



***Facultad
de
Ciencias***

**DISEÑO E IMPLEMENTACIÓN DE UNA
SOLUCIÓN BI PARA EL SECTOR
LOGÍSTICO**

**Design and implementation of a BI solution for
the logistics sector**

**Trabajo de Fin de Grado
para acceder al**

GRADO EN INGENIERÍA INFORMÁTICA

Autor: Nicolás Amigo Sañudo

Directora: Marta Elena Zorrilla Pantaleón

Junio – 2022

Agradecimientos

En primer lugar, me gustaría agradecer a la compañía Fagor Smart Data su colaboración con el Trabajo Fin de Grado, especialmente a D. ^a Rosa María Martínez y D. David Castanedo Santos. Su compromiso ha sido determinante en el desarrollo del proyecto.

Continúo dando las gracias a Dra. Marta Elena Zorrilla Pantaleón, directora del TFG. Este trabajo no habría sido posible sin su incansable dedicación y labor realizada a lo largo de los últimos meses. Su pasión por los datos supone una referencia muy importante para los objetivos que pretendo cumplir en un futuro.

Gracias también a todo el personal docente del Grado en Ingeniería Informática por contribuir positivamente en mi formación y a mis compañeros que han hecho de estos últimos cuatro años una experiencia para toda la vida.

Por último, no me olvido de mi familia que me enseñaron que con esfuerzo y sacrificio se pueden lograr todas las metas que se propongan. Muchas gracias por educarme en estos valores.

Resumen

En la actualidad, existe un conjunto de infraestructuras y tecnologías destinado al análisis de los datos que permite obtener ágilmente los indicadores clave para las organizaciones (KPI). De esta forma, se optimiza la toma de decisiones que llevan a cabo los directivos de las mismas. Este concepto recibe el nombre de *Business Intelligence*.

El objeto del Trabajo de Fin de Grado es el desarrollo de una solución de inteligencia de negocio, basada en el diseño e implementación de un almacén de datos y un cuadro de mandos para una empresa del sector logístico. Entre los KPI se abordan aquellos relativos al mantenimiento y registro de fallos de la flota de vehículos, vida útil del vehículo e informes de conducción. Asimismo, se realiza un análisis de los accesos de los usuarios a la aplicación que actúa como sistema de información.

El *Data Warehouse* se alimenta a partir de la base de datos rut-IA alojada en los servidores de la compañía en el motor relacional SQL Server. Además, diversos ficheros en formato CSV son utilizados con la finalidad de aportar información contextual, así como varias API de geolocalización.

La solución BI se implementa con las herramientas que Microsoft SQL Server proporciona para el almacenamiento, migración y procesamiento de datos. Para la construcción de los informes, se hace uso de Microsoft PowerBI dado que facilita visualizaciones interactivas y otras capacidades de inteligencia empresarial.

Esta solución busca solventar las limitaciones que presenta el sistema de información actual. Estas se resumen principalmente en problemas de rendimiento para la explotación de grandes volúmenes de datos. El sistema desarrollado resuelve satisfactoriamente los requisitos especificados ofreciendo consultas ágiles sin ningún tipo de restricción en el filtrado de información.

Palabras clave

Almacén de datos, cuadro de mandos, análisis de datos, inteligencia de negocio.

Abstract

Currently, there is a set of infrastructures and technologies for data analysis that allows organizations to quickly obtain key performance indicators (KPI). In this way, the decision-making conducted by their managers is optimized. This concept is called *Business Intelligence*.

The purpose of the Final Degree Project is the development of a business intelligence solution, based on the design and implementation of a Data Warehouse and a dashboard for a company in the logistics sector. Among the KPIs are those related to the maintenance and registration of failures of the vehicle fleet, useful life of the vehicle and driving reports. Likewise, an analysis of user access to the application that acts as an information system is performed.

The Data Warehouse is supplied from the rut-IA database hosted on the company's servers in the SQL Server relational engine. In addition, various files in CSV format are used to provide contextual information, as well as several geolocation APIs.

The BI solution is implemented with the tools that Microsoft SQL Server provides for data storage, migration, and processing. For the construction of the reports, Microsoft PowerBI is used since it facilitates interactive visualizations and other business intelligence capabilities.

This solution seeks to solve the limitations presented by the current information system. These are mainly summarized in performance problems for the exploitation of large volumes of data. The developed system satisfactorily solves the specified requirements, offering agile queries without any type of restriction in filtering information.

Keywords

Data warehouse, dashboard, data analytics, business intelligence.

Tabla de contenido

1	Introducción	1
1.1	Contexto	1
1.2	Objeto	1
2	Planificación y desarrollo del proyecto	2
3	Business Intelligence	3
3.1	Introducción	3
3.2	Data Warehouse	4
3.2.1	Data Sources	5
3.2.2	Data Staging Area	5
3.2.3	Data Area	5
3.2.4	Data Access Tools	7
3.3	Tecnologías	7
3.3.1	Microsoft SQL Server	8
3.3.2	Microsoft PowerBI	8
3.3.3	Google Geocoding API	9
4	Diseño e implementación de la solución BI	9
4.1	Arquitectura de la solución	9
4.2	Criterios para la interfaz gráfica del cuadro de mandos	10
4.2.1	Color y tipo de letra	10
4.2.2	Portada	11
4.2.3	Diseño de los informes	11
4.3	Caso de uso: accesos	12
4.3.1	Especificación de requisitos	12
4.3.2	Fuentes externas	13
4.3.3	Base de datos <i>Data Stage Area</i>	14
4.3.4	Base de datos <i>Data Area</i>	14
4.3.5	Cubos OLAP	18
4.3.6	Procesos ETL	19
4.3.7	Informes	22
4.4	Caso de uso: fallos	25
4.4.1	Especificación de requisitos	25
4.4.2	Fuentes externas	25
4.4.3	Base de datos <i>Data Stage Area</i>	26
4.4.4	Base de datos <i>Data Area</i>	26

4.4.5 Cubos OLAP	28
4.4.6 Procesos ETL	29
4.4.7 Informes	31
4.5 Caso de uso: trayectos.....	33
4.5.1 Especificación de requisitos.....	33
4.5.2 Fuentes externas	34
4.5.3 Transformación de datos.....	35
4.5.4 Informes	35
5 Validación del sistema y satisfacción del cliente	38
6 Conclusiones.....	39
7 Líneas futuras	40
Bibliografía.....	41
Anexo.....	42
Informes del caso de uso: accesos	42
Informes del caso de uso: fallos	44
Informes del caso de uso: trayectos.....	45

Tabla de figuras

Figura 1. Calendario de desarrollo del TFG	2
Figura 2. Estructura de un esquema en estrella	6
Figura 3. Ejemplo de un cubo OLAP	7
Figura 4. Componentes de la arquitectura <i>BI</i>	10
Figura 5. Colores de la interfaz gráfica	11
Figura 6. Portada del cuadro de mandos.....	11
Figura 7. Diseño de los informes	12
Figura 8. Vista estándar de la tabla InformeAccesos.....	14
Figura 9. Vista personalizada de la tabla AplicacionDim	15
Figura 10. Vista personalizada de la tabla FechaDim	15
Figura 11. Vista personalizada de la tabla HoraDim	16
Figura 12. Vista personalizada de la tabla UbicaciónDim	16
Figura 13. Vista personalizada de la tabla UsuarioDim	16
Figura 14. Esquema en estrella de accesos	17
Figura 15. Vista del origen de datos de accesos	18
Figura 16. Estructura de Cubo_Accesos	18
Figura 17. Jerarquías de Cubo_Accesos	19
Figura 18. Relación entre la fuente externa y el Data Stage de accesos	20
Figura 19. Carga de la fuente externa al Data Stage de accesos	20
Figura 20. Procedimiento de carga de AplicacionDim.....	21
Figura 21. Llamada al procedimiento de carga de AplicacionDim.....	21
Figura 22. Procedimiento de carga de AccesoFact.....	21
Figura 23. Carga de UbicaciónDim mediante la IP <i>Geolocation</i> API	22
Figura 24. <i>Script</i> para la llamada a la IP <i>Geolocation</i> API	22
Figura 25. Informe Aplicacion_Accesos.....	23
Figura 26. Informe Usuario_Accesos.....	24
Figura 27. Informe Mapa_Accesos	24
Figura 28. Vista estándar de la tabla InformeFallos	26
Figura 29. Vista personalizada de la tabla VehiculoDim.....	27
Figura 30. Vista personalizada de la tabla FalloDim	27
Figura 31. Esquema en estrella de fallos	27
Figura 32. Vista del origen de datos de fallos.....	28
Figura 33. Estructura de Cubo_Eventos	29
Figura 34. Jerarquías de Cubo_Eventos.....	29
Figura 35. Relación entre la fuente externa y el <i>Data Stage</i> de fallos	30

Figura 36. Carga de la fuente externa al <i>Data Stage</i> de fallos	30
Figura 37. Procedimiento de carga de VehiculoDim	31
Figura 38. Procedimiento de carga de EventoFact	31
Figura 39. Informe Modelo_Fallos	32
Figura 40. Informe VIN_Fallos	33
Figura 41. Informe de Tiempo_Trayectos	36
Figura 42. Informe de Consumo_Trayectos	37
Figura 43. Informe de Calificacion_Trayectos	37
Figura 44. Informe de Ubicacion_Accesos	42
Figura 45. Informe de Ubicacion_Accesos_II.....	42
Figura 46. Informe de Hora_Accesos.....	43
Figura 47. Informe de Tabla_Accesos.....	43
Figura 48. Informe de Fecha_Fallos.....	44
Figura 49. Informe de Tabla_Fallos	44
Figura 50. Informe de Velocidad_TRayectos	45
Figura 51. Informe de Distancia_Trayectos	45
Figura 52. Informe de Ubicacion_Trayectos.....	46
Figura 53. Informe de Mapa_Trayectos	46

Índice de tablas

Tabla 1. KPI de accesos y sus perspectivas de observación.....	13
Tabla 2. Relación entre las consultas y las tablas del esquema en estrella de accesos	17
Tabla 3. KPI de fallos bajo perspectivas de observación	25
Tabla 4. Relación entre las consultas y las tablas del esquema en estrella de fallos	28
Tabla 5. KPIs de trayectos bajo perspectivas de observación	33

1 Introducción

En este apartado, se describe el contexto y el objeto del proyecto a realizar cuyo fin es la creación de una solución de inteligencia de negocio para una empresa de logística. Para ello, se cuenta con la colaboración de una compañía líder en el desarrollo de *software* para la gestión de transporte.

1.1 Contexto

En la actualidad, el volumen de información que se genera diariamente es inmenso. Estos datos adecuadamente analizados permiten facilitar a las organizaciones una toma de decisiones acertada. La analítica de datos es una disciplina que está en auge y que se basa en la extracción de conocimientos a partir de los datos propiamente dichos. Esto incluye el análisis, la recopilación, la organización y el almacenamiento de datos, así como las herramientas y las técnicas empleadas para ello, entre las que se encuentran las destinadas al *Reporting* y al desarrollo de informes de inteligencia empresarial.

El proyecto “rut-IA: Investigación industrial de tecnologías de analítica de datos para el tratamiento de información de explotación logística georreferenciada” del Departamento de Ingeniería Informática y Electrónica tiene entre sus objetivos la implementación de un cuadro de mandos para exponer los indicadores clave de rendimiento (KPI) de una empresa de sector logístico. Este proyecto se basa en una colaboración con la compañía Fagor Electrónica y está cofinanciado por la consejería de Industria, Turismo, Innovación, Transporte y Comercio del Gobierno de Cantabria a través de la línea de ayudas INNOVA.

El departamento Smart Data Services de esta empresa proporciona soluciones *software* para la gestión de transporte y logística. Lleva a cabo numerosos proyectos a nivel internacional con presencia directa en países de América del Sur como Colombia y Chile. La firma localizada en Santander se ocupa de impulsar productos *software* que permitan rentabilizar y aumentar la competitividad de compañías de logística.

Sus productos se centran en ofrecer soluciones para la localización de flotas de vehículos, así como para la toma de decisiones a partir de la información relativa a los transportes. La principal plataforma, denominada FlotasNet, es una aplicación web diseñada para el control y optimización de grupos de transporte con localizador GPS. Facilita información de relevancia sobre la actividad y telemetría de los vehículos.

Entre sus clientes se encuentra un conocido fabricante de camiones que hace uso de una aplicación para la visualización de informes sobre los vehículos. Está basada en el producto FlotasNet desarrollado por Fagor. Esta posee una serie de informes sobre telemetría, fallos y calificación de conductores. Sin embargo, la plataforma presenta una serie de limitaciones que merman la experiencia de usuario. Las consultas están acotadas a un periodo de siete días suponiendo un inconveniente muy importante para el análisis de datos. Asimismo, en ocasiones el rendimiento es considerado insuficiente dado el tiempo de carga de la información.

1.2 Objeto

Fagor Smart Data solicita el desarrollo de una solución de inteligencia de negocio para una empresa colombiana del sector logístico que permita computar y mostrar indicadores estratégicos de la compañía mediante informes *Business Intelligence* (BI). Se requiere que la implementación esté basada en el diseño y desarrollo de un almacén de datos y un cuadro de mandos para integrar en su producto comercial. Cabe destacar que se insta a que el repositorio de la información se aloje en Microsoft SQL Server y la generación de los informes se realice mediante la herramienta Microsoft PowerBI.

Entre los indicadores clave de rendimiento, se deben abordar aquellos relativos al acceso a los sistemas de información de la empresa por parte de los usuarios, así como al mantenimiento y registro de fallos de la flota de vehículos, a la vida útil del vehículo y a los informes de conducción.

En resumen, el objeto principal de la solución BI es corregir y paliar las carencias que presenta el sistema de información actual “FlotasNet”, para lo cual debe permitir la realización de consultas de forma ágil sin la limitación de la obtención de información a una única semana. Se persigue también la visualización de los datos a través de diferentes dimensiones con la finalidad de aumentar las perspectivas de observación. Además, se solicita la posibilidad de exportar los resultados a una hoja de cálculo Microsoft Excel.

2 Planificación y desarrollo del proyecto

A principios de octubre de 2021 se propone el desarrollo de un proyecto *Business Intelligence* a la directora del presente Trabajo de Fin de Grado, Marta Elena Zorrilla Pantaleón. Después de recibir una respuesta afirmativa para la realización de este, se lleva a cabo un primer intercambio de ideas sobre la posible temática del trabajo, la estructura del mismo y las tecnologías a emplear. Se concreta una primera reunión de forma presencial para finales de diciembre dada mi presencia en la Universidad de Tulsa en Estados Unidos.

Tras la reflexión y el estudio de distintas fuentes de datos con las que trabajar, se opta por utilizar una serie de casos de uso pertenecientes al proyecto de inteligencia de negocio e inteligencia artificial en el que participa el Departamento de Ingeniería Informática y Electrónica de la Universidad de Cantabria. Asimismo, se establece Microsoft PowerBI como la herramienta de *Reporting* para la transformación y visualización de datos y la creación de informes. Surge también la duda sobre si utilizar exclusivamente una base de datos relacional y fuentes externas diversas o construir un *Data Warehouse* y cubos OLAP como hipotéticos orígenes de datos para el servicio de *Reporting*. Durante el periodo de tiempo previo a la reunión, se llevó a cabo un curso *online* de OpenWebinars denominado “Curso de PowerBI: Introducción” con el objetivo de conocer las potencialidades de la herramienta. Debido a que el trabajo es una colaboración con la sección *Smart Data Services* de la compañía Fagor Electrónica, se avanza con esta la definición de los indicadores para el desarrollo de la solución *Business Intelligence*. Además, se acuerdan las tecnologías a usar para la creación y gestión de los objetos del sistema de información.

En diciembre se realiza una reunión presencial con D. ^a Rosa María Martínez, ingeniera de producción de Fagor Smart Data. En este encuentro se presenta a la compañía el propósito del trabajo y las ideas para su desarrollo. Después de haber sentado las bases del proyecto, se distribuyen las primeras tareas a realizar por parte de la empresa y la nuestra.

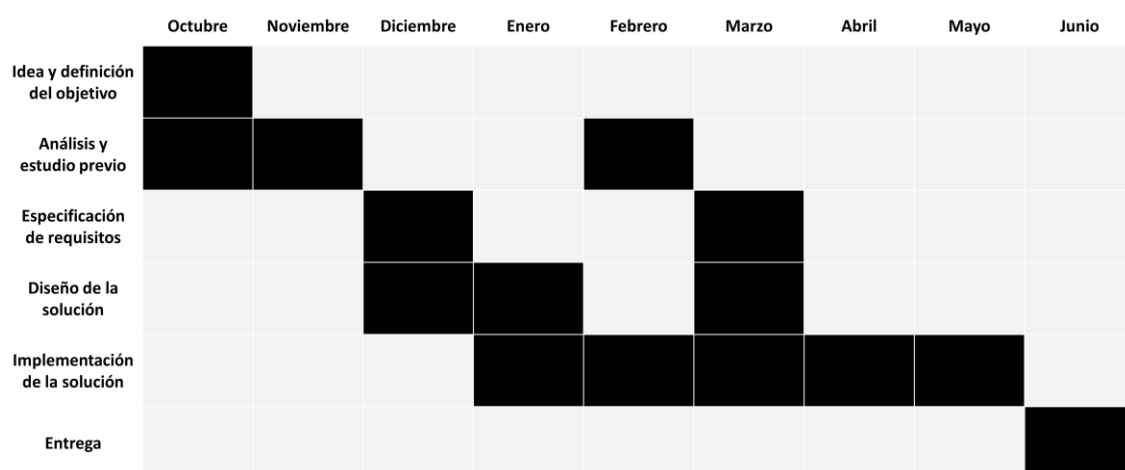


FIGURA 1. CALENDARIO DE DESARROLLO DEL TFG

La siguiente fase del TFG se centra en el inicio del diseño de la solución *BI* de nuevo desde el continente americano en enero de 2022. Para ello, se emplea la documentación elaborada por el departamento donde se estudia las características del sistema de información actual y los requisitos a desarrollar para su mejora. A partir de este momento, el correo electrónico junto con las videollamadas a través de

Microsoft Teams se convierten en las principales vías de comunicación. Después de diversas mejoras incrementales en el diseño, se comienza a implementar el sistema BI con la intención de realizar una primera presentación del producto a Fagor. El 3 de febrero se desarrolla de forma remota y para ello se elaboran un conjunto de diapositivas donde se expone gráficamente la solución explicando los principales componentes de esta. Asimismo, se lleva a cabo una pequeña demostración de la herramienta de *Reporting* para que la empresa conozca sus capacidades. Tras su finalización, se realizan consultas y se plantean dudas para su resolución.

Durante los meses de febrero y marzo, se continúa con la implementación de la solución avanzando en nuevos componentes y perfeccionando los ya desarrollados. Con la finalidad de obtener conocimientos y experiencia en el análisis de datos, se cursa durante este semestre una asignatura de *Business Intelligence & Data Visualization* en la Universidad de Tulsa. Asimismo, se completa un nuevo curso en OpenWebinars llamado “Curso de Trucos de PowerBI”. Este explica cómo diseñar informes de inteligencia de negocio de una forma más atractiva para conseguir un mayor impacto en la audiencia.

A comienzos de abril, se realiza una segunda videoconferencia con Fagor Smart Data donde se obtiene nueva información acerca de los informes externos que suministran. Además, se resuelven nuevas dudas y se solicita ampliar las fuentes de datos. En este mes, la solución BI en su conjunto queda prácticamente consolidada a falta diversas modificaciones o mejoras que surgen en las semanas posteriores.

Después de la finalización del curso académico en Estados Unidos, se fija la planificación para el periodo final del proyecto. El 12 de mayo se lleva a cabo una reunión presencial con la directora del Trabajo de Fin de Grado para establecer el esquema de la memoria y recibir *feedback* acerca del *Reporting* desarrollado. Se comienza a desarrollar la memoria y a completar las sugerencias recogidas sobre el sistema de inteligencia de negocio. En los siguientes días se suceden varias reuniones para seguir la evolución del proyecto. Cabe destacar la reunión presencial con la compañía el 24 de mayo donde se muestra el avance del TFG a la compañía y se recogen las opiniones de los miembros presentes. Para ello, se explica la arquitectura de la solución *Business Intelligence* llevándose a cabo una demostración del uso de los informes interactivos. Se hace especial énfasis en el objetivo de la solución, basado en corregir las deficiencias presentes en el sistema de información actual. También, se produce un diálogo sobre el despliegue y la realización de una jornada formativa para ello.

Durante las dos últimas semanas previas al depósito de la memoria, se procede a su finalización mediante la validación del sistema y el desarrollo de las líneas futuras. Se hace también entrega de esta a la empresa con el objeto de recoger sugerencias y el grado de satisfacción del cliente. Además, se revisa la solución implementada llevando a cabo un proceso de perfeccionamiento y corrección de errores como paso previo al despliegue del producto.

En la segunda quincena de junio, se aborda la fase de transferencia de la solución a un miembro del equipo de Fagor Smart Data. Para ello, se realiza un documento con los pasos necesarios para llevar a otra máquina todos los artefactos implementados. Este informe es proporcionado a la persona destinada a seguir trabajando en el producto para su correcta formación.

3 Business Intelligence

3.1 Introducción

Desde el comienzo de siglo, el volumen de datos generados en todo el mundo se ha multiplicado enormemente. La información producida al inicio de la década del 2010 pasa a ser más 3 veces mayor 10 años después. Durante el futuro próximo, se estima un importante crecimiento anual del *big data* que dé lugar a un todavía mayor incremento (Statista, 2021).

