



**Tecnológico
de Monterrey**

Campus Querétaro

Analítica de datos y herramientas de inteligencia artificial II

Actividad 00 EDA y preparación de datos (recopilación, limpieza, unión)
Análisis exploratorio de datos

Iker Villalobos Gilling	A01710453
Juan Carlos Cárcamo Alarcón	A01710253
José David Castillo Jasso	A01710953
Ana Isabel García Mojica	A01706241

Profesor:

Gustavo Santana

24 de Noviembre del 2025

ÍNDICE

1. Introducción	2
1.1 Propósito y enfoque	2
2. Taxonomía de variables y diccionario de datos	3
2.1 Taxonomía: Telemetría	3
2.2 Taxonomía: Peajes	4
2.3 Taxonomía: Viajes	5
2.4 Taxonomía: Combustible	7
3. Inspección técnica y de calidad	10
3.1 Telemetría	10
3.2 Peajes	10
3.3 Viajes	11
3.4 Combustible	12
4. Análisis exploratorio de datos (EDA): Telemetría	14
4.1 Correlaciones	14
4.2 Boxplots	15
5. Análisis Exploratorio de Datos (EDA): Peajes	16
5.1 Histogramas	16
5.2 Outliers	18
6. Análisis Exploratorio de Datos (EDA): Viajes	18
6.1 Dependencia de Walmart	18
6.2 Operación:	19
6.3 Comportamientos Atípicos en la Flota:	20
6.4 Vehículos y Dollys	20
6.5 FINANCIEROS	21
7. Análisis Exploratorio de Datos (EDA): Combustible	25
7.1 Mapa de calor	25
7.2 Importe Transacción	26
7.3 Monto IVA	27
7.4 Región	29
7.5 Rendimiento Real	31
7.6 Transacción	33
6. KPIs	34
6.1 La Importancia de los KPIs	34
6.2 KPIs Propuestos	34
Obras citadas	38

1. Introducción

La gestión moderna de la cadena de suministro ha dejado atrás los enfoques puramente reactivos (centrados en la corrección tardía de incidencias) para evolucionar hacia modelos predictivos y prescriptivos, donde la toma de decisiones está sustentada en datos confiables, granularidad operativa y analítica de datos. Esta transformación es tanto tecnológica como cultural: hoy, la capacidad de convertir grandes volúmenes de datos crudos en inteligencia operativa constituye uno de los activos estratégicos más determinantes para las organizaciones logísticas.

Por lo que en este ecosistema, caracterizado por la volatilidad de la demanda, la presión inflacionaria en los costos de transporte y la exigencia de niveles de servicio casi perfectos, la integración de fuentes de datos heterogéneas (telemetría satelital, peajes, viajes operativos y registros de combustible) toma gran relevancia para optimizar costos, mejorar la eficiencia y mitigar riesgos operativos y financieros.

El presente documento consolida un Análisis Exploratorio de Datos (EDA) de carácter técnico, exhaustivo y transversal sobre los tres pilares informativos proporcionados por Bepensa:

- Telemetría y Peajes, que permiten observar el comportamiento en ruta, costos asociados y patrones operativos diarios.
- Viajes, que representan la huella digital completa de la operación de transporte: asignación de activos, ejecuciones, tiempos, distancias, facturación y cumplimiento normativo.
- Combustible, históricamente el componente más crítico del costo operativo (30%–40%), analizado aquí con especial énfasis en eficiencia financiera y consumo energético real.

1.1 Propósito y enfoque

Este análisis parte del principio de que los Indicadores Clave de Desempeño (KPIs) deben ser herramientas cuantitativas objetivas que no solo midan el desempeño pasado, sino que permitan anticipar desviaciones futuras y diseñar planes de mejora continua. Para ello, cada dataset fue sometido a un proceso de:

- inspección estadística,
- identificación de anomalías y valores atípicos,
- análisis multivariado

El objetivo central es establecer una base confiable sobre el comportamiento operativo, útil para soportar modelos predictivos posteriores, estrategias de optimización y propuestas de gobernanza de datos.

2. Taxonomía de variables y diccionario de datos

La precisión en el análisis de datos comienza con la definición de cada variable. A continuación, se presenta la taxonomía técnica de los conjuntos de datos, clasificando cada elemento no solo por su tipo de dato informático, sino por su función semántica dentro de la operación logística.

