

Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Negocios y Administración Pública

CARRERA DE ESPECIALIZACIÓN EN
MÉTODOS CUANTITATIVOS PARA LA GESTIÓN Y
ANÁLISIS DE DATOS EN ORGANIZACIONES

TRABAJO FINAL INTEGRADOR DE ESPECIALIZACIÓN

Estudio de la Logística y Distribución a partir del análisis
de datos: Antigüedad de Flota y Costos de
Mantenimiento

AUTOR: LUIS EDUARDO BRITTI

TUTOR: VERÓNICA GRACÍA FRONTI

SEPTIEMBRE 2024

Resumen

Como respuesta a las continuas crisis del mercado argentino, en el transporte de cargas automotor, continuamente se abren mesas de discusión de costos y tarifas entre operadores logísticos y dadores de carga. Este trabajo se centra en respaldar estadísticamente a los primeros, en cómo una mayor antigüedad de flota conlleva mayores costos de mantenimiento correctivo.

Se adopta y analiza para el estudio, el caso de una organización típica del sector. Se describe cómo fue el proceso de consolidación y anonimización de la información, destacando las problemáticas encontradas en la gestión de datos. Luego, se enumeran lineamientos estratégicos para la transformación del ente hacia un modelo data-driven.

Una vez definida la base de trabajo, se ejercitó un análisis de componentes principales (PCA) reduciendo la dimensionalidad y permitiendo destacar una primera correlación entre antigüedad y mantenimiento correctivo.

Finalmente, se llevó a cabo un análisis MANOVA/ANOVA. El resultado permitió identificar la presencia de diferencias significativas entre las medias de las categorías generadas a partir de la antigüedad de los vehículos de este conjunto de datos. Allí, las unidades clasificadas como Viejo (+10 años) gastan en promedio u\$s 3.184 más y experimentan 13 eventos correctivos adicionales a los de una unidad 0 km.

Palabras Clave

big data, mantenimiento, antigüedad de flota, seudonimización, logística, transporte de carga automotor

Índice

Introducción	3
1. Descripción de la organización en estudio	6
1.1. Un operador logístico	6
1.2. Lineamientos y horizonte de una nueva política de gestión de datos	9
Redefiniendo la Gerencia de IT	10
Privacidad de los datos	11
Esquema de gobierno de datos	12
1.3. Base de datos	13
Seudonimización de datos	19
Calidad de la información	20
2. Análisis de Componentes Principales	22
2.1.1. Limpieza de la base de datos	22
2.2. Eliminación de outliers y normalización	23
2.3. Estudio de Componentes Principales	26
3. Influencia de la antigüedad en el Mantenimiento Correctivo	32
3.1. Clasificación de factores	32
3.2. Análisis Multivariante de la varianza	33
3.3. Análisis de la varianza	33
Conclusiones generales	37
Bibliografía	40
Apéndices	41

Introducción

El transporte de cargas automotor desempeña un papel destacado en la economía nacional, siendo uno de los protagonistas de la movilización de mercadería a lo largo del país. Este, fenómeno se acentuó a partir de la década de 1990 cuando la demanda atendida del sector ferroviario cayó un 75%. A partir de allí, en sintonía con la desregulación del sector se ha incrementado su oferta. (Müller, 2004).

Durante las décadas siguientes, el rubro acompañó las fluctuaciones de la economía del país. El hundimiento del volumen transportado durante el 2020 como consecuencia de la pandemia de COVID-19, ha impactado severamente en el rubro. En 2022, cuando parecía que volvería a retomar el ritmo prepandémico, la sequía y la inflación empujaron a las empresas del sector a una tormenta perfecta; una fuerte caída de la actividad y un incremento de los costos por encima del índice de inflación. Mientras ese año se manifestó el mayor IPC anual de los últimos 20 años (94,8%) (Instituto Nacional de Estadística y Censo, 2023), los costos del transporte superaron por varios puntos al primero (121%) (Federación Argentina de Entidades Empresarias del Autotransporte de Cargas, 2022) (FADEEAC).

En octubre del mismo año, la escasez de dólares coyuntural llevó al gobierno argentino disponer de mayores restricciones al control de cambio e importaciones. Por ello, las automotrices pasaron a cotizar las nuevas unidades al tipo de cambio dólar MEP (hasta entonces se tomaba el tipo de cambio dólar Banco Nación). En consecuencia, los costos de renovación de flota se vieron incrementados en aproximadamente un 80%. Esto, sumado al contexto general, empujó al mercado a nuevas discusiones tarifarias entre operadores logísticos y dadores de carga.

La tarifa por servicio de un operador logístico involucra diferentes conceptos:



Imagen 1 - ¿Como se compone a grandes rasgos el valor de un viaje de distribución? - Fuente El Cronista - FADEEAC

En el gráfico se observan seis aspectos: Impuestos, Salarios, Combustibles, Amortización, Costos Operativos y Estructura y la renta o margen para el empresario de transporte. A los fines del presente trabajo, se prescinde del porcentaje de impuestos y se recalcula la proporción destinada a cada término de la ecuación. Se subdividen también, los costos de estructura en costos operativos y de mantenimiento. De este modo, ecuación tarifaria se reescribe del siguiente modo:



Imagen 2 - Estructura tarifaria sin Impuestos – Elaboración propia.

Estos diagramas dejan en evidencia cómo el costo de las unidades impacta directamente en la composición del valor de un flete a partir de la Amortización. El punto abre la discusión por los costos en dos caminos a seguir: si el cliente promueve renovar la flota, el término correspondiente a la amortización de las unidades, se vería incrementado. Ahora bien, ¿qué sucedería si se desistiera del recambio? En ese caso es lógico que se consideraran costos más bajos de amortización porque la de un vehículo más antiguo es menor al de una unidad 0km. Ahora, ¿es esto inocuo para el propietario de los camiones? En un ámbito de negociación, quién paga la tarifa (los clientes) opina que sí. Quién provee el servicio pone una alerta: adoptar unidades más viejas, debiera suponer también, un porcentaje extra de mantenimiento correctivo. Si bien es un planteo en principio lógico; en este ámbito de negociación: ¿Cómo puede el operador logístico conseguir un fundamento estadístico que respalde el pedido de mayor reconocimiento en tarifa por mantenimiento correctivo? Aquí es donde se dispara la primera pregunta del estudio: ¿Existe suficiente evidencia estadística para asegurar que una mayor antigüedad de flota requiere de mayores costos de mantenimiento correctivo?

Para responder este interrogante, se analizará el caso de una organización del rubro logístico, que opera con más de 200 vehículos dedicados al acarreo y distribución de bienes de consumo masivo. La entidad presta servicios a diferentes clientes tanto en el Gran Buenos Aires (GBA) como en localidades del interior del país. Por cuestiones de confidencialidad y tratándose de un mercado relativamente pequeño y al manipularse información sensible relativa a negociaciones de tarifarias, se dará a la organización el seudónimo de Logística de Consumo

(en adelante LogiCon). Esto, tendrá el fin de proteger no solo datos confidenciales de la propia empresa, sino también resguardar la identidad de sus clientes.

En un primer apartado se presentará la organización que será objeto de estudio. Se analizarán qué problemáticas y limitantes se observan hoy en día intrínsecamente relacionadas con la gestión de los datos. Luego, se desarrollarán los lineamientos para guiar la transformación de la organización hacia un enfoque basado en datos. Finalmente, se describirá el proceso de consolidación de la información originada en fuentes dispersas, al tiempo que se evalúa su calidad. Una vez conformada la base de estudio, se someterá a un proceso de anonimización, mediante una técnica de seudonimización sobre variables críticas.

En un segundo apartado se detallará una primera revisión estadística en el software R. Se practicará una descripción inicial, identificando la presencia de valores atípicos y faltantes. Posteriormente, se llevará a cabo una evaluación de los tipos de datos y su variabilidad, incluyendo medidas como medias, medianas, valores máximos y mínimos. Este primer análisis permitirá comprender mejor la naturaleza y su distribución.

En una segunda instancia, se desarrollará un análisis de Componentes Principales (ACP). El ACP busca representar de manera efectiva la información de n observaciones de p variables a través de un conjunto más pequeño de variables no correlacionadas, que son combinaciones lineales de las originales. (Peña, 2002) Esta técnica permitirá simplificar la estructura de los datos y resaltar los patrones más relevantes.

Finalmente, para abordar el disparador planteado se ensayará un análisis de la varianza y un análisis multivariado de la varianza, centrados en la cantidad de eventos y gastos en dólares de mantenimiento correctivo. Estas características serán evaluadas en función de la antigüedad de los vehículos para entender si existen o no, diferencias significativas entre las medias de las categorías descriptas.

Para terminar, se presentarán las conclusiones generales del trabajo, sus aportes y posibles líneas de trabajo a futuro.

1. Descripción de la organización en estudio

En este primer apartado se presentará la organización en estudio y la gestión de los datos que realiza. Tras un diagnóstico inicial, se describirá como es el proceso de seudonimización a aplicar y finalmente, se detallarán los lineamientos para guiar la transformación de la organización hacia una gestión basada en datos.

Finalmente, se describirá el proceso de recopilación de la información disponible y el armado de la base de estudio; se enumerarán y detallarán brevemente las fuentes de datos utilizadas con las posibles deficiencias observadas en cuanto a calidad de los datos.

1.1. Un operador logístico

LogiCon es una empresa de logística y distribución iniciada a principios del siglo XXI dando servicio a una empresa productora de bienes de consumo masivo. En la actualidad, el ente opera en más de 10 ciudades de todo el país, su flota supera los 200 camiones y responde a la demanda de 20 dadores de carga de variados rubros.

Su misión principal es proporcionar soluciones para toda la cadena de suministro. Esto incluye: la gestión de pedidos, la entrega de productos, la facturación y la recaudación de valores. Además, se encarga de la gestión de devoluciones y rechazos, cobros y según el servicio prestado, también de la manipulación de cargas y la gestión de almacenes para asegurar un flujo eficiente de productos.

En una operación típica LogiCon, trabaja para su cliente, un productor de bienes de consumo masivo “X”. Estos clasifican los artículos a transportar a sus propios compradores y en base a esa información se desarrolla un ruteo¹. Se cargan las unidades de transporte según las rutas asignadas. Los vehículos despliegan los recorridos predefinidos y contra entregas se recauda según lo facturado para el productor del bien. La organización trabaja en estrecha colaboración con estos dadores de carga para garantizar la disponibilidad y entrega puntual de los bienes a sus clientes.

La captura de valor se da a través de cobrar por bien entregado o almacenado según corresponda la tarea prestada. Esta tarifa es gravada sobre los dadores de carga. De este modo se persigue generar ingresos sostenibles y rentables como operador logístico al tiempo de contribuir al cumplimiento de los objetivos estratégicos de los clientes en el mercado. Se

¹ Ruteo: estrategia para realizar la distribución adecuada y eficiente de las mercancías, en los diferentes puntos en los cuales lo desee una organización. (Castiblanco, 2015)

concluye entonces: LogiCon es una típica economía tradicional, caracterizada por generar réditos mediante la transformación de insumos en productos y servicios terminados. Como toda organización clásica lineal, operan con jerarquías rígidas y flujos predefinidos de trabajo. (Herrera Pablo, 2023)

El primer paso para determinar el camino seguir para volverse una empresa data driven, es identificar en qué punto de la carrera se encuentra. C. S. Penn propone los estadios en la Imagen 1.1:

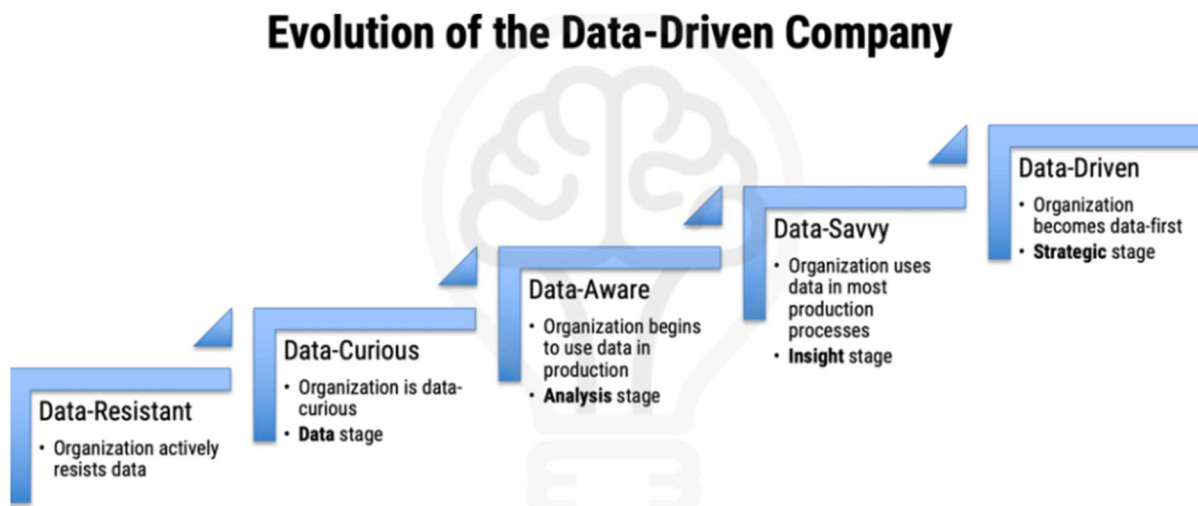


Imagen 1.1 – Evolución hacia compañías Data Driven – Fuente <https://www.christopherspenn.com/>.

Si se tuviera que clasificar, en qué punto está de la evolución según este enfoque propuesto, sería en un punto medio entre las instancias de “Data Resistant” y “Data Curious” (Penn, 2016). El organismo tiene una conciencia incipiente de la importancia de los datos, pero se resiste al cambio. No se ven o no se imaginan aun los beneficios de tener los datos bien gestionados y como pueden aportar a sus operaciones y toma de decisiones.

“En todas las industrias vemos cambios dramáticos en la tecnología que permiten a las personas hacer más con la misma información” (Shapiro, 2000). LogiCon tiene la oportunidad de aprovechar el potencial de sus datos y optimizar operaciones para mejorar la eficiencia de las entregas y potenciar los servicios a sus clientes.

Hasta el momento, la empresa no ha realizado una inversión significativa en la obtención de datos. Principalmente, utiliza un software de gestión logística (TMS) en operaciones locales de menor envergadura.

Otra fuente de interés es provista por el servicio del dispositivo satelital: el módulo de geolocalización no solo señala el posicionamiento; también ofrece sensores de apertura de puertas, encendido y apagado, velocidades máximas, kilómetros recorridos, hora de encendido y horas de ralentí (horas de encendido a velocidad cero). Si bien no se explota, el

sistema ofrece dispositivos personalizados para individualizar el conductor. Toda esta información es registrada instantáneamente en el software del proveedor. A esta aplicación se accede vía web con diferentes perfiles de parametrización y consulta.

El registro en el TMS es esencialmente realizado por parte de los empleados (tanto operadores como personal del taller). De igual modo, el ingreso de consumos de gasoil es importado por acción humana a través de interfaces sobre planillas de cálculo. Finalmente, el módulo satelital realiza lecturas en forma automática por parte de sensores del mismo dispositivo.