Esta situación está originada, en parte, por el auge del denominado Internet de las Cosas (IoT) que se basa en el uso de la tecnología en objetos de uso cotidiano para la obtención y el intercambio de información. En consecuencia, las compañías gestionan cantidades ingentes de datos generados por

dichos elementos, así como sistemas corporativos propios y otras aplicaciones de gestión empresarial. Además de las fuentes internas, existen otros orígenes externos al negocio que pueden llegar a suponer hasta una quinta parte del total de los registros. Todo esto implica una saturación de información para las organizaciones a la que no es sencilla de acceder.

En la actualidad, existe además un entorno empresarial muy competitivo que se caracteriza por una cada vez mayor globalización. Las compañías necesitan buscar nuevos productos o estrategias que supongan anticiparse a sus competidores, siendo el análisis del mercado fundamental para ello. Asimismo, tras las crisis económicas de los últimos años, se busca la optimización de procesos y la reducción de costes para lograr una mayor rentabilidad financiera.

Como solución a dichos problemas se presenta la disciplina denominada *Business Intelligence* o BI (inteligencia de negocio en inglés). Tiene como origen los sistemas de información llamados *Management Information Systems* (MIS) que se comienzan a desarrollar a finales de la década de 1960 (Turban, Sharda, Delen, & King, 2010). No es hasta dos decenios después, cuando surge el concepto *Executive Information Systems* (EIS) que introduce nuevas capacidades como informes dinámicos, análisis predictivo o técnicas de *drill-down*. Con la llegada del nuevo milenio y la incorporación de la inteligencia artificial y otras herramientas de análisis, se implementan nuevos productos bajo el nombre de sistemas *Business Intelligence* que suceden a los anteriores EIS.

Según la consultora estadounidense (Gartner, s.f.), el término *Business Intelligence* es “un concepto general que engloba las aplicaciones, la infraestructura y las herramientas, así como las mejores prácticas para acceder a la información y analizarla con el objetivo de optimizar la toma de decisiones”. De esta forma, los directivos de las organizaciones al igual que todos los usuarios de estas, pueden obtener de forma directa, ágil y sencilla los indicadores clave para el negocio. La interacción con los datos y su manipulación forman parte de las características fundamentales de estos sistemas.

La institución Business Application Research Center (BARC, 2020) establece los principales beneficios que se atribuyen a las soluciones de inteligencia empresarial a través de su Business Benefit Index (BBI). Estos se corresponden esencialmente con el análisis rápido y preciso de la información, la mejora de la toma de decisiones, una mayor satisfacción del empleado y del cliente, así como la eficiencia operacional que da lugar a la reducción de los costes y el aumento de los ingresos.

Una de las principales infraestructuras que pertenece a toda solución *Business Intelligence* es el *Data Warehouse* (almacén de datos en inglés). Este alberga información histórica del negocio de una forma organizada con el objeto de mejorar la accesibilidad y la manipulación de los datos. En la actualidad, existen almacenes que presentan información en tiempo real para dar respuesta a los problemas de forma inmediata.

3.2 Data Warehouse

Existen diversas definiciones y enfoques sobre este sistema de almacenamiento. Bill Inmon es conocido por ser uno de los primeros autores en explicar este concepto. Según este prestigioso científico, un almacén de datos es “un sistema orientado a temas, integrado, no volátil y que recopila datos variables en el tiempo en apoyo a la toma de decisiones” (Inmon, 2005). Su visión del *Data Warehouse* se caracteriza por seguir una estrategia *Top-Down* que busca evitar la redundancia de información y una mayor flexibilidad a los cambios.

Ralph Kimball establece la siguiente definición como alternativa: “una copia de los datos transaccionales específicamente estructurada para la consulta y el análisis” (Kimball, 2002). A diferencia del modelo de Inmon, describe una metodología *Bottom-Up* que conlleva el almacenamiento de información de distintas fuentes en una única unidad. De ahí, determina también que un almacén de datos se puede precisar básicamente como la unión de varios almacenes específicos.

De acuerdo con el modelo de Kimball, un *Data Warehouse* se constituye por cuatro componentes diferentes que presentan una función específica. A continuación, se describe cada uno de ellos.

3.2.1 Data Sources

Se organizan en fuentes de datos internas y externas. Las primeras se corresponden con los registros de múltiples sistemas operacionales independientes que almacenan la actividad diaria de la empresa. Este tipo de fuentes se dividen, asimismo, en tres grupos (compras, contabilidad, recursos humanos) que proporcionan una visión distinta sobre el funcionamiento de la organización. Entre los sistemas de procedencia se encuentran los sistemas ERP, los gestores de cadena de suministro (SCM) y las bases de datos *Online Transaction Processing* (OLTP). Asimismo, se tienen en cuenta los ficheros *log* generados a partir del uso de las aplicaciones y páginas web.

En cuanto a las fuentes de datos externas, provienen principalmente de estadísticas acerca de los clientes y los competidores que son generadas a partir de encuestas, cuestionarios y estudios.

Las prioridades que afectan a los *Data Sources* se basan en el rendimiento y la disponibilidad de su procesamiento (Kimball, 2002). Comúnmente presentan un formato rígido, sin transformación de los datos, que dificulta la obtención de información útil para el negocio. Además, los diferentes orígenes de los que proceden los datos añaden una mayor complejidad a su tratamiento.

3.2.2 Data Staging Area

Según Ralph Kimball, el componente *Data Stage* se corresponde tanto al área de almacenamiento como al conjunto de procesos denominado *extract-transformation-load* (ETL) que se encuentran entre las fuentes de datos y el *Data Area* (Kimball, 2002), como se puede observar en la Figura 4.

En primer lugar, se lleva a cabo la tarea de extracción de datos con el objeto de cargar la información en el elemento de almacenamiento. Para ello, se leen las fuentes de datos y se copian los valores necesarios para un mayor procesamiento posteriormente. Una vez realizado el volcado, existen varios tipos de transformaciones que se pueden realizar como, por ejemplo, la limpieza de datos, la combinación de estos a partir de diferentes fuentes y la asignación de identificadores. Sin embargo, la mayoría de las operaciones que se efectúan están relacionadas con la ordenación y el procesamiento secuencial (Kimball, 2002). Como último paso, se produce la carga de los datos y el indexado de los nuevos registros para la mejora del rendimiento.

3.2.3 Data Area

El *Data Area* se corresponde con el elemento principal del almacén de datos. Representa la entidad de almacenamiento al margen del resto de componentes que forman su entorno. Organiza toda la información de forma coherente para que sea accesible por las herramientas destinadas a su explotación. Kimball hace hincapié en que los datos deben ser almacenados en esquemas dimensionales dado que es la forma más adecuada para su posterior presentación. Su diseño facilita la accesibilidad, aumenta el rendimiento de las consultas y mejora la adaptación a los cambios. Para que los usuarios puedan responder las cuestiones con una mayor precisión, se requiere que la información almacenada presente el máximo nivel de detalle (grano).

Normalmente se hace referencia a este componente como un conjunto de varios *Data Mart* integrados. Este concepto se corresponde con un almacén de datos específico para una determinada área de negocio. Contiene la información de un esquema dimensional en concreto facilitando la segmentación y la portabilidad de los datos.

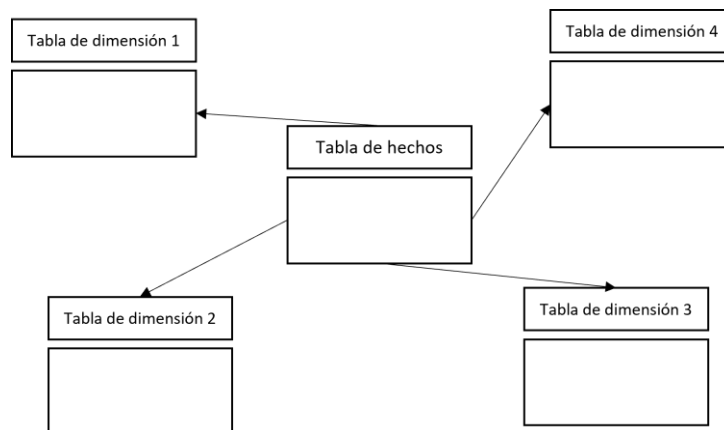


FIGURA 2. ESTRUCTURA DE UN ESQUEMA EN ESTRELLA

Cuando los datos se agrupan en diferentes entidades que son representadas como tablas en una base de datos relacional o sistema *Online Transaction Processing* (OLTP) se habla de “esquema en estrella”. Los sistemas OLTP contienen datos actuales y dinámicos que proceden de las operaciones diarias del negocio. Se dedican al procesamiento simple de la información.

Como se observa en la Figura 2, este tipo de modelo dimensional suele constar de una tabla de hechos y una serie de dimensiones a su alrededor. Se caracterizan por su simplicidad y simetría que dan lugar a una mayor productividad. Dada su estructura, las consultas se realizan eficientemente sin la necesidad de agrupar varias sentencias *join*. Cabe destacar también la extensibilidad que presenta el esquema. Permite añadir nuevas dimensiones, atributos y medidas de forma sencilla.

Una tabla de hechos representa la tabla principal del modelo dimensional. Alberga un conjunto de medidas de negocio de carácter numérico y aditivo. Cada una de estas medidas se denomina mediante el término “hecho” y debe presentar igual grado de detalle que el resto. Cada fila se corresponde con una medida. Asimismo, estas tablas contienen las claves que identifican cada una de las dimensiones que conforman el esquema.

En cuanto a las tablas de dimensión, estas almacenan los descriptores textuales de la información. Cada uno de estos atributos se encargan de describir las filas que forman la dimensión. Son los puntos de entrada a los que se hace referencia desde la tabla de hechos. Por ello, permiten la creación de restricciones y el agrupamiento para las agregaciones de las medidas de la tabla de hechos. Generalmente, existe al menos una dimensión relacionada con el tiempo como la fecha o la hora.

Por otro lado, los datos pueden ser almacenados en “cubos” si la información se organiza en bases de datos multidimensionales *Online Analytic Processing* (OLAP). Este tipo de sistemas almacenan datos históricos de detalle que se agregan a diferentes niveles. Están orientados a la información clave del negocio para la toma de decisiones estratégicas.

Existen varios tipos de cubos que ofrecen distintas características y capacidades. Para un mayor rendimiento, se utiliza el llamado “cubo MOLAP” que se fundamenta en las bases de datos multidimensionales. Se calculan y se indexan los datos con anterioridad. Debido a esto permiten realizar cálculos complejos con un tiempo de respuesta óptimo. Por otro lado, los “cubos ROLAP” contienen la información junto con sus agregaciones en una estructura relacional. Están destinados al manejo de grandes volúmenes de datos. Dada su configuración, emplean las funcionalidades propias de las bases de datos relacionales. Además, es posible el almacenamiento de la información en sistemas híbridos que combinan la estructura relacional con las agregaciones de la opción multidimensional. Estos sistemas reciben el nombre de “cubos HOLAP”.

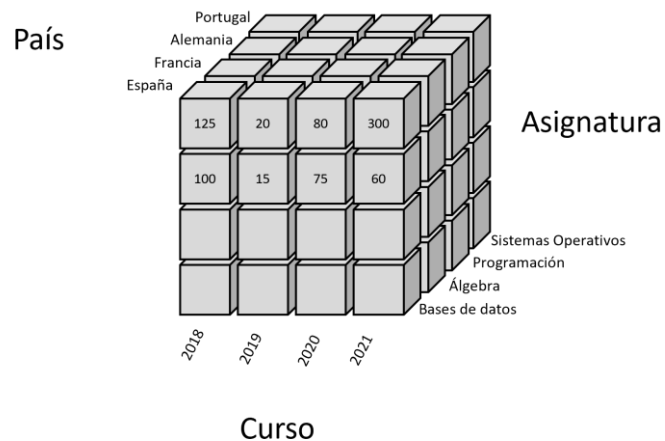


FIGURA 3. EJEMPLO DE UN CUBO OLAP

En la Figura 3, se refleja un ejemplo de cubo OLAP que presenta tres dimensiones. Es posible conocer el número de estudiantes bajo las perspectivas de “país”, “curso académico” y asignatura.

3.2.4 Data Access Tools

Las *Data Access Tools* (herramientas de acceso a datos en inglés) son el último componente que conforma el entorno del almacén de datos. La principal finalidad de estas utilidades son la consulta de la información presente en el elemento anterior. Kimball incluye en este componente tanto herramientas de consulta *ad hoc* como complejas aplicaciones de minería de datos (Kimball, 2002).

De acorde a sus objetivos y capacidades, la tecnología destinada a la explotación de los datos se puede dividir en tres grupos. A continuación, se explica brevemente cada uno de ellos.

- **Reporting.** Responden a cuestiones del tipo “¿Qué está ocurriendo?”. Permiten la representación de los resultados de forma sencilla y visual mediante la creación de informes y cuadros de mandos. Facilitan la evaluación de los indicadores clave de negocio. Actualmente, existen un amplio abanico de este tipo de *software*.
- **Minería de datos.** Responden a cuestiones del tipo “¿Por qué está ocurriendo?”. Se emplean con el fin de identificar patrones y tendencias.
- **Herramientas OLAP.** Llevan a cabo la definición de combinaciones de datos para su análisis desde diferentes perspectivas. Permiten la navegación a través de sus dimensiones haciendo uso de potentes operadores de consulta. Estos habilitan la observación de la información con un mayor detalle.
- **Simulación y optimización.** Responden a cuestiones condicionales. Son empleadas para analizar escenarios futuros con el objetivo de encontrar la estrategia óptima.

3.3 Tecnologías

En la presente sección se describen las tecnologías utilizadas para la implementación de la solución *Business Intelligence*. Para el desarrollo de la base de datos del *Data Stage* como para el *Data Warehouse* y los cubos OLAP, se opta por el empleo de diversos componentes de la suite Microsoft SQL Server 2019. Su elección se basa en la integración de diferentes herramientas para el desarrollo de sistemas inteligentes en un mismo gestor de bases de datos. Además, se corresponde con el gestor que Fagor Smart Data Services emplea para su producto FlotasNet. En cuanto a la construcción del cuadro de mandos, se utiliza el servicio de análisis de datos Microsoft PowerBI que dispone, a su vez, de funcionalidades para la transformación de los mismos.

3.3.1 Microsoft SQL Server

Microsoft SQL Server es un gestor de bases de datos relacional desarrollado por la compañía Microsoft para el almacenamiento y procesamiento de datos que incorpora herramientas para su análisis. Utiliza el lenguaje estándar SQL para la definición, manipulación y control de la integridad de las bases de datos. Todos los componentes SQL son administrados mediante el entorno de desarrollo Microsoft SQL Server Management Studio. Además, la creación de los cubos OLAP y los procesos ETL se llevan a cabo a través de Microsoft Visual Studio haciendo uso de las herramientas correspondientes. A continuación, se exponen los componentes empleados de la versión 2019:

- SQL Server: motor de bases de datos relacional.
- SQL Server Analysis Services: motor de bases de datos multidimensionales. Facilita la definición y gestión de cubos OLAP.
- SQL Server Integration Services: componente para la implementación de tareas de migración de datos. Permite la programación de paquetes ETL (*Extract, Transform and Load*).

3.3.2 Microsoft PowerBI

Microsoft PowerBI es un servicio para el desarrollo de visualizaciones de datos interactivas producido por la compañía Microsoft. Permite el análisis de datos mediante la creación de informes de inteligencia de negocio. Su lanzamiento se produce en el año 2014. Actualmente, es considerada una de las herramientas más potentes del mercado para el uso y explotación de grandes volúmenes de datos.

Este *software* ofrece dos tipos de aplicaciones. Por un lado, PowerBI Desktop es una versión gratuita para su descarga e instalación de forma local que incorpora una serie de herramientas para la transformación y la representación de los datos. Por otro, PowerBI *service* es un servicio de pago basado en la nube o *software as a service* (SaaS) que facilita la gestión de los informes y su exportación. Es empleado para la presentación de los informes ante equipos tanto de dentro como de fuera de la organización.

PowerBI ofrece un conjunto de funcionalidades concretas dependiendo del tipo de licencia. Esta determina, principalmente, el lugar de almacenamiento de los informes y sus correspondientes datos. Los tres tipos de licencias que existen, ordenadas de menor a mayor capacidad, son los siguientes:

- PowerBI (gratuita): acceso a los informes del espacio de trabajo personal impidiendo su compartición otras áreas de trabajo o usuarios.
- PowerBI Pro: permite compartir los informes con todos los usuarios de la organización. Establece una serie de límites de frecuencia de actualización y de tamaño de almacenamiento.
- PowerBI Premium: incorpora funcionalidades destinadas a la Inteligencia Artificial y al Big Data aumentando las limitaciones de las licencias previas. De este modo, es posible desplegar los informes en múltiples regiones y exportarlos fuera de la organización. Se multiplica significativamente el almacenamiento.

Dentro de la aplicación existen decenas de orígenes de datos agrupados en archivos, bases de datos, productos de Azure, servicios en línea y otros tipos. Los más comunes se corresponden con ficheros de texto, Excel, bases de datos SQL y fuentes Web. PowerBI presenta dos formas de acceder a las fuentes externas.

- ImportQuery: se extrae o importa la información del origen de datos. Para que reflejen valores actualizados se debe cargar de nuevo el conjunto de datos completo.
- DirectQuery: se realizan consultas al origen en tiempo real para la obtención de los datos. Se mantiene la información en la fuente externa.

La aplicación ofrece un editor para la transformación de datos denominado PowerQuery. Entre sus funciones principales, se encuentra la sección de modificación. Permite remplazar valores, trasponer

columnas o cambiar el tipo de dato, entre otros. El apartado “Agregar Columna” posibilita añadir nuevos campos de forma condicional o personalizada además de cambiar su formato.

Además de las columnas, existe otro tipo de campo llamado “Medida”. Este no se almacena en memoria, a diferencia de las primeras, calculándose mediante la Unidad Central de Procesamiento (CPU). Las medidas son calculadas al instante siendo útiles para cálculos agregados como una suma. Mientras, las columnas requieren la actualización del informe y se destinan a cálculos por cada fila.

La definición de medidas y columnas calculadas es facilitada por el lenguaje *Data Analysis Expressions* (DAX). Se corresponde con un conjunto de funciones y constantes para la construcción de fórmulas que permitan la devolución de valores.

Asimismo, PowerBI dispone de varias funcionalidades que ayudan a crear informes con un mayor impacto para el usuario. Entre estas, se encuentra el uso de *bookmarks* (marcadores) con el objetivo de capturar información clave para destacarla sobre el resto. Otra interesante característica son los *tooltips* que contribuyen a ampliar la información que se expone. Para ello, aparece un *pop-up* que muestra datos añadidos al situar el cursor sobre una determinada variable de una visualización.

3.3.3 Google Geocoding API

El servicio de *geocoding* de Google permite convertir las direcciones postales en coordenadas geográficas. Mientras, el denominado *reverse geocoding* lleva a cabo el proceso contrario obteniendo una dirección a partir de la longitud y la latitud de una determinada localización. Esta API proporciona un acceso sencillo a estos servicios mediante peticiones HTTP. La respuesta que se obtiene a una solicitud se divide en los componentes de la dirección, así como la dirección completa formateada. Los principales campos retornados son los siguientes: número, calle, barrio, distrito, ciudad, estado, país y código postal (Google, 2022).

Actualmente, esta utilidad presenta un cargo por cada llamada que se realiza al servidor. Existe un periodo de 30 días gratuito de prueba.

4 Diseño e implementación de la solución BI

A continuación, se describe el proceso llevado a cabo para la obtención de la solución *Business Intelligence* en base al contexto y a las tecnologías descritas en los apartados previos. Además, se exponen los criterios para el diseño del cuadro de mandos y se explican los requisitos para la construcción de cada uno de los casos de uso solicitados.

4.1 Arquitectura de la solución

El sistema de inteligencia de negocio construido se organiza en base a las cuatro áreas clásicas como se puede observar en la Figura 4. Por un lado, el área de *Data Sources* lo forman todas las fuentes de datos externas en formato XLS o CSV. El origen es la base de datos (BD) “rut-IA” que registra toda la información relativa a la flota, su mantenimiento y la información en tiempo real de los sensores instalados en los vehículos. Dado que no se tiene acceso a la base de datos de la compañía de forma remota, se trabaja inicialmente con los ficheros Microsoft Excel que se obtienen de los servicios web (WS). Además, se emplean ficheros de texto en formato CSV como información contextual proporcionada por la compañía e interfaces de programación de aplicaciones (API) para la extracción de datos de geolocalización. De dicha base de datos proceden los informes relativos a los fallos de los vehículos (Informe_fallos) y a los trayectos realizados por estos (Informe_tramos). La fuente de “Informe_accesos” que registra las conexiones a la aplicación también procede de “rut-IA”.

Todo el conjunto de datos del componente inicial se almacena en el *Data Stage Area* mediante procesos ETL, a excepción de la información asociada a los trayectos de los vehículos que se interpretará directamente desde Microsoft PowerBI. Estos procesos realizan la carga en la base de datos relacional “DataLogistics” que tiene como objetivo realizar el volcado de datos en los esquemas dimensionales del

componente posterior. Además, se ocupa de facilitar las tareas de sincronización entre las fuentes externas y el *Data Area*.

