corporativas, en particular los sistemas de planificación de recursos empresariales (ERP), para lograr una supervisión integral de las empresas comerciales.
- **Seguridad y Cumplimiento:** mejorar aún más los niveles de seguridad y cumplimiento para garantizar que la información manejada esté en línea con la actual normativa de protección de datos y leyes sobre privacidad, como el RGPD.
- **Análisis en tiempo real:** desarrollar funciones de generación de informes avanzadas para mejorar las capacidades de Power BI para soluciones de análisis de datos en tiempo real que permitan a las empresas responder mejor a los cambios en sus cadenas de suministro y eventos inesperados.

Por último, tener en cuenta el desarrollo de programas educativos y de instrucción para el personal de su empresa para garantizar que haga el máximo uso y alcance el mayor nivel de productividad de esta solución de BI después de su implementación.



7. Bibliografía

- [1] Kaggle. (2022). DataCo Supply Chain Dataset. Retrieved from <https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis>
- [2] Microsoft Power BI Documentation. (2023). Getting started with Power BI. Retrieved from <https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview>
- [3] Python Software Foundation. (2023). Python 3.10 Documentation. Retrieved from <https://docs.python.org/3.10/>
- [4] Pandas Development Team. (2023). Pandas Documentation. Retrieved from <https://pandas.pydata.org/pandas-docs/stable/>
- [5] SharePoint Documentation. (2023). Getting Started with SharePoint. Retrieved from <https://docs.microsoft.com/en-us/sharepoint/introduction>
- [6] OneDrive Documentation. (2023). OneDrive User Guide. Retrieved from <https://docs.microsoft.com/en-us/onedrive/user-guide>
- [7] Azure Blob Storage Documentation. (2023). Introduction to Azure Blob Storage. Retrieved from <https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction>
- [8] Jupyter Notebook Documentation. (2023). Using Jupyter Notebooks. Retrieved from <https://jupyter.org/documentation>
- [9] Postman Documentation. (2023). API Testing with Postman. Retrieved from <https://learning.postman.com/docs/getting-started/introduction/>
- [10] GitHub Documentation. (2023). GitHub Repositories. Retrieved from <https://docs.github.com/en/repositories>



Anexo I – Manuales

Estos manuales son imprescindibles para la correcta instalación y uso del sistema creado en este TFG. Puede encontrar estos manuales también disponibles en el archivo README.md ubicado en nuestro repositorio de GitHub para obtener más detalles.

Manual de instalación:

- Requisitos previos:
 - Python 3.7 o superior.
 - Librería Pandas.
 - Acceso a OneDrive y SharePoint (para ello tiene que ponerse en contacto conmigo por si quisiera comprobarlo, les informe del problema al departamento de IT pero me dijeron que tenían que recurrir a mí para que les diera acceso, igual si quiere ver el Power BI sin descargarse la aplicación).
- Escritorio Power BI:
 - Descargue e instale Power BI Desktop desde el sitio web oficial de Microsoft.
- Instalación de Python y Pandas:
 - Obtenga Python descargándolo de python.org.
 - Puedes instalar Pandas ejecutando “pip install pandas”.
- Configuración del entorno de desarrollo:
 - Recuperar el repositorio de GitHub clonándolo.
 - Encuentre el camino al directorio del proyecto.
 - Establecer un entorno virtual, así como activarlo.

Actualización del Power BI:

- Inicie Power BI Desktop, pero recuerde mantener actualizadas las conexiones entre los archivos almacenados en OneDrive y los ubicados en SharePoint después de una revisión reciente.
- Asegúrese de que la integración de datos se haya realizado correctamente y dé el siguiente paso para actualizar los paneles.



Manual de usuario:

- **Acceso a los paneles:**
 - Power BI Desktop: acceda a los informes directamente desde Power BI Desktop.
 - Enlace web: Abra los informes mediante el enlace generado por Power BI.
 - Página web: vea e interactúe con paneles desde nuestra página (<https://iglobal-store.com/dashboard.html>).
- **Navegación y uso de informes:** para utilizar eficazmente los informes, siga estos pasos.
 - Acceda a las distintas secciones del panel como Descripción general, KPI, Producto, Cliente, etc.
 - Personaliza cómo visualizas los datos mediante el uso de filtros y segmentaciones que están a tu disposición.
 - Cuando sea necesario para informes o presentaciones, extraer visualizaciones junto con los datos.



Glosario de términos

Power BI: esta le permite ver sus datos en colores vivos. Más que eso, le permite compartir información a través de informes y paneles con los que otros pueden interactuar.

ETL: implica extraer, transformar y cargar datos de diferentes orígenes en un sistema de almacenamiento como un almacén de datos.

OneDrive: es un servicio en la nube de Microsoft que permite a los usuarios almacenar y sincronizar archivos entre dispositivos de forma segura.

SharePoint: es otra oferta de Microsoft sirve como plataforma de colaboración dentro de las organizaciones, ayudando a gestionar y compartir información de forma eficaz.

Power Query: es un producto de Microsoft que permite a los usuarios conectar, fusionar y limpiar datos de diferentes orígenes mientras trabajan en Power BI.

Azure SQL Database: es un servicio de base de datos relacional dentro de la nube de Microsoft Azure, una opción para el almacenamiento y la administración de datos.

Azure Blob Storage: es un servicio de almacenamiento de objetos proporcionado por la nube de Microsoft Azure para almacenar cantidades masivas de datos no estructurados.

SAP: significa Sistemas, Aplicaciones y Productos en Procesamiento de Datos: es un sistema de planificación de recursos empresariales que integra varios módulos involucrados en las operaciones comerciales.

API (Interfaz de programación de aplicaciones): puede describirse como un conjunto de definiciones y protocolos que permiten una integración y comunicación perfecta entre diversos sistemas y aplicaciones.

Script: Programa formado por una serie de instrucciones que facilitan tareas de automatización dentro de un entorno informático.

CSV (valores separados por comas): es un formato de archivo diseñado para almacenar datos tabulares en texto sin formato. Cada línea denota un registro con valores separados por comas.

Excel: un software de hoja de cálculo de Microsoft que se emplea principalmente con fines de análisis y visualización de datos.

Python: es más que un simple lenguaje de programación para escribir scripts y desarrollar aplicaciones de análisis de datos.

On-Premise: se refiere a una infraestructura de TI que se encuentra físicamente en el sitio de la propia organización, a diferencia de los servicios en la nube.