En referencia al valor de la Información y efectos de red, hasta el momento no se observan intenciones en la utilización de los datos antes detallados y diferenciarse de sus competidores. El detalle de ingresos al módulo de mantenimiento tiene únicamente fines de imputación de costo. Estas acciones se clasifican hoy día en preventivas y correctivas². No se desarrolla ningún análisis estadístico más riguroso.

Algo similar sucede con el combustible. Sólo se registran los litros consumidos por vehículo. Para evaluar el consumo por kilómetro recorrido, se calcula a partir de ese registro mencionado y el ingreso manual al momento de la carga del odómetro físico. Esto último implica muchos errores o faltantes en las cargas. Los analistas terminan reparando la información mediante repreguntas a la fuente o cálculos estimativos; son pocos los casos, pero dejan entrever fisuras en la confiabilidad de esta información.

En la gestión de las cargas, los clientes más importantes planean sus propias rutas y entregas clasificándolos en sus propios esquemas. Si bien se cuenta con accesos de consulta en pantalla, no hay posibilidades de reportes para el estudio de esa información. Aun tratándose de relaciones comerciales sólidas y de confianza con los dadores de carga; estos no han percibido plenamente la relevancia tanto desde un punto de vista comercial como operativo de establecer este canal de intercambio de datos.

La colaboración en la gestión de información sobre rutas y entregas podría traer beneficios significativos para todos los actores. Una posibilidad es incluir la optimización conjunta de recorridos para reducir costos operativos, mejorar la eficiencia de las entregas y garantizar la satisfacción del cliente. El intercambio de datos en tiempo real puede proporcionar a ambas partes una mayor visibilidad y control sobre las operaciones logísticas, facilitando la toma de decisiones informadas en ambos actores.

² El mantenimiento correctivo está destinado a las reparaciones para enmendar o solucionar averías ya manifestadas. El preventivo en cambio, son acciones futuras que apuntan a prevenir futuros problemas.

Como primera conclusión: LogiCon subutiliza los datos a disposición. A pesar de contar con diversas fuentes de información, como registros de mantenimiento o datos satelitales, la organización no ha asignado recursos significativos ni personal especializado para realizar un análisis detallado. Esta carencia es un factor limitante en el desarrollo de estrategias de diferenciación basadas en la información disponible. Se evidencia una falta de reconocimiento sobre la importancia estratégica de la información en la industria logística. Los esfuerzos solo se concentran en analíticas tradicionales (descripción y diagnóstico) por sobre la analítica avanzada (predictivo y prescriptivo).

1.2. Lineamientos y horizonte de una nueva política de gestión de datos

La optimización de la cadena de suministro es un objetivo fundamental para las empresas que buscan rentabilidad y competitividad. Esto implica mejoras en todos los aspectos del flujo de bienes y servicios, desde la adquisición de materias primas e insumos hasta la entrega final del producto terminado al consumidor. La tarea puede significar: reducir costos, aumentar la velocidad de entrega, mejorar la calidad del servicio, minimizar el riesgo de interrupciones, etc.

La relación entre el análisis de datos y logística presenta múltiples aspectos por explorar. Todos convergen en un punto central común a casi todas las industrias: el aumento de la eficiencia y la mejora de la información disponible para la toma de decisiones. Por poner algunos ejemplos, en materia de gestión de inventarios, la disciplina ha permitido prever la demanda con mayor precisión y evitar tanto excesos como déficits en los stocks. La distribución se ve beneficiada de la optimización de rutas de entrega, reduciendo tiempos y costos de transporte mediante el análisis de datos de tráfico en tiempo real. El Big Data también ha permitido un mejor seguimiento y trazabilidad de los envíos, asegurando la transparencia y la rapidez en la detección de problemas.

Además, el Data mining podría mejorar la eficiencia en la gestión de proveedores. Mediante la evaluación de su desempeño en tiempo real aporta a la toma de decisiones basada para mantener relaciones de suministro eficientes. La optimización también se extiende a la mejora del servicio al cliente, permitiendo una personalización en la entrega y la gestión de devoluciones más eficiente.

En términos de mantenimiento, el Big Data puede prever fallos en los equipos mediante el monitoreo continuo de su rendimiento. Esto permite reducir los costos de tareas correctivas y facilitando la implementación de programas de revisiones proactivas. El mantenimiento

basado en datos se vuelve así en una herramienta para prolongar la vida útil de los activos y reducir los costos. (Cerquitelli, 2021)

En términos del análisis de datos en sí, y tal como se describía en el título anterior, ubicamos a LogiCon en la base de la pirámide de C. S. Penn. “Data Resistant” representa el nivel más bajo de conciencia con respecto a los beneficios que los cambios impulsados por el mundo de los datos pueden aportar. ¿De qué manera se puede fomentar y acompañar esta evolución?

Redefiniendo la Gerencia de IT

LogiCon, actualmente, cuenta con un departamento de sistemas compuesto por un jefe y dos analistas. El sector tiene como función principal proporcionar soporte técnico a todas las áreas y de la empresa. Con el cambio hacia una organización orientada a los datos, se hace necesario de redefinir la estructura del área de IT.

En primer lugar, es necesario transformar el departamento en una gerencia. Esta transición implica un cambio de perspectiva. La gerencia de IT debe ser considerada como un par del resto de los negocios, tomando parte en la mesa de decisiones estratégicas. Este cambio es importante para facilitar la sinergia entre la nueva área y las originales; convirtiéndose así en un nuevo socio.

Esta transformación del espacio de IT permitirá avanzar en el camino hacia una organización data driven. La transición fortalece la colaboración entre las áreas permitiendo: potenciar la capacidad de identificar problemas críticos, generar nuevas oportunidades de negocio y promover la toma de decisiones informadas.

Para la nueva posición gerencial en LogiCon, aunque el conocimiento del sector del transporte es preferible encontrar especialistas locales en big data con experiencia específica en logística puede resultar dificultoso. Una alternativa práctica sería rodear al profesional con los expertos de la misma empresa y así aprovechar los conocimientos combinados. En tal caso, el perfil del CDO convendría ser desarrollado en comunicación y liderazgo. (Deloitte, 2017)

Los equipos de TI deberán centrarse en garantizar la compatibilidad de todas las aplicaciones e implementación efectiva de las políticas de gobierno de datos. En contraste, los usuarios comerciales, colaborarán con el expertise en la materia (Lawton, 2020). La habilidad radica en asignar valor a la utilización de los datos para cumplir objetivos y resolver problemas del ámbito logístico.

En la actualidad, los directores de LogiCon se distinguen por su experiencia a lo largo de toda la cadena de suministro. Sin embargo, es de notar la existencia de una carencia en el ámbito

de la Tecnología de la Información. Para lograr una integración acorde, resultará importante establecer un lenguaje común facilitando la comunicación entre los referentes (propios del rubro del transporte) y las nuevas metas delineadas por el gerente de datos. Esta colaboración se presenta como un norte para alinear la estrategia tecnológica con los objetivos comerciales. Finalmente, se sugiere la participación del Chief Information Security Officer (CISO) para dar solidez al gobierno de la información. El rol tendrá como fin asegurar la protección de la infraestructura, los datos y los sistemas de las amenazas. La tarea debe garantizar la confidencialidad, integridad y disponibilidad de los datos críticos, siendo una preocupación central para esta y cualquier organización.

Una gestión integral de la información propone la colaboración complementaria entre el CDO y el CISO. Mientras el primero se centra en la gestión y garantiza de la calidad de los datos, el segundo, supervisará la protección de la infraestructura y los sistemas de información. Además, abordará amenazas y riesgos derivados del desarrollo del entorno generado por parte del CDO. Esta conexión entre ambos se torna estratégica y completa una gestión integra de toda la generación de valor, desde la calidad hasta la seguridad.

Dada la naturaleza en principio reducida del departamento, se propone inicialmente la contratación de un Gerente (como CDO) especializado en la generación de valor a través del big data. Complementariamente, se podría considerar la tercerización del alojamiento del software y la adopción de servicios de seguridad proporcionados por el CISO. Este enfoque mixto busca optimizar el costo, la eficiencia y la seguridad al combinar habilidades especializadas internas con soluciones externas.

Privacidad de los datos

El avance tecnológico en este siglo ha generado una abundancia de datos a gran escala. Esto dio lugar a la creación de un ecosistema virtual que resulta cada vez más atractivo para las organizaciones. Ese entorno ofrece la oportunidad de obtener información adicional para mejorar su proceso de toma de decisiones. Dentro de esta amplia gama de datos recopilados, están los personales u otros sensibles de suma relevancia. (Salaberri, 2021). El análisis de datos como herramienta, dio lugar a un proceso de acumulación y concentración continuo de información y poder. Ha dejado expuesta la privacidad de los individuos y puso en evidencia la profunda asimetría entre el usuario y las organizaciones de plataforma. (Herrera Pablo, 2023)

Con el fin de asumir la responsabilidad del cuidado ético al respecto; para una eventual transformación de LogiCon en una empresa data driven; se sugiere asumir el concepto de

confidencialidad desde el diseño: la incorporación de barreras de privacidad y protección desde la etapa inicial del desarrollo. La empresa debe llevar a cabo evaluaciones de impacto en la confidencialidad e identificar y abordar sus posibles riesgos.

Es importante como política obtener el consentimiento informado de todas las partes involucradas. Esto abarca desde los operadores, almacenes y transportistas tercerizados, hasta los clientes. El detalle debe incluir una explicación clara sobre el destino de datos adquiridos, cómo y quiénes tendrán acceso a ellos; siendo así, transparentes acerca de sus prácticas. En paralelo, se deben adoptar medidas de seguridad sólidas. Esto incluye medidas como el cifrado de datos, sistemas de autenticación seguros y protección contra accesos no autorizados.

Otro punto por abordar es el principio de minimización de los datos. Según el Reglamento del Parlamento europeo implica coleccionar y retener solo aquellos datos “adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados” (Parlamento Europeo, 2016). En Argentina, la Ley 25.326 de Protección de los Datos Personales también se suma a este principio al establecer que los datos personales deben ser "exactos, adecuados, pertinentes y no excesivos" en relación con el ámbito y finalidad para los que se hubieran obtenido. (República Argentina, 2000) En este caso puntual, se busca asegurar la calidad y eficiencia de las operaciones logísticas. Los datos irrelevantes deben eliminarse.

La privacidad de datos no puede limitarse a un enunciado inicial de normas y prácticas. Para asegurar una defensa eficiente, es necesaria una red de vigilantes activos. Así, tanto empleados como clientes deben estar capacitados en la importancia de la seguridad de los datos y en cómo contribuir a su preservación. Esta instrucción debe completarse no solo en sus normas, sino también en la sensibilización hacia las amenazas a la privacidad y en la habilidad de reconocerlas cuando surjan.

Este no es un objetivo estático, es un proceso en constante evolución. El organismo debe buscar una evaluación continua de sus prácticas. La tecnología avanza con rapidez, al igual que la legislación, y las amenazas se transforman y multiplican. Estas políticas de seguridad deben adaptarse para fortalecer su respuesta a estos cambios.

Esquema de gobierno de datos

En el contexto de una empresa donde la toma de decisiones ya está centralizada, se sugiere optar por un esquema similar de Gobierno de Datos alineado con la estrategia general. Hoy día, las decisiones ya fluyen a través de un punto central. Expandir este enfoque al nuevo

enfoque, ofrece coherencia y sinergia con la forma en que se manejan otros aspectos críticos del negocio.

En términos de riesgos, esta configuración proporciona un marco sólido para la mitigación. Al tener un punto focal para la gestión de datos, se reducen los riesgos asociados con la falta de claridad en las políticas, malentendidos y la inconsistencia en la aplicación de estándares. Además, ofrece la posibilidad de monitorear la transformación desde la dirección, complementando el ingreso a esta filosofía data-driven.

Para medir la proporción de decisiones clave basadas en datos que han contribuido directamente a los objetivos de negocio predefinidos, se pueden evaluar el porcentaje de decisiones basadas en esta información. Una alta contribución indicaría que el gobierno de datos está alineado efectivamente con los objetivos estratégicos de la organización, justificando así la inversión y esfuerzos en el área de datos.

Es importante medir el “porcentaje de adopción de herramientas de gobierno de datos”, permite evaluar la proporción de personal alineadas a las nuevas prácticas. Una alta tasa sugiere que las herramientas están siendo efectivas y aceptadas por el personal. Esto respalda el control directivo al demostrar como las inversiones en tecnologías de gobierno de datos están generando un impacto positivo en la eficacia organizativa.

Desde la sustentabilidad, el Índice de cumplimiento con normativas y procedimientos de gobierno de datos; mide la adherencia de la organización a las normas y procedimientos establecidos en este marco. Un alto porcentaje de cumplimiento sugiere que la organización adopta las nuevas prácticas y está en consonancia con los estándares regulatorios, proporcionando solidez para el control directivo y reduciendo los riesgos relacionados con la gestión de datos.

1.3. Base de datos

Para analizar si existe vínculo entre la antigüedad de la flota y los costos de mantenimiento correctivo es necesario consolidar una base de datos con información dispersa en distintas tablas. Se analizarán: reportes del módulo de mantenimiento y datos proporcionados por el sistema de seguimiento satelital de los vehículos.

El “módulo de mantenimiento” es un software hecho a medida de LogiCon. En la tabla de “equipos” se detallan los siguientes campos: Patente, Operacion , Tipo_Equipo, Condición, Marca, Modelo, Nro Chasis, Nro Motor, Nro Telepase, Nro Póliza, Año, Chofer, Combustible, Observaciones. ([Vínculo a tabla completa “equipos” en Apéndices](#))

Patente	Operación	Tipo_equipo	Condición	MARCA	MODELO	NRO CHASIS	NRO MOTOR	NRO TELEPASE	NRO POLIZA	AÑO	Chofer
CAM0001	OP0020	CHASIS	Propio	MERCEDES BENZ	ATEGO 1418	—	—	—	—	2016	—
CAM0002	OP0020	CHASIS	Propio	MERCEDES BENZ	ATEGO 1418	—	—	—	—	2016	—
CAM0003	OP0003	CHASIS	Propio	MERCEDES BENZ	ATEGO 1418	—	—	—	—	2016	—
CAM0004	OP0003	CHASIS	Propio	MERCEDES BENZ	ATEGO 1418	—	—	—	—	2016	—
CAM0005	OP0020	CHASIS	Propio	MERCEDES BENZ	ATEGO 1418	—	—	—	—	2016	—
CAM0006	OP0005	CHASIS	Propio	MERCEDES BENZ	ATEGO 1418	—	—	—	—	2016	—

Imagen 1.2 – Tabla Equipos (extracto) – Modulo de Mantenimiento.

