UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Diseño y desarrollo de un cuadro de mandos para el análisis de los eventos en cadena de tracción y motor en locomotoras diesel

Trabajo Fin de Grado

Grado en Ciencia de Datos

AUTOR/A: Portilla Alique, Cristina

Tutor/a: Monserrat Aranda, Carlos

Cotutor/a: Escobar Román, Santiago

Cotutor/a externo: NIEVES CORDONES, DAVID

CURSO ACADÉMICO: 2022/2023

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Diseño y Desarrollo de un cuadro de mandos para el análisis de los eventos en cadena de tracción y motor en locomotoras diésel

Trabajo Fin de Grado

Grado en Ciencia de Datos

Autor: Portilla Alique, Cristina

Tutor: Monserrat Aranda, Carlos

Cotutor: Escobar Román, Santiago

Cotutor externo: Nieves Cordones, David

2022/2023

Resumen

Este proyecto consiste en la implementación de un cuadro de mandos en Microsoft Power BI para analizar los eventos más predominantes producidos en la cadena de tracción y motor de las locomotoras diésel fabricadas en Stadler Rail Valencia.

El cuadro de mandos tiene como objetivo dar soporte a la toma de decisiones en el departamento de Software&ICT, y constará de varios gráficos que proporcionen información acerca de los eventos (eg., geolocalización, descripción, categoría, prioridad). Estos gráficos pueden ser filtrados por el usuario por flota, locomotora, tipo de evento y período de tiempo.

Por otro lado, se realiza un análisis de fiabilidad para dos locomotoras, haciendo uso del modelo Mean Cumulative Function (MCF) y el estimador Kaplan-Meier. Este análisis permite estudiar el rendimiento esperado de cada una de las locomotoras. A su vez, se decide implementar el análisis Kaplan-Meier en el cuadro de mandos permitiendo visualizar las expectativas del crecimiento de la probabilidad estimada del evento seleccionado para una locomotora específica.

Palabras clave: Ciencia de datos, cuadro de mandos, locomotoras, análisis de supervivencia, MCF, Kaplan-Meier, SQL, Power BI.

Abstract

This project consists of the implementation of a dashboard in Microsoft Power BI to analyze the most predominant events occurring in the traction chain and engine of the diesel locomotives produced in Stadler Rail Valencia.

The dashboard is intended to support decision making in the Software&ICT department and will consist of several graphs providing information about the events (e.g., geolocation, description, category, priority). These graphs can be filtered by the user by fleet, locomotive, event type and period.

On the other hand, a reliability analysis is performed for two locomotives, making use of the Mean Cumulative Function (MCF) and the Kaplan-Meier estimator. This analysis makes it possible to study the expected performance of each of the locomotives. In turn, it is decided to implement the Kaplan-Meier analysis to the scorecard allowing to visualize the expectations of the growth of the estimated probability of the selected event for a specific locomotive.

Keywords: Data science, dashboard, locomotives, survival analysis, MCF, Kaplan-Meier, SQL, Power BI.

Índice de contenido

Índice de ilustraciones

Índice de tablas

1. Introducción

Stadler Rail Valencia es una empresa industrial perteneciente al grupo Stadler, especializada en la construcción de vehículos ferroviarios. La principal actividad de esta empresa es la fabricación de locomotoras, vehículos ligeros, tranvías y metros de alta calidad, que aportan valor añadido y se adaptan a las necesidades requeridas por los clientes. [1]

Dentro de los vehículos que se construyen, podemos encontrar las locomotoras diésel. Uno de los componentes clave en este tipo de vehículos es la cadena de tracción y motor. Este elemento está sometido a altas cargas y desgaste durante su funcionamiento, lo que puede ocasionar problemas de desgaste prematuro, pérdida de potencia o incluso fallas catastróficas. Es esencial realizar un seguimiento exhaustivo de estos componentes, realizar inspecciones regulares y llevar a cabo el mantenimiento adecuado para prevenir problemas y prolongar su vida útil.

El análisis de los fallos en las locomotoras diésel permite identificar las áreas de mayor vulnerabilidad y diseñar estrategias de mantenimiento preventivo y correctivo más efectivas.

Para cada una de las diferentes locomotoras, se recogen datos históricos relacionados con diferentes eventos que van teniendo lugar en la cadena de tracción y motor. Estos eventos generalmente son registrados porque hay un fallo o más que provoca cada uno de ellos.

En este proyecto se describe el desarrollo de una herramienta que permita visualizar información relacionada con estos eventos mediante la implementación de un cuadro de mandos en Microsoft Power BI. De este modo, se busca agilizar el análisis de los datos de las locomotoras diésel. Por otro lado, se analizará el rendimiento de dos locomotoras mediante el uso del modelo MCF y el estimador Kaplan-Meier.

1.1. Trayectoria de la empresa

Este proyecto ha sido llevado a cabo en la empresa Stadler Rail Valencia, y para contextualizar, este apartado se va a centrar en la trayectoria de la planta y las empresas que han pasado por ella. [1]

La historia comienza en 1897, cuando se forma en Marchalenes (Valencia) la empresa Talleres Devis-Noguera, con el objetivo de fabricar y reparar material ferroviario.

En 1929 se constituye Construcciones Devis S.A., cambiando el enfoque de la empresa hacia la construcción de locomotoras de vapor. En esta época se dedican al suministro de locomotoras para la Compañía Nacional de Ferrocarriles del Oeste, Ferrocarriles Andaluces y Renfe.

En 1947 se crea MACOSA, que consiste en la fusión de Devis con la empresa Barcelona Sociedad Material para Ferrocarriles y Construcciones S.A. Posteriormente, en 1989 el control de la empresa pasa a ser de la multinacional GEG-Alsthom, realizándose en 1997 el traslado a la planta actual de Albuixech (Valencia).

En 2005, debido a la crisis, la planta se pone en venta y es comprada por Vossloh AG, y finalmente, en 2015 se produce el traspaso de la fábrica al fabricante suizo Stadler Rail, siendo el actual dueño de la planta.

Actualmente la planta de Albuixech está compuesta por cerca de 900 empleados, y una superficie de unos 200000m^2 (incluyendo oficinas, plantas de producción, zonas de almacenaje y zonas de inspección de vehículos y ensayos). Stadler Rail Valencia está centrada en locomotoras de maniobras, locomotoras de línea, metros y vehículos de vanguardia para el transporte urbano.

1.2. Motivación

La motivación para llevar a cabo este proyecto es doble. En primer lugar, existe un interés específico por parte de la empresa en encontrar una forma rápida y conveniente de visualizar los eventos que ocurren en las locomotoras diésel, con el fin de obtener información relevante sobre los mismos. Esto complementaría las labores de supervisión y vigilancia de la maquinaria, lo que a su vez contribuiría a mejorar la seguridad y eficiencia de los transportes ferroviarios.

Además, a nivel personal, me motiva la posibilidad de desarrollar un enfoque de diagnóstico preventivo de eventos en las locomotoras. Esto implica la implementación de modelos descriptivos de fiabilidad que permitan identificar y anticipar posibles problemas antes de que se conviertan en eventos importantes. Al aplicar estos modelos, se podría mejorar significativamente la planificación de las tareas de mantenimiento y reducir tanto los tiempos de inactividad como los costos asociados a las reparaciones.

Este proyecto tiene el potencial de brindar soluciones prácticas y efectivas para mejorar la gestión de las locomotoras y optimizar su rendimiento. Mediante la combinación de la visualización de eventos y análisis descriptivos se consigue comprender el comportamiento de dichos eventos y, por tanto, ayudar en la prevención de los fallos que pueden originarlos. Estas contribuciones tanto a nivel empresarial como personal son aspectos clave que me motivan a realizar este trabajo de investigación.

1.3. Marco legal y ético

Antes de comenzar con la explicación del desarrollo del proyecto es importante mencionar algunas consideraciones legales relacionadas con los datos utilizados:

En primer lugar, es importante mencionar que el desarrollo de este proyecto no conlleva el uso ni el tratamiento de ningún tipo de datos de carácter personal.

En segundo lugar, la propiedad legal tanto de los datos como del cuadro de mandos desarrollado pertenece única y exclusivamente a la empresa Stadler. El tratamiento de los datos está sujeto a la política de privacidad y confidencialidad de la misma y a los acuerdos de propiedad intelectual vigentes. Se han respetado los derechos legales de la empresa y se ha trabajado en estrecha colaboración con ellos para garantizar el cumplimiento de las regulaciones correspondientes.

Por este motivo, a lo largo del proyecto no se detallarán los nombres reales de ninguna flota o locomotora. El nombre de las flotas será sustituido por una letra mayúscula y las locomotoras tendrán asignado un número para asegurar la privacidad de los datos.

Finalmente, el desarrollo del proyecto no implica ningún tipo de implicación ética grave a considerar ya que el estudio se centra únicamente en la descripción de los eventos registrados en las locomotoras de la empresa.

1.4. Objetivos

Los análisis sobre los eventos registrados hasta el momento en las locomotoras de Stadler son elaborados de forma individualizada y costosa. La empresa busca desarrollar una herramienta visual que permita al usuario realizar este tipo de análisis sobre eventos de forma eficaz. El objetivo principal de este proyecto es, por tanto, desarrollar e implementar un cuadro de mandos que recopile y visualice la información asociada a los eventos que tienen lugar en la cadena de tracción y motor de las locomotoras diésel.

Este cuadro de mandos actuará como una herramienta para monitorizar las locomotoras, identificar los eventos y tomar decisiones basadas en la información recopilada. Para ello, debe cumplir una serie de requisitos que serán los objetivos secundarios del proyecto:

- Debe ser implementado en el software Power BI.
- Debe recopilar información detallada sobre los eventos que tienen lugar en las locomotoras.
- Debe ofrecer la posibilidad de explorar de forma interactiva la posición geográfica de los eventos históricos registrados.
- Debe presentar de forma clara tanto las distribuciones de los eventos en las locomotoras, como las distribuciones de los códigos de evento y las categorías asociadas a estos.
- Debe mostrar la cantidad de eventos registrados, a la vez que la cantidad de los que estén activos, así como la duración máxima registrada de estos.
- Debe incluir una estimación del rendimiento de la locomotora seleccionada, definiendo éste como la frecuencia con la que se espera que ocurra un evento en dicha locomotora.

Además, se busca realizar un análisis de fiabilidad exhaustivo en las flotas A y B para estudiar el evento 832 en ellas ya que son las más propensas a registrarlo, y estimar los días esperados hasta un nuevo registro de este tipo de evento.

1.5. Estructura

Esta memoria está compuesta por seis apartados importantes, los cuales se describen en detalle a continuación:

- Capítulo 1, Introducción:

En este capítulo se presenta la introducción al trabajo, incluyendo la motivación detrás de la investigación y el contexto en el que se enmarca el proyecto. Se explican las razones que impulsaron la elección del tema y se establecen los objetivos específicos que se pretenden alcanzar con el trabajo.

- Capítulo 2, Marco teórico:

  En este capítulo se realiza un análisis exhaustivo de la literatura existente sobre el tema de estudio. Se revisan documentos relevantes y fuentes de información pertinentes para adquirir una comprensión sólida de los aspectos teóricos y prácticos relacionados con el tema. Se proporciona un marco de referencia teórico que servirá de base para el desarrollo y análisis posterior.

- Capítulo 3, Desarrollo:

  En este capítulo se detalla el problema específico que se aborda en este trabajo y se explica en profundidad la metodología utilizada para obtener los datos necesarios. Se describe la estructura de la base de datos, el preprocesamiento de los mismos y su posterior carga en Power BI para diseñar los gráficos y realizar el análisis de fiabilidad. Finalmente, se explican las técnicas utilizadas y se justifican las decisiones tomadas durante el desarrollo del proyecto.

- Capítulo 4, Resultados:

  En este capítulo por un lado se presentan y analizan los resultados obtenidos a partir del análisis de fiabilidad realizado, y por otro se muestra el cuadro de mandos definitivo, justificando la distribución de los gráficos y proporcionando una interpretación detallada de los hallazgos. Se resaltan los aspectos relevantes y se discute su implicación en relación con los objetivos del trabajo.

- Capítulo 5, Conclusiones:

  En este capítulo se resumen de manera concisa los principales hallazgos y conclusiones del trabajo. Se realiza un análisis de la relación del trabajo con los estudios cursados y se evalúa en qué medida se han alcanzado los objetivos propuestos inicialmente. Se discuten las limitaciones encontradas durante el proceso y se destacan las contribuciones y el valor añadido del trabajo realizado. Además, se describe el legado.

- Capítulo 6, Trabajo futuro:

  En este capítulo se proponen líneas de investigación futuras y se sugieren posibles mejoras o extensiones del trabajo realizado. Se exploran áreas que podrían ser objeto de investigación adicional y se identifican posibles direcciones para el desarrollo y la mejora del análisis de fiabilidad. Se destacan las oportunidades de crecimiento y se plantean recomendaciones para futuros proyectos relacionados.

Además, al final del documento se incluyen las referencias bibliográficas utilizadas para la elaboración del TFG, con el fin de proporcionar un respaldo académico a las afirmaciones y análisis presentados en el informe. También se incluye como anexo la relación del trabajo con los Objetivos de Desarrollo Sostenible de la Agenda 2030, resaltando la contribución del proyecto a la sostenibilidad y los aspectos socioambientales relevantes.

2. Marco teórico

2.1. Estudios de fallos en locomotoras

Las locomotoras diésel poseen un papel muy importante en el transporte ferroviario, ya que brindan potencia y tracción para conseguir el movimiento de los trenes de carga y pasajeros. Sin embargo, como cualquier sistema mecánico complejo, las locomotoras están sujetas a posibles averías y fallos que afectan a su eficiencia, disponibilidad y rendimiento.

El estudio de los fallos en locomotoras diésel busca comprender las causas y reducir las consecuencias de los fallos en este tipo de vehículos. Permite identificar las áreas de mayor vulnerabilidad y diseñar estrategias efectivas de mantenimiento preventivo y correctivo. En este sentido, es crucial examinar cada componente y sistema de la locomotora para comprender su funcionamiento, detectar posibles puntos débiles y desarrollar soluciones que minimicen los fallos y optimicen la disponibilidad operativa.

Para ello, durante la fabricación de las locomotoras diésel, se llevan a cabo diversos procesos buscando garantizar que se cumplan los estándares de rendimiento y calidad requeridos. Podemos encontrar diversos estudios en la literatura sobre la importancia de los componentes de la maquinaria, el artículo [2] propone la elección del boggie, que es la estructura principal que une los ejes con las ruedas, como un elemento esencial en el rendimiento de la locomotora.

La elección del modelo óptimo de *boggie* es primordial, ya que tiene un impacto significativo en el rendimiento, la estabilidad y la seguridad de la locomotora. Para ello, se utilizan métodos de optimización multicriterio que tienen en cuenta diferentes variables, como son la distribución del peso, la capacidad de absorción de impactos, la capacidad de carga y la resistencia a la fatiga. Estos métodos permiten comparar y evaluar diferentes opciones de *boggies* para poder seleccionar el más adecuado para cada aplicación específica.

Además de la elección del boggie, todos los componentes de la locomotora deben ser correctamente seleccionados y diseñados. Según el artículo [3], los ingenieros dedican una gran cantidad de tiempo a resolver problemas relacionados con las locomotoras diésel, especialmente aquellos relacionados con la transmisión durante la construcción. La transmisión en una locomotora diésel es un aspecto crítico, ya que debe ser capaz de transmitir la potencia generada por el motor a las ruedas de manera eficiente y confiable. Se consideran factores como los mecanismos de embrague, la relación de transmisión y la selección de materiales adecuados para garantizar un rendimiento óptimo y una vida útil prolongada.

A medida que avanza el desarrollo tecnológico, el conocimiento y la comprensión de las locomotoras diésel han ido aumentado. Se realizan cada vez más estudios previos y se consideran más variables en el proceso de diseño y construcción de estas máquinas. Se emplean herramientas tanto de simulación como de modelado avanzadas para evaluar el rendimiento de la locomotora en diferentes condiciones y optimizar así su diseño antes de la fabricación. Todos estos avances tienen como objetivo reducir al mínimo posible el número de errores y evitar la necesidad de realizar cambios posteriores a la producción. Evitar la necesidad de realizar este tipo de cambios no sólo ofrece un evidente aumento en la fiabilidad de las locomotoras, sino que también supone un significativo ahorro en costes para las empresas que las diseñan.

En el caso específico del motor diésel, como se menciona en el artículo [4], se deben tener en cuenta las pérdidas magnéticas en el acero del motor de corriente ya que pueden generar calor adicional y reducir la eficiencia del motor. Por lo tanto, estas pérdidas han de ser analizadas y minimizadas por los diseñadores mediante técnicas de diseño y selección de materiales adecuadas.

En resumen, la fabricación de locomotoras diésel implica una serie de procesos y consideraciones, desde la elección del *boggie* y la selección de componentes hasta la resolución de problemas relacionados con la transmisión y la optimización del diseño del motor. Con el avance de la tecnología y el aumento del conocimiento en este campo, se busca mejorar constantemente la calidad, eficiencia y confiabilidad de las locomotoras diésel, teniendo en cuenta factores como la adaptación a requisitos específicos, la optimización multicriterio y la reducción de pérdidas energéticas.

2.2. Cuadro de mandos

Un cuadro de mandos busca el estudio e interpretación de un conjunto de datos mediante el uso de representaciones visuales, como pueden ser gráficos, diagramas o tablas. Estas representaciones ayudan a visualizar de forma más comprensiva e intuitiva patrones, tendencias o relaciones entre los datos.

La raíz de la importancia del cuadro de mandos en el *Business Intelligence* (BI) radica en su capacidad para presentar la información de una manera clara y concisa, permitiendo a los usuarios visualizar y comprender rápidamente los datos con el fin de ayudar a la toma de decisiones.

El BI, según [5], podría definirse como un proceso que consiste en el análisis de grandes volúmenes de datos y su presentación en forma de informes o cuadros de mandos. Este proceso implica sintetizar los conceptos clave implícitos en los datos y traducirlos en conclusiones aplicables para la toma de decisiones con el objetivo de mejorar el rendimiento empresarial.

La tecnología de cuadros de mando, como se menciona en [6], ofrece una solución prometedora para abordar los desafíos actuales en los flujos de trabajo. Al proporcionar una visión clara y comprensible de su estado, los cuadros de mando permiten optimizar su desempeño y ofrecer un servicio de calidad en un entorno cada vez más complejo, abordando todas las deficiencias a la hora de interpretar la información recogida y facilitando la toma de decisiones informadas y optimizadas en relación con el flujo de trabajo, permitiendo a los constructores contar con una herramienta eficaz y confiable que les ayude a gestionar de manera eficiente la complejidad de los sistemas y a tomar decisiones informadas en tiempo real sobre los fallos en las locomotoras.

El desarrollo del sistema de cuadro de mando se basa en el marco de Vercelli [16], que consta de cuatro etapas principales: análisis, diseño, planificación, implementación y control. Uno de los hallazgos destacados es que el proceso de limpieza de datos es crucial para generar información precisa y confiable. Además, se observa que la participación activa de los usuarios, desde el análisis y el diseño hasta la validación de los resultados, mejora significativamente la calidad del cuadro de mando.

Tanto los fabricantes de productos como las organizaciones de mantenimiento de equipos tienen un interés común en comprender el comportamiento de fallos típicos de su maquinaria. En [7] se presenta un enfoque novedoso mediante un cuadro de mando de calidad de datos, el cual tiene como objetivo identificar los problemas de calidad de los datos y proporcionar consejos prácticos para contrarrestarlos. El diseño de este cuadro de mando se basa en una explicación detallada de los problemas típicos relacionados con los datos de fiabilidad. Además, se realiza

una revisión exhaustiva del estado actual de la evaluación de la calidad de los datos en este campo, identificando y contrarrestando los problemas, de manera que se brinde a los fabricantes y a las organizaciones de mantenimiento una ventaja competitiva al mejorar la toma de decisiones y garantizar el funcionamiento óptimo de sus equipos, mejorando su confiabilidad, disponibilidad, mantenibilidad y seguridad, optimizando su desempeño y minimizando los riesgos asociados.

