

**Diseño de un sistema de monitoreo y análisis del consumo de combustible basado en datos
de abastecimiento en la empresa Transpormax**

César Leonardo Ortiz Rovira

Asesor

Julio Eduardo Mejia Manzano

Universidad Nacional Abierta y a Distancia UNAD
Escuela de Ciencias Basicas, Tecnologia e Ingenieria ECBTI
Especialización en Ciencia de Datos y Analítica

2025

Resumen

El proyecto desarrolló un sistema integral de análisis de datos para optimizar el consumo de combustible y detectar anomalías operativas en la empresa Transpormax. A partir de datos históricos de odómetro, kilometraje, rendimientos y registros de abastecimiento, se realizó un proceso exhaustivo de limpieza que depuró inconsistencias, valores imposibles y registros atípicos, obteniendo una base confiable para el análisis. Se construyó un dashboard en Power BI conectado directamente a las bases de datos SQL de la empresa, lo que permitió visualizar en tiempo real los indicadores clave de rendimiento por vehículo y por conductor, así como tendencias históricas del consumo.

El estudio incluyó la generación de gráficos descriptivos, la identificación de patrones operativos y el cálculo de métricas como consumo promedio, eficiencia por galón y dispersión de recorridos. Adicionalmente, se implementó un sistema de alertas basado en reglas de negocio y un modelo de detección de anomalías mediante Isolation Forest, que permitió identificar comportamientos inusuales en el uso de combustible, tanto a nivel de vehículos como de conductores. Los resultados evidenciaron diferencias significativas en la eficiencia entre operadores, así como eventos de posible fraude, errores de reporte o uso inapropiado de la flota.

Este análisis permitió establecer conclusiones orientadas al mejoramiento de la eficiencia operativa y al fortalecimiento de los mecanismos de control. Asimismo, se generaron recomendaciones enfocadas en la capacitación de conductores, la automatización de alertas, la depuración continua de datos y la integración de nuevas fuentes de información que permitan enriquecer el análisis y fortalecer los mecanismos de control operacional, consolidando así un enfoque analítico sustentado en hallazgos clave como la identificación de brechas significativas de rendimiento entre conductores, la detección de 90 anomalías operativas y la estandarización

de un dataset confiable de 1.845 registros que fortalece los mecanismos de control, mejora la eficiencia operativa y aporta evidencia directa para la reducción de los costos asociados al consumo de combustible.

Palabras clave: análisis de datos, consumo de combustible, machine learning, eficiencia operativa, Power BI.

Abstract

This project developed an integrated data analysis system aimed at optimizing fuel consumption and detecting operational anomalies within the company Transpormax. Using historical odometer, mileage, fuel efficiency, and refueling records, an exhaustive data-cleaning process was conducted to remove inconsistencies, impossible values, and atypical records, resulting in a reliable dataset for analysis. A dashboard in Power BI, directly connected to the company's SQL databases, was developed to enable real-time visualization of key performance indicators by vehicle and driver, as well as historical consumption trends.

The study included descriptive visualizations, the identification of operational patterns, and the calculation of metrics such as average consumption, fuel efficiency, and trip variability. Additionally, a rule-based alert system and an anomaly detection model using Isolation Forest were implemented, allowing the identification of unusual fuel-use behaviors at both vehicle and driver levels. The results revealed significant efficiency gaps among operators, as well as events indicative of potential fraud, reporting errors, or improper fleet use.

The analysis led to conclusions focused on improving operational efficiency and strengthening control mechanisms. Recommendations were formulated based on driver training, automated alerts, continuous data cleansing, and the incorporation of new data sources to enrich operational monitoring. This approach, supported by key findings—such as the detection of 90 operational anomalies, notable performance disparities among drivers, and the consolidation of a validated dataset of 1,845 records—provides strong evidence for reducing fuel-related costs and enhancing the transparency and efficiency of fleet management.

Keywords: data analysis, fuel consumption, machine learning, operational efficiency, Power BI.

Tabla de Contenido

Introducción	10
Descripción del Problema	11
Planteamiento del Problema	11
Justificación	13
Objetivos	14
Objetivo General	14
Objetivos Específicos	14
Marco de Referencia	15
Estado del Arte	15
Marco Contextual	17
Marco Teórico	18
Marco Conceptual	19
Marco Normativo	21
Metodología	24
Enfoque de la Investigación	24
Tipo de Investigación	24
Fase 1 Recolección y Limpieza de Datos	24
Fase 2 Integración de Bases de Datos	25
Fase 3 Análisis Exploratorio de Datos	25
Fase 4 Aplicación de Técnicas de Machine Learning	25
Fase 5 Desarrollo del Dashboard Final y Conclusiones	25
Resultados	26

Procesamiento Inicial y Depuración de los Datos	26
Caracterización General de los Datos de Abastecimiento	27
Construcción del Indicador de Rendimiento	28
Análisis del Rendimiento del Combustible	29
Sistema de Alertas Operativas Basado en Reglas de Negocio	30
Galones Abastecidos Superiores al Límite Operativo	31
Kilómetros Recorridos Irreales en un solo Ciclo Operativo	31
Rendimiento Atípico o Sospechoso	32
Kilometraje sin Consumo de Combustible	32
Consumo de Combustible sin Desplazamiento Registrado	32
Registros Duplicados por Vehículo y Fecha.....	32
Registros con Valores Faltantes en Variables Críticas	33
Resultados del Sistema de Alertas	33
Análisis de la Data Limpia	35
Análisis de Correlación Entre Variables Operativas	36
Análisis Descriptivo con Apoyo Gráfico.....	37
Análisis del Consumo y Rendimiento por Conductor	41
Detección de Anomalías Mediante Isolation Forest.....	44
Conductores con Mayor Cantidad de Anomalías	45
Visualización de Anomalías	47
Dashboard en Power BI	48
Conclusiones	51
Recomendaciones	53

Referencias Bibliográficas	55
Apéndices.....	57

Tabla de Figuras

Figura 1 Descripción de las Columnas de la Base de Datos	27
Figura 2 Flags Creadas para la Limpieza de la Base de Datos	31
Figura 3 Distribución de los Registros Inconsistentes.....	33
Figura 4 Descripción de la Base de Datos Limpia	35
Figura 5 Matriz de Correlación entre Variables Operativas	36
Figura 6 Histogramas de Distribución de Galones	37
Figura 7 Histogramas de Distribución de Kilómetros Recorridos	38
Figura 8 Histogramas de Distribución de Rendimientos.....	39
Figura 9 Boxplot Rendimientos	39
Figura 10 Matriz de Correlación	40
Figura 11 Estadísticas por Conductor	41
Figura 12 Estadísticas por Conductor Alto Rendimiento	42
Figura 13 Estadísticas por Conductor Bajo Rendimiento	43
Figura 14 Grafico de Dispersión de Anomalías	47
Figura 15 Vista General del Dashboard en Power Bi	48
Figura 16 Vista por Vehiculo Dashboard en Power Bi	49
Figura 17 Vista por Conductor Dashboard en Power Bi	50

Lista de Apéndices

Apéndice A <i>Enlace de la Presentación</i>	57
Apéndice B <i>Autorización de la Empresa</i>	58
Apéndice C <i>Acuerdo de Confidencialidad</i>	62

Introducción

En el contexto de las empresas de transporte, la eficiencia operativa y el control de los costos asociados al consumo de combustible son factores clave para la sostenibilidad y rentabilidad de la operación (Fernández Romero, 2020). En el caso de Transpormax, encargada del transporte no asistencial de pacientes, el consumo de combustible representa una proporción considerable del gasto logístico, generando presión sobre los márgenes de utilidad y afectando la competitividad de la empresa. A esto se suma la posibilidad de inconsistencias, errores de registro o irregularidades asociadas al abastecimiento, que limitan la capacidad de control y dificultan la identificación de comportamientos ineficientes.

Estudios como los de Castillo et al. (2019) señalan que el sector transporte en Colombia presenta altos niveles de ineficiencia energética, lo que se traduce en costos elevados y pérdidas operativas. En concordancia, análisis como los de Martínez y Marulanda (2019) destacan la necesidad de fortalecer la trazabilidad y el control del combustible mediante herramientas de gestión basadas en datos. En este contexto, se hace necesario diseñar un sistema que permita consolidar la información proveniente de las estaciones de servicio, analizarla de forma integral y generar indicadores que faciliten la toma de decisiones.

El presente proyecto plantea el diseño de un sistema de monitoreo y análisis del consumo de combustible basado exclusivamente en los registros de abastecimiento, con el propósito de identificar consumos atípicos, patrones de comportamiento y oportunidades de optimización. La aplicación de técnicas de machine learning y la implementación de herramientas de visualización en Power BI permitirán fortalecer el control del recurso, reducir pérdidas económicas y mejorar la eficiencia operativa en Transpormax.

Descripción del Problema

El aumento constante en los costos de combustible y la falta de herramientas integradas de control representan un desafío crítico para las empresas transportadoras en Colombia. En Transpormax, los registros de abastecimiento de combustible constituyen la principal fuente de información disponible para el seguimiento del recurso; sin embargo, estos datos no se encuentran consolidados ni sometidos a análisis que permitan identificar irregularidades, tendencias o desviaciones respecto al comportamiento esperado.

La ausencia de un sistema estructurado de monitoreo dificulta el control sobre el consumo real de los vehículos, generando incertidumbre frente a la veracidad de los registros y la eficiencia en la operación. Este escenario limita la capacidad de la empresa para detectar consumos inusualmente altos, errores en los volúmenes registrados, variaciones injustificadas en los costos o posibles prácticas indebidas. Como resultado, la organización enfrenta riesgos económicos significativos, asociados tanto al sobreconsumo como a la falta de trazabilidad del recurso.

En consecuencia, se requiere el diseño de un sistema que permita centralizar los datos de abastecimiento, analizarlos mediante indicadores clave y generar alertas basadas en patrones históricos, con el fin de mejorar la transparencia y el control del combustible en la empresa.

Planteamiento del Problema

El consumo de combustible constituye uno de los principales costos operativos en las empresas de transporte en Colombia, llegando a representar hasta el 40 % del gasto logístico (Fernández Romero, 2020; Rodríguez & Méndez, 2021). En Transpormax, la falta de un sistema de control basado en el análisis estructurado de los registros de abastecimiento ha generado incertidumbre frente a la eficiencia del consumo y la confiabilidad de los datos registrados.

Actualmente, la empresa cuenta con información básica de cada tanqueo, como fecha, volumen, costo y estación de servicio; sin embargo, estos registros se gestionan de manera fragmentada y sin herramientas de análisis, lo que dificulta identificar consumos atípicos, errores de digitación o posibles inconsistencias. Estudios como los de Martínez y Marulanda (2019) advierten que la ausencia de trazabilidad y control en el uso del combustible incrementa los riesgos operativos y genera pérdidas económicas significativas en el sector transporte.

La problemática radica en la incapacidad para analizar de manera integral la información histórica, establecer patrones de consumo por vehículo o conductor, y detectar desviaciones significativas respecto al comportamiento esperado. Esta situación afecta la toma de decisiones, dificulta la implementación de estrategias de optimización y limita el control interno del recurso.

Surge así la necesidad de diseñar un sistema de monitoreo y análisis basado exclusivamente en los datos de abastecimiento, mediante técnicas de análisis estadístico y machine learning, que permitan identificar patrones, detectar anomalías y proporcionar herramientas visuales para fortalecer la gestión operativa de Transpormax.

Justificación

Este proyecto se justifica en la necesidad de abordar una problemática que impacta directamente la sostenibilidad operativa y financiera de las empresas transportadoras. El uso ineficiente del combustible y la falta de trazabilidad en los procesos de abastecimiento no solo representan un desafío interno, sino que generan repercusiones a nivel económico y organizacional.

Diversas investigaciones han evidenciado que la gestión del combustible es un factor crítico en el sector transporte. Castillo et al. (2019) destacan que la optimización del consumo puede generar reducciones significativas en los costos operativos. Asimismo, la ausencia de sistemas integrados de control y monitoreo incrementa la probabilidad de errores, inconsistencias y posibles irregularidades en el proceso de abastecimiento.

Este proyecto cobra relevancia al proponer un sistema que centraliza, analiza y visualiza la información de consumo mediante técnicas avanzadas de análisis de datos y machine learning, lo que permitirá identificar patrones anormales y apoyar la toma de decisiones estratégicas. Además, los hallazgos derivados pueden contribuir al desarrollo de herramientas de control más robustas y aportar bases conceptuales y metodológicas para futuras investigaciones en el campo de la gestión de combustible.

Objetivos

Objetivo General

Diseñar un sistema de monitoreo y análisis del consumo de combustible basado en los datos de abastecimiento de la empresa Transpormax, con el propósito de optimizar la eficiencia operativa y detectar anomalías.