Mediante el atributo Condición, se puede discernir aquellos vehículos propios de aquellos vehículos de terceros que prestan servicio para la compañía (fleteros). En función del año de fabricación, se evalúa la antigüedad mínima, máxima, la media y su desvío:

Condición	Cantidad	%	Min. Antigüedad	Media Antigüedad	Desvío est. Antigüedad	Max. Antigüedad
Fletero	181	39%	0	15,2	6,6	49
Propio	283	61%	0	10,4	6,4	30
Total general	464	100%	0	6,88	6,9	49

Tabla 1.3 – Antigüedad de los vehículos” según propiedad del automotor – Elaboración propia

Para el presente trabajo, conviene centrarse en los automotores propios; aquellos sobre los cuales la gestión es absoluta. En este subconjunto, se identifica una antigüedad media de 10,2 y desvío estándar de 6,4 años, con valores entre 0 y 30. Luego, a partir de esta porción, evaluaremos los tipos de vehículos:

Tipo_equipo	Cuenta de Condición	%
ACOPLADO	6	2%
AUTOELEVADOR	13	5%
BALANCIN	34	12%
CAMIONETA	9	3%
CHASIS	159	55%
ELECTRICO	1	0%
MOTO	1	0%
SEMI	35	12%
TRACTOR	25	9%
Total general	283	100%

Tabla 1.3 – Vehículos propios según frecuencias por “Tipo_equipo” – Elaboración propia

Del extracto de vehículos propios se seleccionan aquellos equipos de calle motorizados; lo que descarta; los autoelevadores, acoplados y semis. Las motos, camionetas y vehículos eléctricos, responden a diferentes operatorias, estándares de consumo y regímenes de mantenimiento con lo cual, también se excluyen del ensayo.

Así, este análisis involucrará tres configuraciones de carga diferentes que se muestran en la Imagen 1.4:

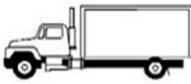
		Capacidad
Chasis		10 pallets / 10 Tn
Balancín		16 pallets / 16 Tn 2 ejes traseros
Tractor (y semi-remolque)		30 pallets / 30 Tn

Imagen 1.4 – Configuraciones de carga – Elaboración propia

De la tabla original de equipos se seleccionan los atributos: Patente, a partir de la cual se vincula con la información de las otras fuentes; Marca, Modelo (ambos datos del fabricante), Tipo, y Año.

En cuanto al elemento “Operación” se permite clasificar cada observación según la actividad practicada. Por ello, en la base de trabajo final consolidada y bajo el nombre “ACTIVIDAD”, se desglosa en acarreo local, de media o larga distancia (ACARREO y LD) y distribución (DISTRIBUCION).

Del mismo módulo mencionado, se extrae el “informe de reparaciones”. Esta consulta ofrece: Año/mes, #_Reparación, Operación, Taller/Proveedor, Tipo_Rep., Patente, Svcs_Externos, Tareas, Repuestos, Total, Observaciones_1, Observaciones_2. Las tareas de mantenimiento son registradas por operadores del taller de la propia organización en el esquema. ([Vínculo a tabla completa “informe de reparaciones” en Apéndices](#))

AÑO/MES	# REPARACIÓN	OPERACIÓN	TALLER / PROVEEDOR	TIPO REP.	PATENTE	SVCIOS EXTERNOS	TAREAS	REPUESTOS	TOTAL
ene-22	23986	OP0007	—	Preventivo	CAM0022	3.660,00	0,00	0,00	3.660,00
ene-22	23987	OP0007	—	Preventivo	CAM0158	3.660,00	0,00	0,00	3.660,00
ene-22	23988	OP0007	—	Preventivo	CAM0156	3.660,00	0,00	0,00	3.660,00
ene-22	23989	OP0007	—	Preventivo	CAM0178	3.660,00	0,00	0,00	3.660,00
ene-22	23990	OP0007	—	Preventivo	CAM0185	3.660,00	0,00	0,00	3.660,00
ene-22	23991	OP0007	—	Preventivo	FLE0014	3.660,00	0,00	0,00	3.660,00

Imagen 1.5 – Tabla Informe de reparaciones (extracto) – Módulo de Mantenimiento.

El costo de reparación tiene la siguiente apertura en función de su naturaleza:

Tipo_rep	Cant	Total Reparación				
		Costo en \$	Min	Media	Desvio est.	Max
Auxilio en Base	165	1.118.186,63	0,00	6.776,89	26.646,46	333.637,68
Auxilio en Ruta	202	1.348.302,18	0,00	6.674,76	10.448,07	58.313,40
Correctivo	4143	73.249.022,73	-34.813,48	17.680,19	62.401,31	1.870.000,00
Correctivo - Sinistro	27	1.677.210,31	0,00	62.118,90	113.484,68	538.009,57
Mejora a Flota	6	1.796.267,88	12.500,00	299.377,98	586.776,30	1.494.500,00
Preventivo	6985	123.169.917,01	0,00	17.633,49	66.404,59	1.900.000,00
Preventivo Semis	1	4.000,00	4.000,00	4.000,00	-	4.000,00
SR - Correctivos	418	18.941.490,29	0,00	45.314,57	82.087,61	654.545,46
SR - Preventivos	326	12.773.272,38	0,00	39.181,82	123.279,17	1.927.628,38
Total general	12273	234.077.669,41	-34.813,48	19.072,57	68.681,51	1.927.628,38

Tabla 1.6 – Cantidad de eventos y costo total por “Tipo Rep.” – Elaboración propia

Se observa un valor mínimo negativo, el cual al ser analizado pertenece a un ajuste sobre partidas mal asignadas que netean el total.

Se hizo una nueva clasificación de los registros en Preventivos y Correctivos. Las líneas indicadas como Auxilio tanto en ruta como en base, fueron reescritos como reparaciones de tipo correctivo, justificado en la base de atender una reacción posterior al siniestro. También se excluyen del análisis los eventos de gomería o reemplazo de cubiertas. El alto costo, la frecuencia, pueden distorsionar los resultados y se tratan de una naturaleza ajena al ejercicio aquí propuesto.

Los valores al estar detallados en pesos argentinos, en un país con una coyuntura inflacionaria, obliga a ser re expresados a una moneda fuerte para que sean mejor montos comparables a lo largo del tiempo. Así, se adopta como referencia el cálculo en dólares al tipo de cambio Banco Nación promedio del mes de registro.

En base a ese reporte, se calculan otras variables de estudio. Según la variable Tipo_Rep., se clasifica la base en eventos correctivo³ y preventivo⁴. Respetando el mismo criterio de temporalidad (total año 2022), se suman en una única observación por equipo, los totales incurridos en eventos correctivos CORRECTIVO_cant (cantidad de eventos) y CORRECTIVO_usd (suma total de gastos). Del mismo modo se calcula para eventos preventivos: PREVENTIVO_usd, PREVENTIVO_cant.

El mismo software aporta para la gestión, un reporte de consumos de los vehículos: Id_Vale, Tipo_combustible, Patente Fecha Petrolera, Litros_carga, Kilometraje, Total_Kilómetros ultima carga. Sobre esta consulta, el área de calidad analiza los consumos y km registrados y ajustan las distorsiones. ([Vínculo a tabla completa “consumos” en Apéndices](#))

³ El mantenimiento correctivo es aquel en el cual se incurre para corregir o solucionar problemas, errores o el mal funcionamiento de la unidad.

⁴ El mantenimiento preventivo responde a aquellas acciones sobre la unidad con el fin de prevenir futuros gastos correctivos. Por ej: cambio de aceite, filtros, lavado y engrase de la unidad, etc.

Registros de Campo							
ID	Tipo Combustible	Patente	Fecha de recarga	Petrolera	Litros Carga (Registro)	Kilometraje	Total de kilómetros desde la última carga
29234	Euro Diesel	CAM0001	7/1/2022	—	118,72	80.416	343,70
29235	Euro Diesel	CAM0001	14/1/2022	—	118,72	80.747	331,90
29236	Euro Diesel	CAM0001	25/1/2022	—	118,72	81.244	496,90
29238	Euro Diesel	CAM0002	8/1/2022	—	97,07	55.771	268,60
29327	Euro Diesel	CAM0002	22/1/2022	—	118,72	56.107	336,60

Imagen 1.7 – Tabla Reporte de consumos (extracto) – Modulo de Mantenimiento y departamento de calidad.

Los campos Lts consumidos (ajustado), Kms recorrido (ajustado) y Consumo (ajustado). Sobre estos últimos 3 campos se calcula un análisis estadístico de mínimos, media, desvío y máximos.

-	Lts Carga (ajustado)	Kms recorrido (ajustado)	Consumo (ajustado)
Min	35,50	31,60	4,63
Media	623,21	2.001,59	32,55
Desvio est.	849,77	2.596,62	12,51
Max	7.386,60	20.219,71	502,31

Tabla 1.8 – Análisis Estadístico de Consumo de Combustible y Kilometraje – Elaboración propia.

Se detectan los valores máximos consumo, como primera carga de una unidad 0km. Los mínimos del mismo índice, se muestran como cargas de unidades menores (camionetas), que manejan rangos diferentes para el ratio. Esto refuerza la decisión de trabajar con chasis, balancines y tractores.

De las variables seleccionadas, mediante el campo EQUIPO o PATENTE se asocia al resto de la información y por este se suman los LITROS consumidos por producto y vehículo a lo largo del año. Por la unidad están clasificados en DIESEL_lts, EURO_lts, y TOTAL_lts; según estos sean diésel, diésel-Premium (euro) y la suma de ambos. Al dividir el total de combustible consumido y los kilómetros recorridos (del campo ODOMETRO), ofrece el consumo promedio (CONSUMO) cada 100 km.

El “sistema de seguimiento satelital” ofrece el reporte de “Actividad periódica”, detallando la actividad por patente (o equipo) y por día con la siguiente apertura de parámetros: Patente, Fecha, Inicio de actividad, Dirección, Área, Conductor, Kilometraje, Fin de actividad, Dirección, Área, Conductor, Kilometraje, Total, Horas de velocidad cero, Total Kilómetros.

Cabe destacar que los km recorridos de esta tabla, son producto de lecturas del satelital.

[\(Vínculo a tabla completa “actividad periódica” en Apéndices\)](#)

Patente	Fecha	Inicio de actividad	Dirección	Área	Conductor	Kilometraje	Fin de actividad
CAM0001	3/1/2022	3/1/2022 06:12	—	—	—	80190,3	3/1/2022 13:19
CAM0001	4/1/2022	4/1/2022 06:09	—	—	—	80225,6	4/1/2022 20:10
CAM0001	5/1/2022	5/1/2022 06:05	—	—	—	80308,3	5/1/2022 13:59
CAM0001	6/1/2022	6/1/2022 06:13	—	—	—	80349,8	6/1/2022 15:08
CAM0001	7/1/2022	7/1/2022 06:00	—	—	—	80386	7/1/2022 15:21
CAM0001	8/1/2022	8/1/2022 05:59	—	—	—	80461,6	8/1/2022 13:03

Imagen 1.9 – Tabla Actividad periódica (extracto) – Sistema de seguimiento satelital.

Sobre las Hs de encendido total, Hs a velocidad cero y Km diarios se calcula un análisis estadístico de mínimos, media, desvío y máximos.

-	Hs de encendido total	Hs a velocidad cero	Km diarios
Min	00:00:01	00:00:00	0,00
Media	03:57:39	01:32:23	96,41
Desvío est.	03:10:52	01:59:31	162,46
Max	19:43:52	18:49:16	1.854,90

Tabla 1.10 – Análisis Estadístico de Hs de encendido, ralenti y Kms diarios recorridos.
– Elaboración propia.

Se observa alta variabilidad entre los mínimos y máximos de kms diarios recorridos. La respuesta deviene de tratarse unidades locales y de Larga Distancia; que exigen trabajos continuos durante varias horas al día a velocidades constantes y extensas distancias recorridas. Lo mismo sucede con las horas de encendido. Por el contrario, las mayores horas de ralenti o de velocidad cero, se destacan en unidades de distribución local.

En forma similar a los otros informes, se opta por sumar la información aquí contenida, en un único registro por patente; dando como resultado el total del año 2022 de cada atributo: KM (Total de kilómetros recorridos), HS_encendido (Total de horas de motor), HS_ralenti (Horas de velocidad cero), DIAS_svs (cantidad de días con reporte de encendido). La variable VELOC_prom representa la velocidad promedio en kilómetros por hora. Esta, es calculada a partir de los kilómetros recorridos y las horas de encendido, deducidas de las horas en ralenti. Como conclusión, a partir del vínculo de las placas patentes, se crea una base de datos consolidada con un único registro por patente. Esta base contiene las siguientes variables categóricas: ID, ACTIVIDAD, OPERACION, TIPO DE EQUIPO, MARCA, MODELO, ANTIGÜEDAD (en años), y TIPO DE COMBUSTIBLE. Además, se incluirán características numéricas como el AÑO de fabricación y la ANTIGÜEDAD del vehículo. Por último, se calcularán sumatorias y promedios de los siguientes atributos registrados en 2022: VELOCIDAD PROMEDIO, LITROS DE DIÉSEL CONSUMIDOS, LITROS DE EURO CONSUMIDOS, TOTAL DE LITROS CONSUMIDOS, KILÓMETROS RECORRIDOS,

CONSUMO (en litros por kilómetro), GASTOS EN MANTENIMIENTO CORRECTIVO (en dólares y en cantidad de eventos), GASTOS EN MANTENIMIENTO PREVENTIVO (en dólares y en cantidad de eventos), HORAS DE ENCENDIDO, HORAS DE RALENTÍ y DÍAS EN SERVICIO. ([Base completa en Apéndices](#))

Seudonimización de datos

Como se mencionó anteriormente, LogiCon es un seudónimo empleado para individualizar a la compañía de estudio. La organización, referente de un sector donde las relaciones son estrechas y dada la sensibilidad de la información manejada (abarca aspectos críticos: estructura organizativa, programas de mantenimiento y acuerdos tarifarios), empuja la necesidad de preservar la confidencialidad. Se adopta entonces, una política de no revelar datos que puedan identificar a la empresa dentro de su nicho.

Para lograr esto se implementa un proceso de seudonimización de variables clave. Esto incluye elementos como la placa patente de los vehículos y referencias geográficas, como ciudades y localidades de operación. Este enfoque permite presentar información relevante para el análisis sin comprometer la identidad específica de LogiCon, asegurando el cumplimiento de estándares éticos y de confidencialidad en la gestión de datos.

¿A qué nos referimos como seudonimización? Tanto la ésta como la anonimización, comparten una misma meta. Sin embargo, es importante no confundirlas: “La idea principal es que el método planteado no modifica los valores del conjunto de datos. En cambio, lo que hace es agregar seudónimos que reemplazan las columnas sensibles deseadas en el conjunto de datos original”. (Mena Madrid, 2021)

Para este estudio se adopta una seudonimización un proceso que reemplaza los valores en las columnas sensibles por otros valores como seudónimos. Estos se guardan en un conjunto o base de datos diferente al resguardo. Cabe aclarar: este método, permite recuperar la información original en cualquier momento.

Para este caso de LogiCon, existen dos factores críticos por las cuales se puede individualizar el nombre real de la organización. La placa patente del vehículo y los aspectos geográficos (provincia y localidad) de operación pueden evidenciar fácilmente la entidad en cuestión.

Cada entrada en esta base de datos está identificada mediante una clave única (ID). Donde originalmente estaban los números de placa-patente como individualizador, a los fines de mantener confidencialidad, se reemplazó por las siglas “CAM” seguido de 4 dígitos numéricos (quedando por ejemplo CAM0010, CAM0121, etc).