2.1 Taxonomía: Telemetría

El EDA detalla un conjunto de datos de telemetría compuesto por 7,601 registros y 10 columnas. Este dataset actúa como el registro del comportamiento físico de la flota.

```
• <class 'pandas.core.frame.DataFrame'>
RangeIndex: 7601 entries, 0 to 7600
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Nombre            7601 non-null    object  
 1   Periodo           7601 non-null    int64  
 2   Fecha de inicio del periodo 7601 non-null    datetime64[ns]
 3   Fecha de fin del periodo   7601 non-null    datetime64[ns]
 4   Duración de conducción    7601 non-null    timedelta64[ns]
 5   Porcentaje de tiempo en conducción 7601 non-null    int64  
 6   Distancia          7601 non-null    float64 
 7   Tiempo en ralentí     7601 non-null    timedelta64[ns]
 8   Porcentaje de tiempo en ralentí   7601 non-null    int64  
 9   Velocidad máxima      7601 non-null    int64  
dtypes: datetime64[ns](2), float64(1), int64(4), object(1), timedelta64[ns](2)
memory usage: 594.0+ KB
```

Variable	Tipo de dato	Categoría semántica	Definición operativa y utilidad analítica
Nombre	Object (String)	Llave Primaria / Dimensión	Identificador único del vehículo . Funciona como el ancla para la integración con otros sistemas. Su alta cardinalidad (376 valores únicos) define el tamaño activo de la flota en el periodo analizado.
Periodo	Int64	Dimensión Temporal (Ordinal)	Representación numérica del ciclo operativo. Su correlación perfecta con la duración sugiere que puede ser un artefacto de clasificación (binning).
Fecha de inicio del periodo	Datetime64[ns]	Límite Temporal Inferior	Marca el tiempo exacto del inicio de la captura de datos para el registro. Esencial para la alineación de series temporales.
Fecha de fin del periodo	Datetime64[ns]	Límite Temporal Superior	Marca de tiempo de cierre. La diferencia entre Inicio y Fin establece el "Tiempo Total Disponible" del activo.
Duración de conducción	Timedelta64[ns]	Métrica de Desempeño	Tiempo absoluto de operación del motor con desplazamiento positivo. Es el numerador principal para el cálculo de productividad.
Porcentaje de tiempo en conducción	Int64	KPI de Eficiencia	Ratio derivado que normaliza la duración sobre el tiempo total disponible. Permite comparar la productividad entre períodos de distinta longitud.
Distancia	Float64	Métrica de Producción	Kilómetros totales recorridos. Es el denominador

			universal para métricas de costo (ej. Costo por Km, Desgaste por Km).
Tiempo en ralentí	Timedelta64[ns]	Métrica de Ineficiencia	Tiempo de operación del motor con velocidad cero. Representa el consumo de combustible sin generación de ingresos (desperdicio).
Porcentaje de tiempo en ralentí	Int64	KPI de Desperdicio	Proporción del tiempo total del motor dedicado al ralentí. Valores altos indican fricción operativa (tráfico, esperas en andén no planificadas).
Velocidad máxima	Int64	Métrica de Seguridad	El pico de velocidad registrado en el intervalo. Importante para la gestión de riesgos y cumplimiento normativo.