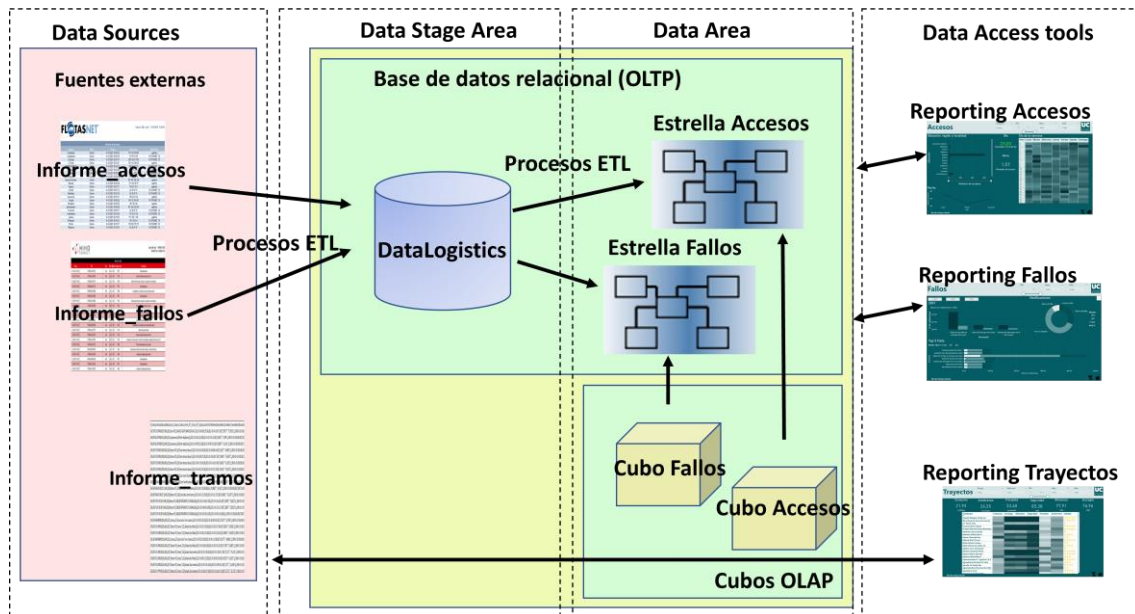


FIGURA 4. COMPONENTES DE LA ARQUITECTURA BI

El *Data Area* consta, en primer lugar, de una nueva base de datos relacional llamada “DWLogistics” que está compuesta por los diagramas en estrella de los accesos a la aplicación y los fallos emitidos por los vehículos. De forma análoga al componente anterior, esta base de datos recibe la información mediante procesos ETL de diversa índole. Por otro lado, se crean dos cubos MOLAP con la finalidad de aumentar el rendimiento de las consultas de selección desde las herramientas de *Reporting*. Estas bases de datos dimensionales se establecen en base a sus respectivos esquemas en estrella que reciben los valores del *Data Stage Area*.

El *Reporting* representa la pieza final de la solución. Se organizan informes en tres grupos distintos de acorde a la semántica de los datos proporcionados. Por cada uno de ellos, se desarrollan cuadros de mandos en base a distintas perspectivas de observación. Los gráficos asociados a los accesos y aquellos relacionados con los fallos de los vehículos presentan como origen de datos los esquemas en estrella y los cubos OLAP. Sin embargo, el *Reporting* de los trayectos y sus características funciona directamente sobre los ficheros externos de los tramos que se suministran como se ha mencionado anteriormente. La herramienta de acceso a los datos empleada Microsoft PowerBI incorpora la opción de transformación y gestión de datos facilitando el empleo de la información. En este caso, el informe de los tramos se trata como una única tabla que contiene todas las variables. Mediante este servicio de Microsoft se realizan todas las tareas ETL limpiando y corrigiendo aquellos campos carentes de significado.

4.2 Criterios para la interfaz gráfica del cuadro de mandos

4.2.1 Color y tipo de letra

Los colores que se emplean en el diseño del *dashboard* se fundamentan en el tono verdoso del logotipo de la Universidad de Cantabria que es ubicado en la esquina superior derecha de los informes. Los colores usados se muestran en la Figura 5. Se derivan pigmentos a partir del color original del emblema para alcanzar un diseño uniforme. Colores más ligeros mediante adición de blanco y otras variantes más oscuras por agregación de negro.

#045C64 4, 92, 100	#045158 4, 81, 88	#034548 3, 69, 75	#03393F 3, 57, 63	#045C64 4, 92, 100	#237077 35, 112, 119	#43858B 67, 133, 139	#62999E 98, 153, 158
#022E32 2, 46, 50	#022226 2, 34, 38	#011719 1, 23, 25	#010B0D 1, 11, 13	#82ADB2 130, 173, 178	#A1C2C5 161, 194, 197	#C0D6D8 192, 214, 216	#E0EBEC 224, 235, 236

FIGURA 5. COLORES DE LA INTERFAZ GRÁFICA

Por otro lado, se emplea la misma tipografía para todo el cuadro de mandos en busca de uniformidad. La fuente es “Segoe UI” desarrollada por Microsoft y empleada frecuentemente en sus interfaces de usuario. Se opta por su uso dado que está optimizada para una mayor legibilidad. Para los rótulos, tanto de la portada como de los informes, se hace uso de su versión en negrita. Estos presentan un tamaño de texto de 40 puntos. Asimismo, las cabeceras de las filas y las columnas de las matrices de datos también se exhiben con un trazado más grueso.

4.2.2 Portada

El cuadro de mandos consta de una página principal desde la que se accede a los diferentes conjuntos de informes. Se busca un *layout* simple y limpio evitando elementos innecesarios que puedan confundir al usuario. Los botones son agrupados en las tres áreas de información. Cada área se indica mediante el título y el icono correspondiente. En la franja superior de la portada, se establece un rótulo empleando el tono más claro de la gama de colores junto a un color verde más oscuro como se puede observar en la Figura 6.



FIGURA 6. PORTADA DEL CUADRO DE MANDOS

4.2.3 Diseño de los informes

En la Figura 7, se puede comprobar como cada informe se divide en dos espacios diferenciados por el color que se emplea de fondo. Por un lado, en la zona de tinte verde claro se establece el encabezado correspondiente al área de información al que pertenece el informe (Accesos, Fallos, Trayectos). A su derecha, se habilitan elementos de segmentación de datos. Todas las páginas presentan este tipo de componentes para el filtrado de los datos por día, mes y año. Estos elementos trabajan sobre una

jerarquía de tiempo formada por dichas unidades junto con el concepto de trimestre. Esta utilidad es ventajosa ya que permitir navegar por su estructura y alcanzar un mayor nivel de detalle temporal.

En el área de color verde oscuro se establecen todas las visualizaciones para la presentación de los datos. Cada una de ellas posee un titular en la parte superior con un tamaño de texto de 20 puntos y alineado a la izquierda. El color del encabezado se corresponde con el más claro en la gama empleada. Así, existe un gran contraste para aumentar su percepción por parte del usuario. Los datos se exponen mediante tonos verdes intermedios prevaleciendo de este modo la consistencia del informe. Los *visuals* (área que recoge cada elemento gráfico) poseen la opción de “Fondo” deshabilitada de forma que se observe una integración de los gráficos en el informe.

Las tablas que muestran datos numéricos son posicionadas en la franja izquierda de todas las secciones del *Reporting*. Cuando se realiza un filtrado de datos mediante la interacción con la tabla se evidencian los resultados en los gráficos de la derecha. De este modo, se pretende hacer uso de la direccionalidad de escritura para elevar el reconocimiento de la interfaz en el usuario.

En los gráficos que contienen datos categorizados, la leyenda informativa siempre es situada en el extremo opuesto a la visualización colindante. Esto tiene la finalidad de evitar confusiones sobre la pertenencia de dicha leyenda dado que los *visuals* carecen de fondo.

Los mapas empleados en el cuadro de mandos ocupan cada uno de ellos un único informe. Se pretende así una mejor visualización de los datos y las zonas geográficas.

Para regresar a la portada, se usa metáfora con la finalidad de facilitar la comprensión de la acción. De ahí, se habilita un botón negro con aspecto de casa situado en la esquina inferior derecha del informe. Además, a su izquierda se establece un icono que facilita el borrado de todos los filtros habilitados en ese momento. Todos los cuadros de mandos indican el nombre y apellidos del desarrollador en el extremo inferior izquierdo.

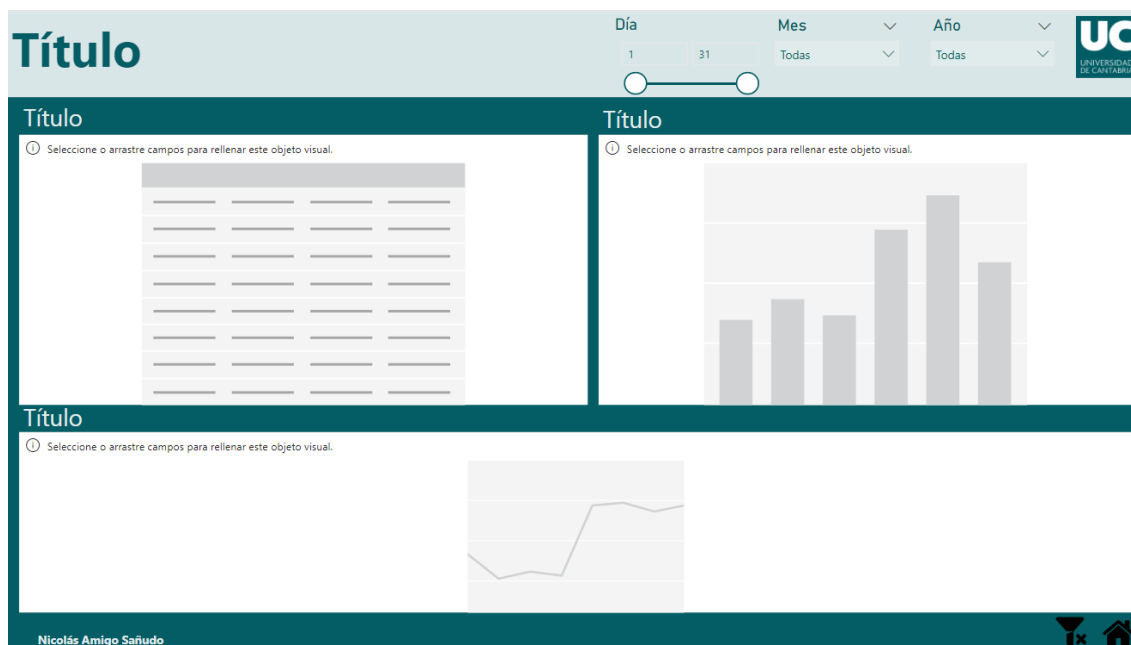


FIGURA 7. DISEÑO DE LOS INFORMES

4.3 Caso de uso: accesos

4.3.1 Especificación de requisitos

Este caso de uso pretende conocer el número de accesos al sistema de información FlotasNet. Los KPI y sus perspectivas de observación se recogen en la Tabla 1.

KPI	Perspectiva
Número de accesos.	Aplicación
	Fecha
	Localización
	Usuario

TABLA 1. KPI DE ACCESOS Y SUS PERSPECTIVAS DE OBSERVACIÓN

Mediante este KPI y las perspectivas asociadas se busca responder a las consultas que se precisan a continuación:

- A001 - Promedio de accesos realizados en un periodo de tiempo.
- A002 - Aplicación con la que se accede con mayor frecuencia.
- A003 - Región desde la que se accede con mayor frecuencia.
- A004 - Localidad desde la que se accede con mayor frecuencia.
- A005 - Uso de una determinada aplicación según el tipo de usuario.
- A006 - Momento del día en el que se produce un mayor número de accesos.
- A007 - Número de accesos por fecha, hora y usuario.

Cada una de las consultas anteriores se deben poder segmentar por tiempo y usuario, así como por el resto de las perspectivas.

4.3.2 Fuentes externas

La fuente de datos que utiliza este informe es la hoja de cálculo Microsoft Excel denominada "Informe_de_accesos.xls". Este fichero contiene 500.000 filas de datos. Cada una de ellas corresponde a un acceso de un usuario al sistema de información a través de una aplicación. Se emplea un procedimiento de randomización mediante Microsoft Excel con el objetivo de obtener una mayor cantidad de información a partir de la existente. Como propósito final de la aleatorización se encuentra la reproducción de un sistema BI real y la comparación de rendimiento entre *Online Transactional Processing* (OLTP) y *Online Analytical Processing* (OLAP). Es por ello por lo que en determinadas visualizaciones se sacrifica su significado y sentido a favor de la construcción de la estructura de análisis de negocio. Para ello, se procede a su aumento mediante la función "Aleatorio()". De esta forma se obtiene un total de aproximadamente 500.000 registros para explotar con el sistema BI.

La hoja de cálculo consta de los siguientes datos:

- Usuario: es el nombre de la cuenta de *login* con el que el usuario accede a la plataforma.
- Nivel: indica el tipo de usuario que se conecta a la aplicación. Existen varios tipos dependiendo su relación con la empresa.
 - Administradores
 - Cliente
 - Servicio
 - Skytracking
 - Ventas
 - Agendamiento
- Fecha: indica el instante de tiempo en el que el usuario accede a la plataforma. Se muestra mediante la fecha y la hora de la acción. No se dispone del momento en el que se abandona o cierra la sesión. Abarca un periodo de tiempo desde 15/12/2021 a 21/02/2022.
- IP: es la dirección IPv4 que identifica la conexión iniciada.

- Aplicación: indica el tipo de aplicación a la que se conecta el usuario.
 - FLOTASNET 3.0: aplicación web.
 - app: aplicación diseñada para dispositivos móviles.

4.3.3 Base de datos *Data Stage Area*

La constitución de la base de datos relacional que soporta el *Data Stage Area* se implementa mediante Microsoft SQL Server. La tabla “InformeAccesos” de la base de datos relacional almacena todos los datos asociados a los accesos a la aplicación. De esta forma, cada vez que se recibe un nuevo paquete de datos se cargan en el *Data Stage* como paso previo a su transformación y almacenamiento en las correspondientes tablas de hechos y dimensiones.

Las columnas de la tabla, mostradas en la Figura 8, se corresponden con los datos de la fuente externa “Informe_de_accesos.xls”. Sin embargo, se añaden dos nuevas columnas con el objetivo controlar la información que ya ha sido archivada en la base de datos del componente *Data Area*. Por un lado, el campo “Cargado” de tipo binario (*bit*) indica mediante el valor “1” si la fila se encuentra presente en el *Data Warehouse*. Por otro lado, “FechaRegistro” muestra el momento en el que se guardaron los datos en el *Data Stage*. Tiene por defecto la fecha y hora actual. La principal ventaja de este primer proceso es evitar filas repetidas en el almacén de datos cuando se trabaja con múltiples informes y realizar tareas de sincronización gracias a la fecha de registro.

InformeAccesos			
	Column Name	Data Type	Allow Nulls
	Usuario	nvarchar(30)	☒
	Nivel	nvarchar(30)	☒
	Fecha	nvarchar(30)	☒
	Dir_IP	nvarchar(30)	☒
	Aplicacion	nvarchar(30)	☒
	Cargado	bit	☒
	FechaRegistro	datetime	☒
			☐

FIGURA 8. VISTA ESTÁNDAR DE LA TABLA INFORMEACCESOS

Cabe matizar que el tipo de dato empleado para las columnas es *nvarchar* con una extensión máxima de 30 caracteres. Así se indica que el dato es un conjunto de símbolos Unicode de logitud variable. Su elección se debe a la compatibilidad con la información que se entrega en los documentos Microsoft Excel evitando problemas durante el proceso de carga. Por otro lado, se permite la introducción de valores nulos ya que en ocasiones ciertos campos se encuentran vacíos. Conviene conservar las columnas restantes para aportar un valor añadido. Posteriormente, en el componente dimensional se almacena la información mediante el tipo de dato adecuado a sus características. Se consigue de esta manera un mayor rendimiento cuando se ejecutan las consultas.

4.3.4 Base de datos *Data Area*

Como modelo dimensional para el *Data Area* se utiliza el esquema en estrella. Se crea una base de datos llamada “DWLogistics” en referencia a las iniciales del concepto *Data Warehouse* (DW). De nuevo, se hace uso de SQL Server para su desarrollo. Se establecen tablas de dimensión que contienen los descriptores textuales que explican conceptos temporales, geográficos y relativos a los accesos a la aplicación. Las dimensiones acerca del tiempo de los hechos, “FechaDim” y “HoraDim”, son compartidas por las dos estrellas.

La estrella definida para construir el modelo dimensional de los accesos a la aplicación consta de un total de 6 tablas. La tabla de hechos denominada “AccesoFact” comprende los identificadores de cada una de las dimensiones junto con la medida “Cuenta”. Esta sirve como hecho aditivo para obtener un cómputo total del número de accesos producidos.

A continuación, se describen los campos que forman las tablas de dimensión:

- **AplicacionDim:** representa las aplicaciones de la compañía a los que puede acceder el usuario. Únicamente se dispone del nombre de cada uno de ellos. Una longitud de 25 caracteres para la columna “Nombre” es considerada suficiente en el caso de que se añadieran más programas. Se expone información adicional sobre esta dimensión en la Figura 9.

AplicacionDim				
	Column Name	Condensed Type	Nullable	Description
🔑	Id_aplicacion	int	No	Identificador de la aplicación.
	Nombre	varchar(25)	No	Nombre de la aplicación.

FIGURA 9. VISTA PERSONALIZADA DE LA TABLA APLICACIONDIM

- **FechaDim:** representa todas las fechas desde cuando existen datos relativos a la empresa. Cada una de ellas se descompone con la finalidad de alcanzar un mayor detalle de información. Permite segmentar los campos asociados al dominio de negocio en base a periodos de tiempo de 24 horas. Se expone información adicional sobre esta dimensión en la Figura 10.

FechaDim				
	Column Name	Condensed Type	Nullable	Description
🔑	Id_fecha	int	No	Identificación de la fecha.
	Fecha	datetime	Yes	Fecha en formato AAAA-MM-DD.
	FechaCompleta	char(10)	Yes	Fecha en formato DD/MM/AAAA.
	DiaMes	varchar(2)	Yes	Día del mes.
	NombreDia	varchar(9)	Yes	Nombre del día de la semana.
	NumDiaSemana	char(1)	Yes	Número del día de la semana.
	NumSemanaMes	varchar(2)	Yes	Número de la semana del mes.
	NumSemanaAño	varchar(2)	Yes	Número de la semana del año.
	NumDiaAño	varchar(3)	Yes	Número del día del año.
	NumMes	varchar(2)	Yes	Número del mes del año.
	Mes	varchar(9)	Yes	Nombre del mes.
	NumMesTrimestre	varchar(2)	Yes	Número del mes del trimestre.
	NumTrimestre	char(1)	Yes	Número del trimestre.
	Trimestre	varchar(9)	Yes	Nombre del trimestre.
	Año	char(4)	Yes	Número del año.
	MesAño	char(10)	Yes	Nombre del mes junto con el año en formato MMM-AAAA.
	MMYYYY	char(6)	Yes	Número del mes y el año.
	PrimerDiaMes	date	Yes	Fecha del primer día del mes.
	UltimoDiaMes	date	Yes	Fecha del último día del mes.
	PrimerDiaTrimestre	date	Yes	Fecha del primer día del trimestre.
	UltimoDiaTrimestre	date	Yes	Fecha del último día del trimestre.
	PrimerDiaAño	date	Yes	Fecha del primer día del año.
	UltimoDiaAño	date	Yes	Fecha del último día del año.

FIGURA 10. VISTA PERSONALIZADA DE LA TABLA FECHADIM

- **HoraDim:** representa todos los instantes de tiempo del día. Cada una de ellas se descompone con la finalidad de alcanzar un mayor detalle de información. Permite segmentar los campos

asociados al dominio de negocio en base a periodos de tiempo de un segundo. Se expone información adicional sobre esta dimensión en la Figura 11.

HoraDim				
	Column Name	Condensed Type	Nullable	Description
🔑	Id_hora	int	No	Identificador de la hora.
	Hora	varchar(8)	No	Hora en formato HH:MM:SS
	NumeroHora	tinyint	No	Número de la hora.
	NumeroMinuto	tinyint	No	Número del minuto de la hora.
	NumeroSegundo	tinyint	No	Número del segundo de la hora.
	HoraSegundos	int	No	Tiempo en segundos desde 00:00:00.
	FranjaHoraria	varchar(15)	No	Periodo de 60 minutos al que pertenece la hora.
	Id_franja	int	No	Identificador de la franja horaria.
	NombreFranja	varchar(100)	No	Nombre de la franja horaria.

FIGURA 11. VISTA PERSONALIZADA DE LA TABLA HORADIM

- **UbicacionDim:** representa las direcciones IPv4 desde donde se realizan las conexiones para el acceso de los usuarios a la aplicación. Se obtiene la ubicación geográfica asociada con el objetivo de responder consultas sobre las zonas con mayor actividad. Todas las variables, exceptuando la dirección IP, pueden contener valores nulos. De esta forma se evitan problemas durante su adquisición mediante la API. Se expone información adicional sobre esta dimensión en la Figura 12.

UbicacionDim				
	Column Name	Condensed Type	Nullable	Description
🔑	Id_ubicacion	int	No	Identificador de la ubicación.
	Localidad	varchar(50)	Yes	Localidad donde se realiza la conexión.
	Provincia_Est...	varchar(50)	Yes	Región donde se realiza la conexión.
	Pais	varchar(50)	Yes	País donde se realiza la conexión.
	Latitud	float	Yes	Latitud donde se realiza la conexión.
	Longitud	float	Yes	Longitud donde se realiza la conexión.
	Dir_IP	varchar(50)	No	Dirección IPv4 de la conexión.

FIGURA 12. VISTA PERSONALIZADA DE LA TABLA UBICACIÓNDIM

- **UsuarioDim:** representa los usuarios que acceden a la aplicación. Solamente se dispone del nombre de la cuenta junto con el tipo de usuario por lo que el resto de los descriptores son nulos. Se expone información adicional sobre esta tabla en la Figura 13.

UsuarioDim				
	Column Name	Condensed Type	Nullable	Description
🔑	Id_usuario	int	No	Identificador del usuario.
	Log_in	varchar(25)	Yes	Nombre de la cuenta del usuario
	Usuario	varchar(100)	No	Nombre del usuario.
	TipoUsuario	varchar(25)	No	Tipo de usuario.
	FechaAlta	datetime	Yes	Fecha de alta de la cuenta del usuario.