El campo y “OPERACION” proporcionan información sobre el cliente y ubicación geográfica en donde, se desempeñan las unidades. Con el fin de anonimizar la base se codificaron las variantes con prefijo “OP” seguido de 4 dígitos numéricos (por ejemplo: OP0013, OP0020).

Calidad de la información

La calidad de la base de datos abordada es analizada en los términos de seis objetivos: completitud, unicidad, oportunidad, validez, precisión y consistencia.

En primer lugar, la completitud de la información es manifiesta de acuerdo con dos perspectivas: se abarcan los vehículos relativos a todas operaciones activas y se ponen a disposición del presente trabajo todas las variables disponibles en las diferentes herramientas susceptibles de ser aprovechadas para los fines descriptos. No se identificaron datos faltantes, aunque sí se avizoraron outliers a subsanar durante la etapa de limpieza de datos.

La unicidad de los registros se manifiesta de manera sólida. Cada vehículo en la base se distingue por una identificación unívoca como es el dominio o patente. Esto elimina la posibilidad de duplicaciones o ambigüedades en el análisis. Este aspecto evita puntos que alterarían la distribución original.

A los fines de respetar el criterio de oportunidad de la calidad de datos, el corte temporal definido se justifica por la regularización de las actividades luego de la pandemia; dando lugar a las dinámicas y cambios experimentados a lo largo del tiempo en la flota de vehículos de LogiCon.

Los datos consolidados sostienen coherencia con respecto a cada una de las definiciones y parámetros establecidos, dando validez a los mismos. Cada variable se alinea de manera consistente con las definiciones preestablecidas de cada una, respetando coherencia de rangos y unidades de medida. Esto refuerza la confianza en la utilidad de los datos para abordar los objetivos planteados en el estudio y da solidez a las conclusiones.

La precisión y consistencia son los puntos más débiles del relevamiento. Se descartan mayores inconvenientes con aquellos puntos capturados de forma automática por parte del módulo satelital. Si bien se pueden interferir con errores de lectura, cualquier procedimiento de limpieza pueden subsanar outliers y faltantes.

No obstante, una atención particular merece el registro del consumo de combustible. Este ratio es producto de la división de litros (detalle a facturar por la petrolera) por kilómetros recorridos (registro manual del odómetro al momento de la carga, por parte del chofer). El segundo punto, puede propiciarse en errores u omisiones.

En la actualidad, toda esta información, proveniente de fuentes automáticas y manuales, se consolida y procesa en hojas de cálculo para evaluar la precisión y consistencia. Todos estos atributos de los datos en conjunto sugieren una base alineada con los estándares, proporcionando solidez para el análisis.

En este primer apartado, se ha detallado la organización bajo estudio y su conexión con la gestión de datos. Luego se proporcionaron pautas estratégicas guiando la transformación hacia un modelo data-driven.

Concluido el proceso, se detalló el proceso de recopilación de información, enumerando las fuentes utilizadas y evaluando tanto las virtudes como las deficiencias en la calidad. Luego, se describió el proceso de seudonimización diseñado para salvaguardar la confidencialidad de la entidad.

A continuación, se llevará a cabo una evaluación sobre la base de estudio para luego efectuar un análisis de componentes principales.

2. Análisis de Componentes Principales

En un primer paso se busca reducir la dimensionalidad de los datos a analizar. Para ello, en el presente apartado se someterá la base de estudio a un Análisis de Componentes Principales. Este método estadístico, permitirá condensar la información contenida en diferentes aspectos, en un conjunto reducido de componentes principales no correlacionados. Los pasos por seguir en esta evaluación incluirán un análisis descriptivo, la limpieza de outliers, el escalado de los datos, el cálculo de la matriz de covarianza o correlación, la obtención de los autovectores y autovalores, y finalmente, la interpretación de los resultados para identificar patrones subyacentes en los datos y coherencia con el objetivo del estudio.

2.1.1. Limpieza de la base de datos

El análisis inicial y la limpieza de datos surgen como un proceso crítico para garantizar la integridad, confiabilidad y utilidad de la información. Esta actividad abarca tanto la identificación de inconsistencias, valores atípicos y datos faltantes; también se exploran estrategias y metodologías para optimizar su calidad y confiabilidad.

En la base no se detectan valores faltantes. Luego, se evalúa la dispersión de los datos, calculando: medias, medianas, cuartiles, máximos y mínimos (ver Tabla 2.1).

-	ANTIG	VELOC_prom	DIESEL_lts	EURO_lts	TOTAL_lts
Mínimo	0,5	6,86	0	0,9	308
1Q	6	14,66	458,4	701,3	2844
Mediana	9	17,86	2210,9	1311,2	3935
Media	10,09	19,75	3480,3	3380,9	6867
3Q	14	22,23	3667,4	3699,9	6056
Máximo	25	59,01	33588,6	50406,4	60327
-	KM	CONSUMO	CORRECTIVO_usd	CORRECTIVO_cant	PREVENTIVO_usd
Mínimo	315,8	17,63	0	0	105,2
1Q	8997,1	28,45	555,7	5	1306,8
Mediana	11859,7	31,64	1692,4	10	2377,9
Media	21947,5	32,17	3022,2	16,63	4374
3Q	20990,8	34,55	3539,3	22,75	5512
Máximo	173957,1	97,53	32051,3	129	28502,8
-	PREVENTIVO_cant	HS_encendido	HS_ralenti	DIAS_svs	
Mínimo	1	22,14	3,81	8	
1Q	13	486,43	102,05	207	
Mediana	25	761,79	197,4	253,5	
Media	25,36	926	366,68	240,9	
3Q	34	1253,68	499,87	290,8	
Máximo	100	4172,7	2041,2	361	

Tabla 2.1 – Resumen – Elaboración propia

El conjunto de datos resulta heterogéneo, compuesto por múltiples valores y diferentes unidades de medida por estudiar. En algunas variables se observa una alta variabilidad. Por

ejemplo, en los litros de combustibles se registran observaciones muy por encima de la media. Al indagar sobre esos casos puntuales se detectan unidades de Larga distancia, lo cual guarda lógica con los valores, a mayor recorrido: mayor consumo. A continuación, se evaluarán medidas a adoptar con los valores anómalos y subsanar inconvenientes.

2.2. Eliminación de outliers y normalización

La calidad y la utilidad de cualquier análisis de datos dependen en gran medida de la integridad de los datos de entrada. Por lo tanto, es importante abordar cuestiones relacionadas con la limpieza y la preparación de los datos antes de sumergirse en el análisis. En este contexto, este subapartado se enfoca en dos procesos: la eliminación de valores atípicos (outliers) y la normalización de datos. La detección de outliers implica también tratar valores extremos o anómalos valores descriptivos del conjunto (Peña, 2002). La normalización, por otro lado, refiere a la transformación de las variables para unificar las escalas, facilitando la comparación e interpretación.

En la imagen 2.2, se observa el conjunto de boxplots de las variables numéricas, dispuestas en un mismo plano. Cada uno de estos gráficos se centra en un atributo en particular y muestra la distribución de sus datos.

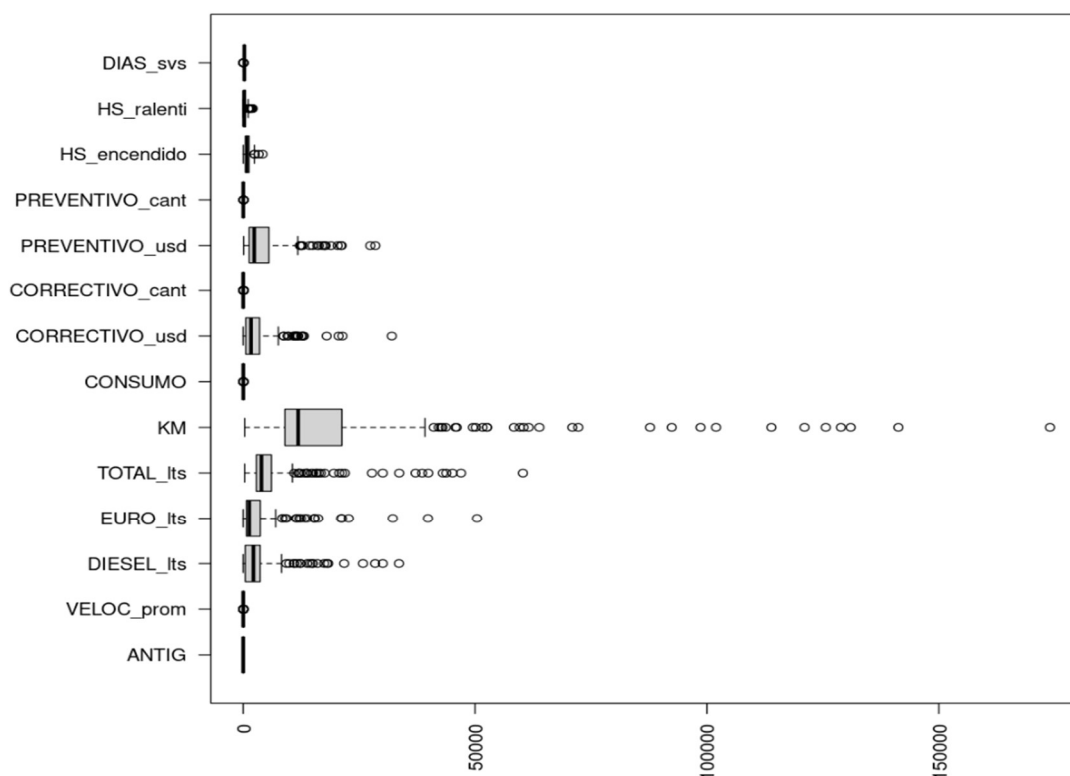


Imagen 2.2 – Boxplot de variables numéricas – Elaboración propia

En el contexto de este análisis de datos, la observación de diferencias en las escalas entre las variables sugiere la posibilidad de como ciertos campos puedan dominar el estudio debido a sus valores absolutos mayores al resto. Un ejemplo de ello son los kilómetros recorridos (KM) o los litros de combustible consumidos. Años, kilómetros por hora, litros, kilómetros o dólares son unidades de medidas diferentes y eso los hace incomparables por definición.

Para abordar este inconveniente, se propone practicar sobre el conjunto de datos, la función `scale` de R. Esta herramienta re-escala el conjunto de atributos restándole la media general del mismo atributo a cada ítem y luego divide el resultado por el desvío estándar. Así, los valores pasan a tener una media de cero y una desviación estándar de uno. Esto asegura tener la misma escala relativa y no deja a ningún atributo domine el estudio según su magnitud.

La Imagen 2.3 presenta una nueva vista de los mismos datos de la Imagen 2.2, con la diferencia de haber sido sometidos al proceso de escalado.

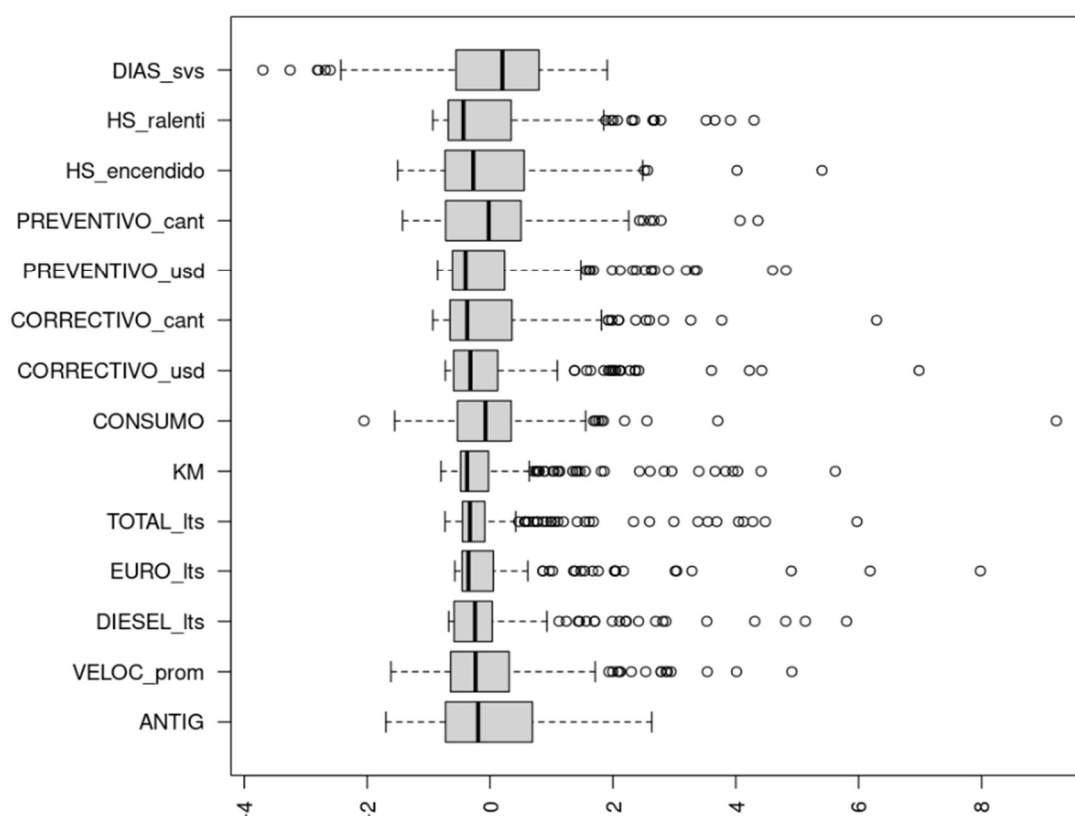


Imagen 2.3 – Boxplot de variables numéricas escaladas – Elaboración propia

La ejecución del procedimiento permite una mejor comparación entre las diferentes características. A través de la Imagen 2.3, se observa la variabilidad en los datos y cómo las diferencias en las escalas se han unificado. Además, esto permite mejorar la convergencia de los algoritmos de aprendizaje automático y mejorar la precisión de eventuales predicciones.

Ante la igualdad de escalas, la existencia de outliers se hace evidente, sin embargo, no se quiere perder la representatividad de los datos. Por ejemplo: hay unidades con gran cantidad de km recorridos. Indagando puntualmente, esos vehículos atienden servicios de acarreo y larga distancia, con lo cual, eliminarlos o truncarlos conlleva quitarle una característica importante a esa actividad.

Para iniciar el análisis detallado de las variables, se aborda un factor crítico como es el "CONSUMO" de combustible. Ese atributo desempeña un rol fundamental en el funcionamiento de las unidades y en la gestión eficiente de recursos. Para comprender en profundidad su distribución y características, se ha optado por esquematizar la variable a través de un histograma. La representación gráfica (Imagen 2.4) permite visualizar cómo se distribuyen los datos y destacar cualquier ocurrencia anómala.