2.2 Taxonomía: Peajes

El conjunto de datos de Peajes contiene 187,036 registros. Este dataset captura la interacción financiera con la infraestructura carretera.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 187036 entries, 0 to 187035
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Tag              187036 non-null   object 
 1   No.Economico     187036 non-null   object 
 2   Fecha            187036 non-null   object 
 3   Hora             187036 non-null   object 
 4   Casetas          187036 non-null   object 
 5   Carril           187036 non-null   object 
 6   Clase            187036 non-null   int64  
 7   Importe          187036 non-null   object 
 8   Fecha Aplicacion 181245 non-null   object 
 9   Hora Aplicacion  181245 non-null   object 
 10  Consecar         187036 non-null   object 
dtypes: int64(1), object(10)
memory usage: 15.7+ MB
```

Variable	Tipo de Dato	Categoría Semántica	Definición Operativa y Utilidad Analítica
Tag	Object	ID de Sensor / Activo	Número de serie del dispositivo RFID (IAVE/PASE). Su cardinalidad (403) excede a la de los vehículos, indicando un problema de gestión de inventario de hardware.
No. Económico	Object	Llave Foránea	Identificador del vehículo asociado a la transacción. Debe coincidir semánticamente con Nombre en Telemetría para permitir el cruce de datos.
Fecha	Datetime	Dimensión Temporal	Fecha calendario del cruce. Variable fundamental para detectar patrones estacionales y semanales.
Hora	Datetime	Dimensión Temporal (Cíclica)	Marca de tiempo precisa. Permite el análisis de ritmos circadianos y la identificación de operaciones en horarios de riesgo.
Casetas	Object	Dimensión Geoespacial	Nombre de la plaza de cobro (ej. "SAN MARCOS",

			"TEHUACAN"). Actúa como proxy de la ruta seleccionada por el operador.
Carril	Object	Detalle de Infraestructura	Información de bajo valor estratégico pero útil para auditorías forenses en caso de disputas de facturación.
Clase	Int64	Driver de Costo	Clasificación tarifaria del vehículo (ej. ejes). Errores aquí derivan en sobrecostos directos recuperables.
Importe	Float64	Métrica Financiera	Costo monetario de la transacción. Variable objetivo para modelos de optimización de rutas.

2.3 Taxonomía: Viajes

La estructura original consta de 108,991 filas y 32 columnas.

El volumen de datos es estadísticamente significativo para modelar el comportamiento anual de la flota, dado que cubre un periodo de 10 meses. Sin embargo, la calidad del dato crudo presenta heterogeneidad, con columnas de alta integridad (como Folio o Cliente) y columnas con deficiencias severas de captura (como Peso Descarga Kgs o Dolly).