Objetivos Específicos

Identificar los factores que influyen en el uso del combustible en la flota vehicular de Transpormax mediante el análisis de los registros de abastecimiento.

Integrar y depurar los datos provenientes de las estaciones de servicio para consolidar una base de información unificada.

Aplicar técnicas de análisis de datos y machine learning para detectar patrones de comportamiento anómalos o ineficientes en el consumo de combustible.

Desarrollar un dashboard interactivo en Power BI que permita monitorear los principales indicadores asociados al consumo de combustible.

Marco de Referencia

Estado del Arte

El análisis del consumo de combustible y la detección de irregularidades en flotas vehiculares se ha convertido en un campo de creciente interés debido a su impacto directo en la eficiencia operativa y en los costos logísticos. En particular, los proyectos que se basan exclusivamente en datos de estaciones de servicio han sido estudiados como una alternativa viable cuando no existe telemetría avanzada o sistemas GPS integrados. Estos enfoques priorizan la calidad de los registros, la identificación de patrones anómalos y el uso de métricas estandarizadas para evaluar la eficiencia vehicular.

Métricas y medición del consumo: La literatura destaca la importancia de utilizar indicadores que permitan comparar el rendimiento de los vehículos sin importar grandes variaciones operativas. Métricas como kilómetros por litro, litros consumidos por 100 km o consumo medio ajustado han sido ampliamente utilizadas para analizar la eficiencia en distintos tipos de flotas. En Colombia, varios estudios han señalado la necesidad de mejorar la trazabilidad del combustible y reforzar los sistemas de control interno derivados de registros de abastecimiento. Esto permite identificar patrones, anomalías y comportamientos poco eficientes en los procesos logísticos (Fernández Romero, 2020; Martínez & Marulanda, 2019).

Limitaciones de trabajar únicamente con estaciones de servicio: El uso exclusivo de registros de tanqueo presenta retos: errores de digitación, variaciones injustificadas en volumen o costo, y la dificultad de relacionar el consumo con el desempeño del vehículo en ruta. Sin embargo, estos datos siguen siendo una fuente confiable para construir modelos estadísticos básicos y avanzados cuando se complementan con técnicas robustas de limpieza y validación. La

literatura coincide en que el éxito del análisis depende en gran parte de la calidad y consistencia previa de la base consolidada.

Técnicas de detección de anomalías: En los últimos años, los métodos de aprendizaje automático han ganado relevancia en estudios sobre consumo de combustible. Algoritmos como Isolation Forest, Local Outlier Factor, DBSCAN y modelos de clustering han demostrado ser efectivos para detectar comportamientos atípicos cuando no se dispone de etiquetas directas de fraude o robo. Su ventaja radica en su capacidad para aprender patrones normales del consumo y resaltar desviaciones significativas, incluso en bases de datos pequeñas o incompletas (Janiesch et al., 2021; Russell & Norvig, 2021).

Dashboards, BI y soluciones prácticas: El uso de herramientas de visualización como Power BI ha sido ampliamente documentado para facilitar la comprensión de grandes volúmenes de datos relacionados con combustible. Estudios aplicados indican que la visualización interactiva gráficos comparativos, tendencias históricas, ranking de vehículos, alertas, etc. complementa y potencia los resultados de modelos estadísticos o de machine learning, mejorando la capacidad de respuesta operacional (Božić & Dimovski, 2019; Ortiz, 2023).

Brecha actual y aporte del proyecto: Pese a los avances, persiste una brecha importante en empresas que solo cuentan con registros de estaciones de servicio y no tienen sistemas GPS o telemáticos. Muchos estudios se enfocan en escenarios con telemetría completa, lo que limita su aplicabilidad en contextos reales como el de Transpormax. Por esta razón, este proyecto contribuye al desarrollo de una metodología ajustada a esta restricción, proponiendo técnicas de limpieza, análisis estadístico y detección de anomalías basadas exclusivamente en registros de tanqueo, complementadas con visualización en Power BI para la toma de decisiones.

Marco Contextual

La empresa Transpormax opera en el sector del transporte no asistencial de pacientes, una actividad que depende de manera directa del desempeño operativo de su flota vehicular. En este tipo de organizaciones, el uso del combustible representa uno de los componentes más significativos dentro de los costos logísticos, debido a la naturaleza continua de los desplazamientos, la variabilidad de las rutas y la necesidad de garantizar el servicio en diferentes zonas urbanas y periurbanas.

El contexto operativo de Transpormax se caracteriza por la ausencia de rutas fijas, ya que los desplazamientos responden a la programación diaria de citas médicas y a la ubicación geográfica de los pacientes. Esta dinámica genera variaciones constantes en la distancia recorrida, la duración de las jornadas y la frecuencia de los trayectos, lo que dificulta establecer patrones de consumo lineales o previsibles. Además, la empresa depende principalmente de los registros de abastecimiento suministrados por las estaciones de servicio, los cuales constituyen la única fuente sistemática de información disponible para analizar el comportamiento del combustible.

A nivel sectorial, el transporte terrestre en Colombia enfrenta múltiples desafíos asociados al incremento del precio de los combustibles, la falta de estandarización de los procesos de control interno y la presencia de riesgos operativos como inconsistencias en los registros, sobreconsumo y posibles prácticas irregulares. Diversos estudios nacionales han resaltado la necesidad de fortalecer los mecanismos de trazabilidad del combustible y adoptar herramientas basadas en datos que permitan identificar pérdidas, consumos atípicos y deficiencias operativas. Sin embargo, muchas empresas pequeñas y medianas como es el caso de

Transpormax no cuentan con sistemas GPS o telemáticos que integren de manera automática la información de recorridos y desempeño, limitando su capacidad de análisis y toma de decisiones.

En este entorno, la consolidación de los datos provenientes de las estaciones de servicio adquiere un papel fundamental para comprender el comportamiento histórico de la flota, detectar tendencias y establecer alertas tempranas. La disponibilidad de registros de fecha, volumen abastecido, valor pagado y tipo de combustible permite construir indicadores clave que pueden ser utilizados para identificar anomalías o comportamientos fuera de lo esperado. Estos registros representan una oportunidad para desarrollar un sistema de análisis que mejore la eficiencia operativa y reduzca las pérdidas económicas asociadas al uso inadecuado del combustible.

Por tanto, el marco contextual de este proyecto se enmarca en un escenario operativo complejo, con restricciones en las fuentes de datos pero con una necesidad urgente de optimizar recursos y mejorar la transparencia en los procesos internos. El desarrollo de un sistema de monitoreo y análisis basado exclusivamente en datos de estaciones de servicio constituye una respuesta innovadora y viable para fortalecer la gestión del combustible en Transpormax, alineándose con las tendencias actuales del sector hacia la toma de decisiones basada en información y análisis estadístico.

Marco Teórico

Consumo de Combustible en el Transporte de Carga en Colombia: El análisis del consumo de combustible es fundamental en proyectos de eficiencia logística, ya que este representa hasta el 40% de los costos operativos en empresas de transporte (Fernández Romero, Universidad Antonio Nariño, 2020). Comprender sus determinantes permite identificar oportunidades de ahorro y sostenibilidad. En este proyecto, el consumo será tratado como variable crítica de análisis, ya que su monitoreo sistemático permitirá evaluar el rendimiento

operativo de la flota y detectar patrones de uso ineficiente o sospechoso, clave para el control del robo de combustible.

Aplicación de Machine Learning en la Detección de Anomalías: Los algoritmos de aprendizaje automático permiten identificar patrones no evidentes a simple vista y detectar comportamientos atípicos. En la Universidad de los Andes (2021), se utilizó machine learning para predecir fallas en la red de transporte de hidrocarburos, logrando anticiparse a eventos costosos. En este proyecto, técnicas similares serán empleadas para detectar anomalías en el consumo de combustible, como aumentos inesperados o inconsistencias entre distancias recorridas y litros consumidos, apoyando así la toma de decisiones proactivas.

Visualización de Datos y Toma de Decisiones: La interpretación eficiente de grandes volúmenes de información depende de la forma en que estos datos se presentan. Los dashboards interactivos, como los desarrollados en Power BI, permiten integrar múltiples fuentes y generar indicadores dinámicos. En un proyecto aplicado por la Universidad del Atlántico (Ortiz, 2023), la visualización facilitó la identificación de vehículos con bajo desempeño. En este trabajo, los dashboards serán diseñados para monitorear el desempeño de la flota y alertar sobre desviaciones operativas, fortaleciendo el control interno y la eficiencia organizacional.

Marco Conceptual

Análisis de Datos: El análisis de datos se refiere al proceso de inspeccionar, limpiar, transformar y modelar datos con el objetivo de descubrir información útil, llegar a conclusiones y apoyar la toma de decisiones. En el contexto del transporte, permite identificar patrones en el consumo de combustible, trayectos realizados y desempeño de los vehículos, lo que facilita el control y la optimización operativa (Provost & Fawcett, 2013).

Consumo de Combustible: Este concepto hace referencia a la cantidad de combustible que un vehículo utiliza para operar durante un periodo o trayecto específico. Es uno de los indicadores más importantes en la gestión de flotas, ya que incide directamente en los costos de operación. Un consumo elevado puede ser indicativo de fallas mecánicas, uso indebido del vehículo o prácticas ineficientes de conducción.

Machine Learning: El aprendizaje automático (Machine Learning) es una rama de la inteligencia artificial que permite a las máquinas aprender a partir de datos, sin necesidad de ser programadas explícitamente. Esta técnica puede aplicarse para detectar anomalías en el consumo de combustible o identificar patrones no evidentes que pueden estar relacionados con el mal uso de los recursos (Russell & Norvig, 2021).

Dashboard Interactivo: Los dashboards interactivos permiten visualizar KPIs y otros indicadores en tiempo real, facilitando la interpretación y la toma de decisiones por parte de los gestores operativos (Božic & Dimovski, 2019). Su uso mejora la reacción ante cambios operacionales.

Eficiencia Operativa: La eficiencia operativa se refiere a la capacidad de una organización para optimizar sus recursos y procesos con el objetivo de maximizar el rendimiento, minimizar desperdicios y reducir costos sin sacrificar calidad (Kaplan & Norton, 1996). En empresas de transporte, este concepto adquiere relevancia al implicar el uso adecuado del combustible, la correcta asignación de rutas y la detección oportuna de comportamientos anómalos que puedan afectar los márgenes de rentabilidad.

- Reduce costos operativos innecesarios.
- Mejora la asignación de recursos.
- Incrementa la productividad general.

Business Intelligence (BI): El Business Intelligence es una herramienta tecnológica que convierte datos operativos en conocimiento útil para la toma de decisiones estratégicas, permitiendo detectar oportunidades, mejorar la eficiencia y reaccionar rápidamente ante amenazas (Božic & Dimovski, 2019).

Indicadores Clave de Rendimiento (KPIs): Los KPIs son métricas cuantitativas utilizadas para medir el rendimiento de procesos. En el sector transporte, indicadores como kilómetros por litro, consumo medio y tiempo en ralentí son esenciales para evaluar y mejorar la eficiencia operativa (Asih et al., 2020).

Logística del Transporte: La logística del transporte comprende la gestión eficiente del flujo de vehículos y recursos asociados al desplazamiento de bienes o personas. Incluye la planificación de rutas, tiempos y costos, buscando maximizar la eficiencia y minimizar los desperdicios (Topolšek et al., 2018).

Marco Normativo

El análisis y control del consumo de combustible en empresas del sector transporte está enmarcado en un conjunto de normas y disposiciones legales que regulan la adquisición, uso, registro, almacenamiento y control de los combustibles líquidos en Colombia. Estas regulaciones buscan garantizar la transparencia en la operación, la correcta trazabilidad del recurso y la prevención de prácticas fraudulentas o riesgos operativos.

En primer lugar, el Decreto 1073 de 2015, compilatorio del sector de Minas y Energía, establece las condiciones para la comercialización, distribución y control de los combustibles líquidos derivados del petróleo. Esta norma define las responsabilidades de las estaciones de servicio en el registro de la información de cada abastecimiento, lo que constituye la principal fuente de datos utilizada en este proyecto. La reglamentación exige que los puntos de venta

registren de manera correcta el volumen abastecido, el precio por galón, tipo de combustible y fecha de suministro, elementos fundamentales para la construcción del sistema de monitoreo propuesto.

De manera complementaria, la Resolución 108 de 1997 del Ministerio de Minas y Energía regula el control de volúmenes y la calidad de los combustibles distribuidos en el país. Esta resolución resalta la importancia de mantener registros claros y verificables que permitan la supervisión del uso adecuado del combustible, coherente con los objetivos de trazabilidad y control planteados en este proyecto.