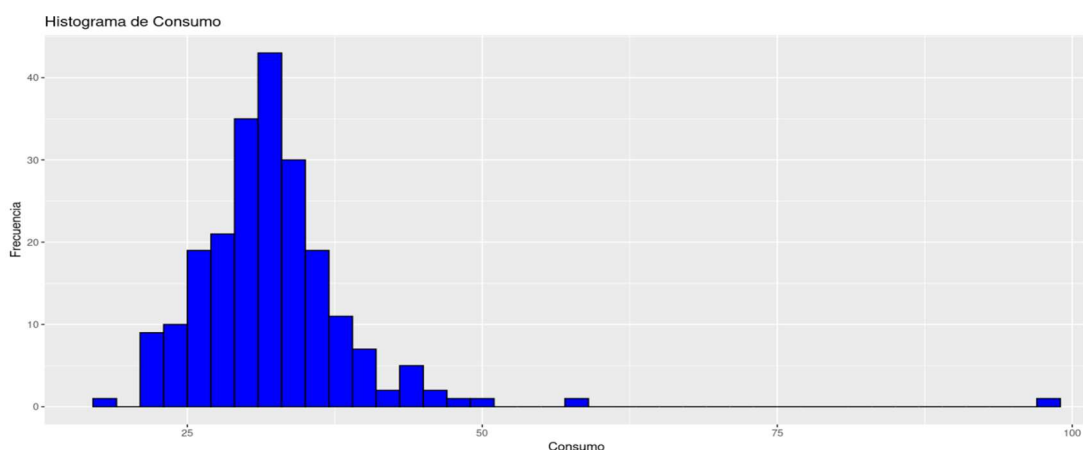


Imagen 2.4 - Histograma de la variable CONSUMO – Elaboración propia

En la figura se advierte la presencia de valores extremos en la variable "CONSUMO". Estos valores atípicos, pueden influir, en cualquier análisis ulterior relacionado con el consumo de combustible. Para los días de servicio, se ejecuta el mismo análisis en la Imagen 2.5

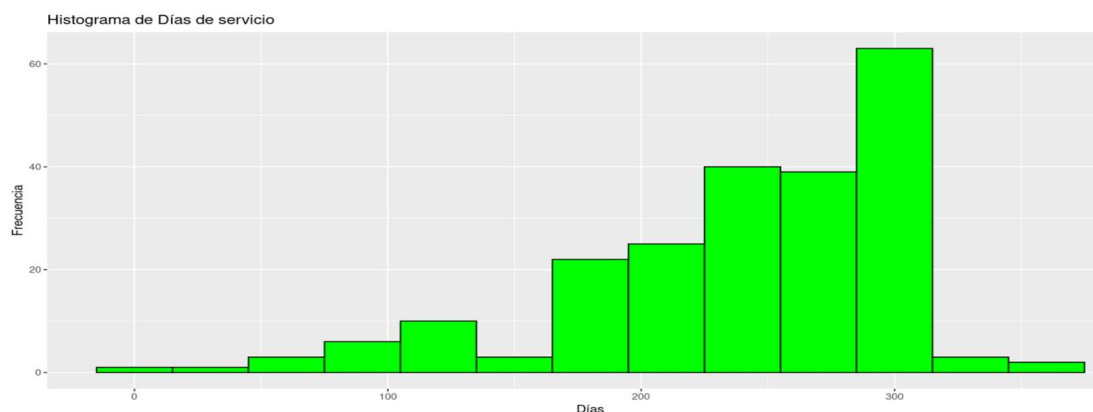


Imagen 2.5 - Histograma de la variable DIAS de servicio – Elaboración propia

En este contexto, se resalta la existencia de observaciones con vehículos de muy pocos días de servicio. Tras un relevamiento detenido, se evidencian como estos registros se corresponden con unidades de reciente incorporación.

En la misma línea, se procede a examinar las variables horas de encendido y horas de ralentí, mediante un análisis gráfico similar (Imagen 2.6)

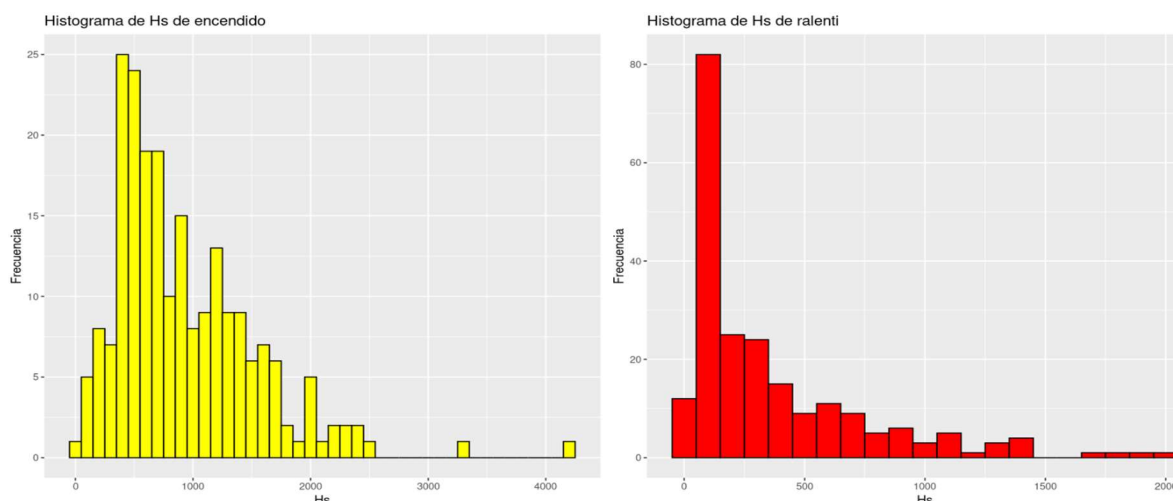


Imagen 2.6 - Histograma HS de encendido y Hs de ralentí. – Elaboración propia

En el gráfico reciente, se pueden identificar en ambos atributos la presencia de outliers con un impacto en la distribución de los datos. Tanto en estas variables como para las anteriores se propone una estrategia para lidiar con estos casos anómalos: sustituirlo por la mediana. Esta decisión es fundamentada por el hecho de que, eliminar la ocurrencia completa implicaría una pérdida de información en un contexto donde los datos son escasos. La elección de la medida, tiene por motivo, la propia resistencia a los valores extremos. Tratándose de distribuciones asimétricas, otra opción como la media, podría ser menos representativa para el conjunto de datos.

Con el reemplazo de outliers y la normalización de la información, se ha realizado un primer paso necesario en la preparación de la base. Al mitigar los efectos de valores extremos y estandarizar las variables, el conjunto de datos es más estable y confiable para el estudio de componentes principales. Todo esto es objeto del próximo subapartado.

2.3. Estudio de Componentes Principales

Actualmente, la base cuenta con nueve variables categóricas y catorce variables numéricas. Mediante un análisis de componentes principales se buscará reducir la dimensionalidad de estos últimos. Así lo explica Peña en su ensayo: “El análisis de componentes principales tiene este objetivo: dadas n observaciones de p variables, se evalúa si es posible representar

adecuadamente esta información con un número menor de variables construidas como combinaciones lineales de las originales.” (Peña, 2002). Como primera acción se revisa la matriz de correlaciones entre los datos en la siguiente imagen:

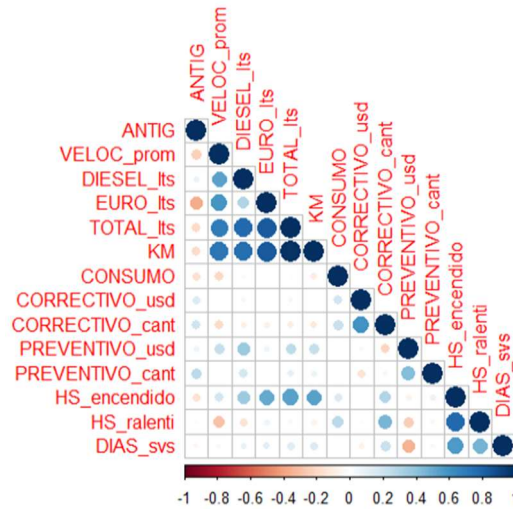


Imagen 2.7 - Matriz de correlación entre variables – Elaboración propia

En la imagen 2.7 se destaca como los km, se relacionan con lógica con los litros de combustible consumidos (ya sea totales, diésel o diésel Premium). A su vez los litros de combustible EURO tienen cierta relación negativa con la antigüedad de las unidades. Esto también tiene coherencia: los motores más viejos, no pueden o no están preparados para aprovechar la funcionalidad de un combustible de mayor octanaje y por tal, trabajan simplemente con diésel. También se lee relación entre las horas de encendido con los km y las horas de ralentí.

Acto seguido, se ejecuta el análisis de componentes principales sobre la matriz de valores escalados, mediante la función PCA. En el gráfico de sedimentación (2.8), se señala cómo cada variable artificial contribuye al total de la varianza explicada.

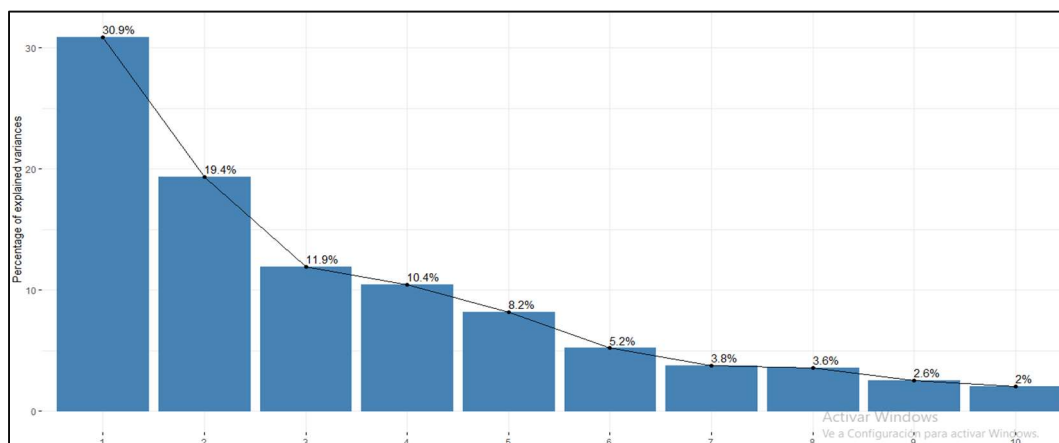


Imagen 2.8 – Grafico del codo – Elaboración propia

El "codo" o "punto de quiebre" se identifica donde la tasa de disminución de la varianza explicada se desacelera significativamente. Ese hito suele ser un método para determinar la cantidad de componentes a adoptar para explicar la variabilidad de los datos. Para este caso no se percibe un quiebre pronunciado de la curva como para inferir a partir del gráfico, la cantidad de componentes para trabajar. Se recurre entonces la tabla de autovalores y a la varianza explicada acumulada Tabla 2.9.

Comp.	Autovalor	% varianza	Varianza acumulada
CP1	4,31960	30,8543%	30,8543%
CP2	2,70930	19,3525%	50,2068%
CP3	1,66890	11,9210%	62,1278%
CP4	1,46200	10,4426%	72,5704%
CP5	1,14720	8,1945%	80,7649%
CP6	0,72980	5,2131%	85,9780%
CP7	0,53140	3,7961%	89,7741%
CP8	0,50320	3,5939%	93,3680%
CP9	0,35790	2,5567%	95,9247%
CP10	0,28610	2,0433%	97,9680%
CP11	0,24940	1,7817%	99,7497%
CP12	0,02870	0,2051%	99,9548%
CP13	0,00630	0,0450%	99,9998%
CP14	0,00000	0,0002%	100,0000%

Tabla 2.9 – Varianza explicada marginal y acumulada - Elaboración propia

En la tabla se destaca el aporte de la CP6, como un equilibrio entre la cantidad de información retenida y la simplicidad del modelo. Para representar esta información visualmente, se recurre al gráfico de varianza acumulada (Imagen 2.10):

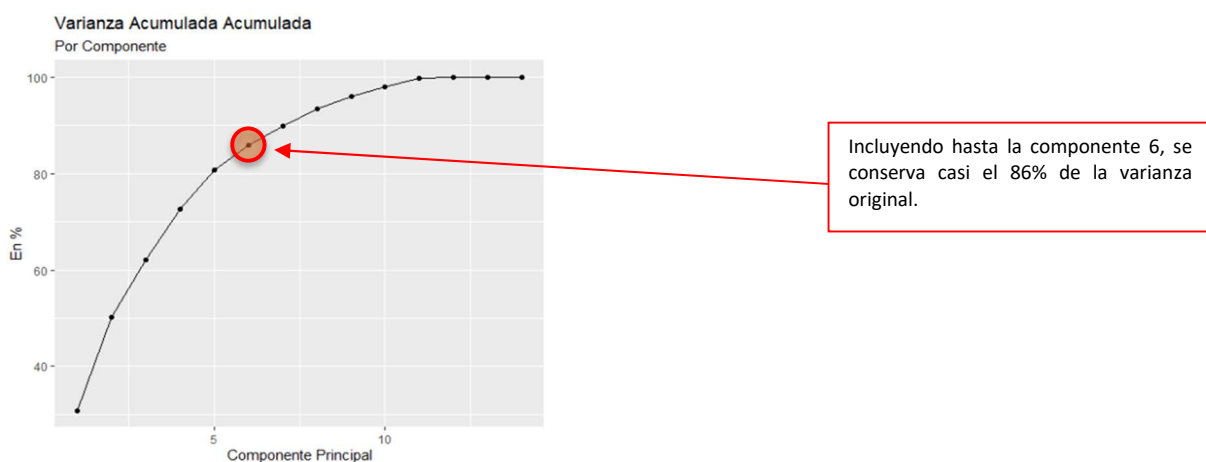


Imagen 2.10 – Varianza Acumulada – Elaboración propia

Visualmente, es más claro evaluar la varianza acumulada (Imagen 2.10) que el aporte marginal de las componentes al observar el gráfico del codo imagen 2.8. A partir del

componente 6, se identifica un punto aceptable de equilibrio entre el aporte de cada componente y la capacidad del modelo para explicar la varianza en los datos.

Para comprender mejor la estructura de las componentes principales y su relación con las variables originales, se estudia las cargas de estas variables sobre las componentes sintéticas creadas y seleccionadas. La Tabla 2.11 proporciona una visión detallada de esta relación.

-	CP1	CP2	CP3	CP4	CP5	CP6
ANTIG	-0,2574	-0,0397	0,6591	0,1548	-0,4999	-0,1757
VELOC_prom	0,7611	-0,3029	-0,0955	-0,2103	-0,1425	0,0931
DIESEL_Its	0,7508	-0,1434	0,3489	0,0291	-0,0955	-0,4568
EURO_Its	0,8272	0,0879	-0,2286	-0,1261	0,0962	0,3096
TOTAL_Its	0,9813	-0,0256	0,0513	-0,0674	0,0085	-0,0598
KM	0,9813	-0,0508	0,0017	-0,0549	-0,0891	-0,017
CONSUMO	-0,0239	0,3532	0,1778	-0,1629	0,7673	-0,384
CORRECTIVO_usd	-0,0063	0,276	0,4763	-0,7389	-0,0576	0,1327
CORRECTIVO_cant	-0,1102	0,6742	0,4405	-0,3701	-0,1092	0,184
PREVENTIVO_usd	0,2771	-0,4807	0,5021	0,2158	0,3367	0,1269
PREVENTIVO_cant	0,1215	-0,1404	0,5487	0,598	0,1324	0,3045
HS_encendido	0,5988	0,7108	0,011	0,2646	0,0106	0,0463
HS_ralenti	0,0168	0,8676	0,0247	0,3141	0,1236	0,1392
DIAS_svs	0,1925	0,6484	-0,2058	0,2951	-0,3166	-0,2295

Tabla 2.11 – Relación entre Variables originales y Componentes principales.