La implementación de una solución portátil, como se describe en [8], se está convirtiendo en una prioridad para la industria, y se están llevando a cabo proyectos de colaboración para desarrollar y evaluar estas tecnologías. Estas soluciones ofrecen un enfoque práctico para el seguimiento de la salud de los equipos, y los resultados preliminares de mantenimiento predictivo han demostrado la capacidad de predecir tiempos de inactividad y fallos clave. Estos avances prometen mejorar la eficiencia, la calidad y los costos en la producción industrial.

En resumen, los cuadros de mandos como soluciones portátiles para la representación de datos ofrecen ventajas significativas en términos de acceso rápido a la información, personalización, visualización intuitiva, interactividad, actualización en tiempo real y capacidad de compartir y colaborar. Estas características los convierten en herramientas muy valiosas a la hora de analizar datos y tomar decisiones efectivas en contextos empresariales.

Aun así, cabe destacar que un mal uso de ellos, como puede ser una complejidad excesiva, falta de contexto o de personalización adecuada, problemas de calidad de datos o falta de actualizaciones y mantenimiento, acaben llevando a malas interpretaciones o a la toma de decisiones erróneas que generan desconfianza y que penalicen a la empresa.

2.2.1. Cuadro de mandos en Power BI

La visualización desempeña un papel fundamental en el análisis de datos al ofrecer una representación más efectiva, interesante y comprensible para todos, sin barreras lingüísticas. Además, permite mostrar una gran cantidad de datos en un espacio reducido de manera fácil y eficiente. En este contexto, como se describe en [9], Microsoft Power BI se destaca como una herramienta ampliamente reconocida para el análisis de datos, ya que permite comprender las tendencias, patrones e impactos de los fallos en las locomotoras de manera clara y accesible. Al aprovechar las funciones de visualización de datos, se pueden encontrar *insights* importantes que ayuden a tomar decisiones informadas.

Como bien indica [10], Power BI permite descubrir, transformar, visualizar y compartir datos, informes y paneles de control con otros usuarios dentro del mismo departamento, organización o incluso con el público en general. Esta aplicación genera automáticamente múltiples informes basados en el análisis de los datos, sin necesidad de intervención humana. Esto no solo agiliza el proceso, sino que también ayuda a reducir errores humanos en cálculos y técnicas estadísticas.

Además, como se describe en [11], Power BI ofrece el lenguaje DAX, a partir del cual se pueden realizar diferentes cálculos en el modelo creado, y la librería *pbiviz*, que nos permite crear nuestros propios objetos visuales. Esta herramienta es óptima a la hora de obtener información de manera rápida y efectiva, y para crear tableros personalizados y realizar cálculos avanzados.

En resumen, como se menciona en [12], la adopción de un cuadro de mando basado en la nube, utilizando Power BI, junto con un enfoque adecuado en la limpieza de datos y la participación activa de los usuarios, puede ser una estrategia efectiva para impulsar el rendimiento y el éxito de una empresa en un mercado altamente competitivo. Es importante destacar que Power BI ofrece capacidades más allá de la simple visualización del cuadro de mando. La

herramienta tiene la capacidad de respaldar el proceso de toma de decisiones al proporcionar análisis más profundos y perspectivas más completas.

2.3. Análisis de fiabilidad

Los ferrocarriles desempeñan un papel crucial en el transporte global debido a su eficiencia, sostenibilidad, conectividad y seguridad. Mantenerlos en buen estado es esencial para garantizar su funcionamiento óptimo y continuar así aprovechando los beneficios que ofrecen en términos económicos, sociales y ambientales. Para ello, se requiere una inversión continua en infraestructura, mantenimiento y modernización. Esto implica mantener las vías férreas en condiciones óptimas, actualizar los sistemas de señalización y comunicación, así como garantizar el correcto mantenimiento de los trenes y equipos. En este sentido, una de las herramientas estadísticas que desempeña un papel clave en el estudio del comportamiento de los fallos en maquinaria en la actualidad es el análisis de fiabilidad.

En la literatura existen diversos estudios sobre fiabilidad y fallos en maquinaria, como es el caso de [13], donde se abordan diversas preguntas relacionadas con este campo, se examinan los aspectos claves de investigación, se identifican y analizan las principales herramientas utilizadas y se investigan los autores y países más influyentes en esta área. Estos estudios son importantes para garantizar un transporte eficiente y seguro.

Además, artículos como [14] describen la importancia de la implementación de políticas de mantenimiento efectivas para asegurar la confiabilidad y disponibilidad de los sistemas, en un entorno competitivo y desafiante. Estas estrategias de mantenimiento son fundamentales, y se utilizan soluciones avanzadas de monitoreo de maquinaria para desarrollar reglas de mantenimiento óptimas, controlando factores intrínsecos, como el desgaste de componentes, y extrínsecos, como las condiciones ambientales.

Para ello, se realizan pruebas en diferentes espacios arbitrarios viendo así como las componentes interactúan, buscando obtener resultados que ayuden a la toma de decisiones en la mejora y mantenimiento de los sistemas, como es el caso del estudio [15], que se centra en abordar el objetivo de fiabilidad en situaciones de concurrentes fallos, llevando a cabo pruebas en diferentes espacios y comparando las probabilidades de fallo del sistema.

Cuando se hacen estos estudios, es importante también identificar los fallos dominantes en los espacios aleatorios, ya que son los que, a priori, serán más propensos en un futuro. Este enfoque es necesario para estudiar y analizar para mejorar la fiabilidad de sistemas con eventos redundantes, como es en el caso de las locomotoras. En [17], para abordar este problema, desarrollan una técnica de búsqueda selectiva basada en la simulación mediante un algoritmo genético, utilizando un método de fiabilidad con matrices multiescala para calcular las probabilidades de fallo.

Por otro lado, en las últimas décadas, se han desarrollado diversas metodologías estadísticas para el análisis y el modelado de la fiabilidad de los sistemas reparables. Estas permiten estimar la intensidad de fallo de un elemento reparable utilizando sucesivos tiempos entre fallos, asumiendo una distribución exponencial para dichos tiempos, como se describe en [18]. Entre estos métodos, los más aceptados y utilizados han sido recopilados por la Comisión Electrotécnica Internacional (IEC) en su Comité Técnico TC56 "Dependability", a través de la emisión de normas para el espacio europeo. Estas normas toman en cuenta los manuales militares (MIL-HDBK) emitidos por el Departamento de Defensa de los Estados Unidos de América.

El estudio [19] aplica estas normas a los sistemas de tracción eléctrica reparables de 23 trenes, donde las pruebas realizadas detectaron períodos con mayor número de fallos seguidos de

períodos sin fallos, para todos los sistemas. Estos fallos recurrentes no consiguen reconocer un patrón estadísticamente significativo para aquellos períodos con mayor número de fallos, ya sea en distancia o tiempo, y no se detecta una asociación aparente entre tipos de fallos, como consecuencia de lo cual no es posible distinguir en la práctica entre fallos primarios y supuestos fallos secundarios. Además, se propone la utilización de "*clusters* de fiabilidad de ítem" para simplificar la presentación de resultados y adaptar las políticas de mantenimiento a las necesidades específicas de cada *cluster*.

Por otro lado, es importante tener en cuenta el problema que surge cuando se trata de múltiples unidades reparables. El estudio [20] propone un análisis con este enfoque, sabiendo que es un reto trabajar con diversas condiciones de funcionamiento. Se sabe que la fiabilidad de unidades idénticas puede variar de una unidad a otra debido a factores como son diferentes conceptos de diseño, procesos de fabricación, materiales o condiciones operativas y medioambientales. Por lo tanto, determinar homogeneidad y la clasificación de las unidades mediante pruebas de tendencia deben considerarse el primer punto de referencia del proceso, y utilizar un único modelo para representar el comportamiento de toda la población puede no ser válido, ya que puede conducir a conclusiones y decisiones erróneas.

Además, estudios como el realizado en [21] también buscan encontrar una metodología sencilla que sirva para estimar el número esperado de fallos de unidades reparables. Esta sencillez sumada al cambio consistente en los fallos a lo largo del ciclo de vida de las unidades hace que la estimación tenga una precisión razonable. Por otro lado, descubren que los modelos son más imprecisos cuando los cambios en los fallos son inconsistentes, siendo muy difícil encontrar una buena estimación paramétrica. En el estudio aplican el modelo MCF (Mean Cumulative Function) para estimar el tiempo hasta el fallo y poder llevar a cabo un mantenimiento preventivo con el que conseguir minimizar el número de fallos operativos, a pesar de que las estimaciones utilizadas no proporcionen una explicación de las causas del fallo.

En el caso de los medios de transporte ferroviario, el estudio [22] presenta un método que combina la simulación de Montecarlo con el análisis del árbol de fallos, buscando evaluar la fiabilidad y disponibilidad de los vehículos. El objetivo principal es analizar las causas y los efectos de los fallos, determinar los índices de fiabilidad relevantes e identificar los componentes más vulnerables que tienen un mayor impacto en el tiempo de inactividad y en la disponibilidad.

En resumen, el análisis de fallos abarca diferentes enfoques, desde análisis gráficos de variables recogidas hasta análisis estadísticos de fiabilidad, con el objetivo de estudiar y predecir el comportamiento de los fallos en diversos sistemas. A lo largo de la literatura, se han desarrollado una amplia variedad de modelos y metodologías que buscan proporcionar soluciones específicas a diferentes problemas, adaptándose al contexto en el que se plantean. Es importante destacar que cada problema y contexto pueden requerir un enfoque específico y una combinación adecuada de técnicas de análisis.

2.3.1. Mean cumulative function (MCF)

El análisis de datos de supervivencia, que representa el tiempo transcurrido hasta que ocurra un evento, requiere considerar dos variables de manera simultánea: el período de tiempo durante el cual se realizó el seguimiento y el estado al final del seguimiento, que especifica si el evento de interés ocurrió, como la muerte o una recaída, o si aún no ha ocurrido al final del seguimiento. Los períodos de seguimiento pueden variar entre individuos debido al reclutamiento progresivo en el estudio, pero el estado puede evaluarse posteriormente para todos los individuos en el mismo momento.

En el artículo [23] se propuso el uso del MCF (Mean Cumulative Function) como un método para resumir el número medio de eventos que ocurren en un individuo dentro de un período de tiempo en una población sujeta a eventos de censura, como pérdidas durante el seguimiento o finalización del estudio. Se asume que el tiempo hasta los eventos es independiente del tiempo de censura. Cuando el suceso estudiado puede darse de forma repetida en un individuo, hablamos de eventos recurrentes. Si un individuo solo puede experimentar un evento, el MCF es equivalente a la proporción de individuos que experimentan algún evento en algún momento.

Para definir el MCF, consideramos una muestra de n individuos que se observan durante un período de tiempo y que pueden experimentar eventos o censuras en los momentos t_1, t_2, t_3...t_N. En el momento t_j en el que ocurre un evento o censura, el MCF dependiente del tiempo (MCF(t_j)) se calcula de la siguiente manera:

$$MCF(t_j) = \sum_{k=1}^{j} \frac{e_k}{n_{k-1}},$$

Donde e_k es el número de eventos que ocurren en el momento t_k y n_{k-1} es el número de individuos en riesgo justo después del momento t_{k-1}. El número de individuos en riesgo en el momento t_{k-1} se obtiene restando del número total de individuos inicialmente en riesgo aquellos que fueron censurados antes del momento t_{k-1}. Además, el conjunto de individuos en riesgo solo disminuye cuando un individuo es retirado del seguimiento.

El artículo [24] recomienda utilizar el MCF para comparar muestras cuando la intensidad de los eventos posteriores varía. Este modelo se adapta de manera natural a los diferentes tiempos de seguimiento de los participantes en un estudio, lo cual es común en ensayos aleatorizados. Además, dado el enfoque en la evaluación económica, el MCF también se puede aplicar a los costos acumulados a lo largo del tiempo. [25]

En el artículo [24], utilizan este estimador para evaluar su utilidad para detectar diferencias entre grupos que experimentan diferentes patrones de intensidad de eventos, con el objetivo de mejorar la comprensión y prevención de las caídas en personas mayores. Concluyeron que el MCF permitía a los investigadores interpretar cuántas caídas podría evitar, en promedio, una intervención en comparación con un grupo de atención habitual durante un período de tiempo específico, a la vez que proporcionaba evidencia sobre cuánto tiempo se necesita para que una intervención comience a tener efecto.

2.3.2. Estimador Kaplan-Meier

El estimador de Kaplan-Meier [26], es un estimador no paramétrico de la función de supervivencia [27]. Sea S(t) la función de supervivencia de una población específica, que representa la probabilidad de que un individuo de esa población sobreviva más allá de un tiempo t. Para una muestra de tamaño N de esta población, consideramos los tiempos en los que ocurren las muertes, ordenados de menor a mayor:

$$t_1 \leq t_2 \leq t_3 \leq \cdots \leq t_N$$

Para cada uno de estos tiempos, definimos:

- d_i como el número de muertes en el momento t_i, y

- n_i como el número de sujetos en riesgo justo antes de t_i. En el caso sin censura, n_i sería el número de sobrevivientes inmediatamente antes del tiempo t_i. Con censura, n_i se calcula restando el número de casos censurados del número de sobrevivientes en ese momento. Es decir, solo se observan los individuos vivos que no han abandonado el estudio en el momento en que ocurre una muerte.

El estimador de Kaplan-Meier de S(t) se calcula mediante el producto:

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}.$$

donde el producto se realiza para todos los tiempos t_i menores que t.

En el estudio [28] se describe que a menudo se utiliza una curva de Kaplan-Meier para resumir visualmente los datos de tiempo hasta el evento. Los pacientes censurados proporcionan información de que no ha ocurrido el evento hasta el momento de la censura, evitando así la exclusión de esta información útil. El período de tiempo se divide en intervalos y se estima la tasa de supervivencia (utilizando la estimación de Kaplan-Meier) en función de las personas en riesgo durante cada intervalo, es decir, aquellas que no habían experimentado el evento al comienzo del intervalo y no fueron censuradas antes o durante dicho intervalo. Las estimaciones se presentan en forma de curva, donde el eje *y* indica la proporción de individuos en riesgo de experimentar el evento, y el eje *x* representa el tiempo.

Las curvas de Kaplan-Meier se caracterizan por escalones, que representan la aparición de uno o más eventos. A menudo se presentan con intervalos de confianza del 95% [29], y las diferencias entre curvas pueden evaluarse estadísticamente, generalmente mediante la prueba de rangos logarítmicos. La curva también puede presentarse invertida, intercambiando el evento y la no ocurrencia del evento. Los errores estándar y los intervalos de confianza de las probabilidades de supervivencia estimadas se pueden calcular utilizando el método de Greenwood [30].

El estudio [31] explica que existen cuatro consideraciones especialmente importantes al interpretar las curvas de Kaplan-Meier. En primer lugar, la validez de la curva depende de la suposición de que todos los participantes en el análisis (incluyendo aquellos censurados y no censurados) tienen el mismo riesgo de experimentar el evento. Sin embargo, esta suposición no se cumple cuando la censura ocurre debido a un evento externo que excluye el evento de interés, como la recurrencia de la enfermedad de interés. En tal caso, la curva de Kaplan-Meier puede ser una representación sesgada de la verdadera curva de supervivencia. Por ejemplo, al utilizar el análisis de Kaplan-Meier para evaluar la supervivencia de un implante en un registro de artroplastia, se tiende a sobreestimar el riesgo de revisión, ya que la muerte impide la revisión del implante, inflando artificialmente el riesgo aparente. Se han desarrollado métodos para abordar estos riesgos concurrentes, pero a menudo requieren supuestos adicionales.

En segundo lugar, la precisión estadística disminuye a medida que aumenta el período de seguimiento y se reduce el número de individuos que contribuyen debido a la ocurrencia del evento o la censura. Esto puede tener un impacto significativo, y un solo evento puede producir un escalón mucho mayor en la curva cuando ocurre más tarde durante el seguimiento. Este aspecto se refleja generalmente en la amplitud de los intervalos de confianza de la curva. Por lo tanto, para interpretar adecuadamente una curva de Kaplan-Meier, es crucial comunicar el número de pacientes en riesgo en momentos clave durante el período de seguimiento.

Finalmente, se asume que en cualquier momento los pacientes censurados tienen las mismas perspectivas de supervivencia que aquellos que siguen en seguimiento. Esta suposición no es fácil de verificar, ya que la censura puede deberse a varias razones.

También se ha de asumir que el evento ocurre en el momento especificado. Esto no representa un problema para los datos de concepción, pero puede ser un desafío si el evento es, por ejemplo, la recurrencia de un tumor que se detectaría en un examen periódico. En ese caso, sólo sabríamos que el evento ocurrió entre dos exámenes. Esta imprecisión sesgaría las probabilidades de supervivencia, incrementándolas artificialmente. Cuando las observaciones se realizan en intervalos regulares, esto se puede tener en cuenta fácilmente utilizando el método actuarial.

En resumen, el estimador de Kaplan-Meier es una herramienta esencial en el análisis de datos de supervivencia, ya que permite estimar de manera adecuada la probabilidad de supervivencia en presencia de datos censurados. Su aplicabilidad se extiende a diversas áreas de investigación, como estudios clínicos, epidemiológicos y de ciencias de la salud, proporcionando información valiosa sobre la supervivencia de los individuos y permitiendo la comparación entre grupos en términos de supervivencia.

3. Desarrollo

En esta sección se presentará un análisis detallado del problema planteado, los datos involucrados y la metodología propuesta para su resolución, así como las herramientas necesarias para llevar a cabo dicho análisis. También se describirán los gráficos implementados en el cuadro de mandos y el análisis de fiabilidad.

3.1. Análisis del problema

Este estudio se centra en las locomotoras de tipo diésel y en los eventos registrados en ellas. Las locomotoras se encuentran agrupadas por flotas, cada una de ellas perteneciente a un cliente diferente, y están repartidas por varios países.

Los datos utilizados provienen de la monitorización de las diferentes locomotoras, pertenecientes a 17 flotas, desde el año 2018 hasta la actualidad. Estas locomotoras se desplazan por numerosos países, y poseen un sistema de telemetría que permite un monitoreo eficiente para poder analizar los datos recopilados, contribuyendo a un mejor mantenimiento y rendimiento de las locomotoras en servicio.

En cada locomotora fabricada por Stadler, se encuentra instalado un ordenador denominado VCU (*Vehicle Controller Unit*), cuya función principal es controlar todos los aspectos del vehículo. Para el sistema de telemetría, se utilizan variables específicas de las cuales se registran sus valores, siendo la VCU responsable de almacenar estas mediciones. Destacar que la programación y el mantenimiento de la VCU, así como la selección de las variables a medir, son realizados por Stadler Rail Valencia.

Stadler utiliza su sistema de telemetría para recoger datos de cada vehículo y almacenarlos en una base de datos. Para ello, se reconoce a cada vehículo con un identificador propio, y se registran de él valores como la velocidad, tiempo en marcha, potencia, dirección, temperatura o posición entre otras muchas, y para cada una de ellas se registra el momento en el que se toma la medida en formato *"día-mes-año hh24:mi:ss.ms"*. El VCU se encarga de leer y escribir en su memoria interna el valor que recogen los sensores del vehículo, cada 100 milisegundos (frecuencia especificada por Stadler), denominándose registro continuo. Destacar que se trata de una memoria no acumulativa, lo que implica que solo almacena un registro por variable, actualizándose en cada tiempo.