FIGURA 13. VISTA PERSONALIZADA DE LA TABLA USUARIODIM

Tras la creación de las tablas de dimensión y la tabla de hechos queda constituido el diagrama en estrella como se muestra en la Figura 14.

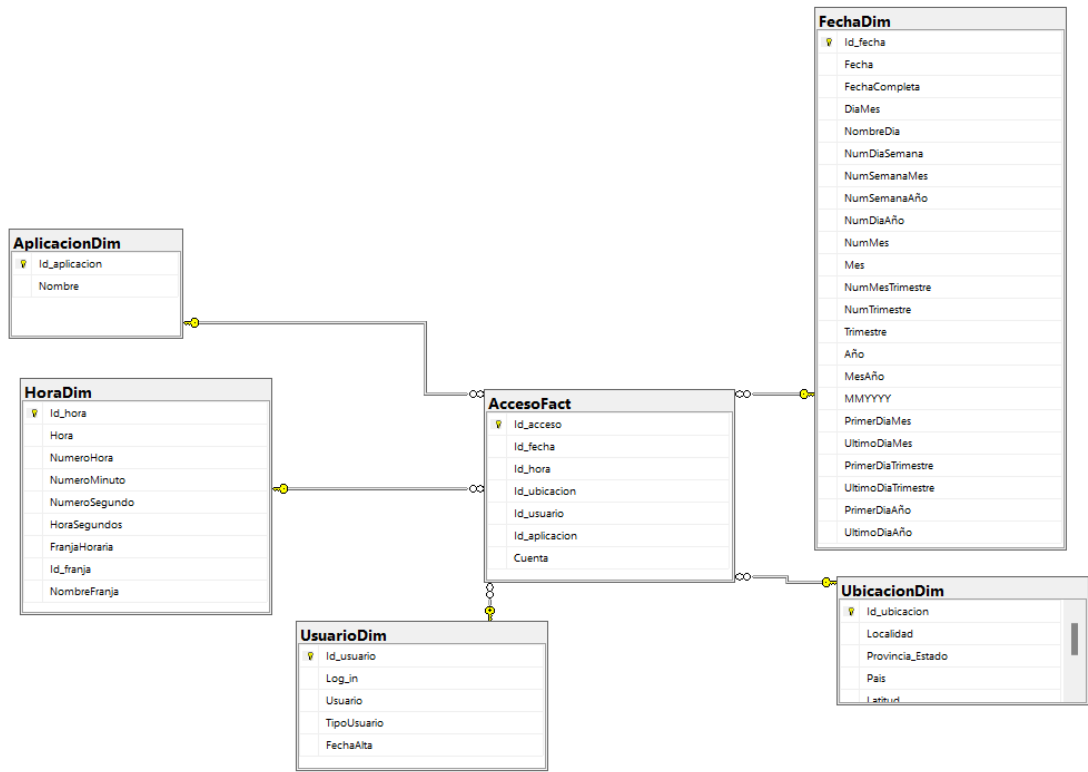


FIGURA 14. ESQUEMA EN ESTRELLA DE ACCESOS

Dadas las consultas indicadas en la sección de “Especificación de requisitos” y los campos de las tablas de dimensión y de hechos que conforman el esquema dimensional, se define una matriz que establece las correspondencias entre ellos. Esta se expone mediante la Tabla 2.

	AplicacionDim	HoraDim	UsuarioDim	UbicacionDim	FechaDim	AccesoFact
A001		NumeroHora, FranjaHoraria			Año, Mes, Trimestre, DiaMes	Cuenta
A002	Nombre		Usuario			Cuenta
A003			Usuario	Provincia_Estado		Cuenta
A004			Usuario	Localidad		Cuenta
A005	Nombre		TipoUsuario			Cuenta
A006		NumeroHora, FranjaHoraria				Cuenta
A007		NumeroHora, FranjaHoraria	Usuario		Año, Mes, Trimestre, DiaMes	Cuenta

TABLA 2. RELACIÓN ENTRE LAS CONSULTAS Y LAS TABLAS DEL ESQUEMA EN ESTRELLA DE ACCESOS

4.3.5 Cubos OLAP

Como paso posterior al diseño del esquema dimensional, se procede a la construcción del cubo OLAP con el objetivo de obtener un mayor rendimiento y eficiencia durante las agregaciones de datos. Para ello, se crea un proyecto SQL Server Analysis Services en Microsoft Visual Studio y se emplea como origen la base de datos “DWLogistics” creada anteriormente.

En segundo lugar, se establece una vista del origen de datos (Figura 15) con todas las tablas que forman la estrella y que estarán incluidas en el cubo de accesos.

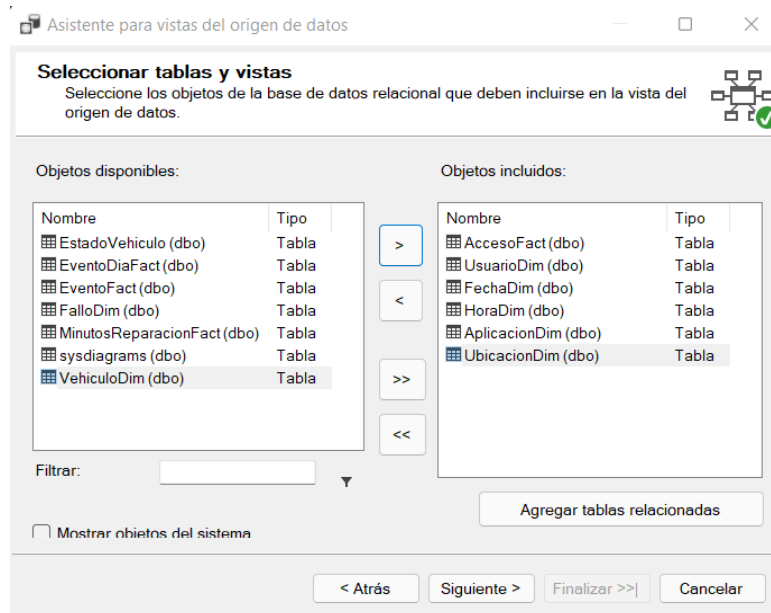


FIGURA 15. VISTA DEL ORIGEN DE DATOS DE ACCESOS

Tras la definición de la vista, se establecen aquellas tablas que presentan un grupo de medidas y cuáles son las medidas que se van a incluir en el cubo. En este caso la única tabla de hechos de la estrella (AccesoFact) y su única medida de nombre “Cuenta”. Como último paso antes de la construcción y ejecución del cubo, el resto de las tablas son seleccionadas como dimensiones quedando el cubo con la estructura que se muestra en la Figura 16.

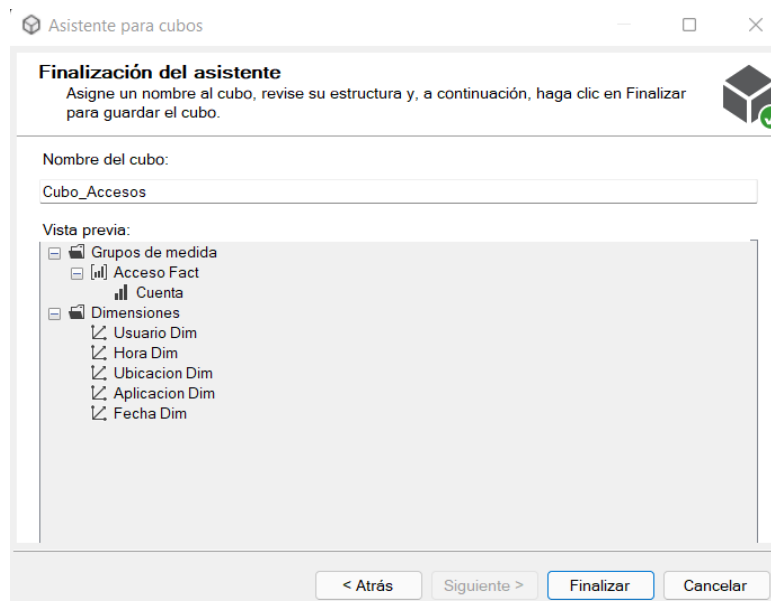


FIGURA 16. ESTRUCTURA DE CUBO_ACCESOS

Antes de llevar a cabo las consultas necesarias, se realizan modificaciones de los objetos mediante la interfaz de diseño para mejorar el cubo. Varias jerarquías (Figura 17) son establecidas en las tablas de dimensión exceptuando “AplicacionDim” dada su escasa información. Conforme a la explicación de Ralph Kimball, este tipo de estructura mejora la navegación entre los datos de la dimensión y aumenta el rendimiento de las consultas. Asimismo, facilita el filtrado de los datos a únicamente estos grupos (Mundy, Thornthwaite, & Kimball, 2011). Se concretan jerarquías relacionadas con las fechas y el tiempo en “FechaDim” y “HoraDim”, respectivamente. En el caso de la segunda dimensión, ambas son prácticamente idénticas diferenciándose en que una de ellas contiene, además, la hora.

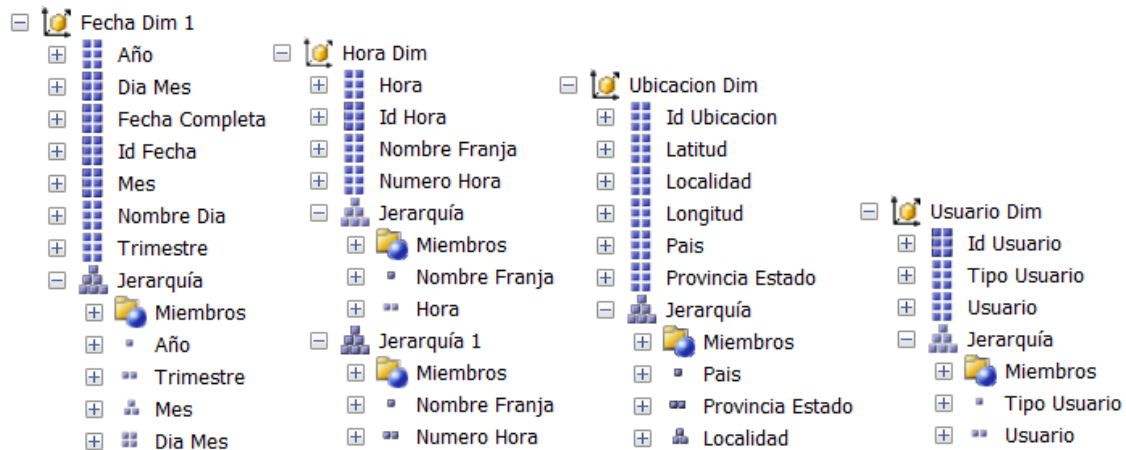


FIGURA 17. JERARQUÍAS DE CUBO_ACCESOS

La jerarquía de “UbicaciónDim” presenta una organización geográfica tradicional sobre la localización desde donde se produce la conexión a la aplicación. Por otro lado, la correspondiente a “UsuarioDim” agrupa al usuario con su tipo.

4.3.6 Procesos ETL

El proceso de extracción, transformación y carga de datos afecta a los tres primeros componentes del sistema *Business Intelligence*. Durante este procedimiento se emplean diferentes técnicas dependiendo del objetivo y los datos a tratar.

Para la extracción de los datos de las fuentes externas y su volcado en la base de datos relacional que conforma el *Data Staging Area*, se hace uso del paquete SQL Server Integration Services (SSIS). Este se utiliza a través de Microsoft Visual Studio al igual que en la creación de los cubos. En primer lugar, se crea un proyecto de nombre “ETL_DWLogistics” desde donde gestionar todos los procesos. Luego se crean y administran todas las conexiones tanto a ficheros Microsoft Excel como a bases de datos. Para la primera extracción se especifica la ruta de acceso a la fuente “Informe_de_accesos.xls” fijando como nombre de la conexión “Excel_accesos”. De forma análoga se instaura un vínculo del tipo OLEDB con el destino de los datos, es decir, “DataLogistics”. Seguidamente, un paquete SSIS es creado estableciendo una tarea de flujo de datos que permita trasladar toda la información de la fuente al *Data Stage*. Dentro de esta tarea se seleccionan los campos que se muestran en la Figura 18. Dado que en el informe los datos comienzan varias filas después a la inicial es necesario modificar la propiedad *OpenRowSet* del origen. Esta permite indicar donde se encuentran los registros que se quieren obtener. El valor que adquiere es “Hoja1\$A4:E” al haber 5 columnas (A a E) y comenzar desde la fila 4. Posteriormente, los objetos en el flujo de datos son enlazados y se configuran las asignaciones entre columnas. Como se aprecia en la Figura 18, existe una relación 1 a 1 entre los campos exceptuando “Cargado”, que no recibe ningún dato al tener por defecto valor 0. Además, se establece el tipo de dato de las columnas de salida como “cadena Unicode” para que coincidan con las de entrada de la base de datos.

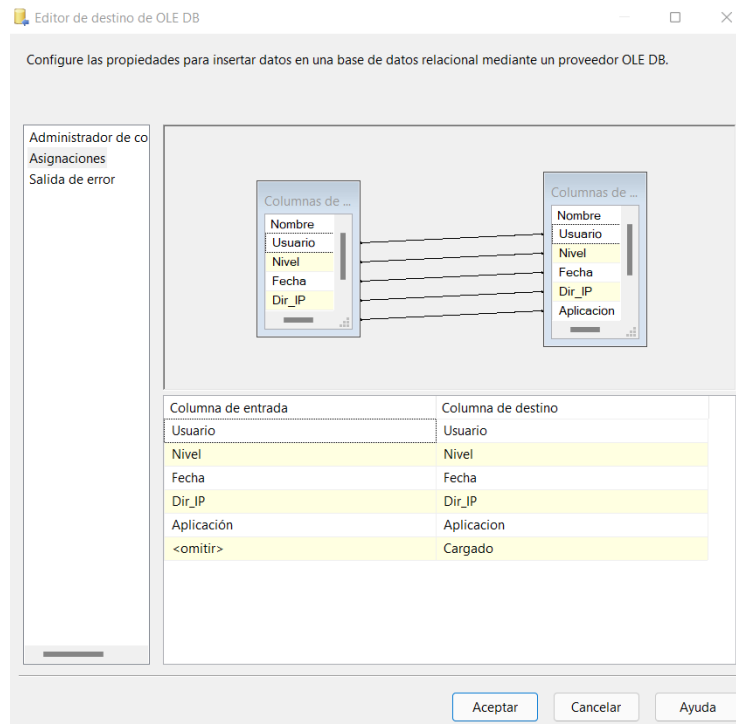


FIGURA 18. RELACIÓN ENTRE LA FUENTE EXTERNA Y EL DATA STAGE DE ACCESOS

Como último paso se inicia el proceso para la transferencia de todos los datos hasta el *Data Stage* como se refleja en el esquema de la Figura 19.

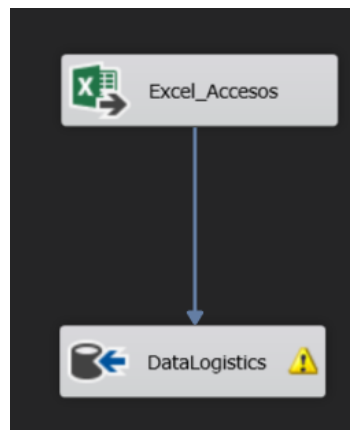


FIGURA 19. CARGA DE LA FUENTE EXTERNA AL DATA STAGE DE ACCESOS

Tras la carga de la base de datos relacional se continua con el volcado de los datos en el esquema dimensional. Para ello, se emplean procedimientos almacenados (Figura 20) con instrucciones Transact-SQL de manipulación de datos. Estos se basan en sentencias "INSERT" y "SELECT" para recoger los valores del *Data Stage* (DWLogistics) e introducirlos en la estrella. Los procedimientos siguen un determinado tipo de estructura dependiendo si están destinados a la tabla de hechos o a las dimensiones. En ambas se comprueba que cada tupla tiene la columna "Cargado" con valor 0 para evitar almacenar registros duplicados. Además, se verifica que los campos únicos no estén ya presentes en el caso de las tablas de dimensión.

```

CREATE OR ALTER PROCEDURE cargaAplicacionDim AS BEGIN
INSERT INTO DWLogistics.dbo.AplicacionDim(Nombre)
SELECT DISTINCT Aplicacion
FROM DataLogistics.dbo.InformeAccesos
WHERE Cargado = 0
AND Aplicacion NOT IN (SELECT Nombre FROM AplicacionDim);
END
GO

EXEC cargaAplicacionDim
GO

```

FIGURA 20. PROCEDIMIENTO DE CARGA DE APLICACIONDIM

En cuanto a la carga de las dimensiones relativas al tiempo, se usan procedimientos almacenados ya predefinidos que se encargan de generar las fechas o horas existentes entre periodos. La llamada a uno de estos procedimientos se muestra en la Figura 21. Además, forman los descriptores textuales reflejados en la Figura 14 como la franja horaria o el trimestre.

```

SET DATEFORMAT dmy
EXEC cargaFechaDim '01/04/2021', '31/05/2022'
GO

```

FIGURA 21. LLAMADA AL PROCEDIMIENTO DE CARGA DE APLICACIONDIM

Por otra parte, el procedimiento “cargaAccesoFact”, que se exhibe en la Figura 22, inserta los identificadores de cada una de las dimensiones junto con la medida “Cuenta” en la tabla de hechos. Esta emplea los campos únicos para ello. Como último paso actualiza el valor de “Cargado” a 1 indicando la presencia de la fila en el *Data Area*.

```

CREATE PROCEDURE cargaAccesoFact AS BEGIN
INSERT INTO DWLogistics.dbo.AccesoFact(Id_fecha, Id_hora, Id_ubicacion, Id_usuario, Id_aplicacion)
SELECT fechas.Id_fecha
, horas.Id_hora
, ubicaciones.Id_ubicacion
, usuarios.Id_usuario
, aplicaciones.Id_aplicacion
FROM DataLogistics.dbo.InformeAccesos informe,
DWHinoV2.dbo.UsuarioDim usuarios,
DWHinoV2.dbo.UbicacionDim ubicaciones,
DWHinoV2.dbo.FechaDim fechas,
DWHinoV2.dbo.HoraDim horas,
DWHinoV2.dbo.AplicacionDim aplicaciones
WHERE informe.Cargado = 0 AND informe.Usuario=usuarios.Usuario AND
informe.Dir_IP=ubicaciones.Dir_IP AND
convert(date, informe.Fecha, 103)=fechas.FechaCompleta AND
convert(time, informe.Fecha)=horas.Hora AND
informe.Aplicacion=aplicaciones.Nombre;
UPDATE DataHino.dbo.InformeFallos SET Cargado = 1 WHERE Cargado = 0;
END
GO

EXEC cargaAccesoFact
GO

```

FIGURA 22. PROCEDIMIENTO DE CARGA DE ACCESOFACT

A diferencia del resto de tablas, en la dimensión “UbicacionDim” solo se utiliza el procedimiento almacenado con el objeto de introducir las direcciones IP quedando el resto de los descriptores vacíos. Esto se debe a la necesidad de emplear una API que permita obtener los datos geográficos asociados a dicha IP. Dado que la herramienta SQL Server Integration Services facilita la creación de *scripts* se ha optado por realizar este proceso mediante su uso como se muestra en la Figura 23. En primer lugar, se especifica una conexión del tipo OLEDB a la base de datos “DWLogistics” para extraer las direcciones

guardadas con las instrucciones SQL (Obtener IPs). Esta conexión es empleada posteriormente para toda la información en la tabla “UbicacionDim” (DWLogistics).



FIGURA 23. CARGA DE UBICACIÓNDIM MEDIANTE LA IP GEOLOCATION API

Como se puede observar en la Figura 24, por cada uno de los identificadores IP se ejecuta una secuencia de comandos en Microsoft Visual C# que hace uso de la *IP Geolocation API* del portal (ipwhois.io, 2022). El motivo de su elección reside en el rendimiento que presenta y el escaso número de fallos que genera siendo, además, gratuita. Permite obtener la ciudad, la región, el país, la latitud y la longitud ligada a la dirección. Si existe una respuesta satisfactoria por parte del servidor, las variables son almacenadas en *buffers* de salida para su inserción en la base de datos de destino (DWLogistics).

```

public override void CreateNewOutputRows()
{
    var ip = Variables.IP;
    string servicio = "http://ipwhois.app/json/" + ip + "?objects=ip,city,region,country,latitude,longitude";
    System.Net.Http.HttpClient cliente = new System.Net.Http.HttpClient();
    cliente.BaseAddress = new Uri(servicio);
    cliente.DefaultRequestHeaders.Accept.Add(new System.Net.Http.Headers.MediaTypeWithQualityHeaderValue("application/json"));
    string APIUrl = string.Format(servicio);

    var respuesta = cliente.GetAsync(APIUrl).Result;
    if (respuesta.IsSuccessStatusCode)
    {
        var resultado = respuesta.Content.ReadAsStringAsync().Result;
        var serializer = new System.Web.Script.Serialization.JavaScriptSerializer();
        var datos = serializer.Deserialize<ResultadosAPI>(resultado);
        Salida0Buffer.AddRow();
        Salida0Buffer.Pais = datos.country;
        Salida0Buffer.Localidad = datos.city;
        Salida0Buffer.Region = datos.region;
        Salida0Buffer.Latitud = datos.latitude;
        Salida0Buffer.Longitud = datos.longitude;
        Salida0Buffer.IP = datos.ip;
    }
}
  
```

FIGURA 24. SCRIPT PARA LA LLAMADA A LA IP GEOLOCATION API

4.3.7 Informes

La sección de los accesos al sistema de información se organiza en 6 informes. Cada uno de ellos abarca una perspectiva de observación con el objeto de mostrar el KPI de forma sencilla y accesible. Debido a que la mayoría de los datos se han generado de forma aleatoria a partir de la información suministrada, los resultados que se muestran no permiten obtener ninguna conclusión real.