Por otra parte, la Superintendencia de Puertos y Transporte (hoy bajo el Ministerio de Transporte) establece lineamientos de control operativo para las empresas de transporte terrestre automotor, enfatizando la responsabilidad de mantener datos confiables sobre costos, mantenimientos y consumos. Aunque no existe una obligación legal de implementar sistemas avanzados de monitoreo, sí se exige a las empresas contar con mecanismos internos que garanticen la adecuada gestión de sus recursos operacionales, entre ellos el combustible, considerado un insumo crítico.

En materia contable y administrativa, la Ley 87 de 1993, que reglamenta el Sistema de Control Interno en las entidades públicas y que sirve como referencia general para empresas privadas, establece la obligación de implementar procedimientos de verificación, registro y seguimiento de los recursos. Esta ley respalda la necesidad de contar con herramientas que permitan identificar inconsistencias, sobreconsumos y riesgos asociados a la operación, tal como lo aborda el modelo de análisis planteado para Transpormax.

Otro marco relevante es la Ley 1581 de 2012 sobre protección de datos personales, dado que los registros de abastecimiento pueden contener información asociada a vehículos,

conductores o usuarios internos. Esta ley obliga a garantizar el tratamiento seguro, confidencial y autorizado de los datos utilizados en los análisis, especialmente en procesos que involucren almacenamiento, transformación o transferencia de información.

Finalmente, aunque no existe una norma específica que regule el uso de modelos de machine learning para la detección de anomalías, las directrices de buenas prácticas del Departamento Administrativo de la Función Pública y las recomendaciones de la NTC ISO 9001:2015 sobre gestión de procesos sugieren la adopción de metodologías basadas en análisis estadístico y mejora continua para optimizar el desempeño operativo. Estas referencias respaldan la implementación de sistemas de monitoreo mediante indicadores y visualizaciones, tal como se propone en este proyecto mediante el uso de Power BI.

En conjunto, el marco normativo vigente respalda la importancia de consolidar, verificar y analizar los datos de abastecimiento de combustible como mecanismo para garantizar eficiencia, transparencia y control en el uso de recursos operativos. El desarrollo de un sistema de monitoreo analítico basado en los registros de estaciones de servicio se encuentra plenamente alineado con estas disposiciones, aportando una solución práctica que cumple con los estándares legales y administrativos aplicables al sector transporte.

Metodología

Enfoque de la Investigación

El proyecto utilizará un enfoque cuantitativo, dado que se trabajará con datos numéricos provenientes de los registros de abastecimiento de combustible suministrados por las estaciones de servicio. Este enfoque permitirá medir, analizar y modelar patrones de consumo mediante indicadores objetivos y reproducibles, facilitando la identificación de comportamientos atípicos y la evaluación de la eficiencia operativa del uso del combustible en la flota de Transpormax (Sampieri et al., 2021).

Tipo de Investigación

El estudio es de tipo aplicado, ya que busca dar solución a un problema real en la operación de la empresa Transpormax, mediante el uso de herramientas tecnológicas y analíticas. Además, se clasifica como una investigación no experimental y longitudinal, al analizar datos históricos recopilados durante un periodo de tiempo para evaluar tendencias y comportamientos en el uso de combustible (Sampieri et al., 2021).

Diseño metodológico: Se seguirá un diseño metodológico en cinco fases:

Fase 1 Recolección y Limpieza de Datos

- Extracción de los registros de consumo de combustible provenientes de las estaciones de servicio, incluyendo variables como fecha, volumen abastecido, costo, tipo de combustible, vehículo y conductor.
- Revisión de la calidad de los datos para identificar inconsistencias, valores faltantes o errores de digitación.

- Limpieza y transformación de la información utilizando Python, Power Query y consultas SQL, con el fin de normalizar campos, estandarizar formatos de fechas, corregir valores atípicos y garantizar la integridad referencial de los datos.

Fase 2 Integración de Bases de Datos

- Consolidación de los registros de abastecimiento en una base de datos unificada, empleando identificadores comunes como placa del vehículo, fecha y estación de servicio.
- Validación estructural mediante procesos SQL para garantizar consistencia entre tablas, claves primarias y relaciones lógicas entre registros.
- Eliminación, mediante scripts SQL, de duplicados, inconsistencias y registros que no cumplan con criterios mínimos de calidad.

Fase 3 Análisis Exploratorio de Datos

- Cálculo de KPIs como: kilómetros por galon, consumo promedio diario, entre otros.
- Visualización de datos mediante dashboards interactivos en Power
- Estadísticas descriptivas para establecer patrones y variabilidad.

Fase 4 Aplicación de Técnicas de Machine Learning

- Implementación de modelos de detección de anomalías (e.g., Isolation Forest, DBSCAN) para identificar posibles casos de robo o uso irregular del combustible (Janiesch et al., 2021).

Fase 5 Desarrollo del Dashboard Final y Conclusiones

- Diseño de dashboard final que integre visualizaciones, alertas automáticas y análisis de eficiencia operativa.
- Recomendaciones operativas basadas en los hallazgos obtenidos.

Resultados

Procesamiento Inicial y Depuración de los Datos

Para el análisis del consumo de combustible se trabajó con una base de datos suministrada por la empresa, la cual contiene los registros históricos de abastecimiento realizados en las estaciones de servicio. En una primera etapa, se realizó un proceso de carga, estandarización y limpieza de la información utilizando Python en Google Colab, con el fin de garantizar la coherencia y calidad de los datos antes de calcular los indicadores de consumo.

Inicialmente, se importó el archivo en formato CSV y se renombraron algunas columnas para facilitar su interpretación durante el análisis. Específicamente, las variables `Total_Galones_Aprobados`, `kilometraje`, `Total_Km_Recorridos` y `Total_Aprobaciones` fueron transformadas a nombres más descriptivos: `galones`, `km_odometro`, `km_recorrido` y `registros`. De esta manera, se homogenizó la estructura del dataset y se estandarizó la nomenclatura utilizada en los diferentes cálculos posteriores.

Posteriormente, se realizó la conversión de tipos de datos para garantizar su consistencia. La columna `fecha` fue transformada al tipo `datetime`, permitiendo su uso en análisis temporales como tendencias, frecuencia de abastecimientos y patrones por mes o por año. Las columnas `galones` y `km_recorrido` fueron convertidas a variables numéricas, corrigiendo posibles errores de digitación o formateo que pudieran afectar los cálculos de consumo y eficiencia operativa.

Este proceso de depuración inicial permitió obtener un conjunto de datos limpio, estructurado y listo para el cálculo de indicadores clave como consumo promedio, variación por vehículo, detección de outliers y análisis descriptivo general. A partir de esta base depurada se desarrollan los análisis presentados en las siguientes secciones.

Caracterización General de los Datos de Abastecimiento

Figura 1

Descripción de las Columnas de la Base de Datos

	fecha	id_conductor	id_vehiculo	km_odometro	km_anterior	km_recorrido	Rendimiento	galones	registros
count	2382	2382.000000	2382.000000	2382.000000	2306.000000	2306.000000	2306.000000	2382.000000	2382.0
mean	2025-11-13 15:27:57.581864192	359.328715	382.612930	183895.724601	183855.131830	332.483955	40.456612	8.916037	1.0
min	2025-11-01 00:00:00	161.000000	180.000000	8.000000	8.000000	1.000000	0.100000	1.000000	1.0
25%	2025-11-07 00:00:00	277.250000	275.000000	77746.000000	77733.000000	193.000000	25.055558	6.000000	1.0
50%	2025-11-13 00:00:00	369.000000	382.000000	171060.500000	171060.500000	276.500000	34.200000	8.000000	1.0
75%	2025-11-20 00:00:00	436.750000	491.000000	238152.500000	237278.000000	398.750000	46.137236	10.000000	1.0
max	2025-11-26 00:00:00	499.000000	646.000000	682820.000000	682638.000000	1469.000000	293.800000	40.000000	1.0
std	NaN	85.908441	127.446974	139063.806530	139032.243715	219.788478	25.908791	4.411013	0.0

El conjunto de datos analizado está compuesto por 2.382 registros de abastecimiento correspondientes al mes de noviembre de 2025. Estos registros contienen información relacionada con la fecha de tanqueo, el conductor asignado, el vehículo abastecido, los valores del odómetro, los kilómetros recorridos desde el último abastecimiento y el volumen de combustible suministrado. La caracterización estadística permite obtener una primera aproximación al comportamiento general del consumo en la flota de Transpormax.

En términos temporales, los abastecimientos se distribuyen entre el 1 y el 26 de noviembre de 2025, con una mediana ubicada el 13 de noviembre, lo que evidencia una distribución homogénea de tanques a lo largo del mes. La presencia de registros diarios de abastecimiento indica una flota con operación continua y de alta rotación.

Respecto al componente operativo, los vehículos presentan una media de 332,48 km recorridos entre abastecimientos, aunque con una elevada variabilidad (desviación estándar de 219,78 km). El rango va desde recorridos mínimos de 1 km hasta máximos de 1.469 km, lo que

sugiere diferencias significativas entre vehículos según sus rutas, frecuencia de operación o programación diaria.

El valor del odómetro muestra una gran dispersión, con un mínimo de 8 km y un máximo de 682.820 km, reflejando la coexistencia de vehículos con niveles de desgaste muy diferentes dentro de la flota. Esta variabilidad es coherente con flotas mixtas donde coexisten vehículos antiguos y modernos.

En cuanto al consumo de combustible, los abastecimientos presentan una mediana de 8 galones, un valor promedio de 8,91 galones, y un máximo registrado de 40 galones. Esto sugiere que la mayoría de los tanqueos corresponden a recargas parciales frecuentes, mientras que los valores más altos podrían estar asociados a abastecimientos completos o a correcciones operativas específicas. El hecho de que todos los registros tengan el valor “1” en la columna registros confirma que cada fila en la base representa un tanqueo individual validado.

El rendimiento promedio reportado es de 40,45 km por galón, aunque el rango de variación (de 0,1 a 293,8 km/galón) evidencia la presencia de valores atípicos o inconsistentes que requieren verificación en análisis posteriores. Estos outliers pueden deberse a errores de digitación, fallas de registro en campo o situaciones operativas inusuales.

En conjunto, esta caracterización refleja una base de datos robusta, con suficiente variación y niveles operativos diversos para permitir análisis posteriores de consumo, detección de anomalías y evaluación del desempeño por vehículo y conductor.

Construcción del Indicador de Rendimiento

Con el fin de evaluar la eficiencia en el uso del combustible por parte de los vehículos de la empresa, se construyó el indicador `rend_km_gal`, definido como la relación entre los kilómetros recorridos desde el último abastecimiento y la cantidad de galones suministrados en

dicho tanqueo. Este indicador permite estimar cuántos kilómetros recorre un vehículo por cada galón consumido, proporcionando una medida directa de su desempeño operativo.

El cálculo se realizó mediante la siguiente expresión:

$$rend_km_gal = \frac{Km_recorrido}{galones}$$

Para garantizar la consistencia del indicador, se manejaron adecuadamente los casos en los que la división podía generar valores indefinidos o infinitos. En particular, se reemplazaron los valores infinitos ($\pm\infty$) por NaN, evitando que registros erróneos o inconsistentes afectaran los cálculos agregados posteriores, como promedios, tendencias o análisis de outliers.

La creación del indicador `rend_km_gal` resulta esencial para identificar patrones de consumo ineficientes, posibles pérdidas de combustible y comportamientos atípicos tanto a nivel de vehículo como a nivel de conductor. En las secciones siguientes se presentan los análisis descriptivos y comparativos derivados de esta métrica.

Análisis del Rendimiento del Combustible

El indicador `rend_km_gal`, que representa los kilómetros recorridos por cada galón abastecido, permitió evaluar de manera directa la eficiencia operativa de los vehículos durante el mes analizado. A partir de los 2.306 registros válidos, se observa un rendimiento promedio de 40,46 km/galón, lo cual constituye el valor de referencia general del comportamiento de la flota.

Los valores mínimos y máximos del indicador muestran un rango bastante amplio, desde 0,1 km/galón hasta 293,8 km/galón. Mientras que los valores cercanos al mínimo pueden indicar inconsistencias de registro, fallas operativas o situaciones atípicas como desplazamientos muy cortos entre abastecimientos, los valores extremadamente altos suelen asociarse a datos incorrectos o a abastecimientos que no corresponden con el consumo real del trayecto. Esta

dispersión se refleja también en la desviación estándar de 25,98 km/galón, evidencia de una flota con comportamientos muy heterogéneos.