Esta unidad desempeña un papel fundamental a la hora de gestionar el control de errores, ya que cuenta con las condiciones que determinan si se produce o no un error. Esto se diferencia del registro continuo de variables previamente explicado en que en este caso sí que se almacenan todos los eventos recogidos, es decir, cuando se produce un fallo, se registra el momento exacto en el que sucedió *(día-mes-año hh:mm:ss.ms)*, y cuando se produce un nuevo error, se guarda el nuevo instante de tiempo, pero sin reemplazar al previo. Cabe destacar que el registro de los eventos y el registro continuo de variables no tienen relación, lo que implica que no se puede asegurar que los tiempos obtenidos por cada uno de ellos sean iguales, es decir, es posible que los eventos estén recogidos en instantes en los que no se recogen variables.

Los datos que se recogen son leídos por el llamado TWC *(Train To Wayside Communication)*, también llamado MPU *(Multi-Purpose Unit)*. Este sistema lee la memoria de sistemas como el VCU o el GPS (que recoge la geolocalización de los vehículos) y se encarga de empaquetar en

ficheros *.csv* los datos recabados para ser enviados a la empresa. Ahí son recibidos en un servidor, donde son descomprimidos y parseados para su posterior tratamiento de software y almacenamiento en base de datos.

La falta de captura de instantes de GPS por parte del VCU puede generar discrepancias entre los *timestamps* de los eventos y los *timestamps* de las posiciones geográficas registradas. Esta situación puede ocasionar problemas a la hora de localizar con precisión los eventos en el espacio geográfico.

Al no contar con los datos de posición geográfica capturados directamente por el VCU, puede resultar difícil establecer una correlación exacta entre los eventos y su ubicación geográfica, afectando a la capacidad de visualizar y analizar los eventos en un contexto geoespacial preciso.

Como consecuencia, es importante tener en cuenta esta limitación al interpretar y analizar los datos recopilados. Es posible que se requieran enfoques alternativos o métodos de estimación para asociar los eventos con sus ubicaciones geográficas más cercanas. La solución implementada para atajar este problema será descrita en el apartado 3.4. Es importante destacar que esta discrepancia en los *timestamps* no invalida los datos recopilados por el VCU. Sin embargo, es fundamental tener en cuenta esta inconsistencia al realizar análisis y tomar decisiones basadas en los datos geográficos asociados a los eventos.

Cabe destacar que, en situaciones urgentes, es posible que se requiera obtener un archivo específico que aún no ha sido entregado, pero que puede ser descargado desde el vehículo bajo solicitud. Esto permite acceder a la información necesaria de manera rápida y eficiente, evitando retrasos en la obtención de datos críticos.

Datos de eventos:

Para la realización del proyecto, se cuenta con nueve eventos diferentes, siendo los más recurrentes en las locomotoras. Estos se encuentran codificados en base de datos mediante un identificador de tipo entero y son los observados en la tabla 1.

EVENTO	DESCRIPCIÓN
832	[illegible]
836	[illegible]
923	[illegible]
957	[illegible]
1822	[illegible]
1832	[illegible]
1839	[illegible]
1846	[illegible]
2882	[illegible]

Tabla 1: Descripción de los eventos estudiados

Estudiar estos eventos específicos permite identificar rápidamente situaciones problemáticas y tomar medidas correctivas de manera oportuna. La detección temprana de estos eventos contribuirá a mejorar la eficiencia y el rendimiento del sistema, asegurando un funcionamiento más confiable y reduciendo los tiempos de inactividad no planificados.

Estos eventos se clasifican en una escala de color basándose en su prioridad y categoría según el esquema de la tabla 2.

CATEGORÍA	PRIORIDAD	COLOR
A	**X** < 20000	ROJO
A	20000 <= **X** < 40000	NARANJA
A	**X** >= 40000	AMARILLO
B	**X** < 40000	NARANJA
B	**X** >= 40000	AMARILLO
C	**X**	VERDE
D	**X**	VERDE
E	**X**	VERDE

Tabla 2: Escala de color en función de la categoría y prioridad de los eventos

A continuación, se detalla la guía de los colores:

- Rojo: indica que el evento detectado implica la circulación del tren. Este tipo de evento requiere una solución inmediata para que el vehículo pueda reanudar su funcionamiento.
- Naranja: significa que el vehículo ha de ser retirado al final del día para recibir una revisión exhaustiva. Este tipo de evento indica que el problema identificado, aunque no impide que el tren circule en ese momento, requiere una evaluación o reparación más detallada antes de poder volver a operar con normalidad.
- Amarillo: indica que el evento ha de ser revisado en la próxima sesión de mantenimiento programado del vehículo. Este tipo de evento es menos urgente y no requiere de intervención inmediata, pero ha de ser abordado en la próxima oportunidad planificada.
- Verde: representa un evento con poca importancia. Este tipo de evento es meramente informativo y no requiere de acción inmediata. Proporciona información adicional sobre el funcionamiento de la locomotora, pero no afecta a su capacidad de operación.

Esta escala de color se implementará en el cuadro de mandos para facilitar la toma de decisiones informadas en base a la importancia de los eventos y comprenderlos de mejor modo, promoviendo la seguridad y la fiabilidad en el entorno operativo.

3.2. Software utilizado

Para llevar a cabo el estudio, se han utilizado diversas herramientas que desempeñan un papel clave, como son las descritas a continuación:

En primer lugar, se hace uso de Oracle SQL Developer [32], un entorno de desarrollo gratuito que simplifica la gestión de bases de datos de Oracle. Esta herramienta se utiliza para generar las consultas SQL que obtienen los datos necesarios para el proyecto, facilitando así tanto el acceso como la manipulación de la información almacenada en la base de datos.

En segundo lugar, se utiliza Python [33], un lenguaje de programación versátil con el que se llevan a cabo varias tareas del estudio, como son la limpieza de datos, la realización de cálculos complejos, la generación de gráficos y la implementación de modelos de predicción. Python posee una amplia gama de bibliotecas y herramientas que facilitan en gran medida estas tareas analíticas. A continuación, se citan las librerías más importantes utilizadas para el desarrollo de este trabajo:

- *pandas:* es una librería diseñada para análisis y manipulación de datos estructurados. En este caso, ha sido utilizada para la lectura de los archivos de datos y la conversión de las columnas.
- *numpy:* una librería que ofrece soporte para la manipulación de vectores y matrices.
- *reliability:* se trata de una librería que proporciona funcionalidades para realizar análisis de fiabilidad y construir modelos de supervivencia. En este caso, fue utilizada para implementar los modelos MCF y Kaplan-Meier.
- *matplotlib:* es una librería para crear visualizaciones que ofrece una amplia gama de herramientas para crear gráficas. En este caso, se utilizó para dibujar las gráficas asociadas a los modelos de fiabilidad.

En tercer lugar, Microsoft Power BI [34], un conjunto de servicios de software, conectores y aplicaciones para transformar y visualizar los datos recuperados. Esta herramienta se utiliza para la implementación del cuadro de mandos que muestre la información recolectada de forma comprensible y coherente, ya que permite crear visualizaciones interactivas e interesantes, atractivas para el usuario.

Se ha decidido utilizar Power BI como herramienta de visualización de datos debido a las numerosas funcionalidades que ofrece y a su capacidad para satisfacer las necesidades de la empresa, ya que es una plataforma sólida y ampliamente reconocida en el campo de la visualización de datos, lo que la convierte en una elección confiable y respaldada por la comunidad empresarial.

Finalmente, se hace uso de Sysloc Server, una aplicación de visualización de datos perteneciente a Stadler, mediante la cual se comparan los resultados obtenidos en el estudio con los resultados obtenidos en esta herramienta. Su objetivo principal es verificar la precisión y la corrección de los datos analizados, permitiendo de esta manera identificar anomalías y discrepancias y pudiendo así realizar las correcciones necesarias.

En resumen, estas herramientas proporcionan un conjunto completo de capacidades para el estudio, pasando desde la extracción y manipulación de los datos hasta la visualización interactiva de los resultados, asegurando de esta manera un análisis eficiente e integral de la información recabada de las locomotoras.

3.3. Estructura de la base de datos

Para este estudio, se ha utilizado la base de datos de Sysloc, perteneciente a Stadler. Esta contiene información detallada sobre más de 150 locomotoras, distribuidas en diferentes ubicaciones alrededor del mundo.

A continuación, en la ilustración 1, se muestra el esquema del modelo Entidad-Relación de las tablas que se emplean en la consulta de los datos.

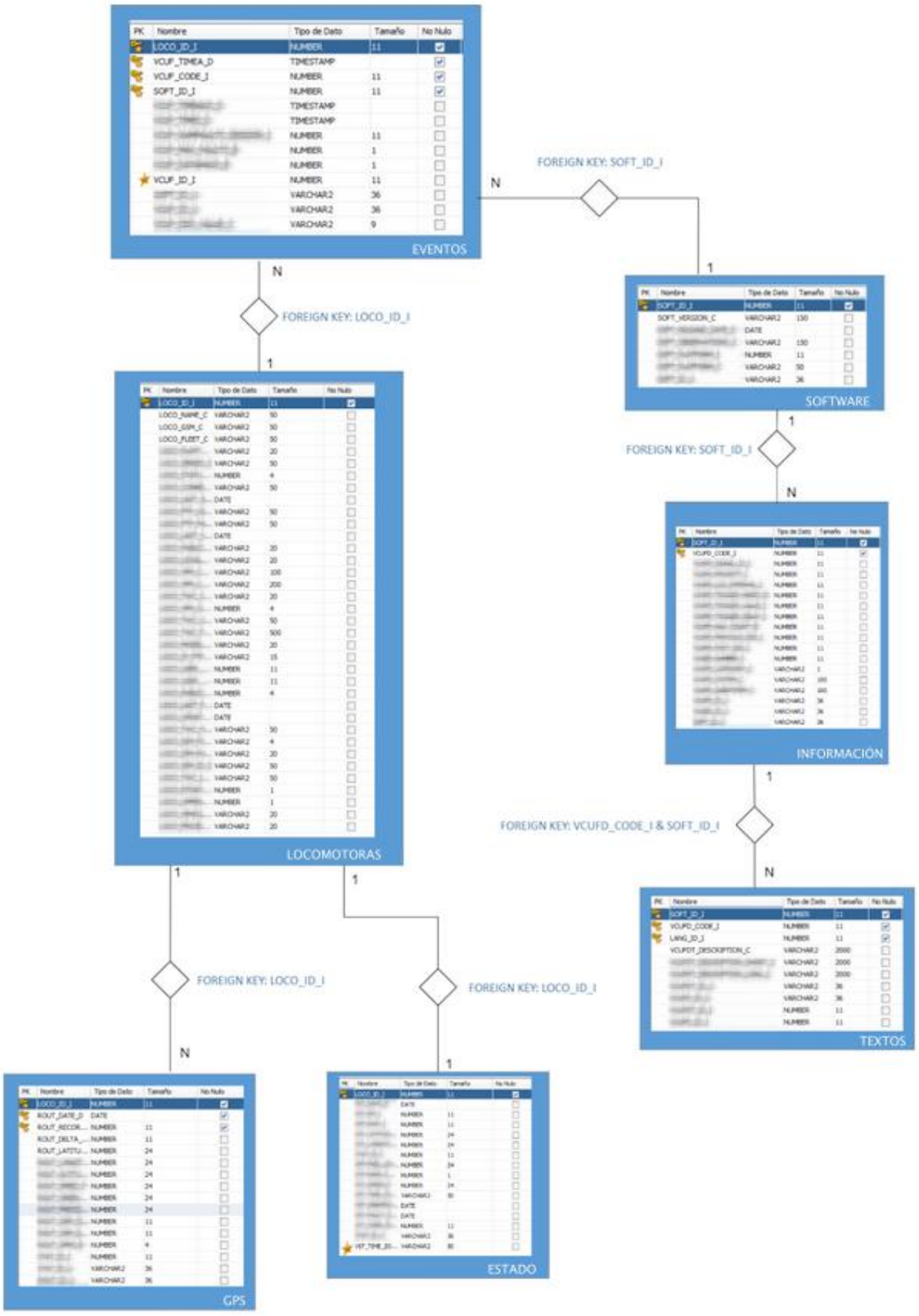

Ilustración 1: Esquema entidad-relación de la base de datos

En este caso, para la obtención de los datos de las locomotoras, se han hecho uso de siete tablas con diferentes informaciones relevantes para el análisis y visualización de los eventos. El

nombre real de las tablas será sustituido por motivos de privacidad, pasando a llamarse LOCOMOTORAS, GPS, ESTADO, EVENTOS, SOFTWARE, INFORMACIÓN y TEXTOS. A continuación, se detallan las relaciones entre las diferentes tablas involucradas.

La tabla central en este esquema es LOCOMOTORAS, que almacena las variables asociadas a las locomotoras, como el nombre de la flota y el de las locomotoras. Esta tabla se relaciona con otras tablas para obtener información adicional.

Una de las relaciones se establece con la tabla GPS, que contiene variables relacionadas con la posición geográfica del vehículo, como la latitud y la longitud. La unión entre estas tablas se realiza mediante el identificador del vehículo, *loco_id_i*, lo que permite establecer una relación 1–N. Esto significa que una fila de la tabla de locomotoras puede corresponder a múltiples filas de la tabla de rutas, ya que cada locomotora tiene varios registros de ruta asociados a ella.

Además, la tabla LOCOMOTORAS se relaciona con la tabla ESTADO, que recopila variables relacionadas con el estado del vehículo, como la zona horaria en la que se encuentra. Esta relación también se establece mediante el identificador *loco_id_i*, generando una relación 1–1. En este caso, cada fila de la tabla de locomotoras se asocia con una sola fila de la tabla de estado del vehículo y viceversa.

La tabla de LOCOMOTORAS se une también con la tabla EVENTOS, que contiene información sobre los eventos registrados, como los códigos de evento y los momentos en que se producen los errores en formato *timestamp*, o los momentos en los que se desactivan, pudiendo esta columna no tener valor ya cabe la posibilidad de que los eventos continúen activos. La unión entre estas tablas se realiza mediante el identificador *loco_id_i*, estableciendo una relación 1–N. Esto implica que una fila de la tabla de locomotoras puede tener múltiples filas asociadas en la tabla de eventos.

Otra relación importante se establece entre la tabla EVENTOS y la tabla SOFTWARE, que almacena información sobre las versiones de software presentes en los vehículos. Esta relación se basa en el identificador de software *soft_id_i* y tiene una cardinalidad N–1, lo que significa que varias filas de la tabla de eventos pueden relacionarse con una sola fila de la tabla de versiones de software.

La tabla SOFTWARE se relaciona, a su vez, con la tabla INFORMACIÓN, que contiene información detallada sobre los eventos , como la prioridad, la categoría y el sistema en el que se produce el evento. Esta relación se establece mediante el identificador de software *soft_id_i* y presenta una cardinalidad 1–N. Esto indica que una fila de la tabla de software puede estar asociada con varias filas de la tabla de descripción de eventos .

Finalmente, se establece una relación entre la tabla de información de eventos (INFORMACIÓN) y la tabla TEXTOS mediante una clave compuesta por el identificador de software *soft_id_i* y el código de evento *vcufd_code_i*. Esta última tabla contiene variables como son las descripciones de los eventos, y mantiene una relación 1–N con la tabla de información de eventos. Esto implica que una fila de la tabla de información de eventos puede corresponder a varias filas de la tabla de textos de descripción de eventos.

Este esquema de relaciones permite obtener una visión completa y detallada de la información relacionada con las locomotoras, sus rutas, estados y eventos registrados. Al unir y consultar estos datos, se pueden realizar análisis exhaustivos y generar visualizaciones significativas para el seguimiento y la toma de decisiones en relación con los eventos de las locomotoras.

Es importante destacar que los nombres de las variables presentes en las tablas se clasifican dependiendo del tipo de datos al que pertenecen. Esta clasificación se indica al final del nombre

de la variable mediante el formato "_X", donde "X" representa una letra que identifica el tipo de dato correspondiente.

Por ejemplo, si la variable es del tipo *character*, se utiliza la letra "c" al final del nombre. Si la variable es un identificador, se utiliza la letra "i". Si la variable es de tipo *bool*, se utiliza la letra "b", y si la variable se corresponde con el tipo *date*, se utiliza la letra "d".

Esta convención de nomenclatura ayuda a identificar rápidamente el tipo de datos al que pertenece cada variable en las tablas del esquema. Proporciona una mayor claridad y consistencia en la estructura de datos, facilitando la comprensión y manipulación de la información para su posterior análisis y visualización en el cuadro de mandos.

3.4. Carga de datos en Power BI

La carga de datos es uno de los desafíos más duros del estudio. Esto se debe a que las consultas requeridas para la recuperación de los datos implican la unión de una gran cantidad de tablas, lo que genera un aumento considerable en los tiempos y costes de carga. La gran complejidad de las consultas y la necesidad de obtención de datos complejos y precisos añaden dificultad adicional al proceso de extracción.

Aun así, a pesar de los desafíos encontrados en la extracción de datos, se consiguen superar consiguiendo la obtención de la información requerida para el estudio.

El proceso de carga de los datos se divide en dos subsecciones. La carga de los datos para el cuadro de mandos y la carga de los datos para el análisis de fiabilidad de las locomotoras.

Cuadro de mandos:

Los datos dirigidos al cuadro de mandos son extraídos mediante SQL Developer, partiendo de las tablas completas, y buscando realizar consultas que condicionen la información a mostrar.

En primer lugar, Stadler tiene datos de todos los vehículos que fabrica, ya sean locomotoras, tranvías, metros, etc. En este caso solo son necesarios los datos de las locomotoras de tipo diésel, por lo que solo interesa recuperar la información de ellas.

En segundo lugar, la empresa también recolecta todos los diferentes tipos de eventos que detectan estas locomotoras, siendo más de 136. En este estudio simplemente interesan los nueve eventos previamente descritos, por lo que se condicionan las consultas para que solo devuelvan la información relacionada con ellos.

Para la obtención de los datos se genera una consulta SQL mediante la que se obtienen los nombres de las locomotoras, sus flotas, plataformas, tiempos de evento tanto en hora local como UTC, categoría y prioridad del evento, duración, descripción, sistema y versión de software entre otros.

Se toma la decisión de realizar una única consulta fusionando las tablas con los datos de las locomotoras, junto a las tablas que resuelven los eventos, ya que el coste estimado disminuía considerablemente, pasando de 1259007 a 833871 unidades. A su vez, el hecho de combinar todas estas tablas en una única, mejora significativamente los tiempos de carga de los gráficos que se realizarán, proporcionando una experiencia de usuario más rápida y fluida. La consulta está incluida en el anexo 1.

En tercer lugar, se encuentran los datos de geolocalización, registrados en la hora local donde se encuentra el vehículo. Es importante destacar que esta hora difiere de la hora internacional (UTC) en la que se reciben el resto de los datos. A su vez, estos registros se miden en unos instantes de tiempo que no tienen por qué coincidir con los instantes en los que se detecta el evento.

Los datos de latitud y longitud, como se ha mencionado previamente, se encuentran en la tabla GPS. Los datos de la zona horaria en la que se encuentra el vehículo están en la tabla ESTADO, y los datos de eventos se encuentran en EVENTOS. Para unir estas tres tablas, es necesario generar una nueva columna en la que la hora del evento en formato internacional se convierte a la hora en la zona horaria en la que se encuentra el vehículo.