El primer apartado de nombre “Aplicación” explica los accesos al sistema de información por parte de los usuarios. Existen dos tipos de aplicaciones. Por un lado, la plataforma web “FLOTASNET 3.0” y por otro, la aplicación móvil “app”. Como refleja la Figura 25, se emplea la visualización “matriz” para

relacionar estas aplicaciones con los tipos de usuario y las regiones o estados desde donde se accede. Los valores se muestran como porcentaje del total de fila. Así es posible conocer qué tipo de sistema de información es más usado por cada tipo de usuario o en qué área geográfica predomina una plataforma u otra. Además, se lleva a cabo un filtrado para las regiones de forma que no se muestren campos en blanco.

En la franja superior derecha de la página, se sitúa un gráfico de anillos que refleja el número total de accesos a través de cada aplicación. La persona que interactúa con el *dashboard* puede saber qué aplicación destaca de forma general o filtrando por usuario. Debajo de este gráfico es mostrado un *visual* de barras horizontales que tiene como eje vertical el campo “Usuario” ordenado alfabéticamente. Las dos aplicaciones son usadas como leyenda. Este permite saber la proporción con la que cada usuario accede a cada sistema de información.

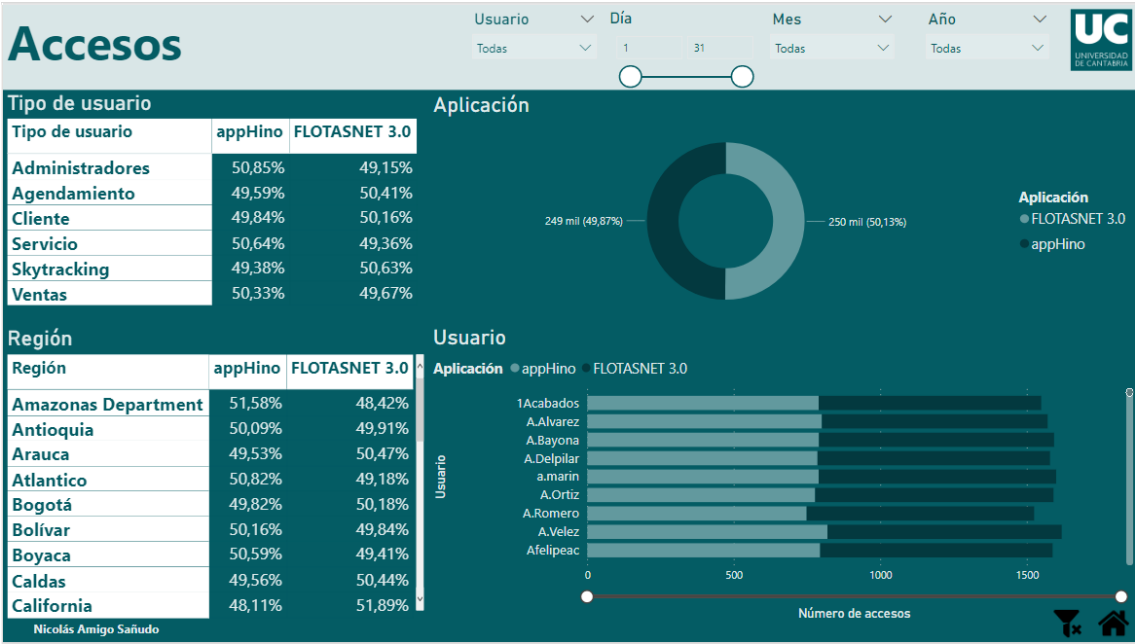


FIGURA 25. INFORME APLICACION_ACCESOS

Además de los elementos de segmentación por fecha, dependiendo de la página existen filtros basados en otros campos como “Usuario” o “Nombre” de la tabla “AplicacionDim”. También es posible gestionar la información que se expone mediante la interacción con las gráficas. Por ejemplo, se muestran los datos de una determinada región al hacer clic en su nombre de la matriz.

El informe de la Figura 26 se corresponde con una visión más global de los datos, pero específica para cada usuario ya que participan varios campos de diferentes tablas de dimensión. El gráfico de barras apiladas ofrece un enfoque geográfico de los accesos mediante la jerarquía de ubicación compuesta por “País”, “Región” y “Localidad”. Puesto que prácticamente todas las conexiones se realizan desde Colombia solo se exponen aquellas realizadas desde los estados y ciudades de dicho país. Los datos de las poblaciones de una determinada región se visualizan en el momento que se hace clic sobre la barra correspondiente. De esta forma se profundiza en la jerarquía mostrando la información con un mayor grado de detalle. Además, se activa un control deslizante de *zoom* con motivo de la diferencia del número de accesos entre las distintas ubicaciones. La gráfica de líneas de título “Fecha” muestra la evolución de los accesos del usuario a lo largo del tiempo. Por lo tanto, el eje X se puede corresponder con los años, meses o días de los accesos. En la matriz se expone la cantidad de *logins* realizados por el usuario cada día de la semana a una determinada hora. Esta figura es útil para conocer los hábitos de horario del usuario según el momento de la semana. Dada la cantidad de filas se opta por mostrar los valores mediante colores. Cuanto más oscura es la tonalidad, mayor número de accesos se realizan en la hora correspondiente. Así el usuario del *dashboard* percibe la información de una manera más rápida y

sencilla en comparación con las cifras numéricas. Si se precisa saber el porcentaje de conexiones a la aplicación en una de las franjas, basta con posicionarse encima de la celda con el cursor.

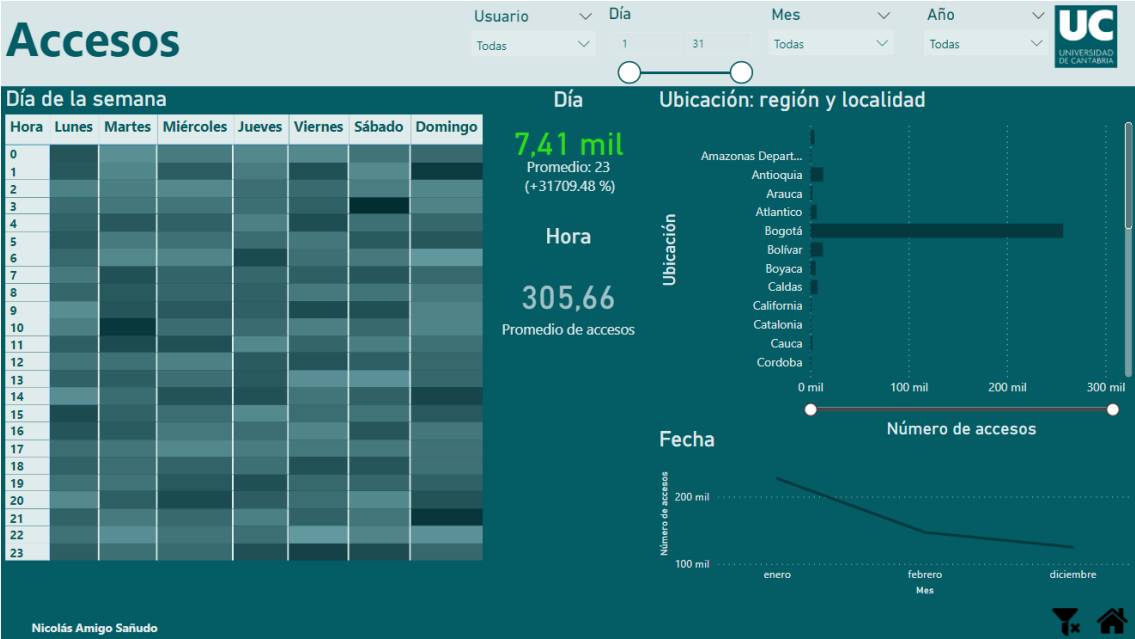


FIGURA 26. INFORME USUARIO_ACCESOS

Asimismo, se establecen dos tarjetas en la zona central del informe relativas a los accesos a la aplicación según una determinada unidad de tiempo. La visualización superior refleja el número promedio de conexiones de un usuario en un día (24 horas). Este se muestra en formato KPI comparando dicho valor con el promedio de accesos calculado entre todos los usuarios. Dependiendo la diferencia porcentual entre ambas cantidades, el valor que se muestra presenta un color u otro. Si la ratio es positiva se indica mediante un verde claro. En el caso de ser neutra, se codifica con un tono amarillo. En otro caso, aparece en color rojo. La otra tarjeta que se expone en el informe indica el promedio de accesos por hora que lleva a cabo el usuario.

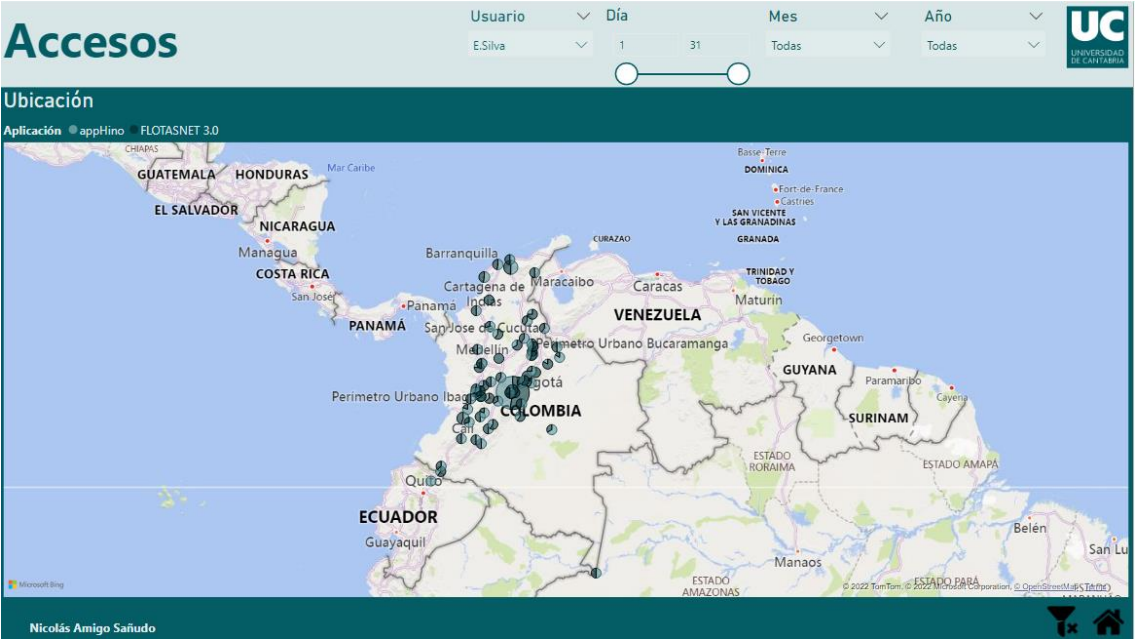


FIGURA 27. INFORME MAPA_ACCESOS

Otro informe relevante está formado exclusivamente por un mapa interactivo. Este muestra la ubicación de desde donde se realizan las conexiones a la aplicación como se puede observar en la Figura 27. Para ello, se utilizan las coordenadas obtenidas mediante la API. El tamaño de cada punto mostrado en el mapa indica la cantidad de accesos efectuados. Asimismo, los objetos que se exponen se corresponden con gráficos circulares cuya leyenda es el tipo de aplicación. Esta visualización permite conocer dónde se agrupan la mayor parte de los accesos para cada usuario y qué aplicación se usa con mayor frecuencia para acceder al sistema de información en cada zona geográfica.

El resto de informes lo componen principalmente gráficas relativas a la localización de los accesos y a las horas y franjas horarias en las que se producen. Se muestran todas sus figuras en la sección “Anexo”.

4.4 Caso de uso: fallos

4.4.1 Especificación de requisitos

Este caso de uso pretende conocer el número de notificaciones recibidas por los conductores desde sus ordenadores de abordo. Los KPI y sus perspectivas de observación se recogen en la Tabla 3.

KPI	Perspectiva
Número de notificaciones emitidas.	Vehículo Fallo Fecha

TABLA 3. KPI DE FALLOS BAJO PERSPECTIVAS DE OBSERVACIÓN

Mediante este KPI y las perspectivas asociadas se busca responder a las consultas que se precisan a continuación:

- F001 - Código de fallo más frecuente en cada modelo de vehículo.
- F002 - Promedio de notificaciones hasta la solución del fallo.
- F003 - Franja horaria en la que se produce un mayor número de notificaciones.
- F004 - Número de notificaciones en una determinada fecha.
- F005 - Número de notificaciones en una determinada hora.
- F006 - Promedio de notificaciones producidas en un periodo de tiempo y por tipo de fallo.

Cada una de las consultas anteriores se deben poder segmentar por tiempo y número de identificación de vehículo (VIN), así como por el resto de las perspectivas.

4.4.2 Fuentes externas

Como ejemplo de fuente de datos externa se hace uso de “Informe_de_fallos.xls” que recoge las notificaciones que se producen en los vehículos de la flota del 01/04/2021 a 01/05/2021. Debido al escaso número de filas suministradas se procede a su aumento mediante la función “Aleatorio()” de Microsoft Excel. De esta forma se obtiene un total de aproximadamente 7.800.000 registros para explotar con el sistema BI.

La hoja de cálculo consta de los siguientes datos:

- Fecha: indica el instante de tiempo en el que el vehículo envía la notificación del fallo que se encuentra activo. Se muestra mediante la fecha y la hora de la acción.
- VIN: del inglés *Vehicle Identification Number*. Es el número de chasis o bastidor único que identifica al vehículo.
- Tipo: indica el tipo de notificación que se ha producido (falla, mantenimiento, preventivo). No se dispone de esta información.

- **Marca:** indica el fabricante del vehículo.
- **Modelo:** identificado por un año más del de fabricación.
- **Código de error:** es una codificación correspondiente al tipo de notificación emitida.
- **Descripción:** es una breve explicación acerca del código de error detectado.

Se ha de tener en cuenta que el equipo envía una notificación cuando se detecta un problema en el vehículo y después repetidamente cada 5 minutos mientras esté activo. Por ejemplo, si falla una bombilla porque se ha fundido, el vehículo estaría enviando cada 5 minutos ese mensaje hasta que se sustituya la bombilla.

4.4.3 Base de datos *Data Stage Area*

La base de datos relacional, que funciona como *Data Staging Area*, destina la tabla “InformeFallos” a almacenar todos los datos relacionados con las notificaciones que emiten los vehículos. Como se expone en la Figura 28, las columnas de la tabla se corresponden con los datos de la fuente externa “Informe_de_fallos.xls”. Al igual que en la tabla de los accesos a la aplicación, se añade un campo “Cargado” y “FechaRegistro” para controlar aquellas filas que se encuentran presentes en el *Data Warehouse* evitando problemas de sincronización.

InformeFallos			
	Column Name	Data Type	Allow Nulls
	Fecha	nvarchar(255)	☒
	VIN	nvarchar(255)	☒
	Tipo	nvarchar(255)	☒
	Marca	nvarchar(255)	☒
	Modelo	nvarchar(255)	☒
	[Codigo de error]	nvarchar(255)	☒
	Descripción	nvarchar(255)	☒
	Cargado	bit	☒
	FechaRegistro	datetime	☒
			☐

FIGURA 28. VISTA ESTÁNDAR DE LA TABLA INFORMEFALLOS

A diferencia del resto de campos, la columna de nombre “Descripción” mantiene una longitud de 255 caracteres debido a que presenta una extensión mayor y desconocida.

4.4.4 Base de datos *Data Area*

La estrella definida para construir el modelo dimensional de las notificaciones emitidas por los vehículos posee un total de 5 tablas. La tabla de hechos denominada “EventoFact” comprende los identificadores de cada una de las dimensiones junto con la medida “Cuenta” al igual que la estrella anterior. Esta sirve como hecho aditivo para obtener un cómputo total del número de notificaciones difundidas. Las tablas de dimensión y sus descriptores se derivan a partir de las variables presentes en los informes.

A continuación, se describen los campos que forman las tablas de dimensión. Las dimensiones de tiempo se omiten al ser idénticas a las que forman la estrella de accesos.

- **VehiculoDim:** representa los vehículos que componen las diferentes flotas. Permite conocer qué modelos y vehículos experimentan un mayor número de fallos y de qué tipo. Se expone información adicional sobre esta dimensión en la Figura 29.

VehiculoDim				
	Column Name	Condensed Type	Nullable	Description
	Id_vehiculo	int	No	Identificador del vehículo.
	VIN	varchar(25)	Yes	Vehicle Identification Number.
	Marca	varchar(25)	Yes	Nombre del fabricante del vehículo.
	Modelo	varchar(25)	Yes	Año de fabricación del vehículo.

FIGURA 29. VISTA PERSONALIZADA DE LA TABLA VEHICULODIM

- **FalloDim:** representa los tipos de notificaciones que emiten los vehículos. Únicamente se dispone del nombre por lo que el resto de los campos son nulos. Se expone información adicional sobre esta dimensión en la Figura 30.

FalloDim				
	Column Name	Condensed Type	Nullable	Description
	Id_fallo	int	No	Identificador del fallo/notificación.
	Codigo_fallo	varchar(50)	No	Codificación del fallo/notificación.
	Descripcion	varchar(250)	No	Descripción del fallo/notificación.
	Tipo	varchar(50)	Yes	Tipo de fallo/notificación.

FIGURA 30. VISTA PERSONALIZADA DE LA TABLA FALLODIM

El diagrama en estrella queda constituido una vez se crean todas las tablas de dimensión y la tabla de hechos como se refleja en la Figura 31.

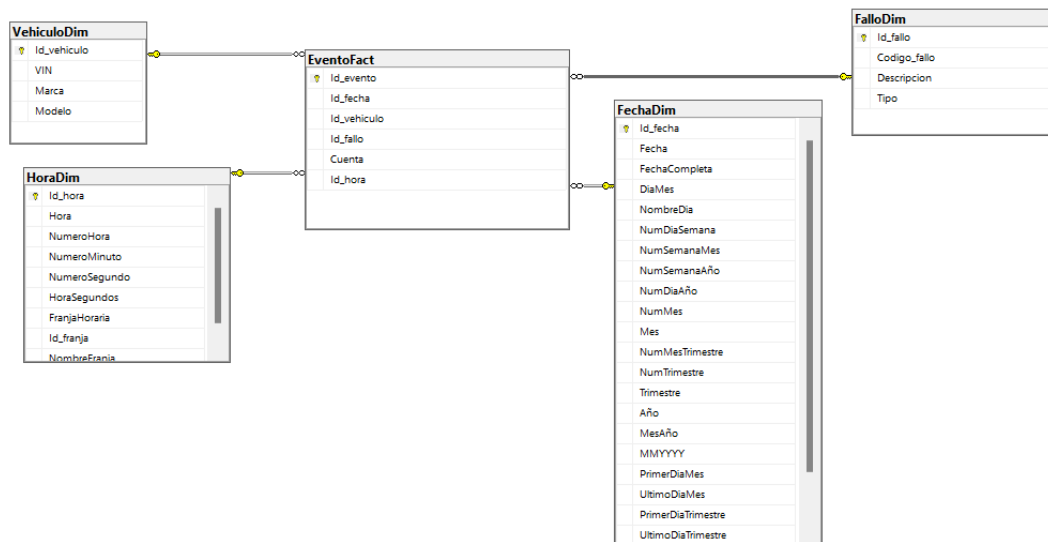


FIGURA 31. ESQUEMA EN ESTRELLA DE FALLOS

A continuación, en la Tabla 4 se establece una matriz con las correspondencias entre las consultas definidas y los campos de la tabla de hechos y las dimensiones del esquema en estrella.

	VehiculoDim	HoraDim	FechaDim	FalloDim	EventoFact
F001	Modelo			Codigo_fallo	Cuenta
F002			FechaCompleta		Cuenta

F003		FranjaHoraria			Cuenta
F004			Año, Trimestre, Mes, DiaMes		Cuenta
F005		NumeroHora			Cuenta
F006		Hora, NumeroHora	Año, Trimestre, Mes, DiaMes	Codigo_fallo	Cuenta

TABLA 4. RELACIÓN ENTRE LAS CONSULTAS Y LAS TABLAS DEL ESQUEMA EN ESTRELLA DE FALLOS

4.4.5 Cubos OLAP

Al igual que en el caso anterior de los accesos a la aplicación, se crea un proyecto SQL Server Analysis Services en Microsoft Visual Studio y se emplea como origen la base de datos “DWLogistics” creada anteriormente.

En segundo lugar, se establece una vista del origen de datos con todas las tablas que forman la estrella y que estarán incluidas en el cubo de fallos. En comparación con el cubo anterior, se mantienen las tablas de tiempo reduciéndose a 5 el número de total como se puede observar en la Figura 32.

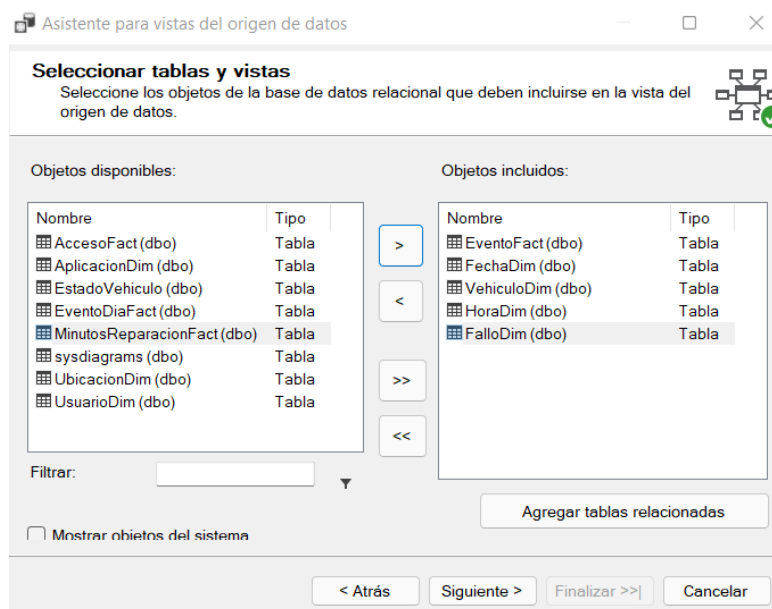


FIGURA 32. VISTA DEL ORIGEN DE DATOS DE FALLOS

Como último paso antes de la construcción y ejecución del cubo, se establecen aquellas tablas que presentan un grupo de medidas y cuáles son las medidas que se van a incluir en el cubo. La tabla de hechos se corresponde con “EventoFact” y su medida con “Cuenta”. Después, el resto de las tablas son seleccionadas como dimensiones quedando el cubo con la estructura que se muestra en la Figura 33.