Se destacan en colores las correlaciones más preponderantes. Siguiendo la misma paleta, se estudia el aporte de las variables sobre cada componente en un gráfico de barras. Las primeras dos componentes retienen el 50% de la varianza explicada:

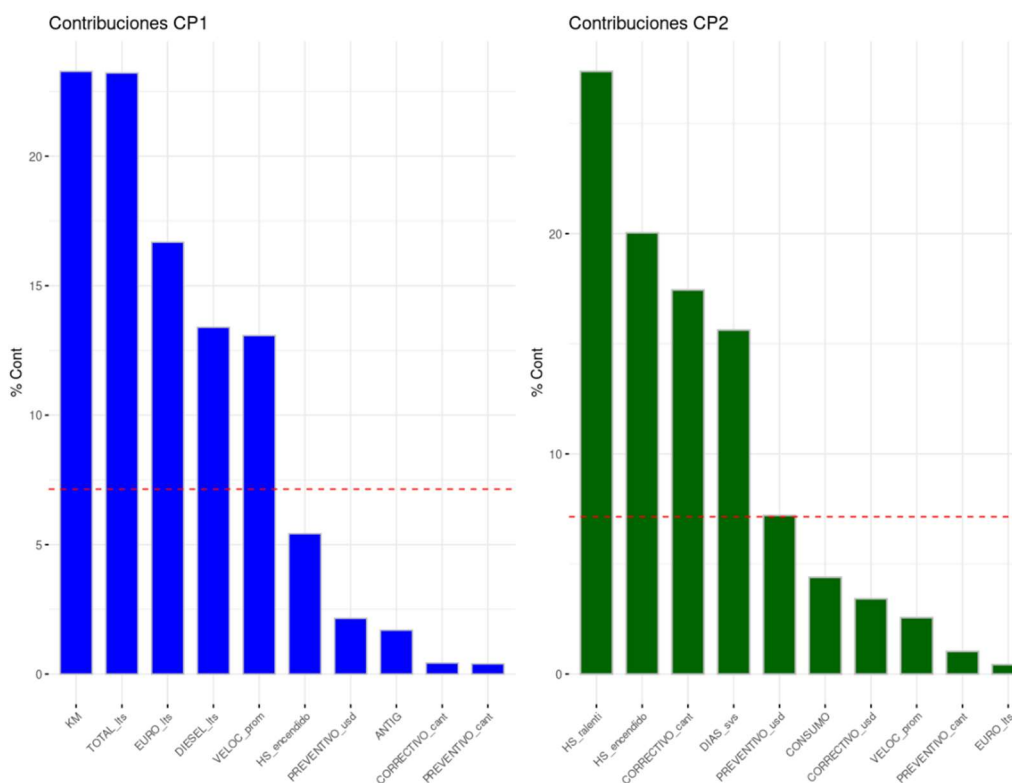


Imagen 2.12 – Contribución de variables a CP1 y CP2 – Elaboración propia

La Componente Principal 1 (CP1) asocia aspectos relacionados con el funcionamiento de las unidades. Kilómetros recorridos y los litros totales de combustible consumidos, atienden con el mayor aporte a la nueva variable. Le siguen en menor proporción: el consumo por tipo de combustible y la velocidad promedio. Esta componente parece resumir la eficiencia operativa general de los vehículos.

Por otro lado, la Componente Principal 2 (CP2) está compuesta por las horas de ralentí, las horas de encendido y los eventos de mantenimiento correctivo. Las Componentes 3 y 4 retienen conjuntamente el 22% de la varianza explicada total ilustradas en la Imagen 2.12.

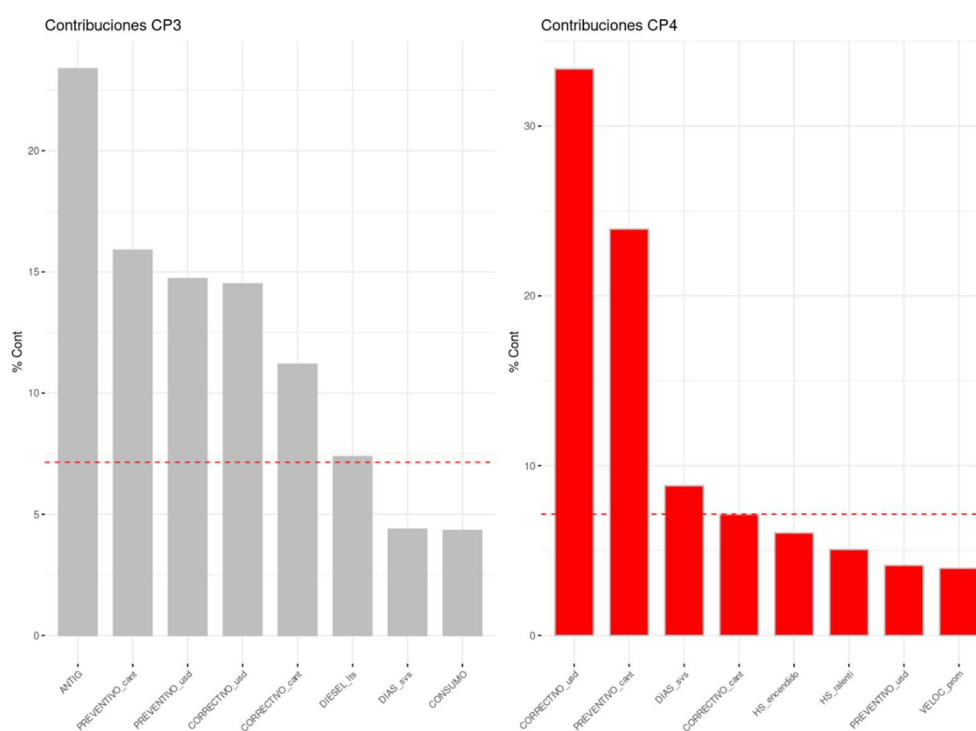


Imagen 2.13 – Contribución de variables a CP3 y CP4 – Elaboración propia

La Componente Principal 3 (CP3) está vinculada a la antigüedad de los vehículos, y a diferentes atributos vinculados al mantenimiento tanto preventivos como correctivo. Un hallazgo es especialmente revelador; se relaciona estrechamente con uno de los objetivos centrales de este trabajo: la antigüedad se correlaciona estadísticamente con el mantenimiento de vehículos.

Por último, la Componente Principal 4 (CP4) muestra una fuerte asociación con los indicadores de mantenimiento correctivo y preventivo. En menor medida, también se puede subrayar un aporte menor por parte de la variable Días de servicio. Finalmente, en el gráfico siguiente se expondrán las últimas componentes por estudiar 5 y 6 con casi un 14% de aporte a la varianza total explicada.

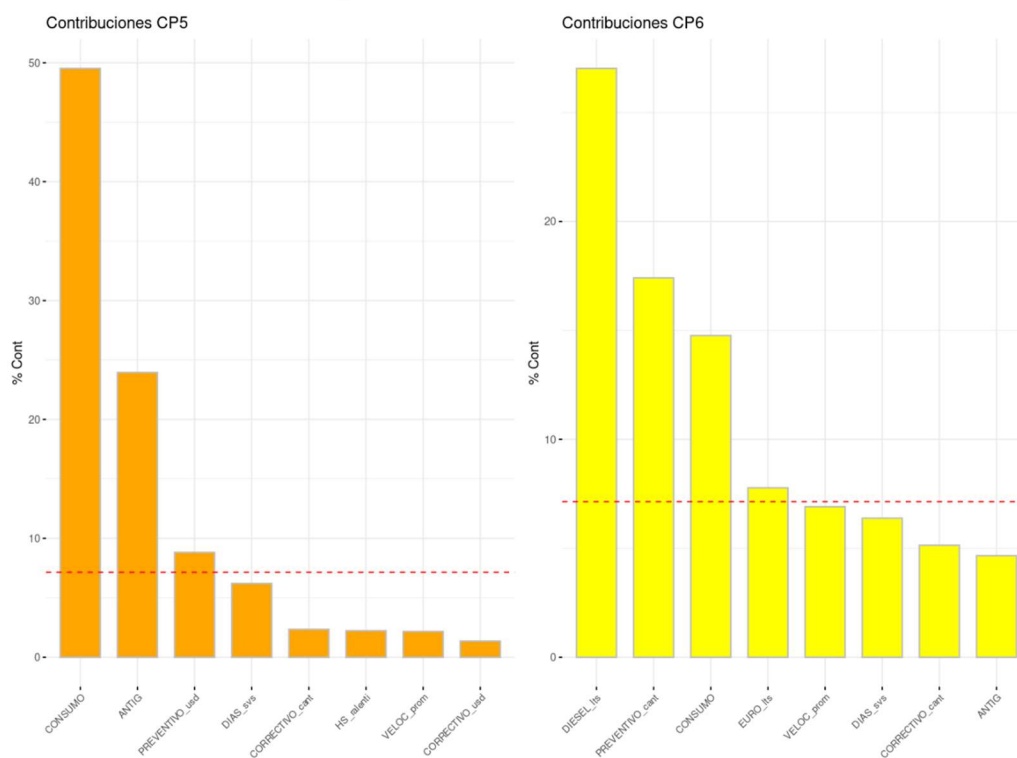


Imagen 2.14 – Contribución de variables a CP5 y CP6 – Elaboración propia

La Componente Principal 5 (CP5) vincula aspectos como el consumo, la antigüedad y el mantenimiento preventivo. La Componente Principal 6 (CP6) contribuye relativamente poco a la varianza acumulada. Está asociada con el aspecto del combustible, incluyendo el consumo promedio, los litros de diésel y los litros de diésel-Premium (euro).

Esta sección, comenzó con un análisis descriptivo de la base de estudio, compuesta por 218 observaciones y dividida en 9 variables categóricas y 13 numéricas. Se abordó una limpieza de datos, eliminando valores atípicos y normalizando las variables, asegurando una escala homogénea para un estudio más efectivo.

A continuación, se aplicó un PCA para reducir la dimensionalidad de la base, identificar patrones y seleccionar los atributos más significativos. Entre las conclusiones se llegó hasta la adopción de la CP6 destacándose como un punto óptimo de equilibrio entre la cantidad de información retenida y la simplicidad del modelo. De esa exploración de las cargas de las variables en las componentes principales, se destaca que la CP3 y CP4 señalan una relación estadística entre la antigüedad de los vehículos y diversos atributos de mantenimiento. Esto resulta interesante resaltarlo a los fines del objetivo general del estudio.

3. Influencia de la antigüedad en el Mantenimiento Correctivo

A continuación, para abordar el disparador planteado, se llevará a cabo un análisis multivariado de la varianza. Estos, se centrarán específicamente en la cantidad de eventos y los gastos en dólares relacionados con el mantenimiento correctivo incurridos por cada unidad.

El Análisis de Varianza Multivariante (MANOVA) es una extensión del ANOVA. Permite, a diferencia de este último, analizar más de una variable dependiente. En este caso, el MANOVA podría proporcionar una comprensión más completa de cómo la variable categórica "Antigüedad" influye en las dos variables dependientes mencionadas. No solo examinaría si hay diferencias en las medias de cada variable por separado; también permitiría inferir en qué sentido se da esa diferencia. Si el MANOVA indica diferencias significativas, se podría concluir que los diferentes niveles de antigüedad tienen un impacto conjunto en las dos variables de mantenimiento correctivo, proporcionando una visión integral de la relación entre estas variables.

3.1. Clasificación de factores

El objetivo de este trabajo fue verificar la influencia de la antigüedad de flota en los costos de mantenimiento correctivo. A los fines del análisis, se practicó una clasificación adicional de la base de datos en función de la variable ANTIG, dando como resultado:

ANTIG_c	Antigüedad
1-0km	0 a 3 años
2-MEDIO	4 a 9 años
3-VIEJO	10 a 15 años
4-MUY VIEJO	+16 años

Tabla 3.1 – Conversión Antigüedad en variable categórica - Elaboración propia

La tabla 3.1 presenta la clasificación de la antigüedad de los camiones de carga en cuatro categorías distintas. La primera categoría, denominada "1-0km", abarca camiones nuevos con una antigüedad de 0 a 3 años. La segunda categoría, "2-MEDIO", comprende vehículos con una antigüedad de 4 a 9 años. La tercera categoría, "3-VIEJO", incluye camiones con una antigüedad de 10 a 15 años. La última categoría, "4-MUY VIEJO", engloba aquellos camiones con una antigüedad de más de 16 años.

3.2. Análisis Multivariante de la varianza

Con los atributos de interés normalizados para el proceso de Componentes principales se efectúa el análisis multivariado de la varianza. El análisis estadístico, se ejecutará mediante la biblioteca statsmodels de R, donde se emplea el ajuste “manova”. Este tipo de estudio es eficaz cuando se tienen múltiples variables dependientes y se busca evaluar las diferencias entre los grupos de una o más variables categóricas. Más tarde, la función “aov” es accionada para realizar un ANOVA, proporcionando estadísticas detalladas sobre la variabilidad entre los grupos. Por último, se llevará a cabo comparaciones entre los grupos, permitiendo realizar pruebas post hoc e identificar así diferencias significativas después del análisis de la varianza. En la siguiente planilla se observa un resumen del MANOVA sobre las variables dependientes Cantidad de eventos correctivos y Costo en dólares de mantenimiento correctivo.

Test	Df	Value	approx_F	num_Df	den_Df	Pr(>F)
Wilks	3	0,933340	2,4919	6	426	0,022180
Hotelling-Lawley	3	0,070719	2,4987	6	424	0,021850
Roy	3	0,058671	4,1852	3	214	0,006637
Pillai	3	0,067324	2,4849	6	428	0,022520
Residuals	214					

Tabla 3.2 – Test MANOVA sobre variables CORRECTIVO_usd y CORRECTIVO_cant – Elaboración propia

Según la tabla 3.2, un valor p bajo en las pruebas de Wilks’ Lambda, Pillai’s Trace, Hotelling-Lawley Trace y Roy’s Greatest Root sugieren que hay una diferencia estadísticamente significativa entre las medias de los grupos conformados a partir de la antigüedad de los vehículos.

Es decir, es muy poco probable que las diferencias observadas en el conjunto de datos hayan ocurrido por casualidad. Por lo tanto, se rechaza la hipótesis nula de que no hay diferencia entre los grupos en las variables dependientes.

3.3. Análisis de la varianza

La prueba MANOVA evalúa las diferencias multivariadas; considera todas las variables dependientes juntas. Si el resultado del estudio es significativo, revela la existencia de diferencias en el vector de medias de las variables dependientes. Sin embargo, este estudio no sugiere qué variables dependientes específicas son responsables de las diferencias.