Una vez se tiene en la hora correcta, se busca la posición GPS *post-evento* más cercana a él, es decir, el siguiente *timestamp* más cercano al *timestamp* del error. Se decidió sólo buscar la posición posterior en vez de la mínima entre la anterior y posterior debido a que esa consulta triplica el coste esperado.

Además, se establece un límite máximo de 34 minutos ya que se busca una combinación precisión-cantidad de datos, es decir, si se busca la posición GPS más próxima pasados los 34 minutos del evento, no será precisa la localización de la locomotora. Se decide ese tiempo en base al propuesto para la aplicación Sysloc Server de Stadler, y calculando tiempos de costes (cuanto más se aumente el rango, más costoso será y más precisión se pierde). Con esta solución se dispondrá de algunos eventos para los cuáles no haya sido posible extraer su geolocalización; no obstante, se asegura una mayor precisión para aquellos casos en los que sí es posible.

Por tanto, se genera una consulta SQL para obtener los datos de localización de los eventos (siendo, como se ha descrito anteriormente, las ubicaciones más cercanas posteriores a los mismos, en un intervalo máximo de 34 minutos) de los dos últimos meses. Es importante destacar que Power BI tiene una limitación de carga de datos de 10 minutos. Se puede observar la consulta realizada en el anexo 1.

En la tabla 3 se muestra el tiempo de carga asociado a diferentes períodos de días previos para todas las flotas.

DÍAS CARGADOS	TIEMPO DE CARGA
30 días previos	2.76 min
60 días previos	5.49 min
90 días previos	11.08 min

Tabla 3: Tiempos de carga de datos asociados a GPS según el histórico consultado

Como se observa, el tiempo de carga es bastante elevado. Ante esta situación, se toma la decisión de cargar únicamente los datos de 60 días previos para evitar largos períodos de espera. La gran complejidad de la consulta y su elevado coste hacen que sea poco factible cargar más datos en Power BI.

Una vez planteadas las consultas que rescatan los datos necesarios para la generación de los informes, se procede a la carga de ellos en Microsoft Power BI.

Esta aplicación permite la importación de datos desde diferentes fuentes, incluyendo las bases de datos de Oracle. Tiene tres tipos diferentes de carga de datos: "importación", "*DirectQuery"* y "composición".

En este caso, se hace uso de la carga de datos mediante *DirectQuery*. De este modo, cuando se consulta el modelo, se utilizan consultas nativas que recuperan los datos directamente de su origen. Se elige esta opción debido al gran volumen de datos que se maneja, y los informes que se buscan obtener necesitan datos en tiempo real, continuamente actualizados.

La carga de datos mediante *DirectQuery* ofrece varias ventajas, como son:

- No tiene límite de tamaño: no hay restricciones en el tamaño de los datos que se pueden manejar.
- No requiere una actualización programada de datos: los datos se actualizan automáticamente desde el origen en tiempo real.
- No se aplican los límites de tamaño de los modelos de importación: no hay restricciones en la cantidad de datos que se pueden importar.
- Los usuarios del informe ven los datos más recientes al tener la capacidad de interactuar con filtros y segmentaciones.
- Los iconos del panel se pueden actualizar automáticamente cada 15 minutos.

Sin embargo, existen algunas limitaciones asociadas a estos modelos:

- Las expresiones de *mashup* y *Power Query* solo pueden ser funciones que se pueden traducir en consultas nativas que el origen de datos comprenda.
- Las fórmulas de *DAX* (lenguaje de programación de Power BI) están limitadas a utilizar solo las funciones que se pueden traducir en consultas nativas del origen de datos. No se admiten las tablas calculadas.

En términos de recursos del servicio Power BI, los modelos de *DirectQuery* requieren una memoria mínima para cargar el modelo (solo los metadatos) al realizar una consulta. En algunas ocasiones, el servicio Power BI puede utilizar recursos significativos de CPU para generar y procesar las consultas enviadas al origen de datos. Esto puede afectar al rendimiento, especialmente cuando varios usuarios consultan el modelo al mismo tiempo.

Tras realizar la importación de datos mediante las consultas descritas, el modelo final de Power BI se representa en la ilustración 2.

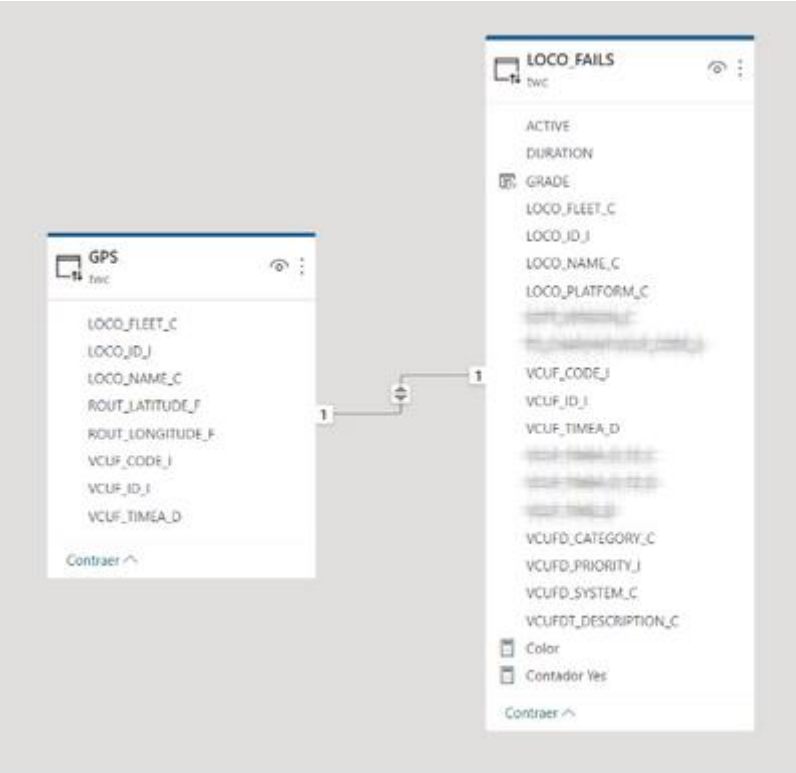

Ilustración 2: Modelo relacional cargado a Power BI

Estas dos tablas se relacionan entre sí mediante el atributo *vcuf_id_i*, presente en ambas tablas, y que corresponde con el identificador del evento. Este atributo es único, y posee una cardinalidad 1–1. Esto significa que un evento en la tabla de eventos se corresponde a una única posición GPS, y a su vez, una posición GPS está asociada a un único evento de la tabla de eventos.

Esta relación entre las tablas permite establecer una conexión directa entre los datos de eventos y las ubicaciones geográficas correspondientes. El modelo final de Power BI proporciona una representación visual de estos datos interrelacionados, lo que permite una mejor comprensión y exploración de la información relacionada con los eventos y las ubicaciones asociadas.

Análisis de fiabilidad:

En el cuadro de mandos también se va a mostrar el análisis de fiabilidad realizado para cada locomotora mediante el estimador Kaplan-Meier. Para lograr esto, se recopilan bajo demanda mediante Power Bi los datos del evento y locomotora deseada para calcular el análisis en tiempo real, garantizando de este modo que esté siempre actualizado. La consulta generada para obtener estos datos es la representada en la ilustración 3.

```
SELECT hvf.loco_id_i, hvf.VCUF_TimeA_d, hvf.vcuf_code_i, tl.loco_name_c, tl.loco_fleet_c
FROM [redacted] hvf
LEFT JOIN [redacted] tl ON (tl.loco_id_i = hvf.loco_id_i)
WHERE  hvf.vcuf_code_i = '"+str(dataset['VCUF_CODE_I'].iloc[0])+"'
    AND hvf.VCUF_TimeA_d  <= SYSDATE
    AND tl.loco_name_c = '"+(dataset['LOCO_NAME_C'].iloc[0])+"'
```

Ilustración 3: Consulta de SQL para la obtención de datos para el análisis de fiabilidad Kaplan-Meier en Power BI

Una vez recopilados los datos, se almacenan en un *dataframe*. Este se encuentra formado por las columnas *loco_id_i*, *vcuf_timea_d*, *vcuf_code_i*, *loco_name_c* y *loco_fleet_c*.

Tras obtener el *dataframe*, los datos se ordenan por tiempo y se crea una columna llamada *diferencia*. Esta columna indica el tiempo transcurrido entre el evento en cuestión y el anterior evento registrado para ese código de evento en esa locomotora. De este modo, se obtiene el tiempo entre cada uno de los eventos para el análisis de Kaplan-Meier. Es decir, el seguimiento de cada evento comprende desde el momento en el que se da el evento anterior hasta que se vuelve a originar dicho evento.

El período de estudio para el análisis de los eventos en cada locomotora comienza en el momento en el que se registra el primer evento, termina en el último evento registrado y consta de tantas muestras como eventos se hayan originado durante dicho periodo de tiempo.

Cabe destacar que se incluirá otro tipo de análisis de fiabilidad en el apartado 4.1.1., el MCF. En este caso, el procesamiento de los datos será muy similar, pero introduce un cambio relevante: el valor de la columna *diferencia* indica el tiempo transcurrido desde el inicio del estudio hasta la fecha de ocurrencia del evento. De este modo, se obtiene el tiempo total transcurrido desde el inicio del estudio hasta el registro de cada evento. Este cambio se introduce ya que el MCF es una herramienta que permite el modelado de eventos recurrentes como ya se mencionó en el apartado 2.4.1.

3.5. Análisis exploratorio de los datos

A continuación, se describe el proceso de exploración de los datos. Una vez concluida la carga de los datos en Power BI, se realiza un análisis exploratorio de los mismos, con el objetivo de comprender de una forma más completa los eventos registrados en las locomotoras y su distribución. Se evaluará la calidad de los datos y se plantearán los posibles preprocesados necesarios para los análisis posteriores.

En primer lugar, se decide examinar la distribución total de los códigos de evento, sin considerar la locomotora en la que se produjeron. Este enfoque, como se observa en la ilustración 4, permite identificar los tipos de eventos más comunes y comprender la frecuencia con la que ocurren en el conjunto de datos.

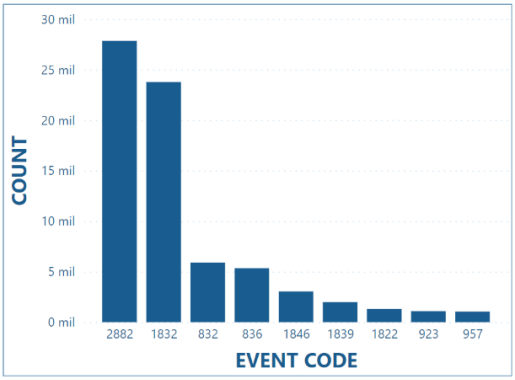

Ilustración 4: Gráfico de barras de la frecuencia por código de evento

La media de errores general es de 7,78 mil. Destacan los eventos 2882 y el 1832 con más de 20 mil registros, muy por encima del resto que se encuentran por debajo de la media.

Además, se decide realizar un análisis de eventos por flota, sin tener en cuenta el código de evento específico, como se observa en la ilustración 5. De este modo se puede obtener una visión más amplia de la cantidad de eventos que se registran en cada flota, permitiendo así realizar comparaciones.

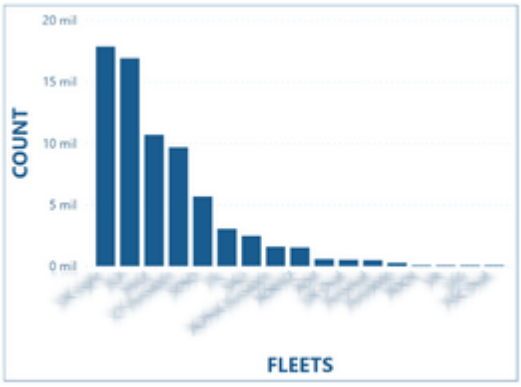

Ilustración 5: Gráfico de barras de la cantidad de eventos registrados por flota

En este caso, la media de errores por flota es de 4.32 mil. Destacan las dos primeras flotas, ya que son en las que más eventos se producen, con más de 16 mil registros cada una, así como las tres últimas, las cuales no llegan a los mil. Esto ocurre ya que esas flotas son las que cuentan con más locomotoras, por lo que es lógico que tengan más eventos en total.

Por otro lado, debido a la gran cantidad de locomotoras de las que se dispone de datos, llevar a cabo un análisis individual y detallado de cada una de ellas resultaría poco práctico y consumiría un tiempo considerable. En lugar de eso, se opta por realizar un análisis más exhaustivo de las flotas A y B, que son las que se utilizarán para explicar en profundidad el análisis de fiabilidad.

En primer lugar, se explora la flota A, compuesta por 34 locomotoras. Se muestra tanto la distribución de eventos por locomotora, sin tener en cuenta el código de evento, como la distribución de códigos de evento dentro de la flota, para comprobar qué eventos son más comunes, pudiéndose ver en la ilustración 6. Estas gráficas pretenden ofrecer una visualización de la distribución de los diferentes eventos registrados en esta flota.

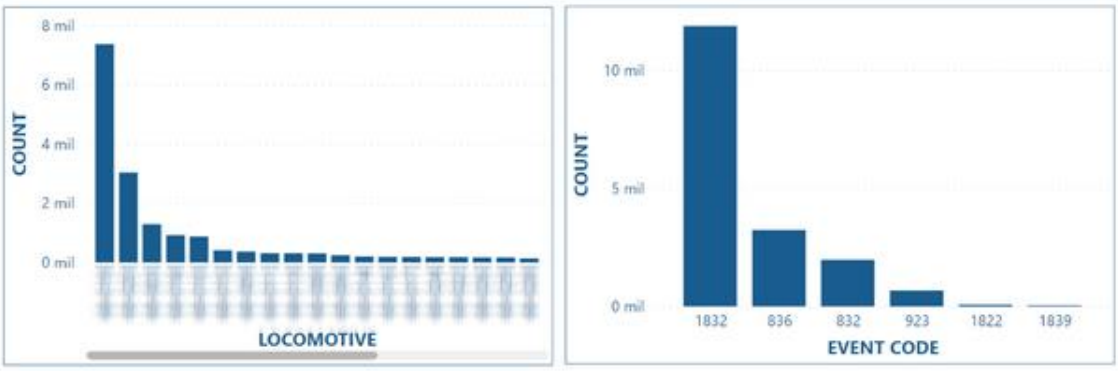

Ilustración 6: Gráficos de barras para el recuento de eventos registrados por locomotora y eventos totales registrados en la flota A por código de evento

En este caso, la media de errores por locomotora es de 0.34 mil, por ello destaca la primera locomotora, ya que tiene 7.2 mil registros, siendo también más del doble de registros que la segunda más frecuente (3 mil). Por otro lado, la media de errores por código de evento es 2.99

mil, y por ello destaca el 1832, con casi 12 mil registros. Se trata de una flota con un gran registro de eventos, llegando a los 17.85 mil registros diferentes.

En segundo lugar, se analiza del mismo modo la flota B, compuesta por 10 locomotoras, mostrándose en la ilustración 7.

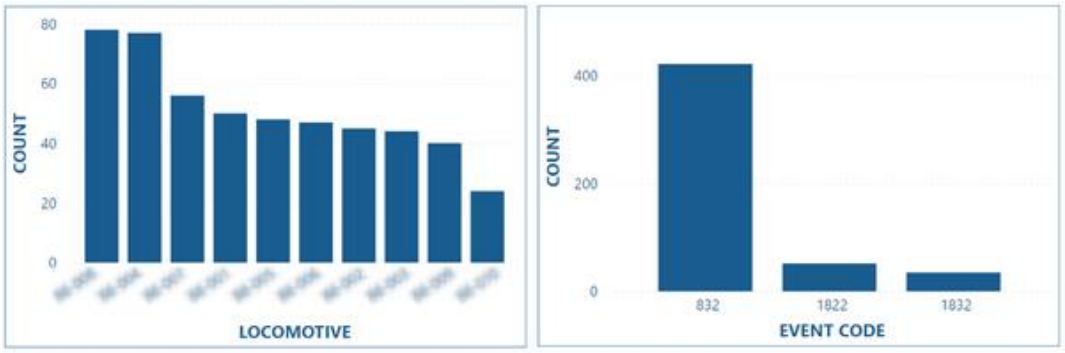

Ilustración 7: Gráficos de barras para el recuento de eventos registrados por locomotora y eventos totales registrados en la flota B por código de evento

En este caso, la distribución de eventos dentro de la flota es bastante parecida para todas las locomotoras, siendo la media 50.8. Se puede observar como el número de registros es bastante más bajo que en la flota anterior, siendo aquí únicamente 509 eventos detectados. En cuanto a la distribución de códigos de evento, se trata de una flota en la que se dan pocos registros, teniendo solo tres de los nueve diferentes. Destaca el 832 que se da 422 veces, superando en gran medida el valor medio.

Por otro lado, se realizará un estudio del tiempo transcurrido entre el registro de los eventos en una locomotora de cada una de las flotas previamente analizadas, la 1 de la flota A con 133 eventos registrados, y la 2 de la flota B con 41.

Este estudio persigue el objetivo de averiguar la calidad de los datos que se usarán en el análisis y explorar posibles transformaciones previas necesarias. Para ello se estudiarán los periodos de tiempo entre eventos y su distribución.

Primeramente, se decide seleccionar un único valor de evento por día, el primero, ya que cuando se registra un evento, este mismo se vuelve a registrar en repetidas ocasiones durante un corto período de tiempo, haciendo referencia al mismo fallo, como se puede observar en la ilustración 8. Este hecho introduciría un sesgo importante en el análisis. Para evitar este suceso, se ha decidido considerar que un evento de cualquiera de los incluidos en el análisis sólo puede surgir una vez por día. De este modo, cuando se registra el mismo evento en un período muy corto de tiempo sólo se tomará en cuenta el primer registro de ellos.

TIMESTAMP (UTC)	LOCAL TIME	FLEET	LOCOMOTIVE	EVENT CODE	CATEGORY	GRADE	EVENT DESCRIPTION
16/05/2023 20:05:03	16/05/23 22:05:03 EUROPE/BERLIN	[illegible]	[illegible]	1846	B		[illegible]
16/05/2023 20:07:06	16/05/23 22:07:06 EUROPE/BERLIN	[illegible]	[illegible]	836	A		[illegible]
16/05/2023 20:07:06	16/05/23 22:07:06 EUROPE/BERLIN	[illegible]	[illegible]	832	B		[illegible]
16/05/2023 20:08:13	16/05/23 22:08:13 EUROPE/BERLIN	[illegible]	[illegible]	836	A		[illegible]
16/05/2023 20:08:13	16/05/23 22:08:13 EUROPE/BERLIN	[illegible]	[illegible]	832	B		[illegible]
16/05/2023 20:08:52	16/05/23 22:08:52 EUROPE/BERLIN	[illegible]	[illegible]	836	A		[illegible]
16/05/2023 20:08:52	16/05/23 22:08:52 EUROPE/BERLIN	[illegible]	[illegible]	832	B		[illegible]
16/05/2023 20:10:36	16/05/23 22:10:36 EUROPE/BERLIN	[illegible]	[illegible]	1846	B		[illegible]
16/05/2023 20:12:11	16/05/23 22:12:11 EUROPE/BERLIN	[illegible]	[illegible]	1846	B		[illegible]

Ilustración 8: Timestamps de los eventos registrados

El objetivo del análisis será, por tanto, identificar cada cuánto suceden estos eventos con una separación más amplia de forma que se asegure que el registro del evento sea propiciado por un fallo diferente.

Una vez realizada esta aclaración se procede con el análisis exploratorio. En primer lugar, se muestra la distribución de las diferencias de tiempo para las locomotoras 1 de la flota A.