La mediana del rendimiento se ubicó en 34,0 km/galón, lo que significa que la mitad de los vehículos presentan rendimientos iguales o inferiores a este valor, indicando que la distribución se encuentra sesgada hacia algunos registros con rendimientos inusualmente altos. El rango intercuartílico (25% = 25,33 km/gal, 75% = 46,25 km/gal) representa el comportamiento típico de la flota y es útil para identificar vehículos con desempeño anómalo fuera de este intervalo.

Este análisis inicial indica que, aunque la mayoría de los vehículos operan dentro de un rango razonable de eficiencia, existen valores extremos que deben ser investigados con mayor detalle para determinar si se deben a errores en el registro de datos, a eventos operativos excepcionales o a posibles irregularidades en el consumo de combustible. En secciones posteriores, estos outliers se explorarán para evaluar su impacto en la eficiencia global y su posible relación con patrones de consumo anómalos.

Sistema de Alertas Operativas Basado en Reglas de Negocio

Con el fin de identificar comportamientos atípicos en el consumo de combustible y detectar posibles inconsistencias o irregularidades en los datos, se implementó un sistema de alertas basado en reglas de negocio. Estas alertas permiten resaltar registros que pueden requerir validación adicional, ya sea por errores en el reporte, comportamientos operativos extremos o casos que podrían sugerir un uso ineficiente o indebido del combustible.

Figura 2

Flags Creadas para la Limpieza de la Base de Datos

```
# Flag 1: Galones imposibles (>40 galones)
df["flag_galones_altos"] = df["galones"] > 40

# Flag 2: Km recorridos irreales (>350 km/día)
df["flag_km_irreal"] = df["km_recorrido"] > 800

# Flag 3: Rendimiento absurdo (<5 km/gal o >80 km/gal)
df["flag_rendimiento_raro"] = (df["rend_km_gal"] < 5) | (df["rend_km_gal"] > 80)

# Flag 4: Galones = 0 pero hay km recorridos (caso sospechoso)
df["flag_km_sin_galones"] = (df["galones"] == 0) & (df["km_recorrido"] > 0)

# Flag 5: km_recorrido = 0 pero hay galones (tanques sin ruta)
df["flag_galones_sin_km"] = (df["km_recorrido"] == 0) & (df["galones"] > 0)

# Flag 6: Registro duplicado (mismo vehículo y misma fecha)
df["flag_duplicado"] = df.duplicated(subset=["placa", "fecha"], keep=False)

# Flag 7: Registros con valores faltantes en columnas clave
df["flag_missing"] = df[["galones", "km_recorrido", "rend_km_gal"]].isnull().any(axis=1)
```

El sistema se compone de siete banderas (flags) diseñadas para capturar diferentes tipos de anomalías:

Galones Abastecidos Superiores al Límite Operativo

Se marcó una alerta para todos los registros con más de 40 galones abastecidos. Este umbral representa un volumen inusualmente alto para los vehículos de la empresa, lo cual puede deberse a errores de digitación, tanques llenados fuera de la estación autorizada o posibles irregularidades en el reporte.

Flag generada: flag_galones_altos.

Kilómetros Recorridos Irreales en un solo Ciclo Operativo

Se marcaron todos los casos donde se reportaron más de 800 km recorridos entre abastecimientos. Valores por encima de este límite son improbables dadas las rutas y distancias habituales de los vehículos de Transpormax.

Flag generada: flag_km_irreal.

Rendimiento Atípico o Sospechoso

Se identificaron dos tipos de comportamiento anómalos:

- Rendimientos extremadamente bajos (< 5 km/gal): pueden indicar fugas, sobreconsumo, errores en el registro del odómetro o situaciones operativas anormales.
- Rendimientos extremadamente altos (> 80 km/gal): suelen asociarse a fallos de registro o a inconsistencias entre el kilometraje y los galones ingresados.

Ambos casos fueron marcados bajo el indicador `flag_rendimiento_raro`.

Kilometraje sin Consumo de Combustible

Se marcaron registros donde se reportan kilómetros recorridos pero el abastecimiento aparece con 0 galones, lo cual no es posible físicamente y sugiere errores operativos o inconsistencia en la captura de datos.

Flag: `flag_km_sin_galones`.

Consumo de Combustible sin Desplazamiento Registrado

Se generó una alerta para los casos donde se reportan galones abastecidos pero el kilometraje recorrido es 0 km, lo cual podría indicar tanqueos injustificados o registros incompletos.

Flag: `flag_galones_sin_km`.

Registros Duplicados por Vehículo y Fecha

Se marcaron los registros donde un mismo vehículo aparece con dos abastecimientos en la misma fecha exacta. Este comportamiento puede ser válido en algunos casos, pero en la mayoría de operaciones representa una inconsistencia que requiere revisión.

Flag: `flag_duplicado`.

Registros con Valores Faltantes en Variables Críticas

Se detectaron casos donde una o más de las variables esenciales para el análisis (galones, km_recorrido, rend_km_gal) presentan valores faltantes. Estos registros fueron marcados con flag_missing para su posterior depuración.

Flag: flag_missing.

Resultados del Sistema de Alertas

Figura 3

Distribución de los Registros Inconsistentes

```
def motivo(row):
    if row["galones"] > 40:
        return "Galones imposibles (>40)"
    if row["km_recorrido"] > 350:
        return "Km recorrido irreal (>350)"
    if (row["rend_km_gal"] < 5) or (row["rend_km_gal"] > 80):
        return "Rendimiento absurdo"
    return "Otro"

df_eliminados["motivo Eliminacion"] = df_eliminados.apply(motivo, axis=1)

df_eliminados["motivo Eliminacion"].value_counts()
```

	count
motivo Eliminacion	
Otro	273
Km recorrido irreal (>350)	225
Rendimiento absurdo	39

dtype: int64

Una vez aplicadas las reglas de negocio para la detección de anomalías, se identificaron 537 registros inconsistentes dentro del conjunto de datos original. Estas observaciones fueron clasificadas según el tipo de alerta generada, lo que permitió determinar la naturaleza y magnitud de las inconsistencias presentes en la información de consumo.

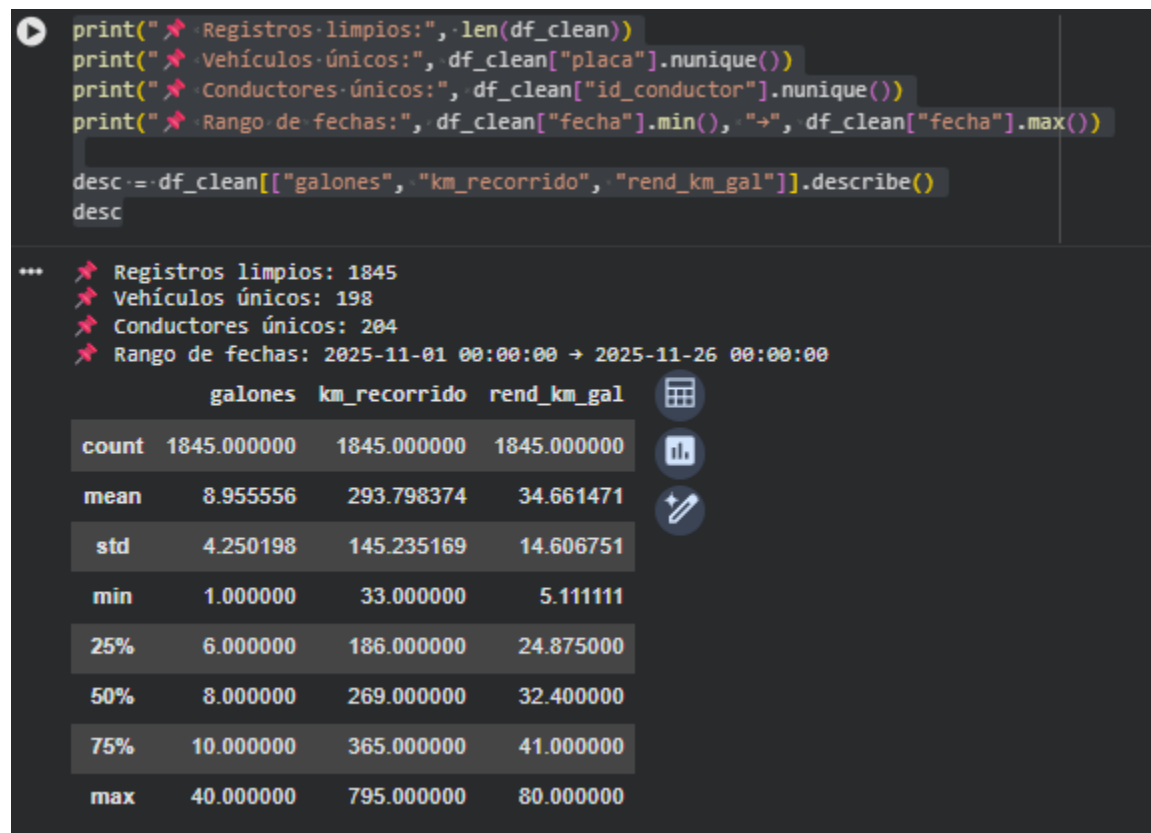
Los resultados muestran que la categoría más frecuente corresponde a “Otro”, con 273 registros, que agrupa inconsistencias generales como duplicados, valores faltantes o combinaciones atípicas entre variables operativas. El segundo grupo más representativo corresponde a los casos de “Kilómetros recorridos irreales (>350 km)”, con 225 registros, los cuales exceden significativamente los patrones esperados para la operación diaria de los vehículos de Transpormax y sugieren errores en el registro del odómetro o inconsistencias en el reporte de abastecimientos. Finalmente, se identificaron 39 registros clasificados como “Rendimiento absurdo”, caracterizados por valores menores a 5 km/galón o superiores a 80 km/galón, considerados físicamente improbables y potencialmente asociados a errores de captura o eventos de sobreconsumo no justificados.

En conjunto, estos hallazgos evidencian que una proporción importante de la información original presenta inconsistencias que afectan directamente la confiabilidad de los análisis operativos. La identificación temprana de estas anomalías permite depurar el conjunto de datos y generar una base más precisa para el análisis de eficiencia y comportamiento del consumo de combustible en la flota.

Análisis de la Data Limpia

Figura 4

Descripción de la Base de Datos Limpia



Luego de aplicar el sistema de alertas y depurar todas las observaciones inconsistentes, el conjunto de datos limpio quedó conformado por 1.845 registros válidos, correspondientes a 198 vehículos y 204 conductores, con un rango temporal que abarca del 1 al 26 de noviembre de 2025. En esta base depurada, los valores muestran un comportamiento mucho más coherente con la operación real de la flota. El volumen de combustible abastecido presenta una media de 8,96 galones, mientras que los kilómetros recorridos entre tanques se estabilizan alrededor de 293,8 km, con una variabilidad moderada que refleja diferentes niveles de uso operativo. El rendimiento promedio alcanzó 34,66 km por galón, dentro de un rango plausible entre 5,11 y 80

km/galón, lo que confirma la eliminación de valores extremos y registros improbables presentes en la base original. Además, los cuartiles muestran una distribución consistente, con el 50% de los registros situados entre 24,87 y 41 km/galón, intervalo que representa el comportamiento típico de la flota. En conjunto, estos resultados indican que la depuración mejoró significativamente la calidad del dato y permitió obtener una visión más precisa y confiable del consumo real de combustible.

Análisis de Correlación Entre Variables Operativas

Figura 5

Matriz de Correlación entre Variables Operativas



La matriz de correlación calculada a partir de la data limpia permite comprender la relación estadística entre los principales indicadores de consumo. Se observa una correlación positiva moderada entre los kilómetros recorridos y los galones abastecidos ($r = 0.492$), lo cual es coherente con la dinámica operativa: a mayor distancia recorrida, mayor es la cantidad de combustible requerida. Asimismo, se identificó una correlación positiva más fuerte entre los kilómetros recorridos y el rendimiento del vehículo ($r = 0.632$), indicando que trayectos más largos tienden a producir valores de rendimiento más estables y eficientes, probablemente debido a una menor frecuencia de detenciones y a condiciones de operación más constantes. Por otro lado, la correlación negativa entre galones abastecidos y rendimiento ($r = -0.268$) sugiere que los

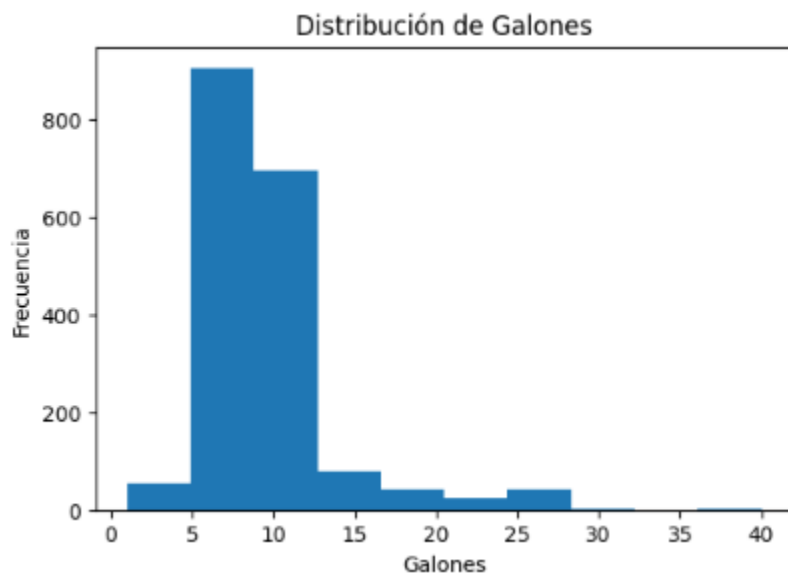
abastecimientos de mayor volumen están asociados con una eficiencia ligeramente menor, lo que podría deberse a variaciones en el tipo de recorrido, condiciones del tráfico o estilos de conducción. En conjunto, estas correlaciones reflejan patrones coherentes y contribuyen a comprender los factores que influyen en el desempeño del combustible dentro de la flota.