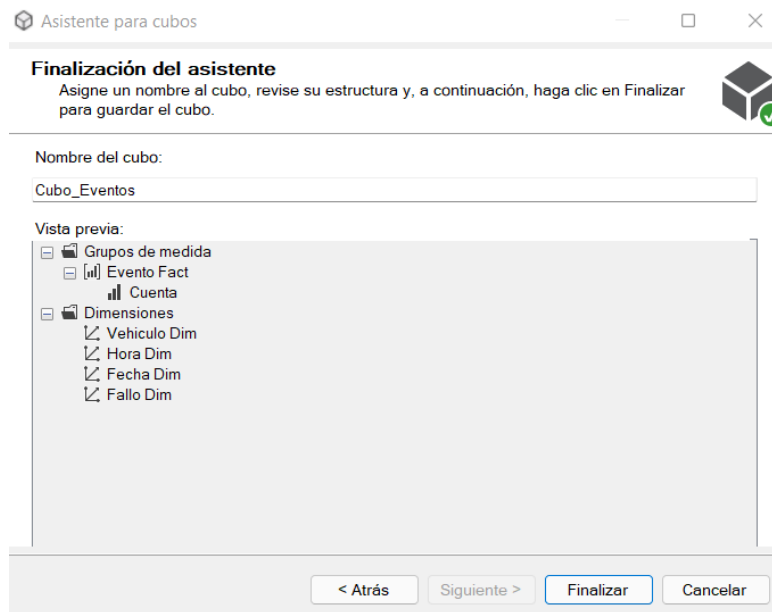


FIGURA 33. ESTRUCTURA DE CUBO_EVENTOS

De nuevo varias jerarquías se establecen mediante la interfaz de diseño para mejorar la accesibilidad al cubo. Estas se pueden observar en la Figura 34. En este caso todas las dimensiones cuentan con al menos una jerarquía. Se concretan jerarquías relacionadas con las fechas y el tiempo en “FechaDim” y “HoraDim”, respectivamente (ya mostradas en la sección previa).

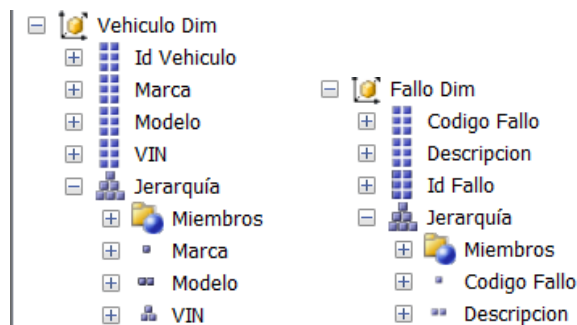


FIGURA 34. JERARQUÍAS DE CUBO_EVENTOS

La jerarquía de “VehiculoDim” organiza la estructura de identificación de los automóviles desde su marca a su número de bastidor (VIN). En cuanto a la tabla “FalloDim”, se agrupa cada código con su descripción.

4.4.6 Procesos ETL

De igual manera que en el proceso ETL de “Accesos” se emplea el paquete SQL Server Integration Services (SSIS) para el volcado de datos desde la fuente externa a la base de datos relacional. Desde el proyecto de nombre “ETL_DWLogistics” creado anteriormente, se especifica la ruta de acceso a la fuente “Informe_de_fallos.xls” fijando como nombre de la conexión “Excel_fallos”. Seguidamente, un paquete SSIS es creado y se establece una tarea de flujo de datos que permita trasladar toda la información de la fuente a “DataLogistics”. Dentro de esta tarea se seleccionan los campos que se muestran en la Figura 35.

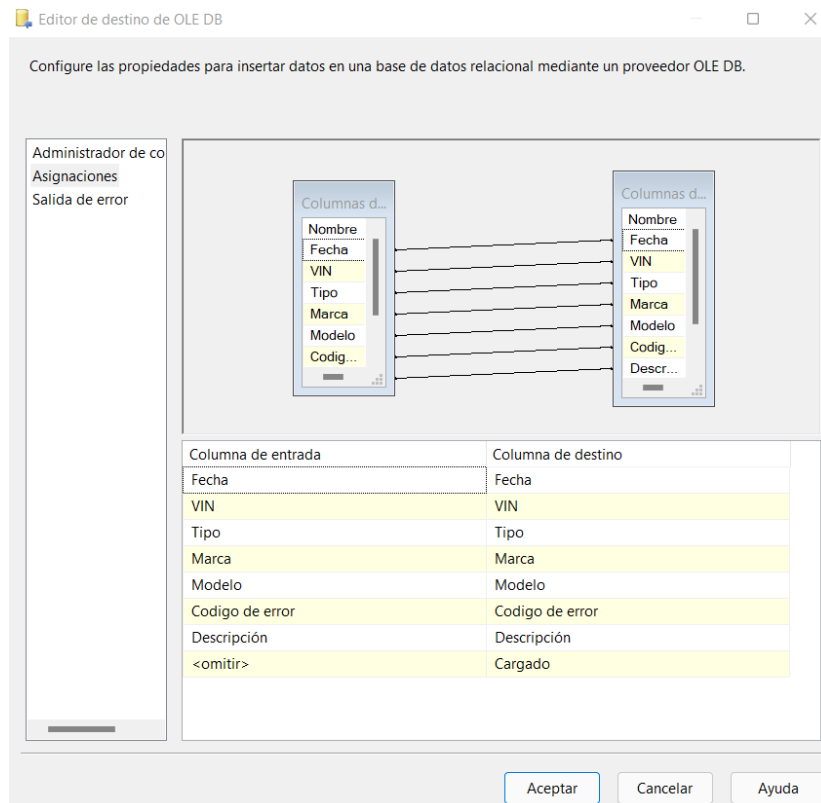


FIGURA 35. RELACIÓN ENTRE LA FUENTE EXTERNA Y EL *DATA STAGE* DE FALLOS

La propiedad *OpenRowSet* del origen de datos se modifica a "Hoja1\$A3:G" al haber 7 columnas (A a G) y comenzar desde la fila 3. Posteriormente, los objetos en el flujo de datos son enlazados y se configuran las asignaciones entre columnas. De la misma manera que en el caso anterior, existe una relación 1 a 1 entre los campos exceptuando "Cargado" con valor 0.

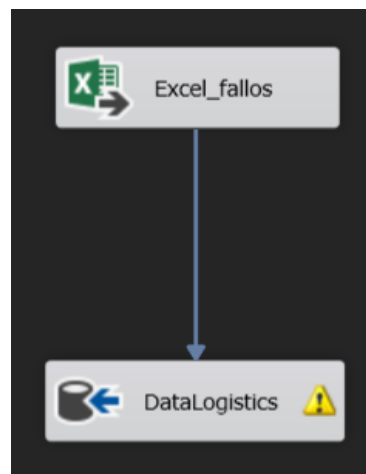


FIGURA 36. CARGA DE LA FUENTE EXTERNA AL *DATA STAGE* DE FALLOS

Como último paso se inicia el proceso para la transferencia de todos los datos hasta el *Data Stage* como se refleja en el esquema de la Figura 36.

Tras la carga de la base de datos relacional se continua con el volcado de los datos en el esquema dimensional mediante los procedimientos SQL. Solamente es necesaria la carga de 3 de las 5 tablas que componen la estrella ya que dos de ellas (FechaDim y HoraDim) son comunes a "Accesos". La estructura de los procedimientos es idéntica a la empleada antes como se puede observar en la Figura 20 y 22 en comparación con la Figura 37 y 38.

```

CREATE OR ALTER PROCEDURE cargaVehiculoDim AS BEGIN
INSERT INTO DWLogistics.dbo.VehiculoDim(VIN,Marca,Modelo)
SELECT DISTINCT VIN,Marca,Modelo
FROM DataLogistics.dbo.InformeFallos
WHERE Cargado = 0
AND VIN NOT IN (SELECT VIN FROM VehiculoDim);
END
GO

EXEC cargaVehiculoDim
GO

```

FIGURA 37. PROCEDIMIENTO DE CARGA DE VEHICULO DIM

```

CREATE OR ALTER PROCEDURE cargaEventoFact AS BEGIN
INSERT INTO DWLogistics.dbo.EventoFact(Id_fecha, Id_vehiculo, Id_fallo, Id_hora)
SELECT fechas.Id_fecha
, vehiculos.Id_vehiculo
, fallos.Id_fallo
, horas.Id_hora
FROM DataLogistics.dbo.InformeFallos informe,
DWHinoV2.dbo.FechaDim fechas,
DWHinoV2.dbo.VehiculoDim vehiculos,
DWHinoV2.dbo.FalloDim fallos,
DWHinoV2.dbo.HoraDim horas
WHERE informe.Cargado = 0 AND informe.VIN=vehiculos.VIN AND
convert(datetime,convert(date,convert(datetime,informe.Fecha),112))=fechas.Fecha AND
informe.[Codigo de error]=fallos.Codigo_fallo AND convert(time(0),informe.Fecha)=horas.Hora;
UPDATE DataHino.dbo.InformeFallos SET Cargado = 1 WHERE Cargado = 0;
END
GO

```

FIGURA 38. PROCEDIMIENTO DE CARGA DE EVENTO FACT

4.4.7 Informes

Se han creado 4 informes para visualizar el KPI correspondiente. Tres de ellos están basados en una determinada perspectiva de observación: modelo del vehículo, fecha del fallo y VIN del vehículo. El cuarto muestra una matriz junto con una gráfica de barras que permite interpretar visualmente las cifras que se muestran en la tabla.

Como se observa en la Figura 39, el primer informe se asocia a los modelos de los automóviles y es constituido por 3 gráficas. En la parte superior izquierda se sitúa un gráfico de columnas apiladas que exhibe el número de notificaciones para los 3 fallos más frecuentes por modelo de vehículo. Las barras son agrupadas por las descripciones de dichos fallos. La que presenta un color más oscuro se corresponde con el total de avisos emitidos y la otra con la cantidad ligada al modelo que se muestra. Se puede cambiar de modelo mediante los botones que se encuentran encima de la gráfica.

En la zona inferior del informe se encuentra una figura de barras apiladas horizontales que muestra los fallos más comunes en los modelos. Cada columna se divide por las notificaciones por los vehículos correspondientes a cada modelo. Además, se hace uso para esta gráfica de un recurso denominado *tooltip* o etiqueta. Esta función permite mostrar un conjunto de datos personalizados cuando se ubica el cursor sobre el campo deseado. Como se puede percibir en la Figura 39, esta característica se presenta mediante una pequeña ventana a modo de *pop-up*. Esta enseña el número de notificaciones asociado al código de fallo de la barra. La cifra se corresponde al modelo sobre el que se sitúa el puntero.

El *visual* restante refleja el número de fallos pertenecientes a cada año de fabricación del vehículo. Para ello, se utiliza un gráfico de anillos.

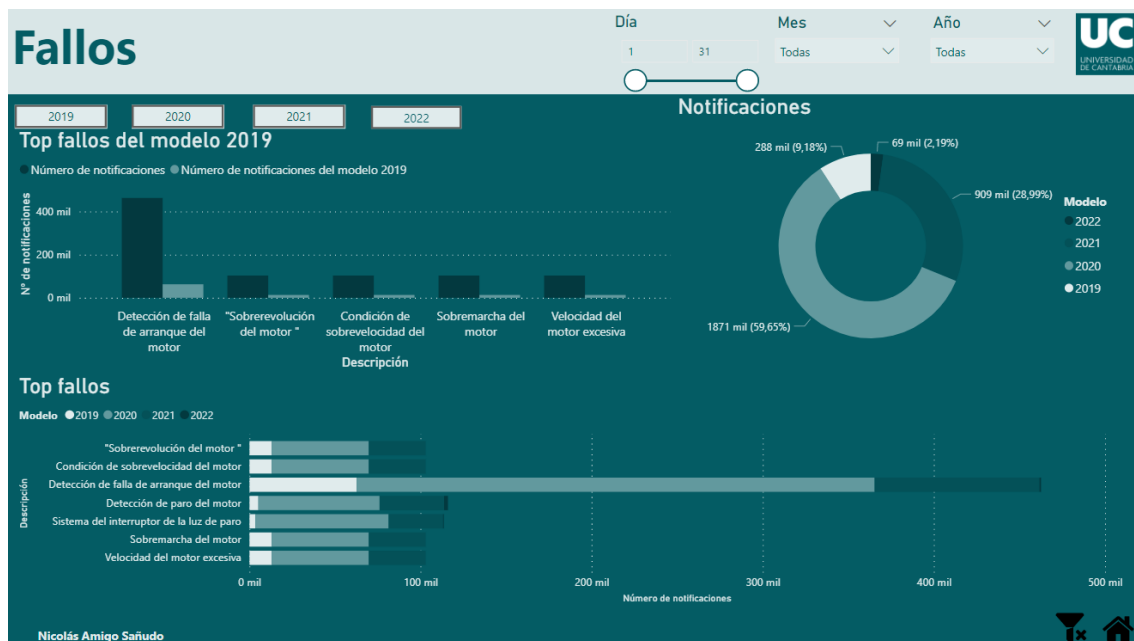


FIGURA 39. INFORME MODELO_FALLOS

El informe en su conjunto ayuda a responder consultas tales como los fallos más habituales para cada modelo, los modelos en los que se generan un mayor número de notificaciones y la distribución porcentual del fallo en cada modelo.

El *dashboard* que se expone en la Figura 40 se centra en la visualización de los fallos desde diferentes perspectivas de observación de forma específica para cada vehículo. Para ello, un elemento de segmentación de datos basado en el campo "VIN" es establecido en la franja superior del informe.

En la zona inmediatamente inferior, se exponen dos gráficas relacionadas con la fecha y el periodo de tiempo en el que se emiten las notificaciones. La primera es una gráfica de líneas que relaciona el número de fallos con la jerarquía de fecha. Esta es constituida por el año, el mes y el día en el que se produce. Además, se añade el campo "Hora". De esta manera, se observa la evolución de los fallos en diferentes unidades de tiempo. Es posible realizar la exploración de las fechas haciendo clic en cada una de ellas. Así se profundiza en la jerarquía pasando de observar, por ejemplo, el día de un mes al conjunto de horas que presenta. La segunda gráfica del área superior consiste en un *treemap* que muestra las notificaciones agrupadas en las franjas horarias en las que se han emitido. Cuánto más espacio ocupa una sección en el diagrama de árbol, mayor número de fallos reproducidos. Se han definido tres franjas horarias: mañana (06:00 a 13:59), tarde (14:00 a 21:59) y noche (22:00 a 05:59). Cada una de ellas presenta igual número de horas.

En la sección inferior del informe, se emplea un gráfico circular que relaciona los tipos de fallos que se desarrollan en el vehículo seleccionado. Se indica el código correspondiente a cada uno de ellos. Este *visual* permite conocer qué tipo de fallo se produce en un determinado vehículo y cuáles predominan. A su derecha, se ubica una tarjeta que facilita la visualización del promedio de notificaciones por día que presenta el vehículo. El informe favorece la resolución de diferentes consultas ligadas a un automóvil dado la variedad de perspectivas. Es posible responder cuestiones tales como las fechas y las franjas donde se originan más fallos, así como relacionar dichos momentos de tiempo a un determinado código de fallo.

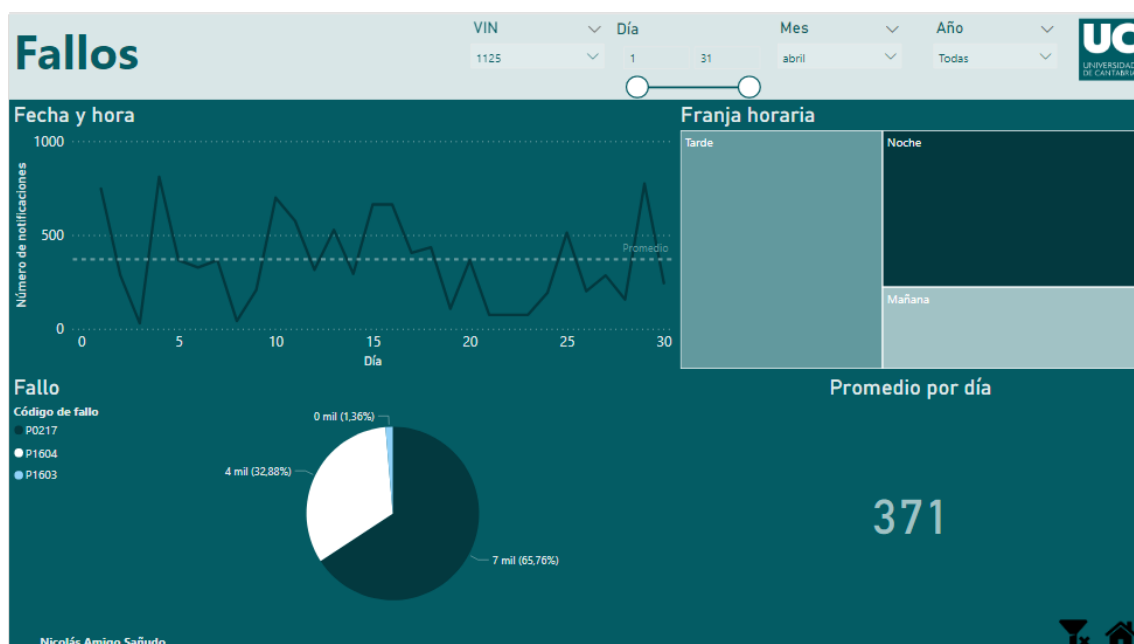


FIGURA 40. INFORME VIN_FALLOS

En cuanto al resto de informes, se basan en las fechas de las notificaciones y su relación con el año de fabricación de los vehículos y el tipo de fallo. Para ello, se emplean de nuevo gráficas que facilitan la interpretación de los resultados.

4.5 Caso de uso: trayectos

4.5.1 Especificación de requisitos

Este caso de uso pretende conocer las diferentes métricas relativas a los tramos de conducción realizados por conductores en sus rutas. Los KPI y sus perspectivas de observación se recogen en la Tabla 5.

KPI	Perspectiva
Distancia recorrida	Vehículo
Velocidad media	Perfil
Tiempo recorrido	Conductor
Calificación	Fecha
Consumo	
Ubicación	

TABLA 5. KPIs DE TRAYECTOS BAJO PERSPECTIVAS DE OBSERVACIÓN

Mediante estos KPI y las perspectivas asociadas se busca responder, entre otras, a las consultas de los tramos que se precisan a continuación:

- Tiempo promedio de un conductor por fecha.
- Tiempo en ralentí de los trayectos de un conductor.
- Velocidad media de los vehículos de un perfil.
- Distancia recorrida por un vehículo en una fecha.
- Protección del medio ambiente (ecología) de los vehículos de un perfil.
- Relación entre el consumo y los acelerones/frenadas.

- Zona geográfica desde donde se realizan más tramos por vehículo.
- Distancia media recorrida desde una localidad.

4.5.2 Fuentes externas

Se emplea el fichero “tramos_muestra.csv” como fuente de datos externa. Este contiene información relativa a la conducción de los vehículos de una serie de talleres realizada por los conductores. Los datos se agrupan en tres áreas principales. Por un lado, “Consumo” abarca registros como los litros consumidos o el tiempo del motor en ralentí. El área de “Productividad” se corresponde con campos asociados a la velocidad media, las horas de conducción y la distancia recorrida. En cuanto al grupo “Uso”, este comprende información acerca del uso del embrague y el número de inicios de marcha, entre otros. Además, existen también registros que valoran el manejo del automóvil en términos de eficiencia y sostenibilidad.

El fichero alberga un total de 20.631 filas y 147 columnas. Tras un análisis realizado por un compañero del grupo de investigación y que se recoge en su informe (Dintén Herrero, 2022), se descartan los campos que no son versión 3 y se seleccionan aquellos cuyo valor “error1” es nulo.

A continuación, se describen las variables que se emplean de la fuente externa:

- TiempoMotor3: tiempo en que el motor se encuentra arrancado.
- TiempoConduccion3: tiempo en que el vehículo está en movimiento (velocidad distinta a 0) y con el motor arrancado.
- TiempoRal3: tiempo durante el cual el vehículo está con el motor arrancado pero parado (velocidad a 0).
- Fecha_EVT_Ini: fecha y hora en la que inicia el tramo.
- Fecha_EVT_Fin: fecha y hora en la que finaliza el tramo.
- VelocidadMedia: velocidad media, en kilómetros por hora, entre el primer arranque de motor del tramo y la última parada de motor.
- Conductor: nombre del conductor que realiza el tramo haciendo uso del vehículo indicado.
- Distancia: distancia, en kilómetros, recorrida durante el tramo.
- Ecología: calificación sobre 100 que valora los aspectos de conducción relacionados con la protección del medio ambiente y contaminación.
- Eficiencia: calificación sobre 100 que valora la eficiencia energética, el consumo de combustible y el desgaste de elementos fungibles en el tramo.
- Seguridad: calificación sobre 100 que valora los aspectos de conducción relacionados con la prevención de riesgos para el conductor y para la seguridad vial en el tramo.
- Calidad: calificación sobre 100 que valora una conducción de calidad, respetando las normas de tráfico teniéndose en cuenta el buen mantenimiento del vehículo.
- Zona: taller al que pertenece el vehículo que realiza el tramo.
- AceleracionesBruscasTOut: número de frenadas bruscas que se han producido en el tramo.
- Consumo100: litros de combustible consumidos cada 100 kilómetros.
- FrenadasBruscasOut: número de frenadas bruscas que se han producido en el trayecto.
- NEmbrague3: número de veces que se acciona el embrague del vehículo.
- LatIni: latitud geográfica desde donde se inicia el tramo.