Como se observa en la ilustración 9, la mayoría de las diferencias se concentran en torno a un máximo de 25 días, con algunos datos extremos. Debido a que estos períodos tan largos podrían indicar que la locomotora estuvo inactiva durante ese tiempo, ya sea debido a otros errores u otras causas, se decide eliminar dichos datos. Para lograr esto, se aplica una condición que permite conservar únicamente aquellos valores de diferencia que son menores que la media más tres veces la desviación estándar.

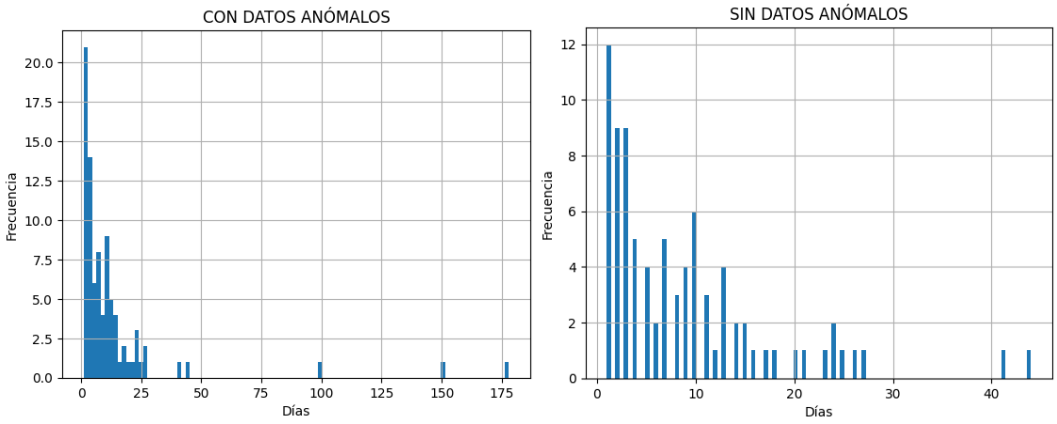

Ilustración 9: Histograma de las diferencias de tiempos entre los eventos registrados antes y después de eliminar datos anómalos para la locomotora 1

Una vez aplicado este filtro, los datos se vuelven más óptimos. Sin embargo, como se puede comprobar en la gráfica, todavía hay algunos valores para periodos de tiempo muy cortos que sobresalen, pero no se puede asegurar que estos registros de eventos se deban al mismo fallo. Por lo tanto, no pueden ser tratados y serán incluidos en el análisis.

Por otro lado, se observa la locomotora 2 de la flota B. De nuevo, en la ilustración 10, se nota la presencia de varios datos extremos, por lo que se repite el mismo proceso tratando de eliminar aquellos valores que podrían introducir sesgo en el estudio.

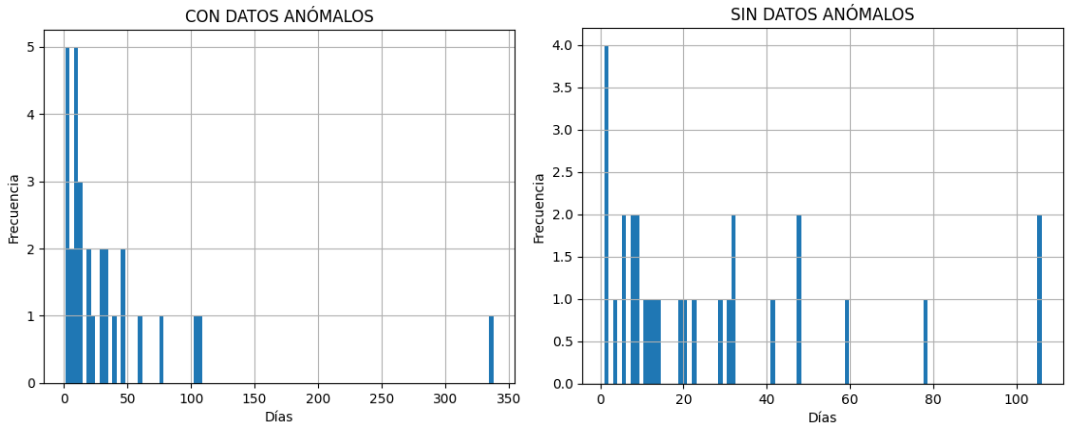

Ilustración 10: Histograma de las diferencias de tiempos entre los eventos registrados antes y después de eliminar datos anómalos para la locomotora 2

Una vez aplicado el filtro, se observa una distribución de los datos más uniforme. Sin embargo, nuevamente destacan los valores cercanos a cero, y se presenta la misma situación mencionada anteriormente, no se puede asegurar que estos registros correspondan al mismo fallo por lo que se mantienen.

Es fundamental tener en cuenta que se observan notables diferencias entre las cantidades de eventos registrados en las distintas flotas. Además, existe una variabilidad significativa tanto en la cantidad total de eventos registrados como en la naturaleza de los mismos. Esta variabilidad ha de ser tomada en cuenta a la hora de realizar el estudio sobre la fiabilidad, ya que cada evento puede presentar características únicas y contribuir de manera diferente al análisis.

Estas diferencias destacadas sugieren la presencia de patrones de comportamiento distintos entre las flotas y los eventos específicos. Además, para evitar la inclusión de datos extremos que puedan condicionar al estudio, este proceso se aplicará también para el análisis de fiabilidad presente en el cuadro de mandos.

3.6. Filtros y segmentadores

Para la implementación del cuadro de mandos, uno de los aspectos más importantes es determinar qué filtros y segmentadores estarán disponibles para que el usuario pueda elegir qué eventos desea visualizar. En este caso, se han incluido varios elementos que brindan flexibilidad y opciones de personalización al usuario:

Un selector de flota (ilustración 11): Este filtro permite al usuario seleccionar una flota específica. Una vez seleccionada la flota, los siguientes filtros se ajustarán automáticamente para mostrar únicamente las locomotoras pertenecientes a esa flota, y los eventos que se han dado en ella. Esto garantiza que solo se muestren opciones válidas y relevantes para la flota seleccionada. Cabe destacar que tiene la opción de selección múltiple, es decir, el usuario puede seleccionar varias flotas en las que esté interesado.

Ilustración 11: Segmentador de flotas

Un selector de locomotoras (ilustración 12): Este filtro permite al usuario elegir una o varias locomotoras. El conjunto de locomotoras disponibles en el selector se basa en la flota seleccionada previamente. De esta manera, se restringe la selección de locomotoras únicamente a las que pertenecen a la flota elegida, evitando opciones incorrectas o irrelevantes.

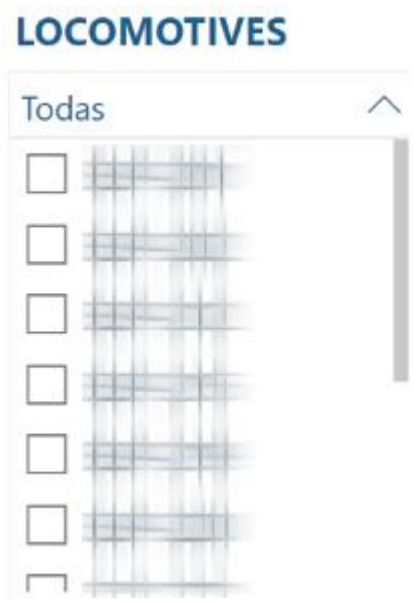

Ilustración 12: Segmentador de locomotoras

Un selector de eventos (ilustración 13): Este filtro permite al usuario elegir uno o varios eventos específicos. Los eventos disponibles en el selector se basan en la flota y las locomotoras seleccionadas. Sólo se mostrarán los eventos que se hayan registrado en la flota o las locomotoras seleccionadas, lo que garantiza que solo se puedan elegir opciones relevantes y aplicables.

Ilustración 13: Segmentador de códigos de evento

Un segmentador de fecha (ilustración 14): Este segmentador permite al usuario seleccionar el intervalo de tiempo en el que desea visualizar los eventos. Puede ajustar la fecha de inicio y fin para definir el período exacto que le interesa analizar. Esto permite una exploración temporal específica y un enfoque en los eventos ocurridos dentro de ese intervalo.

Ilustración 14: Segmentador de fecha

Todos estos filtros y segmentadores están condicionados a la disponibilidad de datos en el período seleccionado. Sólo se mostrarán opciones válidas que correspondan a las flotas, locomotoras y eventos registrados durante el intervalo de tiempo seleccionado. Esta configuración garantiza una experiencia interactiva y personalizada para el usuario, al permitir seleccionar los eventos de su interés y explorar datos relevantes de manera eficiente.

3.7. Diseño de los gráficos

Una vez que se han definido y configurado los aspectos mencionados anteriormente, se procede a la implementación de los gráficos en el cuadro de mandos. El diseño de los gráficos está pensado para cumplir con los requisitos establecidos en el proyecto, buscando proporcionar una representación visual efectiva de los datos seleccionados por el usuario y ofrecer una comprensión clara de los eventos ferroviarios.

Se utilizan diferentes tipos de gráficos con diferentes funcionalidades según la naturaleza de la información que se desea visualizar. En este caso, se diseñan nueve diferentes que se mostrarán en el cuadro de mandos:

En primer lugar, se decide implementar una tabla (ilustración 15) en la que el usuario podrá encontrar información precisa sobre cada uno de los eventos de la locomotora o flota seleccionadas. Esta tabla presenta varios campos para que el usuario pueda realizar un exhaustivo análisis de forma rápida y sencilla, incluyendo información como:

- *Timestamp* del evento
- *Timestamp* en hora local
- Nombre de la flota
- Nombre de la locomotora
- Código de evento
- Categoría
- Prioridad
- Grado (color)
- Descripción del evento
- Sistema
- Booleana que indica si el evento sigue activo
- Duración del evento
- Versión de software

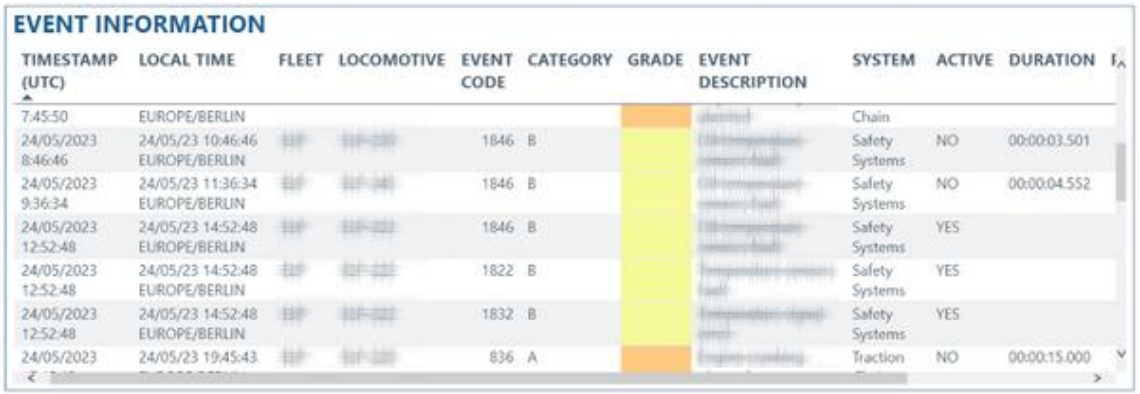

Ilustración 15: Tabla descripción de registros de eventos

En esta tabla, el usuario podrá seleccionar la fila que desee y los demás gráficos serán filtrados en base a esta selección, mostrando así las características asociadas a dicha fila. Esta funcionalidad permite al usuario explorar y analizar los datos de manera interactiva, centrándose en el evento de su interés, y obteniendo información más detallada y específica sobre él. Además, para esta visualización se ha creado una medida, representada en la columna *GRADE*, que asigna a cada evento su color definido previamente en base a su categoría y prioridad. Los colores

utilizados son rojo, naranja, amarillo o verde, lo cual proporciona una gran ayuda visual para el usuario.

Se elige una tabla como visualización para cumplir el segundo requisito (“recopilar información detallada sobre los eventos que tienen lugar en las locomotoras”) ya que permite una representación exacta y condensada de la información. Otra opción planteada es la inclusión de gráficos más simples para cada una de las variables que se muestran, pero se descarta ya que se considera que no logran cumplir el requisito de manera tan satisfactoria como la opción propuesta.

El siguiente gráfico busca proporcionar al usuario información sobre la geolocalización de los eventos. Para ello, se muestran en un mapa de coordenadas (ilustración 16) sus latitudes y longitudes, así como información como es el nombre de la locomotora, la flota, el tipo de evento y el instante de tiempo en formato local e internacional (UTC). De este modo, si el usuario coloca el cursor sobre un punto, podrá ver todos los detalles asociados a él.

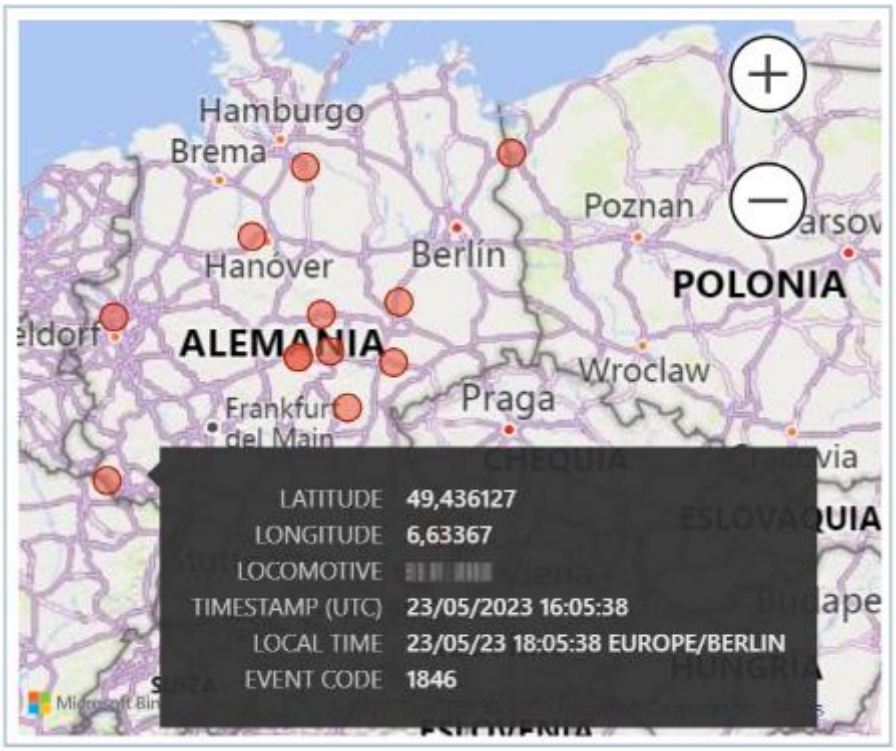

Ilustración 16: Mapa interactivo con geolocalización e información de los eventos

Este mapa es interactivo y proporciona al usuario la capacidad de acercar y alejar la zona según sus preferencias. Esta funcionalidad permite explorar con mayor detalle áreas específicas y obtener una vista más amplia cuando sea necesario. Al utilizar los controles de *zoom* disponibles en el mapa, el usuario puede ajustar la escala de visualización para enfocarse en regiones específicas o ampliar el panorama para tener una visión más general. Esta flexibilidad de *zoom* brinda una experiencia personalizada al usuario, permitiéndole adaptar el nivel de detalle a sus necesidades y facilitando la exploración y comprensión de la información geográfica presentada en el mapa.

En este caso, se elige un mapa para la visualización de las coordenadas GPS de los eventos ya que es la representación gráfica más idónea para representar datos de geolocalización y cumplir así con el tercer requisito (“explorar de forma interactiva la posición geográfica de los eventos históricos registrados”).

Por otro lado, se implementan tres gráficos de anillos, que muestran la distribución de diferentes elementos dentro de los datos.

El primer y segundo gráfico enseñan la proporción de cada uno de los códigos de evento registrados en la locomotora o flota seleccionada por el usuario, y la distribución de categorías de ellos respectivamente (ilustración 17). Permiten visualizar la frecuencia de cada código y comprender la distribución de los eventos registrados y sus categorías en relación con la selección específica realizada. Esto proporciona una visión general de los tipos de eventos más comunes y las categorías más predominantes en el contexto seleccionado.

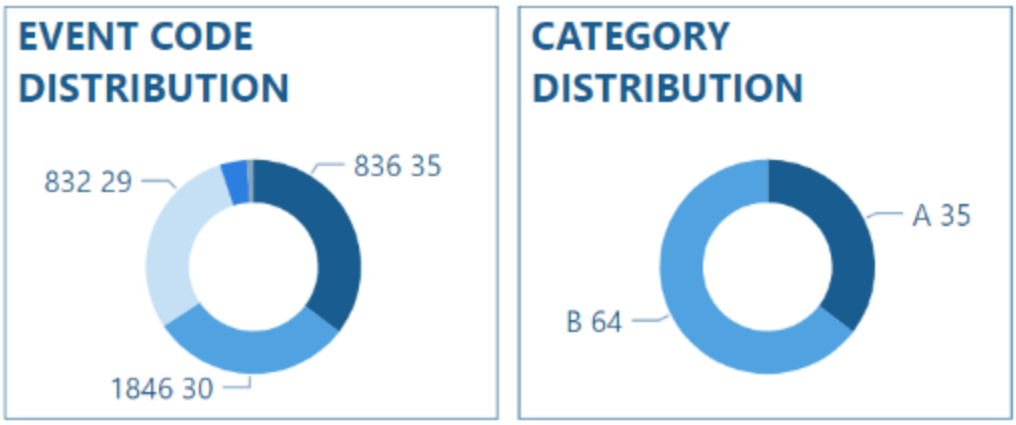

Ilustración 17: Gráficos de anillo para la distribución de códigos de evento y categorías

El tercer gráfico muestra la distribución del número total de eventos registrados entre las locomotoras seleccionadas (ilustración 18). Este gráfico permite distinguir aquellas locomotoras que tienen mayor cantidad de eventos respecto al resto.

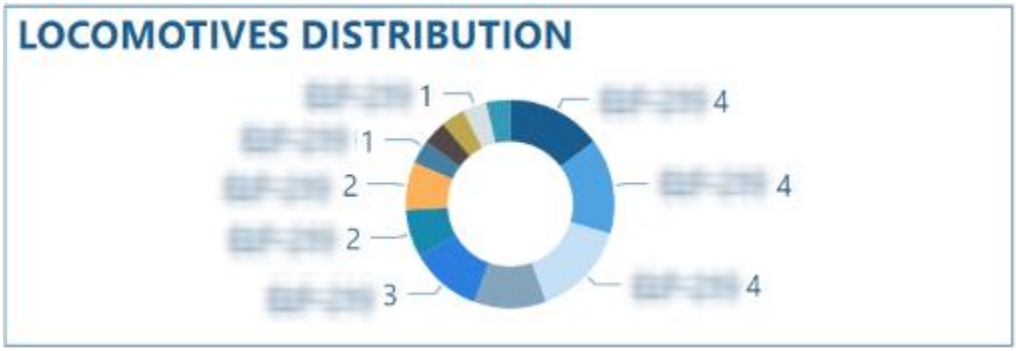

Ilustración 18: Gráfico de anillo para la distribución de eventos registrados por locomotora

En los tres gráficos, en caso de seleccionar cualquiera de los sectores, se aplicarán filtros cruzados entre ellos y se resaltarán las filas de la tabla que correspondan a dicha selección, brindando al usuario una forma conveniente de filtrar y visualizar los eventos específicos que deseen explorar.

Por ejemplo, si el usuario selecciona un sector en el primer gráfico que representa un código de evento específico, los otros dos gráficos y la tabla se actualizarán automáticamente para mostrar únicamente los datos relacionados con ese código de evento. De esta manera, el usuario

puede examinar detalladamente los eventos asociados a esa categoría particular y comprender mejor su distribución en las otras dimensiones presentadas.