Análisis Descriptivo con Apoyo Gráfico

El análisis descriptivo de la data limpia se complementó mediante visualizaciones que permiten comprender el comportamiento de las variables operativas clave: galones abastecidos, kilómetros recorridos y rendimiento del combustible. Estas representaciones gráficas facilitan la identificación de patrones centrales, dispersión, sesgos y presencia de valores atípicos dentro del conjunto de datos.

Figura 6

Histogramas de Distribución de Galones

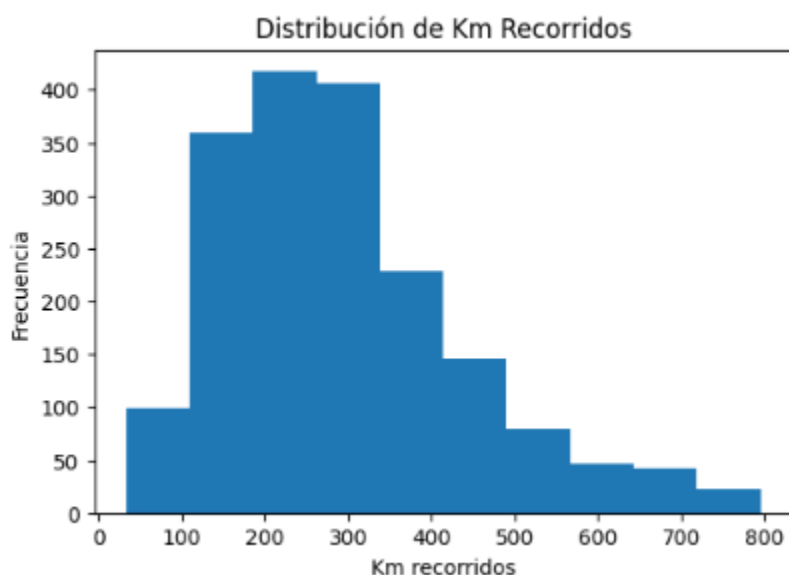


En primer lugar, el histograma de galones abastecidos muestra una distribución claramente concentrada entre 6 y 10 galones, lo que confirma la predominancia de recargas

parciales como parte del patrón operativo habitual de la flota. Si bien existen algunos casos aislados que alcanzan los 30 o 40 galones, estos corresponden a escenarios menos frecuentes y se ubican como posibles abastecimientos completos o situaciones operativas específicas. Esta concentración también se aprecia en el boxplot, donde se observa un rango intercuartílico estrecho y la presencia de valores atípicos hacia el extremo superior.

Figura 7

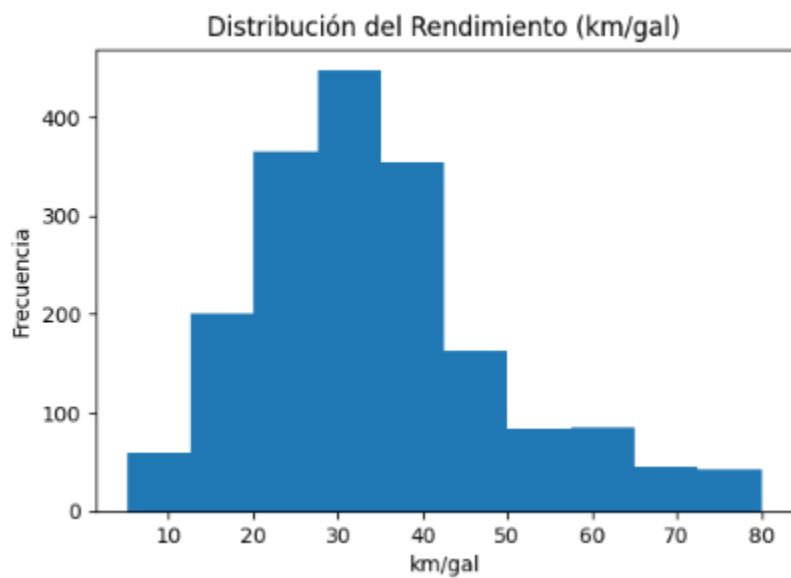
Histogramas de Distribución de Kilómetros Recorridos



El histograma de kilómetros recorridos evidencia que la mayoría de los trayectos se encuentran entre 150 y 350 km por ciclo operativo, lo cual coincide con los recorridos diarios estimados para los vehículos utilizados en el transporte de pacientes. A medida que los valores se alejan del centro de la distribución, la frecuencia disminuye de forma progresiva, hasta alcanzar registros máximos cercanos a 800 km, que aunque menos comunes, todavía se mantienen dentro de los límites plausibles establecidos tras la depuración.

Figura 8

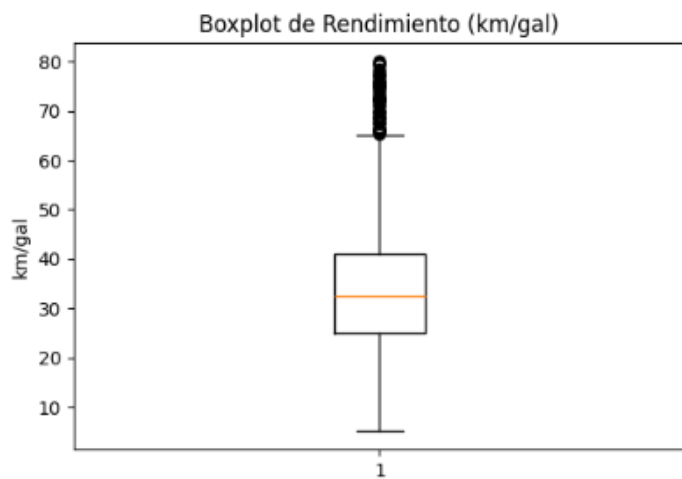
Histogramas de Distribución de Rendimientos



Respecto al rendimiento (km/gal), se observa una distribución centrada entre 25 y 45 km/gal, con una concentración destacada alrededor de los 32 km/gal, que coincide con la mediana calculada en el análisis estadístico.

Figura 9

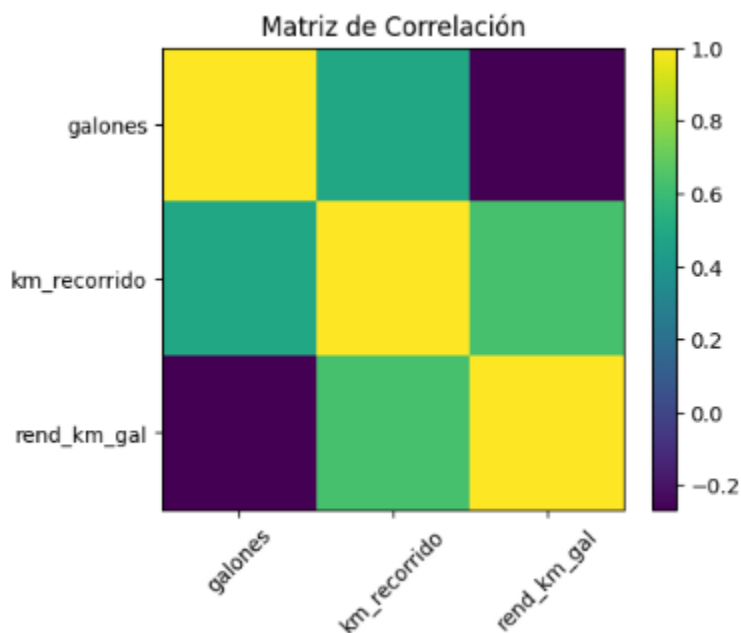
Boxplot Rendimientos



El boxplot confirma este patrón, mostrando un comportamiento relativamente estable, pero con presencia de algunos valores altos aislados que, aunque ya no corresponden a anomalías extremas, reflejan variaciones operativas derivadas del tipo de ruta, el estado del tráfico o el estilo de conducción de cada operador.

Figura 10

Matriz de Correlación



Finalmente, la matriz de correlación visual confirma los patrones identificados previamente: una relación positiva entre kilómetros recorridos y rendimiento, y una asociación negativa entre galones abastecidos y eficiencia del vehículo. Esta representación gráfica refuerza la coherencia interna del dataset depurado y valida la calidad del dato para los análisis posteriores de optimización y detección de anomalías.

En conjunto, los gráficos descriptivos permiten visualizar de manera clara la estructura del consumo en la flota de Transpormax, proporcionando evidencia visual que respalda los

hallazgos estadísticos y facilitando la interpretación operativa del comportamiento de los vehículos.

Análisis del Consumo y Rendimiento por Conductor

Figura 11

Estadísticas por Conductor

Estadísticas por conductor:								
id_conductor	galones_promedio	galones_totales	registros	km_promedio	km_totales	rend_promedio	rend_min	rend_max
161	9.000000	9	1	366.000000	366.0	40.666667	40.666667	40.666667
175	10.500000	21	2	324.000000	648.0	30.454545	22.000000	38.909091
206	21.500000	43	2	364.000000	728.0	16.361991	13.647059	19.076923
208	18.000000	36	2	633.000000	1266.0	44.894545	19.880000	69.909091
209	8.666667	52	6	333.833333	2003.0	37.614352	16.750000	63.700000

El análisis por conductor permite identificar diferencias operativas significativas en la forma en que se gestiona el combustible dentro de la flota. A partir de los registros limpios, se calcularon estadísticas por operador, incluyendo el consumo promedio, el rendimiento, el número de abastecimientos y la distancia recorrida. Este enfoque facilita la detección de patrones tanto de eficiencia como de posible uso inadecuado del combustible.

En términos generales, se observó una variabilidad considerable entre conductores, reflejada en diferencias notables en los promedios de galones consumidos, distancias recorridas y rendimiento del combustible. Algunos conductores presentan comportamientos altamente eficientes, mientras que otros muestran patrones que requieren seguimiento operativo.

Figura 12*Estadísticas por Conductor Alto Rendimiento*

Mejores conductores (rendimiento más alto):								
id_conductor	galones_promedio	galones_totales	registros	km_promedio	km_totales	rend_promedio	rend_min	rend_max
395	10.000000	10	1	795.000000	795.0	79.500000	79.500000	79.500000
303	5.000000	5	1	368.000000	368.0	73.600000	73.600000	73.600000
455	10.000000	10	1	682.000000	682.0	68.200000	68.200000	68.200000
428	9.000000	18	2	566.000000	1132.0	62.888889	46.555556	79.222222
460	7.000000	7	1	409.000000	409.0	58.428571	58.428571	58.428571
300	5.666667	17	3	311.333333	934.0	55.733333	44.000000	74.200000
305	4.833333	29	6	262.500000	1575.0	54.125000	37.200000	76.800000
404	12.000000	36	3	636.000000	1908.0	53.000000	33.250000	64.000000
356	6.800000	68	10	362.400000	3624.0	52.857143	38.000000	79.857143
465	8.666667	26	3	497.666667	1493.0	52.255556	18.666667	69.600000

Los conductores con mayor rendimiento promedio destacan por recorrer más kilómetros con cada galón de combustible. Entre ellos se encuentran los operadores con ID 395, 303, 455, 428 y 460, quienes alcanzan rendimientos entre 58 y 79 km/galón, ubicándose muy por encima del promedio general de la flota (34,66 km/galón). Estos valores altos pueden estar asociados a condiciones favorables de ruta o estilos de conducción optimizados, aunque también deben evaluarse para descartar posibles inconsistencias en los registros.