Para corroborar que el efecto de una variable no este camuflada en otra, se realiza entonces un modelo ANOVA sobre la variable de Costos en dólares de mantenimiento correctivo:

Anova	Df	Sum_Sq	Mean_Sq	F_value	Pr(>F)
ANTIG_c	3	9,180000	3,0589	3,15	0,025900
Residuals	214	207,820000	0,9711	NaN	NaN

Tabla 3.3 – Salida estudio ANOVA sobre variables CORRECTIVO_usd

El p-valor menor a 0.05 de significancia (0.0259) de la Tabla 3.3 en la salida, muestra evidencia de una diferencia de medias estadísticamente significativa para la variable dependiente CORRECTIVO_usd (costo en dólares en mantenimiento), entre los grupos de antigüedad. El mismo estudio se ejecuta en forma independiente sobre la variable CORRECTIVO_cant:

Anova	Df	Sum_Sq	Mean_Sq	F_value	Pr(>F)
ANTIG_c	3	10,890000	3,6300	3,769	0,011500
Residuals	214	206,110000	0,9630	NaN	NaN

Tabla 3.4 – Salida estudio ANOVA sobre variables CORRECTIVO_cant

A la salida del estudio (Tabla 3.4) se observa un p-valor bajo (0.0115); sugiriendo también diferencia de medias entre las categorías de la variable.

La presencia de p-valores bajos en ambas pruebas ANOVA nos resalta, no solo las diferencias entre las medias, también se traduce en que no hay un enmascaramiento de tendencias inter-sujeto. Esto refuerza la idea de que la antigüedad tiene un impacto independiente y diferenciado en los dos aspectos evaluados.

A continuación, la prueba de Tukey es utilizada después de un ANOVA para abordar la siguiente pregunta: ¿Cuáles grupos específicos difieren significativamente entre sí?

-	1-0km	2-MEDIO	3-VIEJO
2-MEDIO	0,349	-	-
3-VIEJO	0,016	0,370	-
4-MUY VIEJO	0,273	1,000	1,000

Tabla 3.5 – Prueba TUKEY sobre la variable CORRECTIVO_usd

El resultado de la prueba Tukey para el ANOVA entre antigüedad y el costo de mantenimiento correctivo (Tabla 3.5), resulta en rechazar la hipótesis nula para la comparación entre 1-0km y 3-VIEJO porque el valor p ajustado es menor que el umbral

(0.05). Esto indica la existencia de suficiente evidencia muestral para determinar la diferencia de las medias entre la categoría 3-VIEJO y el resto.

El mismo algoritmo se ejecuta sobre el ANOVA evaluando sobre la cantidad de eventos de mantenimiento correctivo. La salida se expone en la siguiente imagen:

-	1-0km	2-MEDIO	3-VIEJO
2-MEDIO	0,103	-	-
3-VIEJO	0,020	1,000	-
4-MUY VIEJO	0,013	1,000	1,000

Tabla 3.6 – Prueba TUKEY sobre la variable CORRECTIVO_cant

En este informe específico (Tabla 3.6), se observa el rechazo de la hipótesis nula para las comparaciones entre 1-0km y 3-VIEJO, y entre 1-0km y 4-MUY VIEJO (0.020 y 0.013 respectivamente). Habiendo confirmado la diferencia entre las medias de las categorías, se compararán las medias de cada una en la Tabla 3.7:

Antigüedad	Cantidad de vehículos	Prom. Gastos correctivos en usd	Promedio de eventos correctivos
1-0km	24	1.110,8	6,5
2-MEDIO	101	2.882,2	16,1
3-VIEJO	42	4.295,4	19,8
4-MUY VIEJO	51	3.150,5	19,9
Total general	218	3.022,2	16,6

Tabla 3.6– Media en dólares y cantidad de eventos correctivo por antigüedad - Elaboración propia

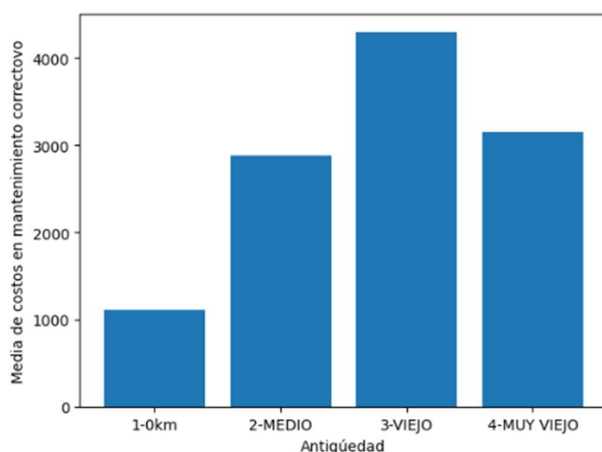


Imagen 3.1 – Media de costos por mantenimiento correctivo por Antigüedad – Elaboración propia

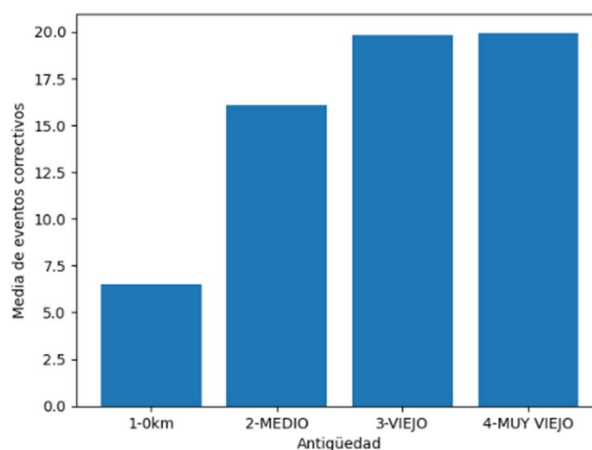


Imagen 3.2 – Media de visitas por mantenimiento correctivo por Antigüedad – Elaboración propia

En la Tabla 3.7, Imagen 3.1 e Imagen 3.2 confirman valores mínimos en la categoría "0km" (vehículos con hasta 2 años de antigüedad) tanto para la cantidad de eventos (6,5 eventos) como para los costos en dólares (u\$s 1110). En marcado contraste, los valores máximos se

encuentran en las categorías "Viejo" (u\$s 4.295 para en promedio de costos en dólares) y "Muy viejo" (19,9 para cantidad de eventos). Estos hallazgos respaldan la hipótesis inicial que postulaba como el aumento de la antigüedad, replica también en mayores costos y una suba en la frecuencia de visitas al taller por mantenimiento correctivo.

Con solo ver la tabla y los gráficos, se puede observar el vínculo directo entre la antigüedad y los costos y la cantidad de eventos de mantenimiento correctivo. Los ejercicios de MANOVA/ANOVA, demostró cómo estas tendencias observadas son estadísticamente significativas en la muestra analizada.

Al pasar a los números de este estudio se afirma: tomando las categorías de interés, se deja ver que las unidades clasificadas como Viejo gastan u\$s 3.184 más que una unidad 0km en mantenimiento correctivo. Además, la lectura nos enseña también, una unidad vieja (entre 10 y 15 años) o muy vieja (+16 años) tienen en promedio 13 visitas al taller más, respecto a una unidad nueva.

Habiendo repasado ya las técnicas planteadas al inicio del trabajo y los resultados obtenidos, en el siguiente apartado pasaremos resumir y presentar las conclusiones generales del estudio.

Conclusiones generales

El trabajo estudia el caso de un operador logístico con sede en el Gran Buenos Aires, pretendiendo demostrar la conexión entre la antigüedad de la flota y los costos de mantenimiento correctivo. En el análisis de la organización, se identificaron diferentes problemáticas y limitaciones relacionadas con la gestión de datos. La primera conclusión extraída es: LogiCon no aprovecha completamente los datos disponibles. Por ello, a pesar de contar con variadas fuentes de información, como registros de mantenimiento o seguimiento satelital, la organización no ha asignado recursos significativos ni personal especializado para ejecutar un análisis profundo. La falta de reconocimiento a su importancia limita el desarrollo de estrategias de diferenciación dentro de la industria. Es decir, los esfuerzos hoy se centran en analíticas tradicionales (descripción y diagnóstico), en lugar de explorar las posibilidades de la analítica avanzada (predictivo y prescriptivo).

Luego se presentaron los lineamientos que conducen la transformación de la organización hacia un enfoque basado en datos. Esto involucra desde la política de privacidad hasta una redefinición de la gerencia de IT y un nuevo esquema de gobierno de datos.

Finalmente, se detalló el proceso de consolidación de información, al tiempo de evaluar su calidad. Completa la construcción de la base de estudio, se llevó a cabo la anonimización de datos críticos utilizando una técnica de seudonimización para preservar la confidencialidad.

En un segundo apartado, se realizó una revisión estadística utilizando el software R. En el proceso se identificaron valores atípicos e información faltante. También se evaluaron los tipos de datos, la variabilidad y la propia distribución; incluyendo medidas como medias, medianas, valores máximos y mínimos. Este análisis inicial proporcionó una comprensión profunda de la naturaleza de la base de estudio.

Con el propósito de reducir la complejidad del dataframe, se aplicó un análisis de Componentes Principales (ACP). El ACP permitió representar de manera efectiva la información de múltiples observaciones y variables mediante un conjunto más reducido de atributos combinados linealmente. Este enfoque simplificó la estructura de la información y resaltó algunos patrones relevantes.

El estudio permitió disminuir una tabla de 14 variables numéricas a 6 dimensiones, rescatando casi el 86% de la varianza acumulada original. La CP1 concentra aspectos de la unidad motora en movimiento con correlación entre las variables con mayor participación: velocidad promedio, kilómetros recorridos, litros de combustible consumidos y horas de encendido. La CP2 agrupa principalmente horas de ralentí y encendido. De la exploración de las cargas

también se observó como la CP3 y CP4 señalan una primera relación estadística entre la antigüedad de los vehículos y diversos atributos de mantenimiento, lo que resulta interesante de indagar a los fines generales del trabajo. La quinta y sexta se asocian más fuertemente a los aspectos de consumos de combustible (tanto promedio como en litros).

En la sección final, como primer paso se clasificó la base en función de la variable antigüedad, en un nuevo atributo de características categóricas. Luego, se describió la ejecución de un estudio estadístico mediante MANOVA y ANOVA.

Este análisis se focalizó en: la cantidad de eventos y los gastos en dólares vinculados al mantenimiento correctivo. Así, las variables fueron evaluadas respecto al factor generado a partir de la antigüedad de vehículos. El resultado permitió discernir la presencia de diferencias significativas entre las medias de las categorías consideradas, arrojando luz sobre la relación entre estos factores y las características dependientes.

Es decir, existe suficiente evidencia estadística para asegurar que las diferencias observadas en el conjunto de datos no son producto de la casualidad. Por lo tanto, se rechaza la hipótesis nula de que no hay diferencia entre los grupos en las variables dependientes.

En relación con el impacto de la antigüedad en los costos y la cantidad de visitas al taller para mantenimiento correctivo, se ha identificado la existencia de una relación positiva entre ambos. Desde la variable medida en dólares, se confirma para este conjunto de datos: las unidades clasificadas como Viejo gastan en promedio u\$s 3.184 más que una unidad 0km en mantenimiento correctivo. En cantidad de situaciones, se observa como las unidades mayores a 10 años tienen en promedio 13 eventos correctivos más que una unidad 0 km.

Este comportamiento puede atribuirse al desgaste natural de los componentes, la obsolescencia tecnológica y la consecuente mayor probabilidad de fallas.

El incremento de situaciones de mantenimiento correctivo, tiene como consecuencia la posibilidad de obstaculizar la eficiencia de la gestión operativa. Desde la perspectiva económica impacta con aumento en los egresos por el gasto directo de la reparación: la mano de obra dedicada, piezas de repuesto y otros conceptos relacionados. Estos costos no previstos ejercen presión sobre el presupuesto afectando la rentabilidad del negocio.

El impacto financiero no se detiene allí: la inmovilización de un vehículo afecta, no solo su capacidad para cumplir con las entregas y rutas programadas; también crea un desajuste en la cadena logística en su conjunto. Los plazos pueden retrasarse, derivando en reprogramaciones y reasignaciones de recursos aumentando los costos. Además, estas interrupciones pueden alterar negativamente los índices de atención al cliente, erosionando la confiabilidad y la reputación de la empresa en el mercado.

Estas conclusiones suman potencia con evidencia empírica, a los argumentos que plantean los operadores logísticos en las discusiones tarifarias con los dadores de carga. Es decir, al considerarse una flota más antigua de vehículos, el cliente debiera reconocer como parte de tarifa, mayores costos destinados al mantenimiento correctivo.

Este trabajo, también deja entrever un camino prometedor hacia futuras investigaciones o implementaciones en materia de datos. Por poner algunos ejemplos, se puede indagar en la propuesta de Oliveros, donde utiliza modelos no paramétricos para analizar la eficiencia en la distribución de productos en una empresa procesadora de alimentos. A partir de allí, describe una herramienta flexible para evaluar el desempeño en redes de suministro, ofreciendo un análisis detallado que puede mejorar la toma de decisiones. (Oliveros, 2015)

En materia de revisiones y cuidado sobre los equipos, Killeen desarrolla “un producto de aprendizaje automático semi supervisado para mejorar la selección de características de sensores [...] Este MVP es la base de nuestra investigación en aprendizaje automático y análisis de datos para el mantenimiento predictivo.” (Killeen Patrick, 2019). Mantener un enfoque proactivo o preventivo en la gestión del mantenimiento, no solo mejora la confiabilidad de las entregas, también sirve como estrategia para reducir los riesgos asociados con fallas inesperadas en la flota. Este paradigma ofrece un camino a la eficiencia operativa y despeja una oportunidad valiosa para optimizar los costos a largo plazo. Nuevas perspectivas que abren avenidas de investigación, donde se promueve la continua evolución y mejora en la gestión de flotas logísticas.