Esta funcionalidad de filtrado cruzado entre gráficos y la tabla facilita al usuario la exploración de datos específicos y permite centrar la atención en los eventos de su interés de manera cómoda y eficiente.

En este caso, para cumplir el cuarto requisito ("presentar de forma clara tanto las distribuciones de los eventos en las locomotoras, como las distribuciones de los códigos de evento y las categorías asociadas a estos"), se eligen los gráficos de anillos en vez de otro tipo de visualización como los gráficos de barras ya que ofrecen una visualización de manera más concisa y fácilmente comprensible.

Se implementan además tres indicadores que muestran información relevante para el usuario. Estos indicadores ofrecen datos clave sobre los eventos registrados en el período seleccionado:

La primera tarjeta muestra la cantidad total de eventos registrados en ese período, y la segunda la cantidad de ellos que se encuentran activos (ilustración 19). Esta información proporciona una visión general de la magnitud de los eventos y permite al usuario tener una idea de la escala de los incidentes ocurridos.

Ilustración 19: Indicador de eventos registrados y eventos activos

La tercera tarjeta muestra la duración máxima entre los eventos inactivos (ilustración 20). Esto proporciona una medida de referencia sobre la duración más larga entre los eventos, lo cual puede ser útil para evaluar la estabilidad del sistema o la efectividad de las acciones de mantenimiento y reparación.

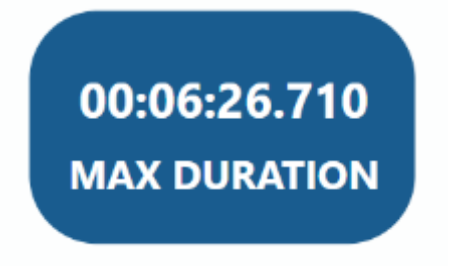

Ilustración 20: Indicador de la duración máxima de los eventos registrados

En este caso, se usan estos indicadores para cumplir el quinto requisito ("mostrar la cantidad de eventos registrados, a la vez que la cantidad de los que estén activos, así como la duración máxima registrada de estos"), ya que brindan al usuario información relevante de manera cómoda y precisa, permitiéndoles obtener una instantánea rápida de los aspectos clave de los eventos

registrados. Esto ayuda a tomar decisiones informadas y a comprender mejor la situación general en relación con los eventos ferroviarios.

Finalmente, se implementa un gráfico de líneas que representa la evolución de la probabilidad de que un evento vuelva a ocurrir a lo largo del tiempo. Para calcular esta probabilidad, se ha utilizado el estimador de Kaplan-Meier, un método comúnmente utilizado en los análisis de supervivencia y fiabilidad.

En el gráfico, se marca el umbral de 0.5, lo que indica la probabilidad de que el evento se repita. Al detectar el punto en el tiempo en el que la probabilidad supera este umbral, se puede estimar los días esperados hasta que el evento se produzca nuevamente, proporcionando una perspectiva valiosa sobre la recurrencia de los eventos y permitiendo a los usuarios anticipar y planificar en consecuencia.

El gráfico de líneas proporciona una representación visual clara de cómo varía la probabilidad de recurrencia a medida que pasa el tiempo. Esto ayuda a identificar patrones y tendencias en la ocurrencia de eventos y proporciona una base sólida para la toma de decisiones en términos de mantenimiento preventivo, programación de tareas y asignación de recursos.

Este gráfico se crea utilizando un objeto visual de Python, herramienta de Power BI. Esta funcionalidad permite realizar visualizaciones utilizando un editor de scripts de Python directamente dentro del cuadro de mandos. De este modo, se puede escribir el código necesario en el mismo entorno para generar la imagen de la predicción.

El proceso de generación de la predicción se basa en la selección de la locomotora y el código de evento específico. Todos los datos necesarios para realizar la predicción se recogen en tiempo real, entrenándose el modelo cada vez que el usuario elige los valores, lo que garantiza que la predicción sea actualizada y refleje las últimas condiciones y factores influyentes relevantes.

El código de generación del gráfico de líneas incorpora tres posibilidades que pueden surgir. En primer lugar, si no se ha seleccionado una locomotora y un código de evento específicos, se mostrará un mensaje indicando que es necesario seleccionar ambos para poder realizar el análisis (ilustración 21). Esta notificación asegura que los criterios de selección sean claros antes de proceder con la generación de la predicción. Es necesaria esta selección ya que el análisis solo se realizará para cada locomotora y código de evento de forma individual.

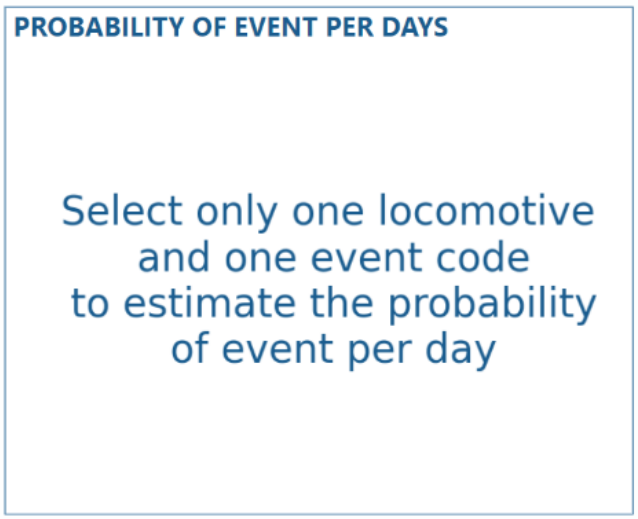

Ilustración 21: Mensaje obtenido del análisis de fiabilidad al no cumplir los criterios de selección

En segundo lugar, si se ha seleccionado una locomotora y un código de evento, pero el código de evento para esa locomotora no tiene al menos dos registros, el modelo no podrá ser ajustado. En este caso, se mostrará un mensaje indicando que no es posible realizar el análisis debido a la falta de datos suficientes (ilustración 22).

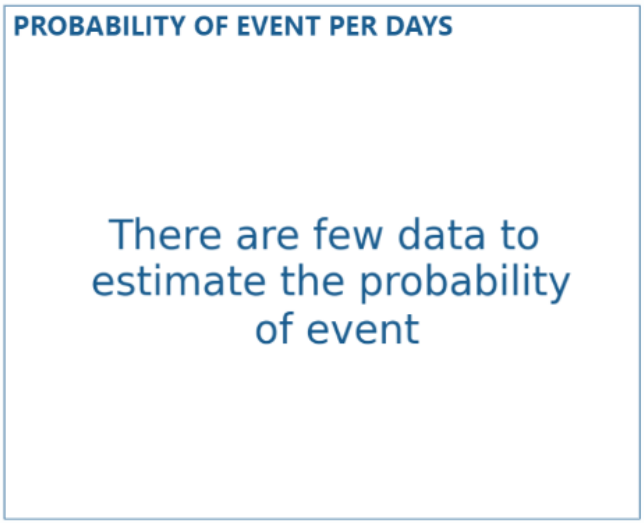

Ilustración 22: Mensaje obtenido del análisis de fiabilidad al contar con pocos datos para ser realizado

Finalmente, si se ha seleccionado una locomotora y un código de evento con valores adecuados, se generará y mostrará la imagen del modelo junto a la estimación del tiempo hasta el siguiente registro (ilustración 23). Esta imagen proporciona una representación visual de la evolución de la probabilidad de recurrencia del evento a lo largo del tiempo basándose en el estimador de Kaplan-Meier. Además, se ha optado por añadir un indicador cuando la probabilidad de supervivencia estimada alcance el 50% a modo informativo como posible fecha de ocurrencia del evento para llevar a cabo un mantenimiento preventivo.

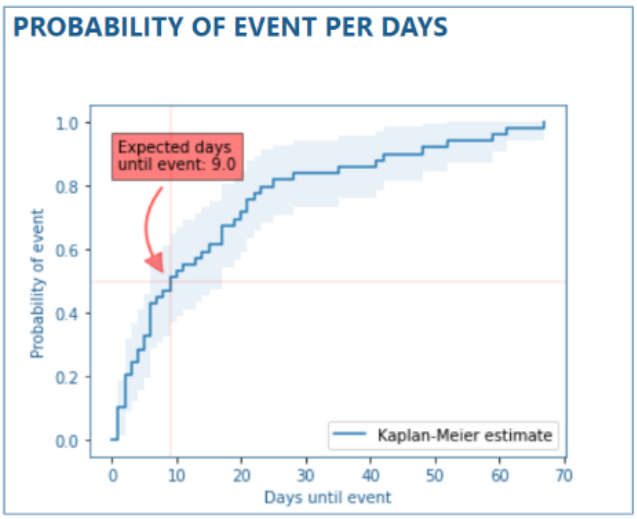

Ilustración 23: Gráfica del análisis de fiabilidad para una locomotora y un código de evento

La interpretación del gráfico de Kaplan Meier es más compleja ya que se deben tener en cuenta diversos factores. Se detallará más adelante en la sección 4.1.2.

El código incluye condiciones para manejar distintas situaciones: la falta de selección de locomotora y código de evento, la falta de datos suficientes para realizar el análisis y, finalmente, la generación de la imagen del modelo si se cumplen los criterios necesarios. Estas condiciones y mensajes proporcionan una experiencia interactiva al usuario y garantizan que los análisis se realicen de manera adecuada y útil.

Es importante destacar que en un cuadro de mandos se busca evitar una gran variedad de gráficos y la inclusión de elementos complejos, ya que esto podría complicar el proceso de análisis en lugar de agilizarlo.

En este caso, se ha optado por diseñar una combinación de gráficos sencillos y complejos para mostrar toda la información necesaria de forma rápida y sencilla dentro del cuadro de mandos, buscando cumplir con los requisitos del proyecto. Esta combinación permite alcanzar un equilibrio entre la claridad de la presentación y la capacidad de transmitir información detallada, ayudando a los usuarios a comprender de manera eficiente el rendimiento y la fiabilidad de las locomotoras. Por otro lado, se ha realizado una exhaustiva descripción de los gráficos para facilitar la interpretación de estos por los usuarios.

4. Resultados

En esta sección de resultados, se presentan tanto los hallazgos obtenidos a través del análisis de fiabilidad de las locomotoras como el cuadro de mandos final en el contexto de la investigación realizada.

4.1. Análisis de fiabilidad

Como se ha descrito en el apartado 2.4., un análisis de fiabilidad es una técnica utilizada para evaluar la capacidad de un sistema para funcionar correctamente durante un período de tiempo determinado.

En el contexto del cuadro de mandos, este análisis se aplica a las locomotoras diésel para evaluar su rendimiento y determinar la probabilidad de que se registre un evento.

En este caso, se han realizado dos modelos para el análisis de fiabilidad: el MCF y el estimador de Kaplan-Meier. La estructura de los datos utilizados para los modelos es la descrita en el apartado 3.4, en una versión congelada hasta el día 25 de junio de 2023.

Es importante destacar que el inicio del estudio de cada locomotora se calcula a partir del momento en el que se produce el primer evento en la locomotora analizada. Esto es debido a que es el único modo fiable de asegurar que el vehículo está en funcionamiento, ya que no se dispone de la fecha exacta en la que comenzó a circular.

Para llevar a cabo los análisis, se han seleccionado como objeto de estudio la flota A, con la locomotora 1, y la flota B, con la locomotora 2, ambas para el evento 832. Se han seleccionado estas dos locomotoras junto a este evento en concreto ya que cuentan con una abundante cantidad de eventos , lo que permitirá realizar un estudio más exhaustivo y preciso.

Cabe destacar que el análisis para otras configuraciones de locomotora-evento puede ser realizado desde el cuadro de mandos del mismo modo que el presentado en este documento, siempre y cuando se dispongan de datos suficientes del evento deseado para esa locomotora.

Para el análisis se va a hacer uso del evento 832. Es importante mencionar que este evento puede tener un impacto significativo en el rendimiento y la seguridad de las locomotoras. Por lo tanto, es crucial realizar un análisis exhaustivo y tomar las medidas necesarias para prevenir y corregir estos fallos, garantizando así el correcto funcionamiento de las flotas.

En la elaboración de este tipo de análisis en locomotoras, idealmente se utiliza la distancia recorrida en kilómetros desde el inicio del estudio en lugar del tiempo en horas, ya que los fallos solo ocurren cuando el vehículo está en funcionamiento. Sin embargo, en este caso específico, se ha utilizado el tiempo en horas debido a la falta de datos confiables sobre los kilómetros recorridos por cada locomotora en el momento de cada fallo. Esto se debe a que la medición de la distancia recorrida se realiza en un sistema diferente que no está disponible para este análisis.

A lo largo de esta sección, se presentarán los resultados obtenidos mediante los modelos de análisis de fiabilidad. Se destacarán las probabilidades esperadas de fallos asociadas a las locomotoras y los eventos seleccionados para su estudio. Estos resultados proporcionarán una

visión más clara de la confiabilidad de la flota y permitirán tomar medidas específicas para abordar los eventos identificados.

4.1.1. Mean cumulative function (MCF)

El MCF es un modelo no paramétrico que permite estudiar el número de eventos esperados para un individuo a lo largo del tiempo sin supuestos previos sobre la distribución de los datos.

En este caso, en primer lugar, se estudiará la influencia del evento 832 sobre las locomotoras de la flota A, compuesta por 34 vehículos. A continuación, en la ilustración 24, se muestra el modelo MCF ajustado.

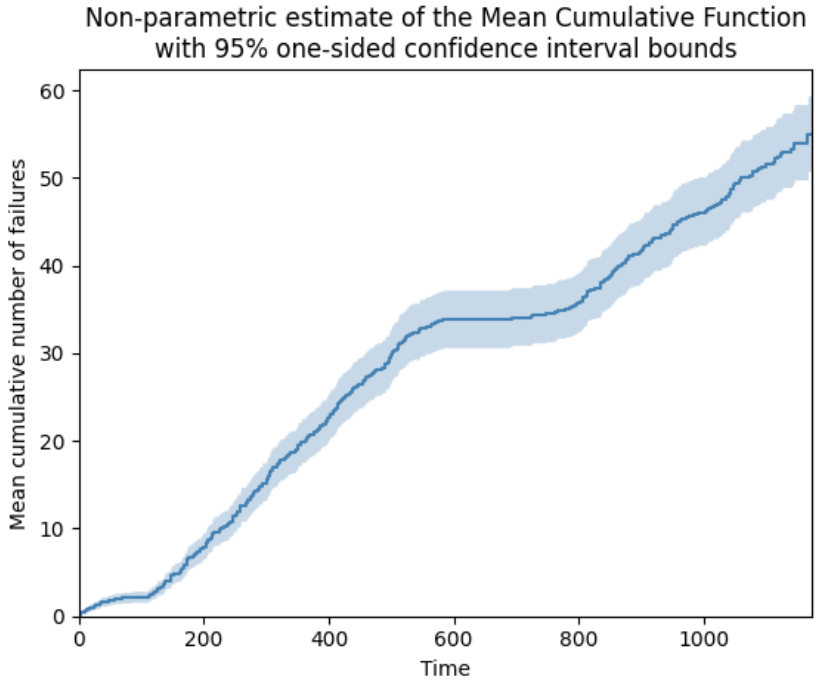

Ilustración 24: Gráfico resultado de MCF para la flota A y el evento 832

El gráfico muestra una estimación del número medio de eventos en función del tiempo para las locomotoras de esa flota, junto con un intervalo de confianza del 95%. En este caso, el modelo muestra una línea recta, seguida de un período de estancamiento, y continuada por otra línea recta algo menos pronunciada que la primera, esto indica que el crecimiento de la media acumulada de evento se ha ralentizado ligeramente. Es decir, podría ser un indicio de una mejora en la fiabilidad de las locomotoras.

Por otro lado, el tramo de estancamiento representa un intervalo de tiempo donde apenas se registraron errores, lo que podría ser debido a que durante ese periodo las locomotoras no estuvieron en funcionamiento. Sería necesario realizar un estudio más en profundidad sobre la flota para encontrar una explicación a este fenómeno.

En la tabla 4 encontramos información sobre los eventos medios acumulados esperados para una locomotora de esa flota en diferentes momentos del tiempo.

DÍAS	MEDIA DE EVENTOS ACUMULADOS
1	0.05
100	2.23
200	8.04
500	29.71
1000	46.21

Tabla 4: Tabla de media de eventos acumulados esperados para la flota A y el evento 832

Resalta un gran aumento en términos relativos de los errores esperados entre los 500 primeros días, llegando a casi cuadruplicar el número de errores esperados. Sin embargo, este crecimiento disminuye rápidamente para los 1000 días donde los errores esperados apenas son el doble de los esperados a los 500 debido, principalmente, al estancamiento que se observaba en la gráfica.

En segundo lugar, se muestra un nuevo ejemplo de estudio del evento 832, en este caso en la flota B, compuesta por 10 locomotoras. A continuación, en la ilustración 25, se muestra el modelo MCF ajustado.

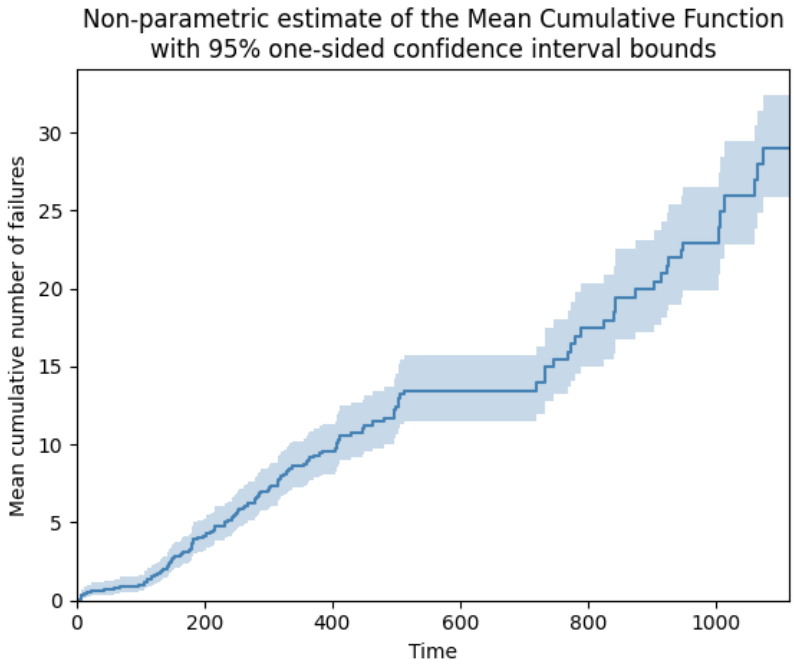

Ilustración 25: Gráfico resultado de MCF para la flota B y el evento 832

En este caso, en el gráfico se observa de nuevo un patrón muy similar al obtenido en el análisis anterior. En este caso, compuesto por una línea recta, seguida de un período de estancamiento y continuado por otro período recto, con pendiente más pronunciada que el anterior. Esto podría ser un indicio de un empeoramiento en la fiabilidad de las locomotoras con el paso del tiempo.

Cabe destacar que los intervalos de confianza del 95% se ensanchan con el paso del tiempo, lo que puede ser debido a la falta de registros de eventos en locomotoras que lleven tanto tiempo en funcionamiento.

En la tabla 5 encontramos de nuevo información sobre los eventos acumulados medios esperados en diferentes momentos del tiempo para una locomotora de esta flota.