Figura 13*Estadísticas por Conductor Bajo Rendimiento*

▲ Peores conductores (rendimiento más bajo):

id_conductor	galones_promedio	galones_totales	registros	km_promedio	km_totales	rend_promedio	rend_min	rend_max
329	17.000000	17	1	193.000000	193.0	11.352941	11.352941	11.352941
411	11.800000	177	15	172.400000	2586.0	14.536303	8.058824	23.400000
394	25.000000	300	12	393.083333	4717.0	15.723333	9.640000	25.240000
391	24.272727	267	11	386.000000	4246.0	15.801558	9.080000	28.360000
206	21.500000	43	2	364.000000	728.0	16.361991	13.647059	19.076923
406	28.555556	257	9	466.555556	4199.0	16.438742	11.571429	22.857143
387	22.000000	22	1	368.000000	368.0	16.727273	16.727273	16.727273
462	8.916667	107	12	152.833333	1834.0	16.970370	12.500000	30.555556
443	17.916667	215	12	319.083333	3829.0	17.935020	13.333333	28.888889
396	19.923077	259	13	383.769231	4989.0	18.970202	9.636364	31.318182

Por otro lado, los conductores con menor rendimiento promedio incluyen los IDs 329, 411, 394, 391 y 206, con valores entre 11 y 17 km/galón, muy por debajo del comportamiento típico. Estos casos son críticos, ya que un rendimiento bajo puede indicar sobreconsumo, fallas mecánicas, rutas ineficientes o incluso eventos de pérdida de combustible. Además, algunos de estos conductores presentan un número elevado de abastecimientos, lo que amplifica el impacto operativo de su ineficiencia.

Al analizar los conductores con mayor número de registros, destacan los IDs 368, 273, 262, 290, entre otros, quienes realizaron entre 16 y 20 abastecimientos durante el periodo. Estos conductores representan una parte importante del comportamiento global de la flota, por lo que su rendimiento promedio influye de manera directa en los indicadores generales. En la mayoría de estos casos, los rendimientos se ubican en rangos medios entre 25 y 40 km/galón, lo cual coincide con el patrón típico observado en el análisis descriptivo.

Finalmente, se identificó un grupo de conductores outliers tanto por rendimiento extremadamente alto como extremadamente bajo. Este grupo incluye operadores con valores superiores a 70 km/galón y otros con rendimientos menores a 15 km/galón, lo que amerita una revisión detallada de sus rutas, condiciones operativas y registros de odómetro. La presencia de estos conductores fuera del rango típico constituye una señal temprana que puede ayudar a la empresa a focalizar auditorías, capacitaciones y ajustes en la operación.

En conjunto, el análisis por conductor evidencia que el desempeño individual tiene un impacto directo sobre el consumo de combustible y que la identificación temprana de patrones extremos permite mejorar la eficiencia operativa de la flota.

Detección de Anomalías Mediante Isolation Forest

Además del sistema de alertas basado en reglas de negocio, se implementó un método estadístico de detección de anomalías mediante el algoritmo Isolation Forest, con el fin de identificar patrones atípicos de consumo que no son evidentes a simple vista. Esta técnica es especialmente útil para detectar registros inusuales dentro de cada grupo de vehículos que comparten características similares.

Para garantizar un análisis más preciso, el modelo se entrenó de manera independiente para cada marca y modelo de vehículo, evitando comparaciones entre vehículos con consumos y rendimientos naturalmente diferentes. En cada grupo se utilizaron las variables clave: galones abastecidos, kilómetros recorridos y rendimiento (km/galón). Se excluyeron los grupos con menos de 20 registros para evitar modelos poco representativos.

El algoritmo se configuró con una tasa de contaminación del 5%, lo que implica que el modelo considera aproximadamente el 5% de los registros como potencialmente anómalos según

su comportamiento estadístico. Tras aplicar esta metodología sobre el conjunto depurado, se identificaron 90 registros anómalos, frente a 1.755 registros considerados normales.

Estos valores representan un 5,1% del total de observaciones limpias, lo cual se alinea con los parámetros establecidos. Las anomalías detectadas incluyen combinaciones inusuales entre galones, recorrido y rendimiento, tales como:

- Consumos atípicamente altos o bajos para el mismo modelo de vehículo.
- Recorridos demasiado cortos o largos en comparación con el patrón normal de la unidad.
- Rendimientos inconsistentes con los valores típicos del grupo.
- Puntos extremos que podrían indicar errores de registro, condiciones operativas anormales o eventos de sobreconsumo.

La identificación de estos 90 registros atípicos es especialmente relevante porque señala casos que no fueron capturados por las reglas tradicionales del sistema de alertas, demostrando el valor agregado de los métodos basados en machine learning para el control operativo. Esta combinación de enfoques reglas explícitas y algoritmos estadísticos ofrece una visión más completa y robusta de las posibles irregularidades presentes en los datos de consumo.

Conductores con Mayor Cantidad de Anomalías

Al analizar los registros identificados como anómalos por el modelo de Isolation Forest, se realizó un conteo por conductor con el fin de determinar qué operadores presentan comportamientos atípicos con mayor frecuencia. En total, se identificaron 69 conductores con al menos una anomalía detectada.

Los casos más relevantes corresponden a aquellos conductores que acumulan más eventos anómalos. Entre ellos destacan:

Conductor 362, con 5 anomalías, siendo el operador con el mayor número de registros atípicos. Este patrón sugiere la necesidad de revisar su operación, condiciones de ruta o consistencia en el registro del consumo.

Conductores 477 y 242, cada uno con 3 anomalías, lo que indica comportamientos irregulares recurrentes.

Conductores 258 y 235, con 2 anomalías cada uno, que aun sin ser los más críticos, podrían estar mostrando desviaciones respecto al patrón normal de consumo de su vehículo.

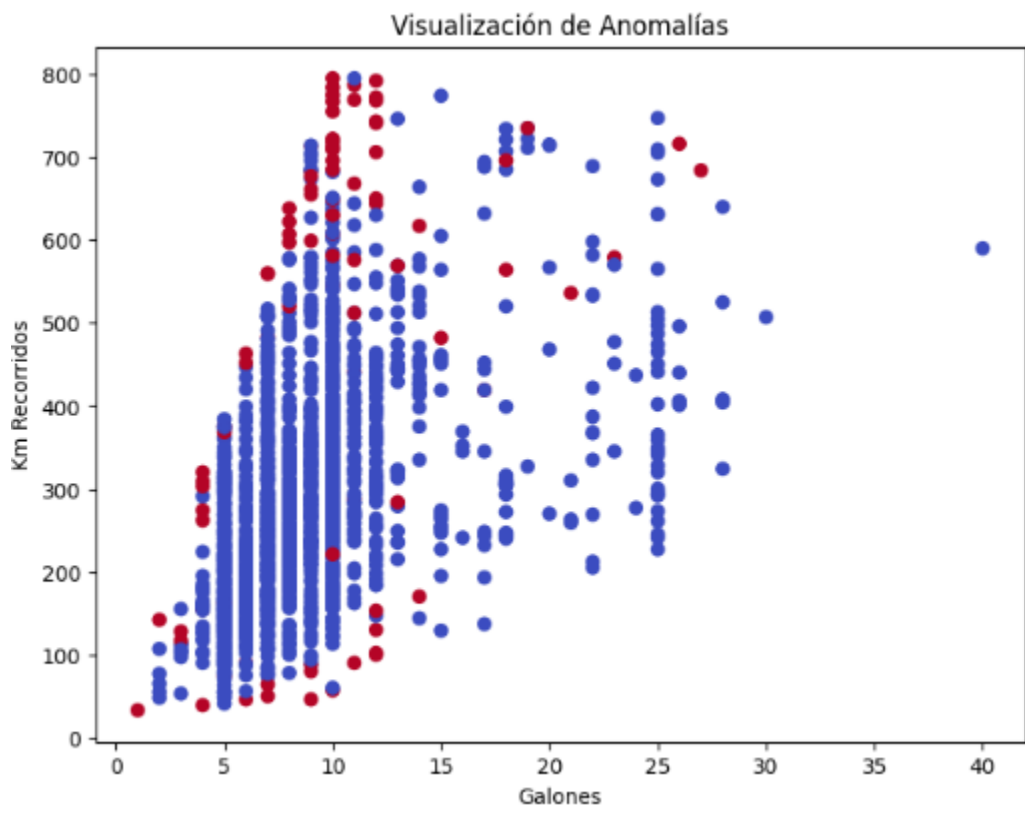
El resto de los conductores (un total de 64) presentan solo una anomalía cada uno, lo cual puede corresponder a eventos aislados, circunstancias operativas particulares o posibles errores puntuales en los datos.

Este análisis permite priorizar acciones de control, capacitación o verificación mecánica, enfocándose primero en los conductores con mayor número de anomalías detectadas, ya que estos casos representan un mayor riesgo de ineficiencia o potencial pérdida de combustible. Además, al correlacionar estas anomalías con información operativa (rutas, horarios, tipo de vehículo, etc.), se pueden obtener insights adicionales para la gestión y optimización del consumo en la flota.

Visualización de Anomalías

Figura 14

Grafico de Dispersión de Anomalías



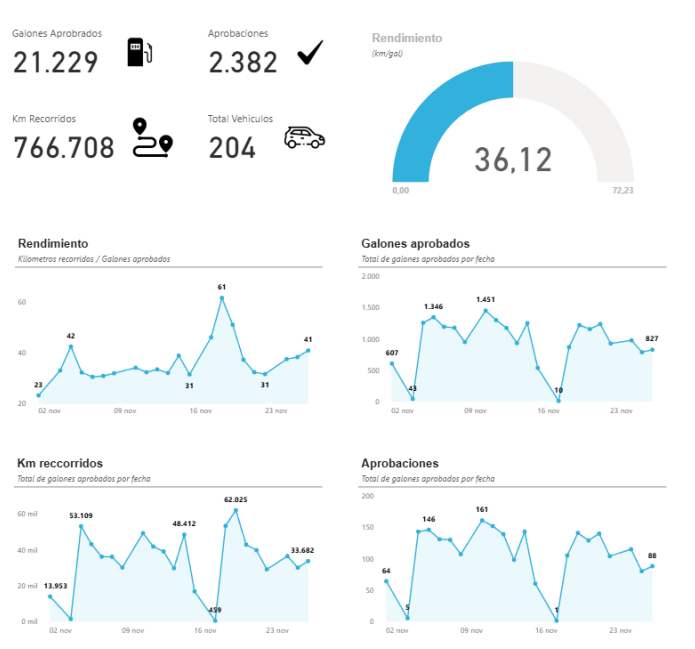
La visualización muestra la relación entre galones cargados y kilómetros recorridos para todos los registros limpios del dataset, resaltando en color rojo aquellos puntos que el modelo Isolation Forest clasificó como anómalos. En general, la nube de puntos azules evidencia un patrón consistente: a mayor cantidad de galones, mayor distancia recorrida, lo que refleja un comportamiento operativo normal. Sin embargo, los puntos rojos evidencian desviaciones significativas de este patrón. Entre las anomalías destacan dos tipos: registros con pocos galones pero muchos kilómetros, lo cual puede indicar inconsistencias de registro o posible transporte sin declarar combustible; y registros con muchos galones pero pocos kilómetros, que suelen

asociarse a ineficiencia, posibles fugas o fallos en el control del consumo. La concentración de anomalías en las zonas de bajo rendimiento sugiere que estas desviaciones deben ser revisadas con prioridad, ya que representan un riesgo potencial de pérdidas de combustible o errores sistemáticos en la operación.

Dashboard en Power BI

Figura 15

Vista General del Dashboard en Power Bi



Se elaboró un dashboard interactivo en Power BI conectado directamente a las bases de datos operativas de la empresa mediante SQL, lo que permite una actualización automática y en tiempo real de la información de combustible y recorrido. El tablero integra visualizaciones clave que facilitan el monitoreo continuo del desempeño de la flota, incluyendo:

- Rendimiento histórico del combustible, con tendencias por día, semana y mes.

- Rendimiento por vehículo, permitiendo identificar unidades con desempeño sobresaliente o con posibles problemas mecánicos o de consumo anormal.
- Rendimiento por conductor, lo que habilita la comparación entre operadores, la detección de patrones de conducción eficientes y la identificación de perfiles de riesgo.