- LatFin: latitud geográfica desde donde se termina el tramo.
- LonIni: longitud geográfica desde donde se inicia el tramo.
- LonFin: longitud geográfica desde donde se termina el tramo.

4.5.3 Transformación de datos

Dado que el informe trabaja directamente sobre la fuente de datos externa, las tareas de transformación de datos se llevan a cabo mediante la herramienta de acceso a datos Microsoft PowerBI. Dichas tareas se fundamentan en el cribado de las columnas y las filas que no aportan valor, la modificación de los tipos de dato, la extracción de unidades temporales a partir de las fechas y la obtención de información geográfica mediante la API de geolocalización de Google.

Siguiendo las indicaciones del informe correspondiente (Dintén Herrero, 2022), se filtran las filas que presentan un valor nulo en el campo “error1”.

Tras la carga inicial, varios campos reflejan un tipo de dato erróneo que impide mostrar los resultados correctamente en las visualizaciones. Concretamente, se cambian algunas columnas sobre información temporal que indicaban texto en lugar de fecha/hora o duración. Se lleva a cabo un proceso de transformación de unidades de tiempo a partir de esta modificación creándose nuevas variables. A continuación, se describen ejemplos de algunas de las nuevas columnas junto con los campos originales de los que proceden:

- TiempoMotor3Min (TiempoMotor3): tiempo, en minutos, en que el motor se encuentra arrancado.
- TiempoMotor3Seg (TiempoMotor3): tiempo, en segundos, en que el motor se encuentra arrancado.
- Horalni (Fecha_EVT_Ini): momento de la fecha en el que se inicia el tramo.
- Horalni_Hora (Fecha_EVT_Ini): tiempo en unidades de 60 minutos del momento de la fecha en el que se inicia el tramo.

Por último, se hace uso de la API *reverse Geocoding* de Google que permite convertir coordenadas geográficas en direcciones legibles por personas. La latitud y la longitud, desde donde se inician y terminan los tramos, son usadas como argumentos en una función que emplea el servicio de geolocalización. Mediante este método se obtiene la calle, el número, la localidad y el país correspondiente a dichas coordenadas. Por ejemplo, se obtienen los siguientes campos para la localización del comienzo del trayecto:

- calle_ini (LatIni/LonIni): calle y número desde donde se inicia el tramo.
- localidad_ini ((LatIni/LonIni)): localidad desde donde se inicia el tramo.
- país_ini (LatIni/LonIni): país desde donde se inicia el tramo.

Las llamadas a la API producen un conjunto de fallos que dan lugar a filas inoperantes que deben ser eliminadas. Se entiende que es producido a la existencia de coordenadas que carecen de una dirección postal reconocida por el servicio de Google. Debido a esto y al cribado de otras filas erróneas, finalmente restan unas 9000 filas de las 20.631 que se encuentran en el fichero externo.

4.5.4 Informes

Los KPI correspondientes a la información de los tramos de los vehículos se visualizan en 7 informes. Cada uno de ellos explica un indicador clave de rendimiento (KPI) bajo las perspectivas de observación de “Perfil”, “Vehículo”, “Fecha” y “Conductor”. Aquellos relacionados con el tiempo del trayecto, la velocidad media y la distancia recorrida presentan la misma estructura. Por ello, se analiza únicamente el primer caso como ejemplo de los demás.

El informe “Tiempo”, que se muestra en la Figura 41, explica la duración de los tramos mediante la variable que mide el tiempo durante el que está activo el motor. Está constituido por un conjunto de tarjetas en la franja superior. Estas exponen el tiempo total recorrido, el tiempo promedio por trayecto y la proporción del tiempo correspondiente a la conducción y al ralenti. Para ello, se crea una nueva medida en formato “Porcentaje” que divide la variable pertinente entre el tiempo de motor. Las dos primeras tarjetas indican dicho tiempo en minutos. Estas variables, junto con el resto de los gráficos, pueden ser filtradas por los elementos de segmentación acordes a las perspectivas de observación citadas anteriormente.

La visualización de la izquierda expone la evolución del tiempo total recorrido por día. Se emplea un gráfico de columnas apiladas y de líneas que agrupa el tiempo correspondiente a la conducción y al ralenti usando la línea para la duración del motor activo. A su derecha se sitúa un *visual* que exhibe el tiempo total recorrido por los conductores. Estos se encuentran ordenados alfabéticamente de forma ascendente. Mediante este informe se pueden responder consultas como el conductor con mayor tiempo recorrido en un mes, el tiempo promedio por trayecto para un determinado taller (perfil) o el porcentaje del tiempo empleado por un vehículo en la conducción.

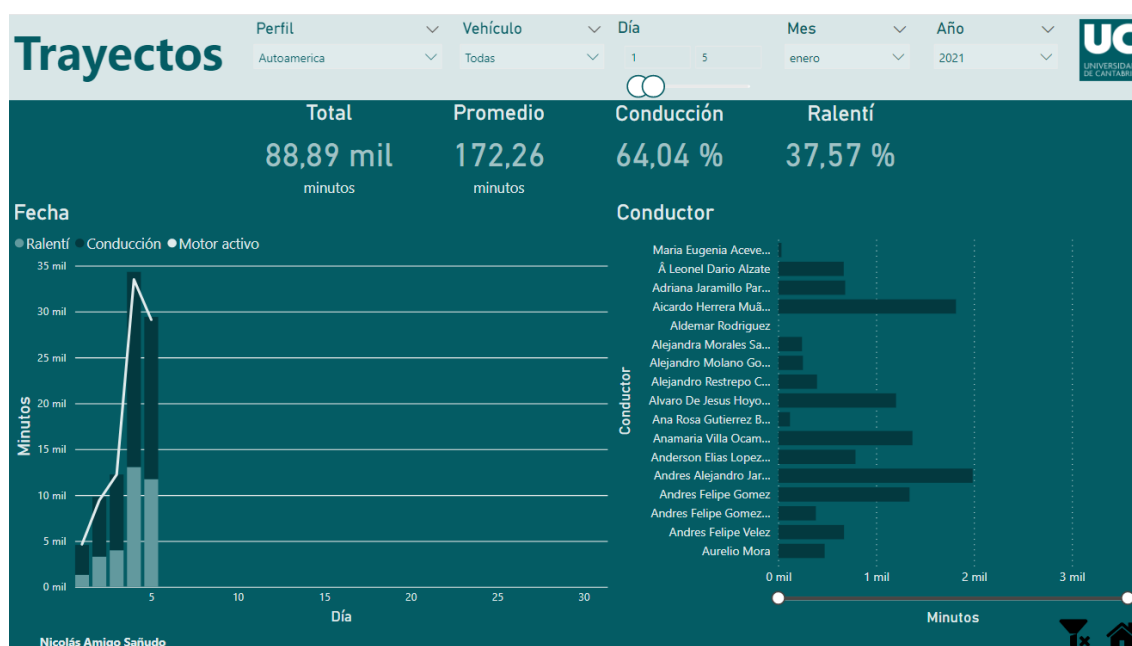


FIGURA 41. INFORME DE TIEMPO_TRAYECTOS

Como se puede observar en la Figura 42, el informe “Consumo” pretende relacionar diferentes métricas de conducción con los litros consumidos cada 100 kilómetros. A la izquierda, se muestra un gráfico de dispersión que establece una asociación entre una determinada métrica y el consumo. Las variables que se exponen son el uso del embrague, el número de frenadas y acelerones realizados y la velocidad media. Es posible cambiar la variable presente en el eje de abscisas haciendo uso de los botones de la franja superior. Cada punto de la gráfica se corresponde con una fecha concreta. Se establece una línea de regresión para indicar el tipo de relación. Al lado de dicha visualización se posiciona un gráfico de líneas para observar el progreso del consumo de combustible por día. Asimismo, se utiliza una línea horizontal para indicar el promedio entre las fechas mostradas. Como resultado, es posible conocer qué métrica influye más en el empleo de combustible o cuando se ha producido un mayor consumo para un conjunto de vehículos.

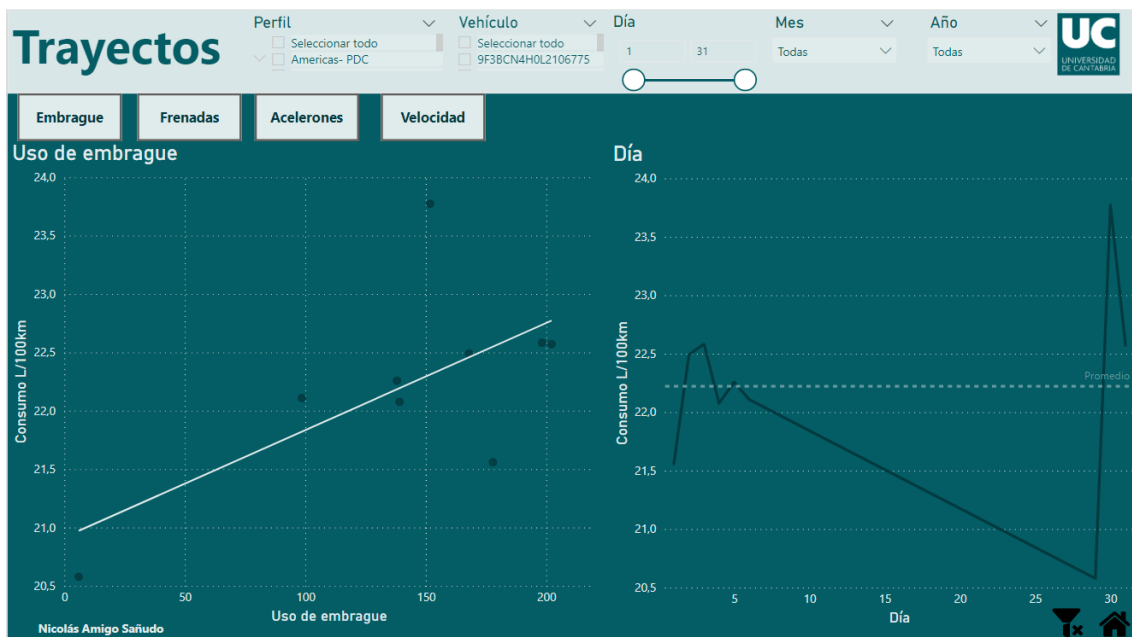


FIGURA 42. INFORME DE CONSUMO_TRAYECTOS

En cuanto a las calificaciones sobre sostenibilidad, seguridad y buen uso del vehículo, se crea un informe, reflejado en la Figura 43, que presenta dos visualizaciones principales y una serie de tarjetas con las métricas en formato numérico. Estas últimas se sitúan en la franja superior. Todas ellas presentan valores sobre un total de 100 puntos. Permiten conocer las puntuaciones de forma rápida y sencilla dado que el resto de información se expone de forma gráfica. La matriz de la izquierda establece como elementos cruzados la lista de conductores junto con todas las calificaciones. Se hace uso de tonalidades condicionales para que sea posible visualizar de una forma gráfica las diferentes métricas. El color del campo es más oscuro según lo positiva que es una nota para un determinado conductor. Además, se añade una nueva medida denominada “Valoración” que consiste en la media de las cuatro variables empleadas. A la izquierda de cada cifra, se sitúa un icono indicando la idoneidad de la calificación. Si esta es menor de 85, se exhibe una figura amarilla con un signo de exclamación. El símbolo es una cruz sobre un fondo rojo en caso de ser inferior a 60 puntos. Cuando la valoración es considerada mayor o igual que 85, se muestra un icono verde con una marca de verificación.

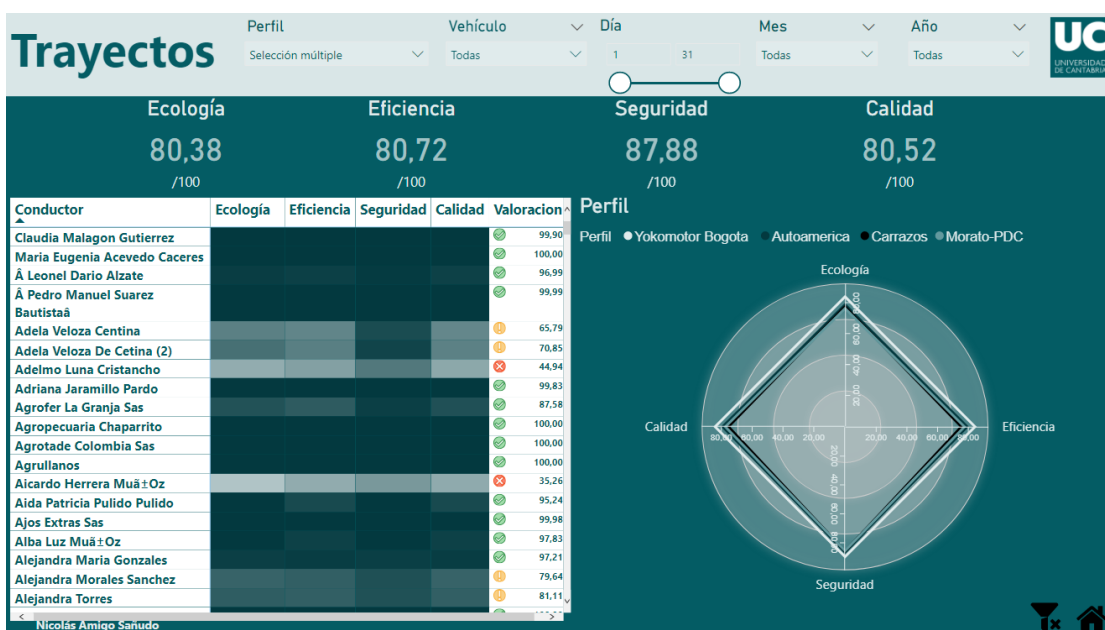


FIGURA 43. INFORME DE CALIFICACION_TRAYECTOS

La otra gráfica principal es un *radar chart* descargado de la tienda de elementos de Microsoft PowerBI. Utiliza las diferentes métricas como valores que forman los vértices. Estos se agrupan por perfil. Solo se muestran los principales talleres en la Figura 43 para una mejor visualización. El criterio principal de la elección de esta gráfica se debe a su presencia en el sistema de información original de la compañía. Asimismo, permite contrastar varias variables de forma sencilla. Gracias al informe se responden cuestiones sobre si un determinado conductor o conjunto de vehículos (taller) debe corregir sus hábitos de conducción o el estado actual del vehículo.

5 Validación del sistema y satisfacción del cliente

Tras el diseño y la implementación de la solución *Business Intelligence*, se procede a la validación del sistema desarrollado con el objeto de certificar que cumple con los criterios que se plantean al inicio del proyecto. Los requisitos que tiene que cumplir la solución se basan principalmente en el rendimiento de las consultas. Estas deben caracterizarse por la agilidad presentando un tiempo de respuesta óptimo. Asimismo, la información que suministra la solución no debe estar limitada a un periodo de tiempo concreto, de forma que sea posible realizar un análisis global de los datos desde los primeros registros almacenados. Esto conlleva que el *Data Warehouse* sea capaz de gestionar grandes volúmenes de datos para ser mostrados en el cuadro de mandos de forma dinámica y rápida.

En primer lugar, se lleva a cabo la comprobación del rendimiento de la solución BI para un conjunto de datos. Estos se acotan a una semana debido a que es la limitación que presenta el sistema de información original. Con el objeto de constatar la productividad de los informes que trabajan sobre el almacén de datos y los cubos OLAP, se realiza una consulta de las notificaciones de los fallos de los vehículos. Además de que no expone la obligación de seleccionar un automóvil específico, el nuevo sistema implementado presenta un mejor tiempo de respuesta para el mismo volumen de datos. La actualización de la información se realiza de forma casi instantánea siendo considerable la diferencia de rendimiento con la plataforma inicial. Si se aumenta la cantidad de datos a gestionar por la solución BI a un periodo mayor al limitado por la original (7 días), se puede comprobar que apenas existen diferencias con el control anterior aun trabajando con un volumen superior.

Por otro lado, se realiza un análisis del rendimiento de las consultas de la actividad de los vehículos. Los informes asociados no presentan como origen las bases de datos que conforman el *Data Warehouse*. Trabajan directamente sobre la hoja de cálculo suministrada. Teniendo esto en cuenta, se ejecuta una comprobación más exhaustiva ya que se desconoce el funcionamiento del *reporting* sin la infraestructura desarrollada para los otros casos de uso. Por ello, se comprueba el tiempo de respuesta tanto para conjunto de datos restringido a una semana como a un periodo mayor. Además, se hace uso de todos los elementos de segmentación de datos presentes en dichos informes. Como resultado, se puede afirmar que presentan un rendimiento más eficiente en comparación con el sistema de información original. A pesar de la arquitectura de este caso, es posible realizar consultas en un tiempo óptimo sin limitaciones temporales y mediante diferentes perspectivas de observación.

La responsable del proyecto “rut-IA” en Fagor Smart Data realiza la siguiente valoración del TFG:

“El trabajo desarrollado por Nicolás nos ha facilitado el acercamiento a las tecnologías BI de aplicabilidad directa sobre nuestro producto y que tienen un impacto directo sobre el negocio. Gracias a la solución diseñada e implementada se ha conseguido dotar a nuestra plataforma Flotasnet de una capacidad de análisis de datos que hasta ahora no era posible por medios tradicionales, poniendo especial cuidado en ofrecer un excelente nivel de usabilidad mediante reportes y cuadros de mandos adaptables y configurables. Valoramos especialmente la agilidad con la que se pueden consultar datos de distintos márgenes temporales que se muestran en tiempo real y que, junto con los indicadores implementados, nos permite ofrecer a nuestros clientes información de mayor valor para sus negocios. Estamos enormemente satisfechos y esperamos poder seguir evolucionando en esta línea, aprovechando estas herramientas para conseguir una analítica de datos un paso superior al que veníamos aplicando además

de tender hacia la minería e inteligencia artificial de los datos con el fin de anticiparnos a los hechos futuros.”

6 Conclusiones

La solución BI pretende corregir y paliar las carencias que presenta el sistema de información actual “FlotasNet”. Por ello, se persigue la realización de consultas de forma ágil y sin limitaciones de volumen de datos, así como la visualización de la información a través de diferentes dimensiones con la finalidad de aumentar las perspectivas de observación. Además, se solicita la posibilidad de exportar los resultados a una hoja de cálculo Microsoft Excel.

Se puede señalar que estos objetivos planteados al comienzo del desarrollo del Trabajo de Fin de Grado se han cumplido en el plazo de tiempo acordado y mediante las expectativas de calidad previstas. Se parte de la idea de la realización de un cuadro de mandos haciendo uso de la herramienta de *reporting* Microsoft PowerBI. Al ser esta una utilidad reciente, ofrece soluciones interesantes para la industria.

Tanto la aportación de los indicadores clave de rendimiento como de las fuentes de datos suministradas por Fagor Smart Data suponen un salto de calidad en el desarrollo del TFG. La colaboración con la firma santanderina permite “profesionalizar” el proyecto al tener que seguir una serie de directrices para adaptarse a los requisitos que solicitan. Asimismo, implica la necesidad de mostrar la evolución de la solución y recibir *feedback* por parte de los interesados.

El desarrollo del Trabajo de Fin de Grado me ha permitido poner en práctica los conocimientos adquiridos principalmente en las asignaturas impartidas por la directora del TFG (Bases de Datos, Desarrollo de Sistemas Información). Otros cursos estudiados a lo largo del grado presentan especial importancia en la implementación del proyecto. Entre ellos destacan los relacionados con el módulo de programación y el diseño de interfaces gráficas. Debido a este proyecto también se realizan asignaturas asociadas a la Ciencia de Datos y la inteligencia de negocio durante el curso académico realizado en la Universidad de Tulsa. Además, con motivo del empleo de las tecnologías descritas en el apartado 3.3, se realizan diversos cursos *online* para comprender su funcionamiento y dominar las utilidades necesarias para la implementación del sistema.

En relación con las dificultades encontradas en la realización del trabajo, se encuentran principalmente la programación de los procesos ETL para la extracción, transformación y carga de los datos en los diferentes componentes. Dada la variedad de los métodos para su realización, se produce un pequeño cuello de botella en la planificación del proyecto en busca de la mejor solución. Se opta por el uso de procedimientos SQL, así como de paquetes de la herramienta SQL Integration Services. El empleo de las API expuestas anteriormente sumaron más complicaciones a dichos procesos debido a problemas en las llamadas al servidor.

Por el contrario, la herramienta Microsoft PowerBI ha resultado sencilla de manejar gracias a la facilidad de su interfaz para la creación de los informes. Presenta un modo de uso muy intuitivo que permite llevar a cabo las tareas de *reporting* de forma sencilla y rápida. El lenguaje DAX, que incluye para el análisis de los datos, no requiere un tiempo de aprendizaje elevado dado que se basa en otros utilizados en hojas de cálculo. Sin embargo, se presentan problemas relacionados con la exportación y el despliegue de la solución BI. Se deben a que no se cuenta con las licencias necesarias y el coste económico para ello es elevado.