DÍAS	MEDIA DE EVENTOS ACUMULADOS
1	0.03
100	1
200	4.15
500	12.44
750	15.53
1000	23.91

Tabla 5: Tabla de media de eventos acumulados esperados para la flota B y el evento 832

En esta ocasión se observa un comportamiento similar al análisis anterior como ya se mencionó. Sin embargo, destaca el crecimiento entre 500 y 1000, donde se observan dos tramos, en el primero apenas se esperan nuevos errores y en el segundo con más del 150%.

En resumen, se ha aplicado el modelo MCF a flotas para estudiar el patrón de errores observable en cada conjunto de vehículos a lo largo del tiempo, destacando un estancamiento en ambos gráficos, por lo que sería interesante estudiar las posibles causas del mismo. Esta visión general puede ayudar a la empresa en diversas tareas, como puede ser por ejemplo, la creación de previsiones sobre reparaciones de fallos en trenes.

Sin embargo, este análisis no será incluido en el cuadro de mandos, debido a que se ha preferido incluir un análisis Kaplan-Meier que estudiará el comportamiento de cada una de las locomotoras para cada evento de manera independiente, ya que la intención del cuadro de mandos es presentar la información de forma individualizada para el vehículo seleccionado por el usuario.

4.1.2. Estimador Kaplan-Meier

El estimador de Kaplan-Meier es otro modelo no paramétrico. En este caso, este modelo parte de varias suposiciones que han de cumplirse, como se mencionó previamente en la sección 2.3.2. Por lo tanto, es preciso mencionar algunas consideraciones previas:

En primer lugar, se consideran todos los eventos de una máquina como supuestos independientes, por lo que no hay factores que puedan influir de forma sistemática y todos los sucesos tienen la misma probabilidad de ocurrir. Por otro lado, la precisión estadística podría verse condicionada según el número de eventos registrados en la locomotora. No obstante, se ha tomado la decisión de incluir los intervalos de confianza al 95% en el análisis con el fin de

apaciguar los posibles sesgos que podrían aparecer en locomotoras con pocos sucesos registrados. Mediante la amplitud de estos intervalos se puede visualizar la incertidumbre de la predicción. Finalmente, cabe destacar que para el estudio se cuenta únicamente con datos no censurados.

El análisis de Kaplan-Meier estima la probabilidad de supervivencia de un grupo de individuos para diferentes instantes de tiempo. No obstante, se ha decidido por motivos de comprensión visual, a la hora de ser implementado en el cuadro de mandos, representar la probabilidad de que se origine el evento en lugar de la probabilidad de supervivencia.

La probabilidad de que se origine el evento en un instante *t* es calculada como uno menos la probabilidad de supervivencia estimada por el modelo para ese instante puesto que ambos sucesos son complementarios.

El análisis será realizado en primer lugar sobre la locomotora 1, perteneciente a la flota A, para el estudio de la incidencia del evento 832 sobre la misma. La elección de esta locomotora viene motivada por la alta frecuencia de aparición de eventos de este tipo en ella, llegando a ser la más susceptible de entre toda la flota.

Cabe destacar que el Kaplan-Meier no es un modelo de predicción, sino que estima la función de supervivencia del grupo de individuos a lo largo de un periodo de tiempo. No obstante, a modo de ampliación se ha decidido incluir en el gráfico una alerta en la que se indica el periodo de tiempo para el cuál la estimación de la probabilidad de supervivencia alcanza un 50%. Esta alerta pretende informar de forma visual de una posible fecha a partir de la cuál la probabilidad de registrar un evento es lo suficientemente grande como para implementar las medidas preventivas necesarias. Con esto se pretende dar apoyo a la toma de decisiones relacionadas con la planificación de labores de mantenimiento.

A continuación, en la ilustración 26, se muestra la gráfica de la probabilidad esperada de evento por días, junto a los correspondientes intervalos de 95% de confianza.

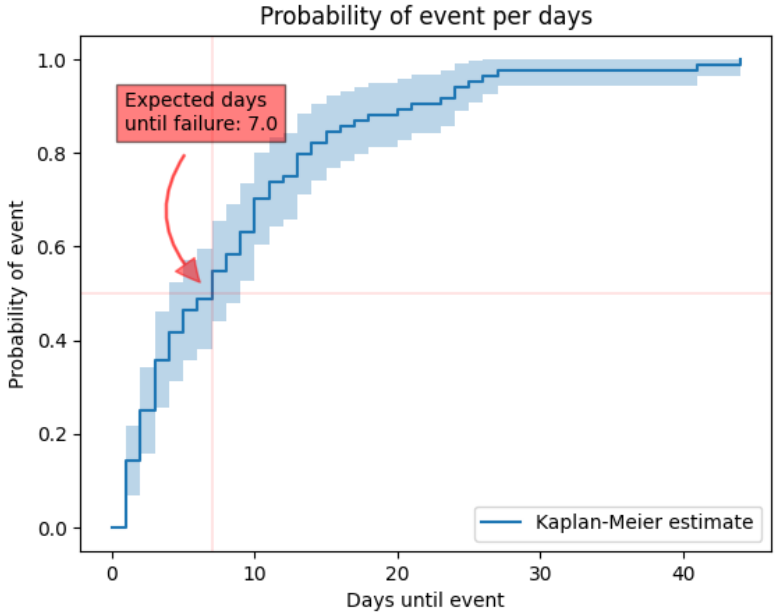

Ilustración 26: Resultado del análisis Kaplan-Meier para la locomotora 1 y el evento 832

Este gráfico muestra la evolución de la probabilidad estimada de que se origine el evento 832 en la locomotora 1 a lo largo de 45 días. Se observa cómo esta probabilidad aumenta de forma pronunciada durante los 20 primeros días, y se va ralentizando con el tiempo tendiendo a casi

estabilizarse sobre un 95% de probabilidad. Finalmente, se observa cómo el modelo ajustado pronostica que la probabilidad de registrar un evento superará el 50% a los siete días del último evento registrado.

Para complementar los resultados obtenidos del gráfico, se han registrado los días esperados hasta el evento para valores de probabilidad 0.25, 0.5, 0.75 y 0.95 en la tabla 6.

PROBABILIDAD ESTIMADA	DÍAS
0.25	2
0.5	7
0.75	12
0.95	25

Tabla 6: Tabla de probabilidad estimada por tiempo para la locomotora 1 y el evento 832

Los resultados muestran cómo la probabilidad estimada aumenta rápidamente, llegando a triplicarse desde el 25% en 10 días, para posteriormente ralentizarse hasta alcanzar 95% en los 13 días posteriores.

Por otro lado, se analiza el mismo evento en la locomotora 2, vehículo con la mayor frecuencia de registros de este evento en la flota B. De nuevo, en la ilustración 27, se muestra la gráfica de probabilidad de evento por días, junto a los correspondientes intervalos de confianza del 95%, con la probabilidad de evento marcada en 0.5.

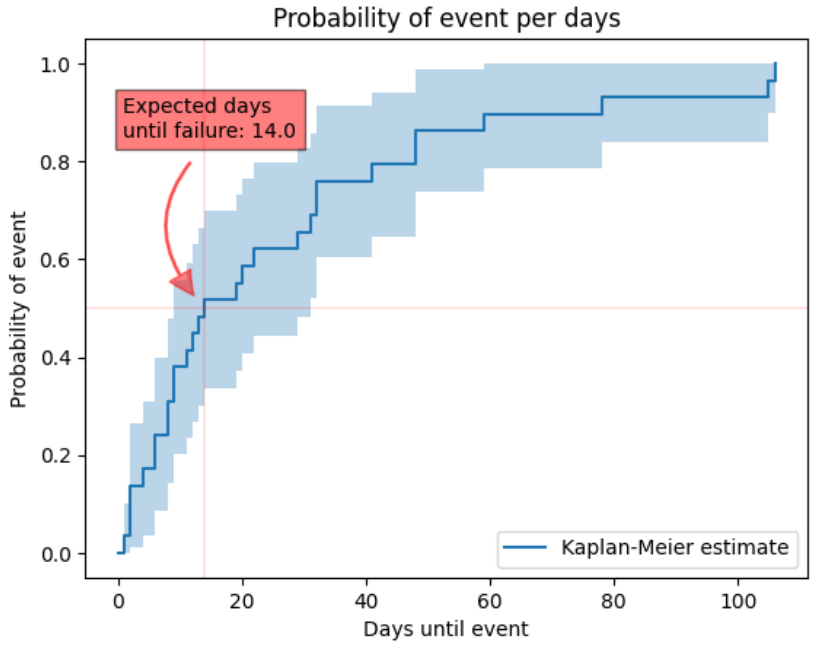

Ilustración 27: Resultado del análisis Kaplan-Meier para la locomotora 2 y el evento 832

El gráfico muestra la evolución de la probabilidad estimada para el evento 832 en la locomotora 2, a lo largo de 110 días. En esta ocasión el modelo comprende un periodo de tiempo

más extenso por la existencia de varios eventos que han tenido una mayor demora en originarse, lo que implica que hay mayor variabilidad en el tiempo esperado hasta el registro de un evento en esta locomotora. El modelo Kaplan-Meier tiene dificultades al lidiar con este problema, pudiendo observarse en la gran amplitud de los intervalos de confianza, indicando que el pronóstico sobre la probabilidad debe ser tomado con cautela.

Por otro lado, destaca un rápido incremento hasta los 15 días, seguido de una ralentización hasta estabilizarse acerca de los 60 días. Finalmente, se observa cómo el modelo ajustado pronostica que la probabilidad de registrar un evento superará el 50% a los 14 días del último evento registrado.

Se muestran además en la tabla 7 los días esperados hasta el evento para las probabilidades 0.25, 0.5, 0.75 y 0.95.

PROBABILIDAD DE FALLO	DÍAS ESPERADOS
0.25	6
0.5	14
0.75	32
0.95	105

Tabla 7: Tabla de probabilidad estimada por tiempo para la locomotora 2 y el evento 832

En este caso, los resultados muestran como la probabilidad estimada aumenta muy rápidamente, esperando 14 días con un 50% de probabilidad, pasando a ralentizarse en gran medida hasta llegar a esperar un 95% a los 105 días.

En conclusión, se ha aplicado el estimador Kaplan-Meier a locomotoras para estudiar el patrón de errores observable en cada conjunto de vehículos a lo largo del tiempo. En este caso, se busca hacer un análisis por locomotora para capturar la influencia de cada uno de los eventos en cada vehículo de una forma más individualizada.

Por este motivo, se ha decidido introducir este modelo en el cuadro de mandos. En este caso, el usuario seleccionará la locomotora y el código de evento de interés, y el sistema calculará y mostrará dinámicamente el modelo asociado a dicha combinación. Este análisis podrá servir de guía para una correcta interpretación de los resultados obtenidos, permitiendo así que cualquier empleado pueda utilizar la información de manera efectiva.

Este modelo se complementa con el modelo MCF, permitiendo un análisis detallado de los eventos en las locomotoras, facilitando la interpretación de los resultados y buscando proporcionar ayuda a la empresa en tareas como la creación de previsiones sobre reparaciones de fallos en trenes.

4.2. Cuadro de mandos

El cuadro de mandos desarrollado en este proyecto tiene como intención principal dar soporte a la toma de decisiones en la empresa. Ofrece una visualización rápida y efectiva de los eventos de interés en las locomotoras diésel, presentando tanto información general como distribuciones específicas que ayuden al usuario a contextualizar y comprender los eventos estudiados.

En el cuadro de mandos (ilustración 28), se han implementado todos los gráficos previamente descritos, dando lugar a una interfaz intuitiva y fácil de usar, que permite al usuario interactuar con los datos, explorar visualmente los eventos de interés y obtener información relevante de manera eficiente.

Ilustración 28: Cuadro de mandos definitivo

La decisión de esta distribución se toma después de realizar un pequeño análisis de usabilidad con los interesados. A partir de este análisis, se evaluaron diversos aspectos, como la factibilidad de uso, la claridad de la información presentada y la accesibilidad de las funciones. Se recopilaron comentarios y sugerencias de los usuarios para mejorar la experiencia de uso.

Con base en los resultados obtenidos, se realizaron ajustes y mejoras en el diseño del cuadro de mandos. Se prioriza la visualización clara y concisa de los datos relevantes, así como la inclusión de gráficos y métricas que facilitarán la toma de decisiones.

Además, se implementaron funcionalidades interactivas para permitir a los usuarios explorar y profundizar en los detalles de los datos. Esto incluye la capacidad de filtrar y desglosar la información según sus necesidades y de obtener informes personalizados.

El análisis de usabilidad sirvió para garantizar que el cuadro de mandos cumpliera con los requisitos y expectativas de los usuarios, mejorando su experiencia y permitiéndoles obtener información precisa y relevante de manera más eficiente.

Cabe destacar que el tiempo de carga del modelo dentro del cuadro de mandos es algo elevado, llegando incluso a dos minutos. Esta demora es comprensible, ya que en el momento se obtienen los datos de bases de datos, se preprocesan y se genera el modelo solicitado. De este período de tiempo, el modelo en sí mismo tarda menos de un segundo, por lo que se ha decidido mantener la configuración originalmente planteada de entrenar el modelo cada vez que el usuario elige los datos, consiguiendo así su actualización constante. Esto es posible ya que la cantidad de datos con los que se espera contar no alcanzará dimensiones tan grandes como para suponer un problema a tener en cuenta a corto plazo.

Se ha diseñado un pequeño estudio para comprobar el tiempo de carga del estimador Kaplan-Meier para diferentes cantidades posibles de eventos registrados como se puede observar en la tabla 8.

EVENTOS REGISTRADOS	TIEMPO (s)
10	0.5
50	0.5
100	0.5
150	0.5
200	0.6

Tabla 8: Tiempos de ejecución por tamaño muestral para el Kaplan-Meier

Cabe destacar que el despliegue del cuadro de mandos no ha sido un proceso trascendental en el proyecto ya que se ha realizado con el software de Microsoft Power BI. Una vez terminado de diseñar el cuadro de mandos, Power BI permite a cualquier usuario de la empresa tener acceso al informe de una manera rápida y sencilla.

El cuadro de mandos desarrollado cumple los requisitos descritos al inicio de este trabajo. En primer lugar, ofrece la posibilidad de observar la geolocalización de los eventos en el mapa. En segundo lugar, proporciona información detallada sobre los eventos en la tabla. Además, muestra distribuciones sobre eventos y los registros de estos para las distintas locomotoras, así como los eventos totales registrados, los activos y la duración máxima de los inactivos. Finalmente, incluye una estimación del rendimiento de la locomotora y evento deseados.

5. Conclusiones

A lo largo de este proyecto se ha descrito el proceso de diseño de un cuadro de mandos para el análisis de nueve tipos de eventos en más de 150 locomotoras diésel, repartidas en 17 flotas.

Tras la descripción del problema y la estructura de la base de datos, se han presentado las consultas necesarias para cargar los datos en Power BI, herramienta utilizada para la generación del cuadro de mandos.

Posteriormente, se realizó un análisis exploratorio donde se descubrió que los eventos 2882 y 1832 se dan con una frecuencia muy superior al resto, con la flota A entre las más propensas a registrar eventos. Además, se encontró una gran variabilidad en la cantidad de eventos registrados entre las diferentes locomotoras de las flotas A y B. Una vez realizado el análisis exploratorio de los datos, se procedió con el diseño de los gráficos y su implementación en el cuadro de mandos para cumplir con los requisitos propuestos.

Por otro lado, otro de los objetivos de este trabajo era estudiar el evento 832 en las flotas A y B, ya que son las más propensas a registrarlo, mediante un exhaustivo análisis de fiabilidad en el que se utilizaron dos modelos. Por un lado, se aplicó el modelo MCF y se descubrió que la flota A presenta un indicio de mejora en la fiabilidad, y la flota B presenta un indicio de empeoramiento. Por otro lado, se aplicó el modelo Kaplan-Meier a las locomotoras 1 (A) y 2 (B) y se observó que ambas presentan un comportamiento similar, con un crecimiento muy rápido de la probabilidad de fallo durante los primeros días y un posterior ralentizamiento de este crecimiento.

Se consideró que una locomotora tendría una probabilidad significativa de registrar un evento al alcanzar esta el 50% para estimar el tiempo esperado hasta un nuevo registro del evento de código 832 a fin de tomar decisiones preventivas. En el caso de la primera locomotora estudiada (1 de la flota A) se estimó que este tiempo era de siete días y en el caso de la segunda (2 de la flota B) de 14 días.

Finalmente se presentó el cuadro de mandos ya implementado en el que se incluían todos los gráficos y análisis previamente descritos. La distribución de estos fue decidida en base a un pequeño estudio de usabilidad realizado entre diferentes usuarios de la empresa, solicitando opiniones y propuestas de mejora. Se concluye, por tanto, que el cuadro de mandos presentado logra cumplir satisfactoriamente con todos los requisitos descritos al inicio del proyecto.

5.1. Limitaciones

Durante el desarrollo del proyecto, se han enfrentado algunas limitaciones que han impactado en la calidad y la precisión de los análisis realizados. Una de las principales limitaciones fue la falta de información precisa sobre las posiciones GPS donde se producían los eventos en las locomotoras. Por ello, fue necesario realizar estimaciones, lo que introdujo cierta incertidumbre en los datos.

Otra limitación significativa fue la ausencia de datos sobre los kilómetros recorridos, lo cual condicionó a tener que realizar los análisis basados en la variable tiempo, lo cual puede no reflejar completamente la relación entre el tiempo y los eventos producidos en las locomotoras.

Por otro lado, puesto que los datos y las herramientas las ofrecía la empresa, únicamente se podía trabajar en la oficina durante el período de prácticas en el horario laboral, lo que condicionó en gran medida el tiempo disponible tanto para la realización del análisis como para la implementación del cuadro de mandos.

Una última limitación es el hecho de que los eventos se registran muchas veces en un período de tiempo corto haciendo referencia a un mismo fallo, y por ello no siempre es posible establecer una clara relación entre un evento registrado y el fallo que lo ha podido provocar.

Sin embargo, a pesar de estas limitaciones, se considera que los resultados obtenidos son significativos y proporcionan información valiosa para la toma de decisiones.

5.2. Legado

El legado de este TFG reside en tres ámbitos. A nivel empresarial este trabajo ofrece una herramienta con potencial para agilizar los procesos de análisis y toma de decisiones. Las descripciones de los gráficos y los análisis de fiabilidad realizados esperan servir de referencia para el estudio e interpretación del cuadro de mandos desarrollado. Además, quedan a disposición de la empresa tanto el código como la herramienta para la continuación del desarrollo de esta, a fin de optimizarla y refinarla.

Por otro lado, a nivel académico, se han mostrado diferentes enfoques para realizar un análisis de fiabilidad sobre datos de eventos registrados en locomotoras. Además, se propone un ejemplo de cuadro de mandos que permite analizar de forma visual la influencia de cada uno de los eventos posibles sobre las diferentes flotas con sus correspondientes locomotoras.

Finalmente, a nivel personal, el proyecto me ha permitido fortalecer y desarrollar habilidades como científica de datos. A su vez me ha permitido adquirir nuevos conocimientos y mantenerme en constante evolución. También me ha ayudado a fortalecer la confianza y la autonomía en la toma de decisiones, ya que he tenido que asumir responsabilidades y tomar decisiones importantes para el correcto desarrollo del mismo. Finalmente, me ha permitido obtener una visión más clara del campo del análisis de datos, lo que contribuye a mi enriquecimiento personal y puede ser aplicable en futuros proyectos.

5.3. Relación con los estudios cursados

El desarrollo de este proyecto ha sido posible gracias a los conocimientos adquiridos durante el grado en Ciencia de Datos. La habilidad desarrollada para analizar datos y crear visualizaciones de estos junto con los conocimientos adquiridos sobre Python, SQL, Power BI y modelización han sido fundamentales para llevar a cabo las diferentes etapas del proyecto.