Figura 16

Vista por Vehiculo Dashboard en Power Bi

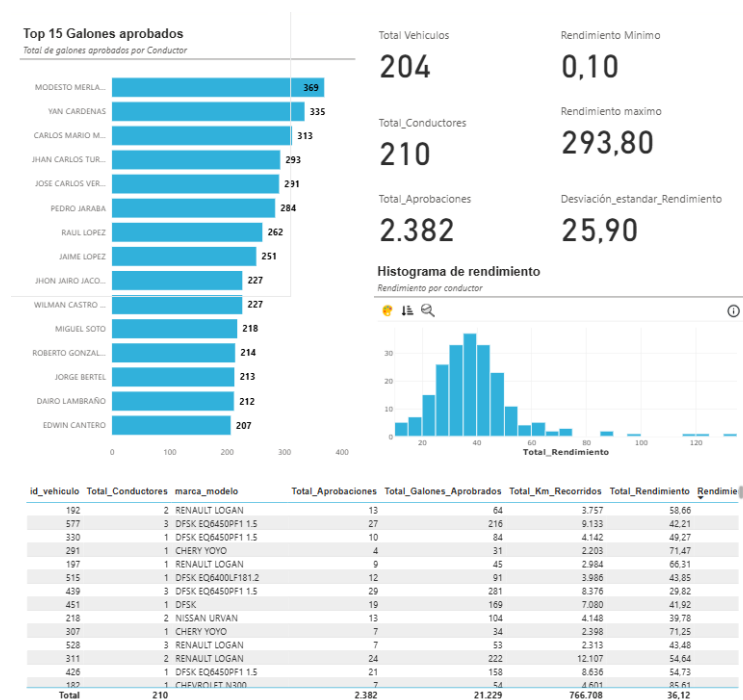
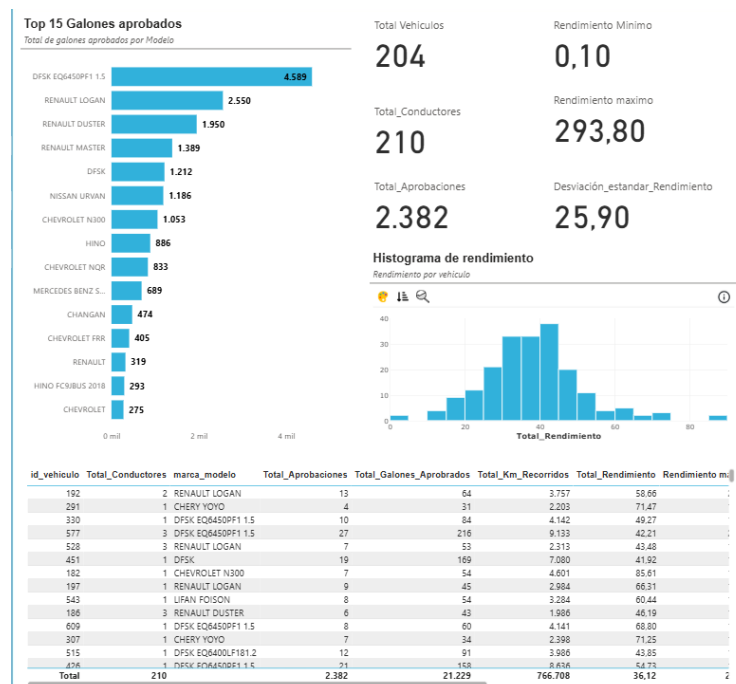


Figura 17

Vista por Conductor Dashboard en Power Bi



Este dashboard se convierte en una herramienta central de apoyo para la toma de decisiones, permitiendo detectar comportamientos atípicos, gestionar acciones correctivas oportunas y mejorar la eficiencia operativa de manera continua.

Conclusiones

La limpieza de datos representó un componente fundamental para garantizar la fiabilidad del análisis. La eliminación de registros con galones imposibles, recorridos irreales o rendimientos físicamente inconsistentes permitió depurar el conjunto inicial y trabajar únicamente con información verificada, lo que redujo significativamente el ruido estadístico y aumentó la calidad de las conclusiones. Este proceso derivó en un dataset de 1.845 registros limpios que sirvió como base sólida para el análisis posterior.

El comportamiento general del consumo evidencia una operación relativamente estable dentro de la flota, con promedios de 8,9 galones por tanqueo y recorridos que rondan los 294 km. El rendimiento promedio de 34,6 km/galón se ubica en un rango aceptable, aunque con variaciones importantes entre conductores y vehículos. Las correlaciones obtenidas revelan patrones operativos relevantes: recorridos más largos suelen asociarse con un mejor rendimiento, mientras que cargas de combustible más altas tienden a reducir la eficiencia, posiblemente por el peso adicional o diferencias en las condiciones de operación.

El análisis por conductor evidenció una dispersión significativa en los niveles de rendimiento. Algunos operadores mantienen desempeños consistentemente altos, alcanzando valores superiores a 70 km/galón, mientras que otros presentan rendimientos preocupantemente bajos, incluso por debajo de los 15 km/galón. Estas diferencias resaltan la influencia del estilo de conducción, la planificación de rutas y los hábitos operativos, abriendo una oportunidad clara para intervenciones dirigidas, procesos de retroalimentación y programas de capacitación.

La incorporación del modelo de detección de anomalías mediante Isolation Forest añadió una capa avanzada de control operativo. La identificación de 90 anomalías concentradas en un grupo reducido de conductores permite focalizar las revisiones e investigaciones internas sin

depender de revisiones manuales extensas. Este mecanismo automatizado constituye un insumo clave para un futuro sistema de alertas tempranas que ayude a la empresa a reaccionar rápidamente frente a comportamientos inusuales o potencialmente fraudulentos. Finalmente, el desarrollo del dashboard en Power BI, conectado directamente a las bases de datos SQL de la empresa, complementó el análisis estadístico y permitió visualizar el rendimiento histórico, por vehículo y por conductor, facilitando la toma de decisiones informadas. En conjunto, los resultados obtenidos demuestran que la empresa cuenta con oportunidades claras para mejorar la eficiencia operativa, reducir costos de combustible, estandarizar buenas prácticas y fortalecer los mecanismos de control. La integración entre análisis avanzado, visualización y modelos de detección ofrece una base estratégica para alcanzar el objetivo de reducir los costos asociados al combustible y mejorar la gestión de la flota de manera sostenible.

Recomendaciones

Los hallazgos del análisis sugieren la necesidad de implementar un programa de capacitación continua para los conductores, orientado a promover prácticas de conducción eficiente. Dado que se identificaron diferencias marcadas en el rendimiento entre operadores, estandarizar buenas prácticas podría reducir significativamente la variabilidad y mejorar el consumo promedio de la flota. Esta estrategia permitiría trasladar los hábitos de los conductores con mejor desempeño hacia aquellos que presentan rendimientos bajos o inestables.

Es recomendable establecer un sistema automatizado de alertas basado en reglas de negocio y modelos de detección de anomalías como los aplicados en este estudio. La integración de un módulo de alertas permitiría identificar en tiempo real comportamientos atípicos, como cargas de combustible anómalas, recorridos inconsistentes o rendimientos improbables. Esto no solo fortalecerá los mecanismos de control operativo, sino que reducirá el tiempo de reacción ante posibles casos de fraude o uso ineficiente del combustible.

Los resultados también evidencian la importancia de mejorar los procesos de registro y calidad de datos. Mantener registros consistentes en odómetro, galones y rutas permitirá disminuir la necesidad de depuraciones extensivas y garantizar análisis más precisos. La empresa puede implementar controles automáticos de validación en el punto de carga de combustible y en la captura del kilometraje, evitando errores humanos o manipulaciones involuntarias.

El análisis sugiere igualmente que la empresa debería avanzar hacia la integración de fuentes adicionales de información, como los datos GPS, lo que permitiría enriquecer el modelo actual con variables como tiempos de detención, patrones de ruta, velocidad promedio y comportamiento del motor. Esta integración permitiría realizar análisis más robustos y construir modelos predictivos más precisos sobre consumo, desgaste y anomalías operativas.

Finalmente, se recomienda implementar un esquema de seguimiento continuo mediante dashboards operativos en Power BI, basados en conexiones en tiempo real a las bases de datos corporativas. Estas visualizaciones permiten supervisar el comportamiento de la flota día a día, identificar tendencias, visualizar comparativos por vehículo y conductor, y evaluar el impacto de estrategias implementadas para mejorar la eficiencia. El uso sistemático de estos dashboards contribuirá directamente a la toma de decisiones informadas y al cumplimiento del objetivo corporativo de reducir costos y optimizar el uso del combustible.

Referencias Bibliográficas

- Asih, I. A., Purba, H. H., & Sitorus, T. M. (2020). Key Performance Indicators: A Systematic Literature Review. *Journal of Strategic Performance Management*, 8(4).
<https://www.researchgate.net/publication/344493860>
- Božic, K., & Dimovski, V. (2019). Business Intelligence Capabilities and Firm Performance: A Study in China. *Information & Management*, 56(6), 103207.
<https://doi.org/10.1016/j.im.2018.12.002>
- Castillo, J., Restrepo, Á., Tibaquirá, J., & Quirama, L. (2019). Estrategias de eficiencia energética en vehículos livianos del transporte por carretera en Colombia. *Revista UIS Ingenierías*, 18(3), 129–140. <https://doi.org/10.18273/revuin.v18n3-2019013>
- Fernández Romero, Y. A. (2020). Análisis de consumo de combustible de vehículos de carga al aplicar técnicas de conducción eficiente. Universidad Antonio Nariño. Recuperado de <https://repositorio.uan.edu.co/bitstream/123456789/2806/2/2020YennyAndreaFernandezRomero.pdf>
- Hernández-Sampieri, R., Mendoza, C., & Fernández, C. (2021). Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta (7.^a ed.). McGraw-Hill Education.
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning.
<https://arxiv.org/abs/2104.05314>
- Kaplan, R. S., & Norton, D. P. (1996). *The Balanced Scorecard: Translating Strategy into Action*. Harvard Business Press.
- Martínez, A., & Marulanda, M. (2019). Riesgos de la distribución minorista de combustibles líquidos de uso automotor en Colombia.
<https://www.repository.fedesarrollo.org.co/handle/11445/3896>

- Ortiz, R. (2023). Análisis de datos para la optimización de la gestión de flotas vehiculares en el sector del transporte de mercancía. Universidad de Antioquia. Recuperado de https://bibliotecadigital.udea.edu.co/dspace/bitstream/10495/35743/1/OrtizRonald_2023_AnalisisDatosFlotas.pdf
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.
- Rodríguez, J., & Méndez, L. (2021). Análisis integral del consumo de combustible en el transporte terrestre en Colombia. Universidad Nacional de Colombia. <https://repositorio.unal.edu.co/handle/unal/123456>
- Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.
- Topolšek, D., Čičiūnienė, K., & Cvahte Ojsteršek, T. (2018). Defining transport logistics: a literature review and practitioner opinion based approach. *Transport*, 33(5), 1196–1203. <https://doi.org/10.3846/transport.2018.6965>
- Universidad de los Andes. (2021). Predicción de fallas en la red colombiana de transporte de hidrocarburos: Un enfoque Data Driven basado en Machine Learning. Recuperado de <https://repositorio.uniandes.edu.co/entities/publication/7ffa0657-3809-47f7-a102-9bc45d4ae147>
- Universidad Piloto de Colombia. (2019). Sistema de seguimiento vehicular para rutas de transporte escolar. Recuperado de <https://repository.unipiloto.edu.co/bitstream/handle/20.500.12277/5737/00051024.pdf?sequence=1>

Apéndices

Apéndice A

Enlace de la Presentación

<https://www.youtube.com/watch?v=HebZnfCBB34>

Apéndice B

Autorización de la Empresa

v0.1

Barranquilla, 29 de Noviembre de 2025

Señora:

Leidy Tatiana Rosales Castillo
Coordinadora de abastecimiento de combustible

Asunto: Autorización para la ejecución del proyecto titulado: Diseño de un sistema de monitoreo y análisis del consumo de combustible basado en datos de abastecimiento en la empresa Transpormax.

Cordial saludo estimada Coordinadora de abastecimiento de combustible,

Como es de su conocimiento, actualmente me encuentro adelantando estudios de posgrado en la Especialización en Ciencia de datos y analítica ofertado por la Universidad Nacional Abierta y a Distancia "UNAD". Para finalizar mi proceso académico es mi objetivo desarrollar un trabajo de grado aplicado a Transpormax, de manera que pueda aportar mis conocimientos adquiridos y generar un impacto positivo en la empresa, relacionado con los temas de programación para análisis de datos , business intelligence y machine learning, motivo por el cual, muy comedidamente solicito su autorización y aprobación para la ejecución del proyecto titulado: Diseño de un sistema de monitoreo y análisis del consumo de combustible basado en datos de abastecimiento en la empresa Transpormax el cual se encuentra avalado por parte la Institución de educación superior "UNAD".