Bibliografía

- Castiblanco, Y. L. (2015). *Análisis de las Características y aplicaciones de los sistemas de ruteo de vehículos*. Bogotá: Universidad Militar Nueva Granada. Obtenido de <https://repository.unimilitar.edu.co/handle/10654/13308>
- Cerquitelli, T. (2021). Mantenimiento predictivo basado en datos: una introducción a la metodología. En N. N. Tania Cerquitelli, *Negocios y Administración*. Singapur: Springer. doi:https://doi.org/10.1007/978-981-16-2940-2_3
- Deloitte. (2017). La evolución del rol del Chief Data Officer en los Servicios Financieros. *De Guardian y Administrador a Estrategia de Negocios*, 1-14.
- Federación Argentina de Entidades Empresarias del Autotransporte de Cargas. (2022). *Departamento de Estudios Económicos y de Costos*. Obtenido de FADEEAC: <https://www.fadeeac.org.ar/estudios-economicos-y-costos/>
- Herrera Pablo, y. H. (22 de Julio de 2023). Opacidad y nitidez epistémica en la economía de plataformas. Organizaciones que las producen sostienen y recrean. *Ciencias Administrativas*, 1-10. doi:<https://doi.org/10.24215/23143738e118>
- Instituto Nacional de Estadística y Censo. (2023). *Índice de precios al consumidor (IPC)*. Técnico, Indec, Ciudad Autónoma Buenos Aires. Obtenido de https://www.indec.gob.ar/uploads/informesdeprensa/ipc_01_23891D383E4F.pdf
- Killeen Patrick, D. B. (29 de Abril de 2019). IoT-based predictive maintenance for fleet management. *ScienceDirect*, 607-613. doi:<https://doi.org/10.1016/j.procs.2019.04.184>
- Lawton, G. (5 de febrero de 2020). *Funciones y responsabilidades del gobierno de Datos*. Obtenido de TechTarget: <https://www.techtarget.com/>
- Mena Madrid, D. A. (2021). *Servicio Web para la Anonimización y Pseudonimización de Datos Estructurados*. Universidad de San Francisco de Quito (USFQ), Quito.
- Müller, A. (Abril - Junio de 2004). Reforma y privatización en el sector transporte: reseña y balance preliminar. *Desarrollo Económico*, 44(57), 57-83. doi:<https://doi.org/10.2307/3455867>
- Oliveros, C. (2015). *Análisis no paramétrico de la eficiencia de redes de distribución logística*. Obtenido de https://www.researchgate.net/profile/Manuel-Camacho-Oliveros/publication/281685090_Nonparametric_Analysis_of_the_Efficiency_of_Logistic_Distribution_Networks/links/55f449d008ae6a34f6608c2f/Nonparametric-Analysis-of-the-Efficiency-of-Logistic-Distribution-
- Parlamento Europeo. (27 de Abril de 2016). Reglamento (UE) 2016/679. *Relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)*. (D. O. Europea, Ed.) Europa. Obtenido de <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32016R0679>
- Penn, C. S. (8 de Agosto de 2016). *Christopher S. Penn – Marketing AI Keynote Speaker*. Obtenido de The Evolution of the Data-Driven Company: <https://www.christopherspenn.com/2019/08/the-evolution-of-the-data-driven-company/>
- Peña, D. (2002). *Análisis de Datos Multivariantes*. España: McGraw-Hill Interamericana.
- República Argentina. (4 de Octubre de 2000). Ley N°25326. *Protección de datos personales*. Ciudad Autónoma de Buenos Aires, Argentina: Boletín Oficial. Obtenido de <https://www.argentina.gob.ar/normativa/nacional/ley-25326-55604>
- Salaberri, N. (13 de Septiembre de 2021). Gestión de la Privacidad de los datos personales: el modelo de privacidad diferencial. *Revista de Investigación en Modelos Matemáticos Aplicados a la Gestión y la Economía*, 20-37.

Shapiro, C. &. (2000). *El dominio de la información. Una guía estratégica para la economía de la red*. Antoni Bosch Editor.

Apéndices

Acompañan al presente informe los siguientes enlaces:

- Fuente: Módulo de mantenimiento
 - ⇒ Equipos → [Equipos.xlsx](#)
 - ⇒ Informe de reparaciones → [Informe_Reparaciones.xlsx](#)
 - ⇒ Informe de consumos → [Informe_consumos.xlsx](#)
- Fuente: Sistema de seguimiento satelital
 - ⇒ Actividad periódica → [Actividad periódica.xlsx](#)
- Base de datos en formato csv → [BRITTI_TFI_db.csv](#)

A continuación se expone también el código en software R que analizan los datos.

```
#####  
# TRABAJO PRACTICO DE INTEGRACIÓN FINAL #  
# Especialización en Métodos Cuantitativos #  
# para la Gestión y Análisis de Datos en Organizaciones #  
# FACULTAD DE CIENCIAS ECONÓMICAS #  
# UNIVERSIDAD DE BUENOS AIRES #  
#####  
## Actividad: Analisis Empresa Logística  
## Año Lectivo: 2023  
## Tutor: Veronica García Fronti  
## Estudiante: Britti, Luis Eduardo
```

```
install.packages("FactoMineR")  
install.packages("factoextra")  
install.packages("NbClust")  
install.packages("psych")  
install.packages("openxlsx")  
install.packages("cluster")  
install.packages("gridExtra")  
install.packages("car")  
install.packages("MANOVA.RM")  
install.packages("ggdendro")  
install.packages("RColorBrewer")
```

```
library(FactoMineR)  
library(factoextra)  
library(NbClust)  
library(psych)  
library(FactoMineR)  
library(openxlsx)  
library(cluster)  
library(gridExtra)  
library(MANOVA.RM)  
library(car)  
library(corrplot)  
library(ggdendro)  
library(cluster)  
library(ggplot2)  
library(RColorBrewer)
```

```
#-----  
#####  
### SE CARGA LA BASE COMPLETA ###  
#####  
  
setwd("C:/Trabajo Final")  
  
base<-read.table(file.choose(),header=TRUE,sep=";",dec=".",row.names=1)  
s<-summary(base)  
str(base)  
s  
  
names(base)  
  
datos<-base[, 8:21]  
summary(datos)  
names(datos)  
  
grafico_barras <- ggplot(base, aes(x = ANO)) +  
  geom_bar() + # Agrega las barras  
  labs(title = "Gráfico de Barras para la Variable ANO", x = "Año", y =  
"Frecuencia")  
grafico_barras  
  
boxplot(datos, las = 2, horizontal = TRUE, main = "Boxplot original de variables  
numéricas")  
  
sapply(datos, var)  
  
plot1 <- ggplot(datos, aes(x = CONSUMO)) +  
  geom_histogram(binwidth = 2, fill = "blue", color = "black") + # Ajusta el ancho  
de los bins y los colores  
  labs(title = "Histograma de Consumo", x = "Consumo", y = "Frecuencia")  
plot1  
  
plot2 <- ggplot(datos, aes(x = DIAS_svs)) +  
  geom_histogram(binwidth = 30, fill = "green", color = "black") + # Ajusta el  
ancho de los bins y los colores  
  labs(title = "Histograma de Días de servicio", x = "Días", y = "Frecuencia")  
plot2  
  
plot3 <- ggplot(datos, aes(x = HS_encendido)) +  
  geom_histogram(binwidth = 100, fill = "yellow", color = "black") + # Ajusta el  
ancho de los bins y los colores  
  labs(title = "Histograma de Hs de encendido", x = "Hs", y = "Frecuencia")  
  
plot4 <- ggplot(datos, aes(x = HS_ralenti)) +  
  geom_histogram(binwidth = 100, fill = "red", color = "black") + # Ajusta el  
ancho de los bins y los colores  
  labs(title = "Histograma de Hs de ralenti", x = "Hs", y = "Frecuencia")  
  
grid.arrange(plot3, plot4, ncol=2)  
  
### se escala la base para normalizar  
  
datos_s<-as.data.frame(scale(datos))  
boxplot(datos_s, las = 2, horizontal = TRUE, main = "Boxplot original de variables  
escaladas")  
names(datos_s)  
summary(datos_s)  
  
datos
```

```
#-----
#####
### TRATAMIENTO DE VALORES ATÍPICOS###
#####

datos_sc<-datos_s
names(datos_sc)

# Reemplazar los valores atípicos por un valor máximo

median(datos_sc$CONSUMO)
### CONSUMO
datos_sc[datos_sc$CONSUMO>2.5,]
datos_sc$CONSUMO <- ifelse(datos_sc$CONSUMO>2.5, median(datos_sc$CONSUMO)
, datos_sc$CONSUMO) #TRUNCAMOS CONSUMO

### HS ENCENDIDO
datos_sc[datos_sc$HS_encendido>2.5,]
datos_sc$HS_encendido <- ifelse(datos_sc$HS_encendido>2.5,
median(datos_sc$HS_encendido), datos_sc$HS_encendido) #TRUNCAMOS HS_encendido
boxplot(datos_sc)

### HS RALENTI
datos_sc[datos_sc$HS_ralenti>3,]
datos_sc$HS_ralenti <- ifelse(datos_sc$HS_ralenti>3, median(datos_sc$HS_ralenti),
datos_sc$HS_ralenti) #TRUNCAMOS HS_ralenti
boxplot(datos_sc)

datos_sc[datos_sc$DIAS_svs<(-2.5),]
datos_sc$DIAS_svs <- ifelse(datos_sc$DIAS_svs<(-2.5), median(datos_sc$DIAS_svs),
datos_sc$DIAS_svs) #TRUNCAMOS DIAS_svs
boxplot(datos_sc)

datos_sc[datos_sc$VELOC_prom>(3),]
datos_sc$VELOC_prom <- ifelse(datos_sc$VELOC_prom>(3), median(datos_sc$VELOC_prom),
datos_sc$VELOC_prom) #TRUNCAMOS VELOC_prom

names(datos_sc)
boxplot(datos_sc[,1:7])
boxplot(datos_sc[,8:14])

r<-cor(datos_sc)
r

library(corrplot)

corrplot(r,
sig.level = 0.05,
type = "lower") #type dice si grafica todo (full) o la diag sup o inf
(upper/lower)

#-----
#####
### COMPONENTES PRINCIPALES###
#####

names(datos_sc)
pca_s<-PCA(datos_sc, ncp= 6, graph=FALSE) ### calculamos las Componentes
Principales
cargas_pca<-pca_s$var$coord ### vemos las cargas de cada variable sobre las CP
respca_s<-pca_s$ind$coord ### vemos las cargas de cada variable sobre las CPse
eig.val<-as.data.frame(get_eigenvalue(pca_s)) ## autovalores y varianza acumulada
```

```
sedim<-eig.val$cumulative.variance.percent ### vector de varianza acumulada

####grafico del codo#####
fviz_eig(pca_s, addlabels = TRUE, ylim = c(0, 35), main = "Gráfico del Codo",
title_size = 20)

####grafico del sedimentación#####
ggplot(data = data.frame(sedim, pc = 1:14),
  aes(x = pc, y = sedim)) +
  geom_point() +
  geom_line() +
  theme_grey() +
  labs(x = "Componente Principal",
  y = "En %") +
  ggtitle("Varianza Acumulada Acumulada ", subtitle = "Por Componente")

round(eig.val,4)
pca_s$var$cor

### grafico de contribuciones según 1ra y 2da componente
plot.PCA(pca_s,choix="var")
fviz_pca_var(pca_s,col.var="contrib",gradient.cols=c("blue","yellow","red"))

#-----

## grafico de contribuciones de variables originales sobre las componentes
seleccionadas

plot1=fviz_contrib (pca_s, choice = "var", axes = 1, top = 10, color="grey", fill
="blue" ) +
  theme(axis.text.x = element_text(size = 8)) +
  labs(title="Contribuciones CP1",y= "% Cont")

plot2=fviz_contrib (pca_s, choice = "var", axes = 2, top = 10, cex = 0.6,
color="grey", fill= "darkgreen") +
  theme(axis.text.x = element_text(size = 8)) +
  labs(title="Contribuciones CP2",y= "% Cont")
grid.arrange(plot1, plot2, ncol=2)

plot3=fviz_contrib (pca_s, choice = "var", axes = 3, top = 8, cex = 0.6,
color="grey", fill="grey" ) +
  theme(axis.text.x = element_text(size = 8)) +
  labs(title="Contribuciones CP3",y= "% Cont")

plot4=fviz_contrib (pca_s, choice = "var", axes = 4, top = 8, cex = 0.6,
color="grey", fill="red" ) +
  theme(axis.text.x = element_text(size = 8)) +
  labs(title="Contribuciones CP4",y= "% Cont")
grid.arrange(plot3, plot4, ncol=2)

plot5=fviz_contrib (pca_s, choice = "var", axes = 5, top = 8, cex = 0.6,
color="grey", fill="orange" ) +
  theme(axis.text.x = element_text(size = 8)) +
  labs(title="Contribuciones CP5",y= "% Cont")

plot6=fviz_contrib (pca_s, choice = "var", axes = 6, top = 8, cex = 0.6,
color="grey", fill="yellow" ) +
  theme(axis.text.x = element_text(size = 8)) +
  labs(title="Contribuciones CP6",y= "% Cont")

grid.arrange(plot5, plot6, ncol=2)

grid.arrange(plot1, plot2,plot3, plot4, plot5, plot6, ncol=3)

#-----
```

```
#####  
### ANOVA / MANOVA###  
#####
```

```
if(nrow(base) == nrow(datos_sc)) {  
  # Crear un nuevo dataframe con las columnas deseadas  
  datos_ms <- data.frame(  
    ANTIG_c = base$ANTIG_c,  
    CORRECTIVO_cant = datos_sc$CORRECTIVO_cant,  
    CORRECTIVO_usd = datos_sc$CORRECTIVO_usd  
  )  
  
  # Nuevo dataframe  
  print(datos_ms)  
} else {  
  print("Los dataframes no tienen el mismo número de filas.")  
}  
  
names(datos_ms)  
datos_ms$ANTIG_c <- factor(datos_ms$ANTIG_c)  
  
### manova final - combustible Y antigüedad predicen correctivo  
names(datos_ms)  
man_f<-manova(cbind(CORRECTIVO_usd, CORRECTIVO_cant)~ANTIG_c,data=datos_ms)  
  
summary.manova(man_f,intercept=TRUE,test="Wilks")  
summary.manova(man_f,intercept=TRUE,test="Wilks")$SS  
summary(man_f)  
  
summary.manova(man_f, intercept = TRUE, test = "Hotelling-Lawley")  
summary.manova(man_f, intercept = TRUE, test = "Roy")  
  
# Efectos inter-sujetos  
aov_usd<-aov(data=datos_ms,CORRECTIVO_usd~ANTIG_c)  
summary(aov_usd)  
  
aov_cant<-aov(data=datos_ms,CORRECTIVO_cant~ANTIG_c)  
summary(aov_cant)  
  
# Basados en bonferroni  
pairwise.t.test(datos_ms$CORRECTIVO_usd,datos_ms$ANTIG_c,p.adjust="bonferroni")  
pairwise.t.test(datos_ms$CORRECTIVO_cant,datos_ms$ANTIG_c,p.adjust="bonferroni")  
  
# Tukey  
tukey_usd <- TukeyHSD(aov_usd)  
print(tukey_usd)  
plot(tukey_usd)  
  
tukey_cant <- TukeyHSD(aov_cant)  
print(tukey_cant)  
plot(tukey_cant)
```

Buenos Aires, 02 de septiembre de 2024

Reporte Trabajo Final Integrador de Especialización

“Estudio de la logística y distribución a partir del análisis de datos: Antigüedad de Flota y Costos de Mantenimiento”

Autor: Luis Eduardo Britti

En mi carácter de mentora del Trabajo Final Integrador de Especialización de Luis Eduardo Britti, me dirijo a ustedes con el objetivo de transmitirle mi opinión sobre la investigación realizada.

En este trabajo el estudiante analiza la problemática que se le presenta a las empresas de logística cuando su flota de vehículos envejece y requiere mayor costo de mantenimiento correctivo. Para esto analiza las características principales de esta organización y como toman decisiones a partir de la gestión de sus bases de datos.

Considero que el tema elegido, sobre la problemática de una empresa de logística es relevante para la Especialización. El objetivo planteado en el trabajo es claro y coherente con la metodología desarrollada. La bibliografía utilizada es pertinente y actual.

Por otro lado, utiliza correctamente metodologías como el análisis de componentes principales para reducir la dimensionalidad y técnicas de análisis multivariado como ANOVA/MANOVA. A partir de los resultados a los que llega plantea conclusiones acordes.

Todo esto me permite dar mi expresa conformidad a la entrega de este Trabajo Final Integrador de Especialización.

Sin otro particular, saludo cordialmente.



Dra. Verónica García Fronti