En primer lugar, el desarrollo de la capacidad para analizar datos ha permitido realizar una exploración exhaustiva de los conjuntos de datos, identificando patrones y anomalías, y aplicando técnicas de limpieza y preprocesamiento de datos para garantizar la calidad de la información utilizada.

El conocimiento en SQL ha sido esencial para la extracción, manipulación y análisis de datos a partir de bases de datos. Mediante consultas SQL, se han recolectado los datos necesarios para implementar el cuadro de mandos y realizar el análisis.

La visualización de datos ha sido crucial para representar de manera efectiva la información recopilada, permitiendo una comprensión clara y visualmente atractiva de los resultados. Mediante la creación de gráficos y visualizaciones interactivas en Power BI, se ha facilitado la comunicación de hallazgos importantes y la presentación de conclusiones de manera comprensible para los usuarios.

Por último, la modelización de datos mediante Python ha permitido la aplicación de técnicas como el estimador Kaplan-Meier y el MCF, así como el preprocesamiento de los datos para estos análisis.

Cabe destacar que mediante la realización de este proyecto se han requerido y puesto en práctica diversas competencias transversales como son innovación y creatividad, ya que se ha requerido pensar de manera original para diseñar un cuadro de mandos efectivo; responsabilidad y toma de decisiones, ya que ha habido situaciones en las que ha sido necesario asumir responsabilidad y tomar decisiones informadas; comunicación efectiva, ya que se han adaptado las explicaciones a la audiencia que las recibirá; y compromiso social y medioambiental, ya que el estudio de los eventos busca contribuir a la reducción de la contaminación.

6. Trabajo futuro

En el trabajo futuro, se pueden considerar las siguientes acciones:

- Añadir datos GPS. Debido al tiempo de carga, no es posible considerar datos de más de dos meses anteriores a la hora de graficar el mapa. Sería interesante explorar diferentes enfoques que permitan optimizar la carga de datos consiguiendo así representar más información.

- Ampliar la inclusión de eventos importantes en el análisis. Esto implica identificar y agregar más tipos de eventos que puedan ser relevantes para la empresa.

- Considerar la incorporación de flotas o locomotoras nuevas en el estudio. Según transcurra el tiempo, pueden surgir vehículos que la empresa tenga interés en analizar por lo que deberían ser incluidos.

- Realizar modelos de predicción adicionales. Además de los modelos existentes, se pueden desarrollar modelos específicos para ciertos tipos de locomotoras. Al considerar características únicas de las locomotoras individuales, es posible identificar aquellas que podrían ser más propensas a eventos. Esto permitirá una gestión proactiva y específica de los mantenimientos.

- Introducir una funcionalidad para comparar locomotoras. Para ello, se podría utilizar el test estadístico *logrank* [35]. Esta técnica se basa en el método Kaplan-Meier y permite realizar comparaciones estadísticas entre grupos de locomotoras y determinar si existen diferencias significativas en la ocurrencia de eventos.

- Explorar otras variables que puedan influir en los eventos. Además de las variables ya consideradas, es importante investigar otras posibles variables que podrían tener impacto en los eventos de los trenes. Esto puede incluir condiciones ambientales, características de la vía, datos operativos, entre otros.

- Estudiar la posible influencia de otras variables en la ocurrencia de eventos. Para ello, se podría utilizar el modelo de regresión de Cox [36]. Este modelo permite analizar la relación entre las variables y el tiempo hasta que ocurre un evento. La aplicación de este enfoque proporcionará una comprensión más precisa de la influencia de cada variable en los eventos de los trenes. Para ponerlo en práctica, habría que hacer un estudio de la factibilidad de su implementación.

- Implementar los modelos de fiabilidad utilizando la variable de kilometraje en lugar de la variable tiempo. Actualmente esta variable no está disponible porque no se recoge con el mismo sistema que los eventos. En el futuro sería interesante disponer de una forma óptima de recoger el kilometraje puesto que permitiría contar con una medida fiable del funcionamiento de las locomotoras.

En resumen, el trabajo futuro implica la exploración de nuevas vías de análisis y la consideración de variables adicionales para obtener una comprensión más completa de los eventos en los trenes. Con estas acciones, se espera mejorar la capacidad de identificar patrones, realizar predicciones precisas y tomar decisiones informadas en relación con el mantenimiento y la gestión del sistema ferroviario.

7. Referencias

[1] Sobre nosotros - Stadler. (s/f). Stadler Rail. Recuperado el 22 de junio de 2023, de https://www.stadlerrail.com/es/sobre-nosotros/

[2] Myamlin, S., Neduzha, L., & Urbutis, Ž. (2016). Research of innovations of diesel locomotives and bogies. Procedia Engineering, 134, 469–474. https://doi.org/10.1016/j.proeng.2016.01.069

[3] Simpson, T. F. B. (1957). Diesel locomotive building and maintenance. Journal of the Institution of Locomotive Engineers, 47(256), 131–194. https://doi.org/10.1243/jile_proc_1957_047_022_02

[4] Goolak, S., & International Science Group. (2021). Improvement of the model of power losses in the pulsed current traction motor in an electric locomotive. En Theoretical foundations of engineering. Tasks and problems (pp. 135–159). International Science Group.

[5] Negash, S. (2004). Business intelligence. Communications of the association for information systems, 13(1), 15.

[6] Morgan, M. B., Branstetter, B. F., 4th, Mates, J., & Chang, P. J. (2006). Flying blind: using a digital dashboard to navigate a complex PACS environment. Journal of Digital Imaging, 19(1), 69–75. https://doi.org/10.1007/s10278-005-8732-2

[7] Gitzel, R., Turring, S., & Maczey, S. (2015). A data quality dashboard for reliability data. 2015 IEEE 17th Conference on Business Informatics.

[8] Moyne, J., Iskandar, J., Hawkins, P., Furest, A., Pollard, B., Walker, T., & Stark, D. (2013). Deploying an Equipment Health monitoring dashboard and assessing predictive maintenance. ASMC 2013 SEMI Advanced Semiconductor Manufacturing Conference.

[9] Singh, G., Kumar, A., Singh, J., & Kaur, J. (2023). Data visualization for developing effective performance dashboard with power BI. 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), 968–973.

[10] Krishnan, V. (2017). Research Data Analysis with Power BI. INFLIBNET Centre.

[11] Marzá, M., & Jesús, I. (2020). Dashboard mediante tecnología Power BI, lenguaje DAX y librería pbiviz. Universitat Jaume I.

[12] maggiesMSFT. (s/f). Documentación de Power BI - Power BI. Microsoft.com. Recuperado el 22 de junio de 2023, de https://learn.microsoft.com/es-es/power-bi/

[13] Amin, M. T., Khan, F., & Zuo, M. J. (2019). A bibliometric analysis of process system failure and reliability literature. Engineering Failure Analysis, 106(104152), 104152. https://doi.org/10.1016/j.engfailanal.2019.104152

[14] Leite, M., A. Costa, M., Alves, T., Infante, V., & Andrade, A. R. (2022). Reliability and availability assessment of railway locomotive bogies under correlated failures. Engineering Failure Analysis, 135(106104), 106104. https://doi.org/10.1016/j.engfailanal.2022.106104

[15] Arias Velásquez, R. M., Mejía Lara, J. V., & Melgar, A. (2019). Reliability model for switchgear failure analysis applied to ageing. Engineering Failure Analysis, 101, 36–60. https://doi.org/10.1016/j.engfailanal.2019.03.004

[16] vericeli

[17] Kim, D.-S., Ok, S.-Y., Song, J., & Koh, H.-M. (2013). System reliability analysis using dominant failure modes identified by selective searching technique. Reliability Engineering & System Safety, 119, 316–331. https://doi.org/10.1016/j.ress.2013.02.007

[18] Álvarez, M. Á. N. (2017). Estudio de la fiabilidad de los sistemas reparables y desarrollo de un procedimiento de análisis multivariante. UNED. Universidad Nacional de Educación a Distancia.

[19] Alencar, A. R. (2023). Reliability & maintainability strategies for repairable systems in rail transportation fleets. 2023 Annual Reliability and Maintainability Symposium (RAMS).

[20] Garmabaki, A. H. S., Ahmadi, A., Block, J., Pham, H., & Kumar, U. (2016). A reliability decision framework for multiple repairable units. Reliability Engineering & System Safety, 150, 78–88. https://doi.org/10.1016/j.ress.2016.01.020

[21] Block, J., Ahmadi, A., Tyrberg, T., & Kumar, U. (2014). Fleet-level Reliability of Multiple Repairable Units: a Parametric Approach using the Power Law Process. International journal of performability engineering, 10, 239-250.

[22] Szkoda, Maciej & Kaczor, Grzegorz. (2016). Reliability and availability assessment of diesel locomotive using Fault Tree Analysis. Archives of Transport. 40. 65-75. 10.5604/08669546.1225470.

[23] Nelson, W. (1995). Confidence Limits for Recurrence Data: Applied to Cost or Number of Product Repairs. Technometrics, 37(2), 147–157. https://doi.org/10.2307/1269616

[24] Donaldson, M. G., Sobolev, B., Kuramoto, L., Cook, W. L., Khan, K. M., & Janssen, P. A. (2007). Utility of the mean cumulative function in the analysis of fall events. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 62(4), 415–419. https://doi.org/10.1093/gerona/62.4.415

[25] Nelson, W. B. (2003). Recurrent events data analysis for product repairs, disease recurrences, and other applications. Society for Industrial and Applied Mathematics.

[26] Kaplan, E. L., & Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association, 53(282), 457. https://doi.org/10.2307/2281868

[27] Pearson, D. (s/f). Las principales pruebas no paramétricas son las siguientes: Aiu.edu. Recuperado el 22 de junio de 2023, de https://cursos.aiu.edu/METODOS%20CUANTITATIVOS%20DE%20INVESTIGACION/7/Sesi%C3%B3n%207.pdf

[28] Ranstam, J., & Cook, J. A. (2017). Kaplan-Meier curve. The British Journal of Surgery, 104(4), 442. https://doi.org/10.1002/bjs.10238

[29] Barker, C. (2009). The mean, median, and confidence intervals of the Kaplan-Meier survival estimate—computations and applications. The American statistician, 63(1), 78–80. https://doi.org/10.1198/tast.2009.0015

[30] Miettinen, O. S. (2008). Survival analysis: Up from Kaplan-Meier-Greenwood. European Journal of Epidemiology, 23(9), 585–592. https://doi.org/10.1007/s10654-008-9278-7

[31] Bland, J. M., & Altman, D. G. (1998). Statistics Notes: Survival probabilities (the Kaplan-Meier method). BMJ, 317(7172), 1572–1580. https://doi.org/10.1136/bmj.317.7172.1572

[32] SQL Developer. (s/f). Oracle.com. Recuperado el 22 de junio de 2023, de https://www.oracle.com/database/sqldeveloper/

[33] Welcome to. (s/f). Python.org. Recuperado el 22 de junio de 2023, de https://www.python.org/

[34] maggiesMSFT. (s/f). Documentación de Power BI - Power BI. Microsoft.com. Recuperado el 22 de junio de 2023, de https://learn.microsoft.com/es-es/power-bi

[35] Bland, J. M., & Altman, D. G. (2004). The logrank test. BMJ (Clinical Research Ed.), 328(7447), 1073. https://doi.org/10.1136/bmj.328.7447.1073

[36] Lunn, M., & McNeil, D. (1995). Applying Cox regression to competing risks. Biometrics, 51(2), 524–532. https://doi.org/10.2307/2532940

8. Apéndices y anexos

ANEXO 1: CONSULTAS SQL

- Consulta SQL para la recuperación de datos de eventos

```
SELECT hvf.loco_id_i, tl.loco_name_c, tl.loco_fleet_c, tl.loco_platform_c, hvf.VCUF_TimeA_d,
    from_tz(hvf.vcuf_timea_d, 'UTC') at time zone mvs.vst_time_zone_region_c AS vcuf_timea_d_tz_d,
    to_char(from_tz(hvf.vcuf_timea_d, 'UTC') at time zone mvs.vst_time_zone_region_c,
            'DD/MM/YY HH24:MI:SS TZR') AS vcuf_timea_d_tz_c,
    hvf.VCUF_Timei_d,hvf.vcuf_code_i, to_char(hvf.vcuf_code_i), hvf.vcuf_id_i, tvfd.vcufd_category_c,
    DECODE(hvf.VCUF_Timei_d, null, 'YES', 'NO') AS ACTIVE,
    regexp_substr(to_char(hvf.VCUF_Timei_d-hvf.VCUF_TimeA_d),'[^ ]+',1,2) AS DURATION,
    tvfdt.vcufdt_description_c, tvfd.vcufd_system_c, tvfd.vcufd_priority_i, lsv.soft_version_c
FROM [redacted] hvf
LEFT JOIN [redacted] tl ON (tl.loco_id_i = hvf.loco_id_i)
LEFT JOIN [redacted] tls ON (tl.loco_series_c = tls.serie_name_c)
LEFT JOIN [redacted] mvs ON (hvf.loco_id_i = mvs.loco_id_i)
LEFT JOIN [redacted] lsv ON (lsv.soft_id_i = hvf.soft_id_i)
LEFT JOIN [redacted] tvfd ON (tvfd.soft_id_i = lsv.soft_id_i
                              AND tvfd.vcufd_code_i = hvf.vcuf_code_i)
LEFT JOIN [redacted] tvfdt ON (tvfdt.soft_id_i = lsv.soft_id_i
                               AND tvfdt.vcufd_code_i = tvfd.vcufd_code_i
                               AND tvfdt.lang_id_i = 1)
WHERE  hvf.vcuf_code_i IN ('2882','1832','832','836','1846','1839','1822','923','957')
    AND hvf.VCUF_TimeA_d  <= SYSDATE
    AND loco_platform_c IN [redacted]
    AND die_mode_c IS NOT null
    AND tl.loco_fleet_c IN [redacted]
ORDER BY hvf.VCUF_TimeA_d DESC
```

- Consulta SQL para la obtención de datos de localización

```
SELECT ff.vcuf_id_i, ff.loco_fleet_c, ff.loco_id_i, gps.loco_name_c, ff.vcuf_code_i, ff.VCUF_TimeA_d,
    gps.rout_longitude_f, gps.rout_latitude_f
FROM (SELECT fallo.vcuf_id_i, fallo.loco_fleet_c, fallo.loco_id_i, fallo.vcuf_code_i, fallo.VCUF_TimeA_d,
           fallo.local_time, min(mov.rout_date_d) AS FECHA
      FROM (SELECT hvf.vcuf_id_i, tl.loco_fleet_c, hvf.loco_id_i, hvf.VCUF_TimeA_d, hvf.vcuf_code_i,
           to_date(to_char(CAST(FROM_TZ(CAST(CAST(hvf.vcuf_timea_d AS DATE) AS TIMESTAMP),'UTC')
                                AT TIME ZONE mvs.vst_time_zone_region_c AS TIMESTAMP),
                           'yyyy-mm-dd hh24:mi:ss'),'yyyy-mm-dd hh24:mi:ss') AS local_time,
           mvs.vst_time_zone_region_c
           FROM [redacted] hvf
           LEFT JOIN [redacted] tl ON (tl.loco_id_i = hvf.loco_id_i)
           LEFT JOIN [redacted] MVS ON (mvs.loco_id_i = HVF.loco_id_i)
           WHERE hvf.vcuf_code_i IN ('2882','1832','832','836','1846','1839','1822','923','957')
               AND hvf.VCUF_TimeA_d BETWEEN (SYSDATE -  60) AND SYSDATE
               AND tl.loco_fleet_c IN [redacted]
           ) FALLO
       LEFT JOIN [redacted] mov ON (fallo.loco_id_i = mov.loco_id_i)
       WHERE fallo.local_time <= mov.rout_date_d
           AND rout_date_d < (fallo.local_time + 34/1440)
           AND fallo.loco_fleet_c IN [redacted]
       GROUP BY fallo.vcuf_id_i, fallo.loco_fleet_c, fallo.loco_id_i, fallo.vcuf_code_i, fallo.VCUF_TimeA_d,
       fallo.local_time) FF
LEFT JOIN (SELECT mov.loco_id_i, mov.rout_date_d, mov.rout_longitude_f, mov.rout_latitude_f, tl.loco_name_c
           FROM [redacted] mov
           LEFT JOIN [redacted] tl ON (mov.loco_id_i = tl.loco_id_i)
           WHERE tl.loco_fleet_c IN [redacted]
           AND mov.rout_date_d BETWEEN (SYSDATE - 60) AND SYSDATE
           ) GPS ON (ff.loco_id_i = gps.loco_id_i AND ff.FECHA= gps.rout_date_d)
```

ANEXO 2: OBJETIVOS DE DESARROLLO SOSTENIBLE

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

Objetivos de Desarrollo Sostenibles	**Alto**	**Medio**	**Bajo**	**No Procede**
ODS 1. **Fin de la pobreza.**			**X**	
ODS 2. **Hambre cero.**				**X**
ODS 3. **Salud y bienestar.**				**X**
ODS 4. **Educación de calidad.**				**X**
ODS 5. **Igualdad de género.**				**X**
ODS 6. **Agua limpia y saneamiento.**				**X**
ODS 7. **Energía asequible y no contaminante.**				**X**
ODS 8. **Trabajo decente y crecimiento económico.**				**X**
ODS 9. **Industria, innovación e infraestructuras.**	**X**			
ODS 10. **Reducción de las desigualdades**.				**X**
ODS 11. **Ciudades y comunidades sostenibles**.		**X**		
ODS 12. **Producción y consumo responsables.**				**X**
ODS 13. **Acción por el clima.**			**X**	
ODS 14. **Vida submarina.**				**X**
ODS 15. **Vida de ecosistemas terrestres.**				**X**
ODS 16. **Paz, justicia e instituciones sólidas.**				**X**
ODS 17. **Alianzas para lograr objetivos.**				**X**

Reflexión sobre la relación del TFG/TFM con los ODS y con el/los ODS más relacionados.

ODS 1 - Fin de la pobreza: Para lograr la erradicación de la pobreza, es fundamental garantizar que todas las personas tengan acceso equitativo a recursos y servicios. Los vehículos ferroviarios desempeñan un papel clave al proporcionar transporte. Este trabajo de fin de grado contribuye a mejorar los servicios ferroviarios, lo que se traduce en una mayor disponibilidad para los usuarios finales y, en última instancia, en un acceso más equitativo a los recursos y oportunidades.

ODS 9 - Industria, innovación e infraestructuras: Como se ha mencionado previamente, este proyecto contribuye a mejorar el funcionamiento de las locomotoras. Esto puede resultar en una mejora de la disponibilidad y eficiencia de la infraestructura ferroviaria, al permitir el análisis de

los eventos producidos por los fallos, pudiendo así mejorar dicha infraestructura. Esto es un objetivo central de este ODS, al igual que el fomento de tecnologías que permitan analizar los eventos de manera más eficiente, como se aborda en este trabajo.

ODS 11 - Ciudades y comunidades sostenibles: Uno de los objetivos de este ODS es proporcionar sistemas de transporte seguros, accesibles, asequibles y sostenibles para todos, al tiempo que se mejora la seguridad vial. Como se ha mencionado anteriormente, este proyecto puede contribuir a mejorar el funcionamiento de las locomotoras, pudiendo aplicarse también a otros medios de transporte como metros o tranvías, lo que está alineado con el objetivo de promover ciudades y comunidades sostenibles.

ODS 13 - Acción por el clima: Este objetivo se centra en tomar medidas urgentes para combatir el cambio climático y sus impactos. Las locomotoras al igual que los demás medios de transporte contribuyen a la contaminación del medio ambiente. Al mejorar los servicios ferroviarios, este proyecto puede ayudar a reducir las emisiones y mitigar el cambio climático.