El proyecto en su objetivo general describe lo siguiente: Desarrollar un sistema de análisis de datos que permita optimizar el consumo de

V0.1

combustible y detectar comportamientos anómalos en la flota de vehículos, mediante la integración de fuentes como GPS y estaciones de servicio, técnicas de machine learning y la creación de un dashboard interactivo para la toma de decisiones operativas; al mismo tiempo será apoyado por los objetivos específicos:

1. Identificar los principales indicadores que permiten detectar anomalías en el consumo de combustible de los vehículos de la flota.
2. Integrar y transformar datos provenientes de GPS y estaciones de servicio para analizarlos de manera conjunta.
3. Aplicar técnicas de análisis de datos y algoritmos de machine learning para detectar patrones de consumo inusuales o ineficientes.
4. Diseñar y desarrollar un dashboard interactivo en Power BI que permita visualizar en tiempo real el desempeño de la flota y apoyar la toma de decisiones operativas

para obtener como resultado un alto impacto en la organización Transpormax.

De obtener esta autorización, se elaborará un acuerdo de confidencialidad para proteger la identidad de la empresa y sus activos de información; a su vez se destacan los siguientes procesos para ser garantes en la transparencia de la ejecución del proyecto:

- Se prohíbe la ejecución de cualquier tipo de pruebas de seguridad que no estén autorizadas expresamente por Transpormax (Aplica para la especialización en Seguridad Informática).
- La empresa Transpormax deberá establecer qué tipo de información es privada y cuál es pública para delimitar el acceso y aplicación de pruebas en la ejecución del proyecto.
- La solicitud de información al igual que ejecución de pruebas deben quedar por escrito y se genera un informe de resultados

V0.1

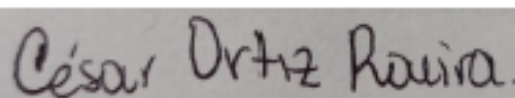
periódicamente el cual será compartido con el gerente de la organización o empresa.

- La persona autorizada siempre debe operar dentro de la ley 1273 de 2009 y de las demás regulaciones establecidas en la empresa.
- Respetar la privacidad de todos los individuos y mantener su privacidad en los reportes. Está prohibida divulgar información personal en tales reportes.

El resultado del proyecto se verá reflejado en un documento el cual será cargado al repositorio institucional de la Universidad Nacional Abierta y a Distancia "UNAD". El documento ampara la confidencialidad y anonimato de la empresa, estos aspectos se encuentran estipulados en el acuerdo de confidencialidad; agradezco el apoyo prestado en esta etapa de mi carrera profesional.

Firman en Barranquilla., a los (29) días del mes de (Noviembre) de 2025

Cordialmente,



Nombre del estudiante
Estudiante UNAD

V0.1

de la rosa

Nombre
Cargo

Apéndice C

Acuerdo de Confidencialidad

ACUERDO DE CONFIDENCIALIDAD ENTRE CESAR ORTIZ ROVIRA Y TRANSPORMAX SAS

Por la parte reveladora

Nombre: TRANSPORMAX SAS

Dirección: CONJ LA BONGUITA BL1 LC8.

Teléfono: _____

E-mail: _____

Por la parte receptora de la información

Nombre: César leonardo Ortiz Rovira

Dirección: cra 12 # 77ª - 138

Teléfono: _____

E-mail: _____

Identificación del proyecto

Entre los firmantes, identificados anteriormente, hemos convenido en celebrar el presente acuerdo de confidencialidad previa las siguientes **CONSIDERACIONES**

1. Que la información compartida en virtud del presente acuerdo pertenece a la *Transpormax*, y la misma es considerada sensible y de carácter restringido en su divulgación, manejo y utilización. Dicha información es compartida en virtud del desarrollo del proyecto aplicado: *Diseño de un sistema de monitoreo y análisis del consumo de combustible basado en datos de abastecimiento en la empresa Transpormax.*
2. Que la información de propiedad de *Transpormax* ha sido desarrollada u obtenida legalmente, como resultado de sus procesos, programas o proyectos y, en consecuencias abarca

documentos, datos, tecnología y/o material que considera único y confidencial, o que es objeto de protección a título de secreto industrial.

3. Que el presente acuerdo se realiza por un lado entre la parte receptora de la información como integrante del proyecto de *Diseño de un sistema de monitoreo y análisis del consumo de combustible basado en datos de abastecimiento en la empresa Transpormax*, César Leonardo Ortiz rovira que, para el presente caso actual como revelador, guarda y administrador de la información de propiedad de *Transpormax*.

En consecuencia, las partes se suscriben a las siguientes cláusulas:

Primera. Objeto: en virtud del presente acuerdo de confidencialidad, la parte receptora, se obliga a no divulgar directa, indirecta, próxima a remotamente, ni a través de ninguna otra persona o de sus subalternos o funcionarios, asesores o cualquier persona relacionada con ella, la información confidencial perteneciente al *Transpormax*, así como también a no utilizar dicha información en beneficio propio ni de terceros, sólo con fines estadísticos y de mejoramiento de la *Transpormax*.

Segunda. Definición de información confidencial: se entiende como Información Confidencial, para los efectos del presente acuerdo:

1. La información que no sea pública y sea conocida por la parte receptora con ocasión del proyecto en desarrollo.
2. Cualquier información societaria, técnica, jurídica, financiera, comercial, de mercado, estratégica, de productos, nuevas tecnologías, patentes, modelos de utilidad, diseños industriales, modelos de negocios, información del personal de la organización, información relacionada con la estructura organizacional y/o cualquier otra relacionada con el proyecto *Diseño de un sistema de monitoreo y análisis del consumo de combustible basado en datos de abastecimiento en la empresa Transpormax*, bien sea escrita, oral o visual, o en cualquier forma

tangible o no, incluidos los mensajes de datos (en la forma definida en la ley), de la cual, la parte receptora tenga conocimiento o a la que tenga acceso por cualquier medio o circunstancia en virtud de las reuniones sostenidas y/o documentos suministrados.

3. La que corresponda o deba considerarse como tal para garantizar el derecho constitucional a la intimidad, la honra y el buen nombre de las personas y deba guardarse la debida diligencia en su discreción y manejo en el desempeño de sus funciones.

Tercera. Origen de la información confidencial: provendrá de documentos suministrados en el desarrollo del proyecto y que tiene que ver con las creaciones del intelecto, a la naturaleza, medios, formas de distribución, comercialización de productos o de prestación de servicios, transmitida verbal, visual o materialmente, por escrito en los documentos, medios electrónicos, discos ópticos, microfilmes, películas, e-mail u otros elementos similares suministrados de manera tangible o intangible, independiente de su fuente o soporte y sin que requiera advertir su carácter confidencial.

Cuarta. Obligaciones de la parte receptora: Se considerará como parte receptora de la información confidencial a la persona que recibe la información, o que tenga acceso a ella. La parte receptora se obliga a:

1. Mantener la información confidencial segura, usarla solamente para los propósitos relacionados con él, en caso de ser solicitada, devolverla toda (incluyendo copias de esta) en el momento en que ya no requiera hacer uso de la misma o cuando termine la relación, caso en el cual, deberá entregar dicha información antes de la terminación de la vinculación.
2. Proteger la información confidencial, sea verbal, escrita, visual, tangible, intangible o que por cualquier otro medio reciba, siendo legítima poseedora de la misma *Transpormax*, restringiendo su uso exclusivamente a las personas que tengan absoluta necesidad de conocerla.

3. Abstenerse de publicar la **información confidencial** que conozca, reciba o intercambie con ocasión de las reuniones sostenidas.
4. Usar la **información confidencial** que se le entregue, únicamente para los efectos señalados al momento de la entrega de dicha información.
5. Mantener la **información confidencial** en reserva hasta tanto adquiriera el carácter de pública.
6. Responder por el mal uso que le den sus representantes a la **información confidencial**.
7. Guardar la reserva de la **información confidencial** como mínimo, con el mismo cuidado con la que protege la **información confidencial**.
8. La **parte receptora** se obliga a no transmitir, comunicar revelar o de cualquier otra forma divulgar total o parcialmente, pública o privadamente, la **información confidencial** sin el previo consentimiento por escrito por parte de *Transpormax*.
9. La **parte receptora** se compromete a establecer que los datos a utilizar son: *información sobre el consumo de gasolina en estaciones de combustible y recorridos realizados por los vehículos.*
10. La información capturada por la **parte receptora** se observará como *cifras para estudio estadístico, comparativo, información cualitativa*, no existirá ningún tipo de ganancia económica, es netamente con fines educativos.
11. La identidad de todo el personal de *Transpormax* no será revelada, dado que no se capturará sus nombres completos ni algún otro tipo de información que revele su identidad física o digital.

12. Las pruebas realizadas por la **parte receptora** nunca pondrán en peligro los activos tecnológicos de **Transpormax**, ni violentará la ley de delitos informáticos Colombiana 1273 de 2009 estando en el margen de las buenas prácticas y los procesos legales pertinentes.
13. El estudiante **César Leonardo Ortiz Rovira** se compromete a difuminar, bloquear y ocultar toda información que revele la identidad de la empresa **Transpormax** para salvaguardar la confidencialidad e identidad de la misma en el documento final del proyecto el cual será publicado en el repositorio institucional y de acceso público.
14. La empresa **Transpormax**, autorizará, SI / NO su nombre aparecerá en el título del proyecto, dado el caso, de que no sea autorizado, este deberá adecuarse.

Parágrafo: Cualquier divulgación autorizada de la **información confidencial** a terceras personas estará sujeta a las mismas obligaciones de confidencialidad derivadas del presente **Acuerdo** y la **parte receptora** deberá informar estas restricciones incluyendo la identificación de la información como confidencial.

Quinta. Obligaciones de la parte reveladora: Son obligaciones de la parte reveladora:

1. Mantener la reserva de la **información confidencial** hasta tanto adquiera el carácter de pública.
2. Documentar toda la **información confidencial** que transmita de manera escrita, oral o visual, mediante documentos, medios electrónicos, discos ópticos, microfilmes, películas, e-mails u otros elementos similares o en cualquier forma tangible o no, incluidos los mensajes de datos, como registro de la misma para la determinación de su alcance, e indicar específicamente y de manera clara e inequívoca el carácter confidencial de la información suministrada de la **parte receptora**.

Sexta. Exclusiones a la confidencialidad: La parte receptora queda relevada o eximida de la obligación de confidencialidad, únicamente en los siguientes casos:

1. Cuando la **información confidencial** haya sido o sea de dominio público. Si la información se hace de dominio público durante el plazo del presente acuerdo, por un hecho ajeno a la **parte receptora**, esta conservará su deber de reserva sobre la información que no haya sido afectada.
2. Cuando la **información confidencial** deba ser revelada por sentencia en firme de un tribunal o autoridades competentes en desarrollo de sus funciones que ordenen el levantamiento de la reserva y soliciten el suministro de esta información. No obstante, en este caso la parte reveladora será la encargada de dar cumplimiento a la orden, restringiendo la divulgación a la información estrictamente necesaria, y en el evento de que la confidencialidad se mantenga, no eximirá a la parte receptora del deber de reserva.
3. Cuando la **parte receptora** pruebe que la **información confidencial** ha sido obtenida por otras fuentes.
4. Cuando la **información confidencial** ya la tenía en su poder la parte receptora antes de la entrega de la información reservada.

Séptima. Responsabilidad: la parte que contravenga el acuerdo será responsable ante la otra parte o ante los terceros de buena fe sobre los cuales se demuestre que se han visto afectados por la inobservancia del presente acuerdo, por los perjuicios morales y económicos que estos puedan sufrir como resultado del incumplimiento de las obligaciones aquí contenidas.

Octava. Solución de controversias: Las partes (*César Leonardo Ortiz Rovira – Transpormax*) se comprometen a esforzarse en resolver mediante los mecanismos alternativos de solución de conflictos

cualquier diferencia que surja con motivo de la ejecución del presente acuerdo. En caso de no llegar a una solución directa para la controversia planteada, someterán la cuestión controvertida a las leyes colombianas y a la jurisdicción competente en el momento de presentarse la diferencia. La Universidad Nacional Abierta y a Distancia como institución educativa no se hace responsable del no cumplimiento de las cláusulas del presente acuerdo de confidencialidad por parte de César Leonardo Ortiz Rovira.

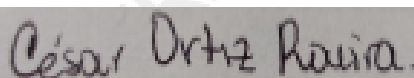
Novena. Legislación aplicable: Este acuerdo se regirá por las leyes de la República de Colombia y se interpretará de acuerdo con las mismas.

Décima. Aceptación del Acuerdo: Las partes han leído y estudiado de manera detenida los términos y el contenido del presente Acuerdo y por tanto manifiestan estar conformes y aceptan todas las condiciones.

Firman en Barranquilla., a los (29) días del mes de (Noviembre) de 2025

Como Parte Receptora:

Por la parte reveladora:



Cesar Leonardo Ortiz Rovira.
Estudiante UNAD.

Leidy Tatiana Rosales Castillo.
Transpormax SAS