Además del aprendizaje adquirido en materia de análisis y visualización de datos, el presente proyecto y todo el estudio realizado para su desarrollo son motivos por los cuales se va a realizar un Máster en Ciencia de Datos para seguir profundizando y aprendiendo técnicas de gestión de datos.

7 Líneas futuras

El diseño y la implementación de la solución *Business Intelligence* solicitada por la compañía Fagor Smart Data se desarrolla dentro del plazo previsto en este periodo de tiempo. Se llevaron a cabo las tareas esenciales para la definición de los componentes del *Data Warehouse* y los informes de visualización de los datos. Partiendo de la premisa de que se ha cumplido el objetivo, existen determinadas tareas y funcionalidades que se pueden incorporar al sistema desarrollado a corto o medio plazo. A continuación, se hace un breve resumen de cada una de ellas ordenadas de mayor a menor prioridad.

1. Diseño e implementación de un esquema dimensional en estrella y un cubo OLAP asociado para el caso de uso Trayectos. De esta forma, es posible descomponer la información en diferentes entidades con el objeto de mejorar la accesibilidad y el rendimiento de las consultas. Además, se facilita los procesos de actualización de datos.
2. Definición de nuevos indicadores clave de rendimiento (KPI) para los casos de uso desarrollados. Para ello, se requiere la obtención de fuentes de datos que proporcionen información contextual.
 - a. Accesos. Respuesta a consultas asociadas con el tiempo promedio por sesión, el patrón de comportamiento de las cuentas y la relación de los datos con los relacionados con los vehículos.
 - b. Fallos. Respuesta a consultas relacionadas con el tiempo de reparación de un determinado fallo.
 - c. Trayectos. Respuesta a consultas centradas en las coordenadas geográficas que componen la totalidad de los tramos.
3. Mejora del diseño de los informes en base a información del uso del cuadro de mandos por parte de los usuarios. Se recoge *feedback* por parte de ellos tras un periodo de tiempo de empleo mediante diferentes estrategias de evaluación propias de la disciplina Interacción Persona-Computador.
4. Desarrollo de la versión del cuadro de mandos destinada a dispositivos móviles. Se añade una mayor portabilidad a la solución *Business Intelligence* para que los usuarios puedan acceder a los datos desde cualquier ubicación a través de la aplicación PowerBI Mobile. Asimismo, se facilita la interacción con los informes mediante las pantallas táctiles.

Bibliografía

- Bakhshi, S. (2016). *Power BI and Google Maps API (Address Lookup)*. Obtenido de bi insight: <https://www.biinsight.com/power-bi-and-google-maps-api-address-lookup/>
- BARC. (2020). *How to maximize benefits from a new BI and analytics tool*.
- Dintén Herrero, R. (2022). *Análisis tramos coltanques*.
- Gartner. (s.f.). *Glosario de Gartner*. Obtenido de Gartner: <https://www.gartner.com/en/glossary>
- Google. (2022). *Geocoding API*. Obtenido de Google Developers: <https://developers.google.com/maps/documentation/geocoding/>
- Inmon, W. (2005). *Building the Data Warehouse*. Wiley Publishing.
- ipwhois.io. (2022). *API Documentation*. Obtenido de ipwhois.io: <https://ipwhois.io/documentation>
- Kimball, R. (2002). *The Data Warehouse Toolkit*. Wiley Publishing.
- Microsoft. (2022). *Documentación de Analysis Services*. Obtenido de Microsoft Docs: <https://docs.microsoft.com/es-es/analysis-services>
- Microsoft. (2022). *Documentación de Power BI*. Obtenido de Microsoft Docs: <https://docs.microsoft.com/es-es/power-bi/>
- Microsoft. (2022). *Documentación de SQL Integration Services*. Obtenido de Microsoft Docs: <https://docs.microsoft.com/es-es/sql/integration-services>
- Mundy, J., Thornthwaite, W., & Kimball, R. (2011). *The Microsoft Data Warehouse Toolkit*. Wiley Publishing.
- OpenWebinars. (s.f.). *OpenWebinars*. Obtenido de <https://openwebinars.net/>
- Statista. (2022). Obtenido de <https://www.statista.com/statistics/871513/worldwide-data-created/>
- Turban, E., Sharda, R., Delen, D., & King, D. (2010). *Business Intelligence: A Managerial Approach* (2 ed.). Prentice Hall.
- Zorrilla Pantaleón, M. E. (2010). *Tutorial. Analysis Services 2008*.
- Zorrilla Pantaleón, M. E. (s.f.). *Tema 8. Introducción a los almacenes de datos*.
- Zorrilla Pantaleón, M. E. (s.f.). *Tema 9. Diseño multidimensional*.

Anexo

Informes del caso de uso: accesos

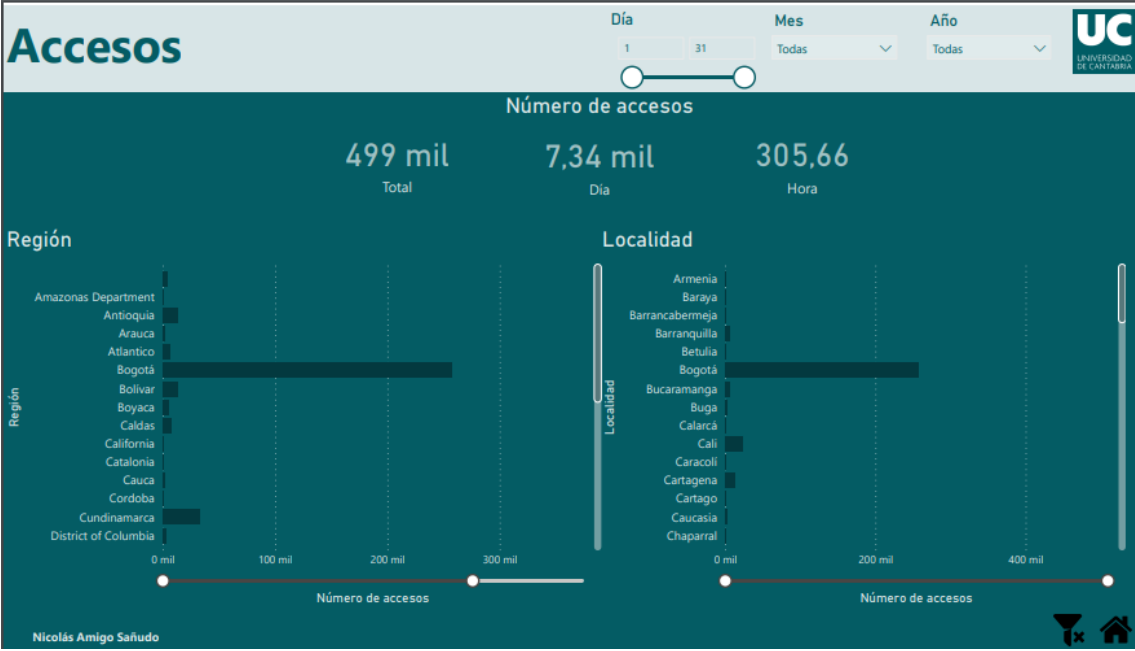


FIGURA 44. INFORME DE UBICACION_ACCESOS



FIGURA 45. INFORME DE UBICACION_ACCESOS_II

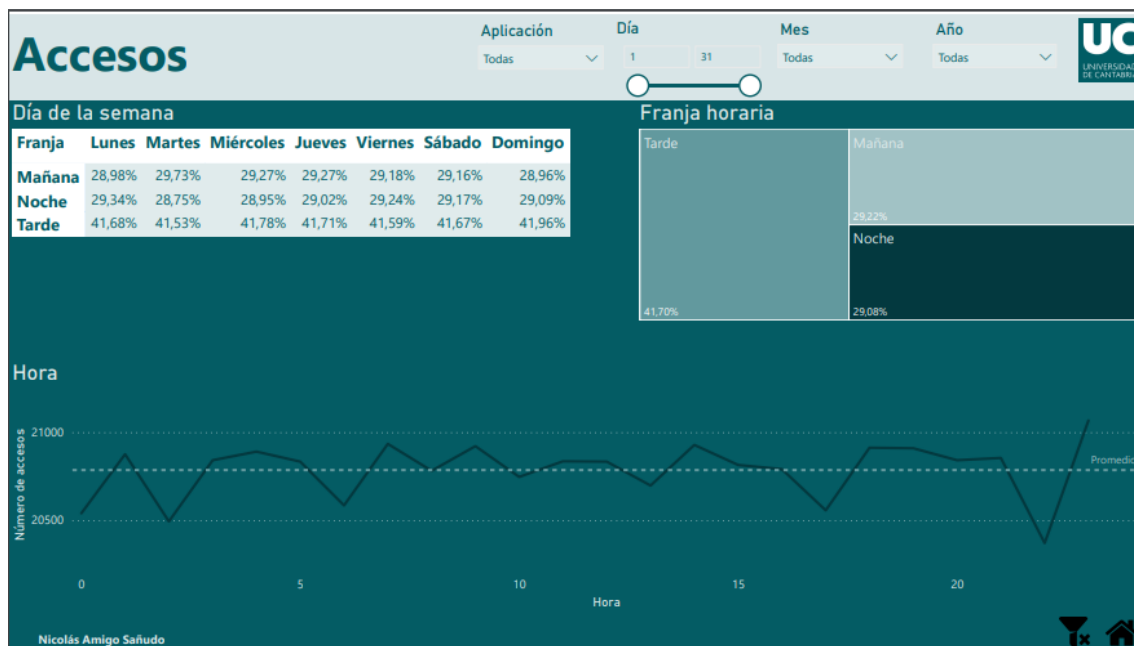


FIGURA 46. INFORME DE HORA_ACCESOS

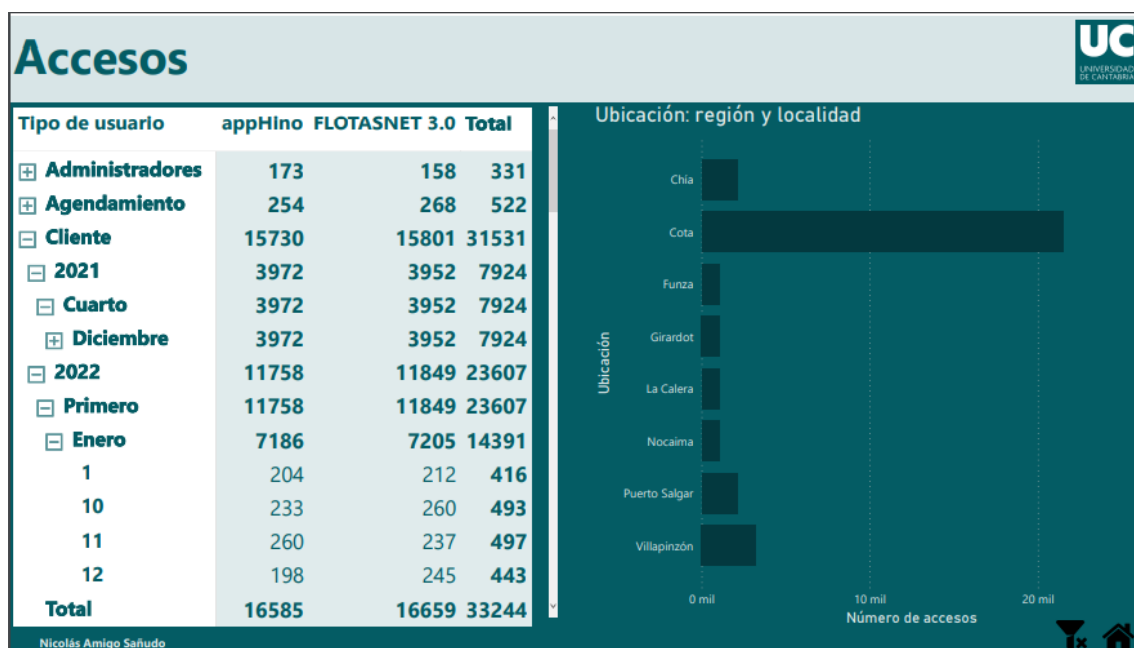


FIGURA 47. INFORME DE TABLA_ACCESOS

Informes del caso de uso: fallos



FIGURA 48. INFORME DE FECHA_FALLOS

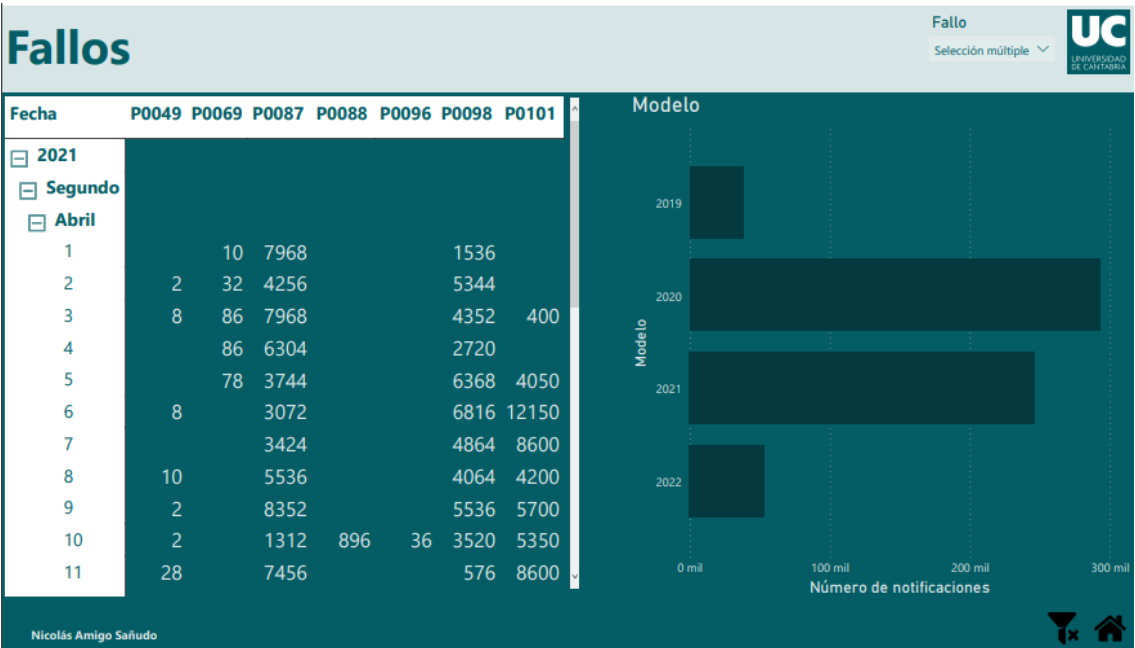


FIGURA 49. INFORME DE TABLA_FALLOS

Informes del caso de uso: trayectos



FIGURA 50. INFORME DE VELOCIDAD_TRAYECTOS

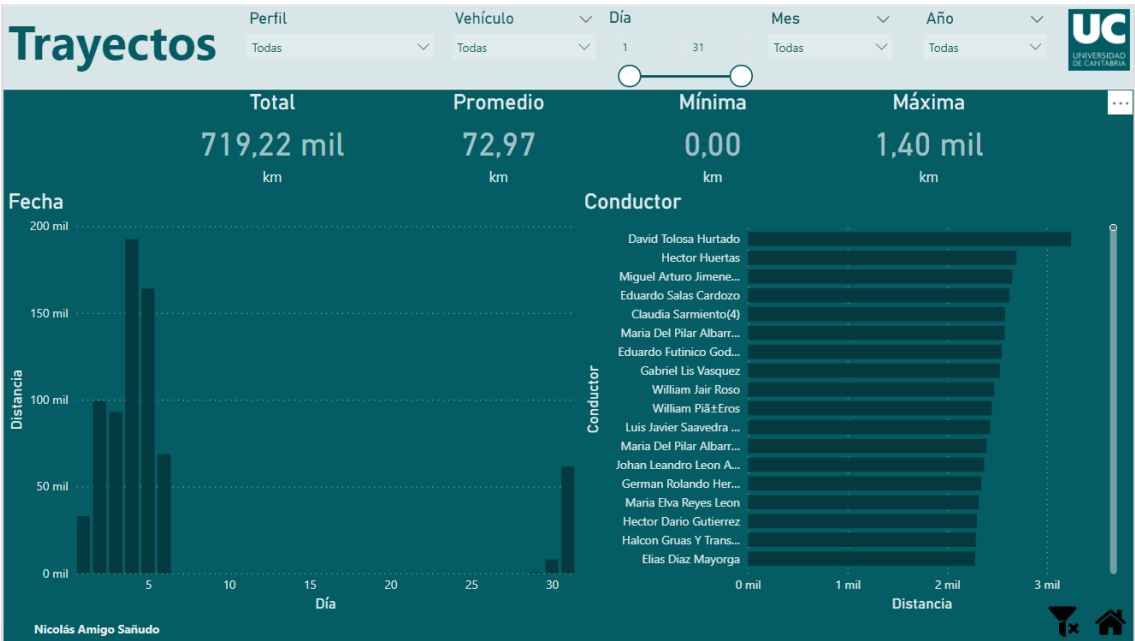


FIGURA 51. INFORME DE DISTANCIA_TRAYECTOS

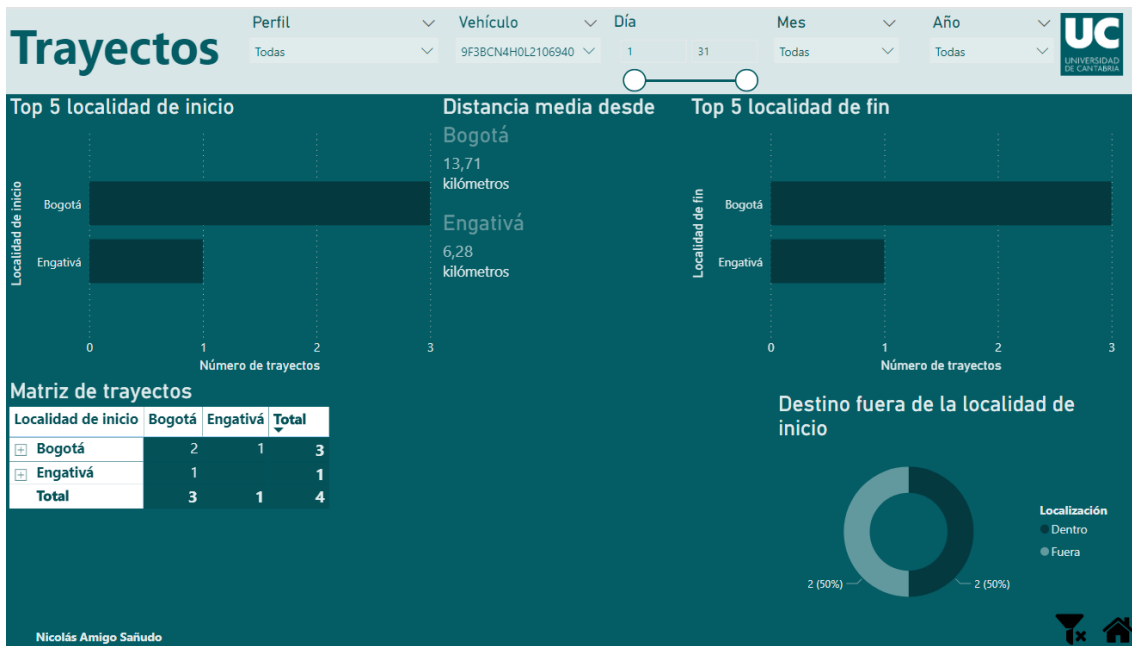


FIGURA 52. INFORME DE UBICACION_TRAYECTOS

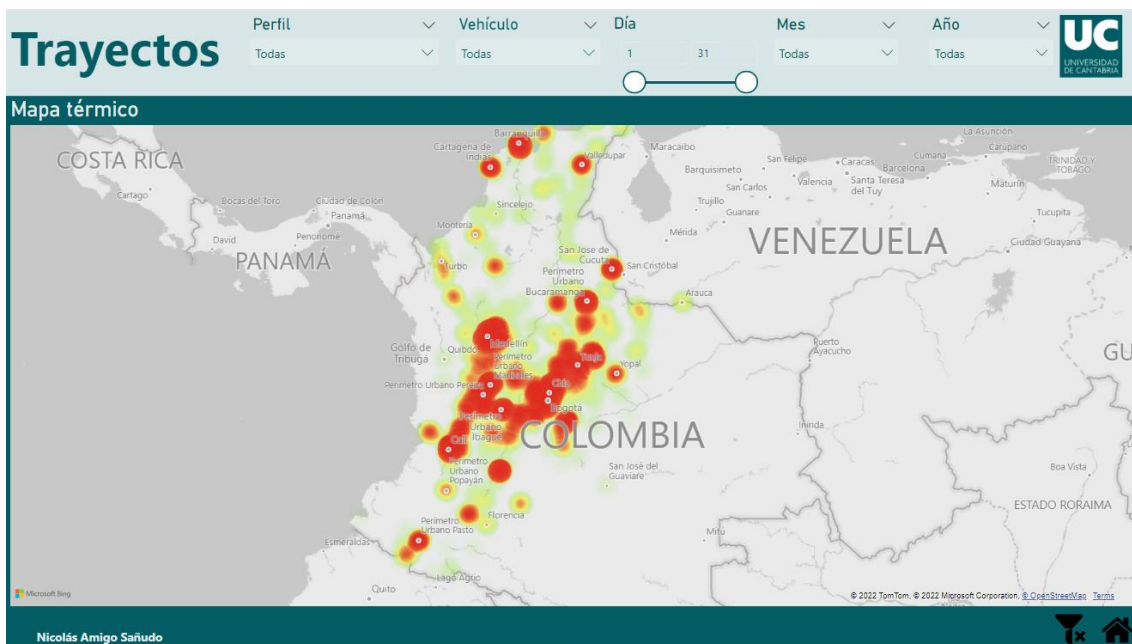


FIGURA 53. INFORME DE MAPA_TRAYECTOS