```
Data columns (total 32 columns):
 #   Column           Non-Null Count  Dtype  
 --- 
 0   Numero          108991 non-null   int64  
 1   Nombre Cliente  108991 non-null   object  
 2   Folio           108991 non-null   int64  
 3   Ruta            108991 non-null   object  
 4   Viaje           108991 non-null   object  
 5   Viaje Docto    108991 non-null   object  
 6   Sucursal        108991 non-null   object  
 7   No. Viaje Cte  108991 non-null   object  
 8   Fecha...9       108990 non-null   datetime64[ns]
 9   Moneda          108991 non-null   object  
 10  SubTotal        108991 non-null   float64 
 11  IVA             108991 non-null   float64 
 12  Retención       108991 non-null   float64 
 13  Total           108991 non-null   float64 
 14  Factura         105298 non-null   object  
 15  Liquidación    72998 non-null   float64 
 16  Fecha...17      105298 non-null   datetime64[ns]
 17  Fecha Vencimiento 105298 non-null   object  
 18  Tractocamión   108094 non-null   object  
 19  Remolque 1     99679 non-null   object  
 20  Dolly           14485 non-null   object  
 21  Remolque 2     42909 non-null   object  
 22  Nro Ope         108094 non-null   float64 
 23  Operador        108094 non-null   object  
 24  Fecha Salida   82862 non-null   object  
 25  Fecha Llegada  82825 non-null   object  
 26  Peso Kgs        78647 non-null   float64 
 27  Peso Descarga Kgs 8765 non-null   float64 
 28  Diferencia      69968 non-null   float64 
 29  Documentos      0 non-null      float64 
 30  UUID CP         0 non-null      float64 
 31  Estatus de Viaje 106167 non-null   object  
 dtypes: datetime64[ns](2), float64(11), int64(2), object(17)
```

Variable	Tipo de dato	Categoría semántica	Definición operativa y utilidad analítica
Número	int64	Índice	Identificador secuencial técnico. Sin valor analítico intrínseco.
Nombre Cliente	object	Dimensión (Entidad)	Identifica al receptor del servicio (ej. "NUEVA WAL MART DE MEXICO"). Es la variable crítica para el análisis de concentración de cartera (Pareto) y riesgo comercial.
Folio	int64	Llave primaria	Identificador único del viaje en el TMS (Transportation Management System). Fundamental para la trazabilidad,

			auditoría y cruce con otras bases de datos como Telemetría o Peajes. ²
Ruta	object	Dimensión (Geoespacial)	Descripción textual del origen y destino (ej. "WM CEDIS CHALCO/WM BODEGA"). Permite el análisis de rentabilidad por corredor logístico y optimización de redes.
Viaje	object	Identificador operativo	Código alfanumérico interno.
Sucursal	object	Dimensión (Organizacional)	Centro de costos o base operativa de origen (ej. "MÉRIDA/YUC001"). Clave para evaluar el desempeño regional y la asignación de recursos geográficos.
Fecha...9	datetime	Dimensión temporal	Fecha de creación del registro. Sirve como ancla para series de tiempo de demanda y estacionalidad.
SubTotal	float64	Métrica financiera	Monto base del ingreso antes de impuestos. Variable objetivo para análisis de rentabilidad bruta.
IVA	float64	Métrica fiscal	Impuesto al Valor Agregado (16%). Su consistencia debe validarse contra el SubTotal para asegurar la integridad contable.
Retención	float64	Métrica fiscal	Monto retenido por el cliente. En México, por ley (Art. 1-A LIVA), corresponde al 4% del valor del flete en autotransporte. ³
Total	float64	Métrica financiera	Monto neto a cobrar (Cash Flow). Impacta directamente en la liquidez.
Factura	object	Estado administrativo	Identificador fiscal (UUID). Los valores nulos indican servicios "Work in Progress" (WIP) o ingresos devengados no facturados.
Liquidación	float64	Estado administrativo	Monto pagado al operador (viáticos/sueldo). Los nulos sugieren procesos administrativos inconclusos con la fuerza laboral.
Tractocamión	object	Dimensión (Activo)	Identificador de la unidad de potencia (ej. "T541"). Permite calcular la utilización de la flota (Asset Utilization) y costos de mantenimiento.
Remolque 1/2	object	Dimensión (Activo)	Identificadores de los equipos de arrastre. Su presencia simultánea indica una configuración "Full" (T3-S2-R4).
Dolly	object	Dimensión (Activo)	Equipo de conexión para configuraciones de doble remolque. Su ausencia en viajes "Full" es un indicador crítico de falta de control de activos.
Operador	object	Dimensión (Personal)	Nombre del conductor. Vital para análisis de desempeño, seguridad y rotación de personal.
Fecha Salida	datetime	Hito operativo (Inicio)	Marca de tiempo real del inicio del viaje.
Fecha Llegada	datetime	Hito operativo (Fin)	Marca de tiempo real de la conclusión. La diferencia con la Salida determina el Lead Time o Cycle Time.
Peso Kgs	float64	Métrica operativa	Peso declarado de la carga. Factor determinante para el consumo de combustible y desgaste de activos.
Estatus de Viaje	object	Estado Operativo	Condición actual (ej. "TERMINADO", "EN RUTA"). Permite filtrar el universo de análisis a viajes cerrados para cálculos de tiempos.

2.4 Taxonomía: Combustible

Los conjuntos de datos analizados (Enero-Febrero, Agosto, Septiembre 2025) presentan una estructura columnar consistente, lo que permite su tratamiento homologado. A continuación, se detalla la taxonomía técnica de las variables, enriquecida con contexto operativo y normativo específico para el mercado mexicano.

Data columns (total 48 columns):			
#	Column	Non-Null Count	Dtype
0	Id Grupo Región	13894	non-null float64
1	Grupo Región	13894	non-null object
2	Id Región	13894	non-null float64
3	Región	13894	non-null object
4	Id Centro costos	13894	non-null object
5	Centro Costos	13894	non-null object
6	Num Tarjeta	13894	non-null float64
7	Id Vehículo	13894	non-null object
8	Descripción Vehículo	13894	non-null object
9	No Comprobante	4328	non-null float64
10	Fecha transacción	13894	non-null datetime64[ns]
11	Tipo Transacción	13894	non-null object
12	Código Pemex	13894	non-null float64
13	Razón social Afiliado	13894	non-null object
14	Certificación Cualli	13894	non-null float64
15	Estado/Ciudad de Transacción	13894	non-null object
16	Ticket de Bomba	0	non-null float64
17	Km Ant Transacción	4324	non-null float64
18	Km Transacción	13894	non-null float64
19	Recorrido	4324	non-null float64
20	Rendimiento Vehículo	4328	non-null float64
21	Rendimiento Real	13894	non-null float64
22	Capacidad de Tanque	4328	non-null float64
23	Mercancía	13894	non-null float64
24	Descripción Mercancía	13894	non-null object
25	Cantidad Mercancía	13895	non-null float64
26	Precio Unitario Merc	13894	non-null float64
27	Saldo Ant Transacción	4054	non-null float64
28	Saldo Actual Después de Transacción	4055	non-null float64
29	Importe Neto	13895	non-null float64
30	Monto IVA	13895	non-null float64
31	% IVA	13894	non-null float64
32	Importe Transacción	13895	non-null float64
33	TAR	0	non-null float64
34	Monto descuento	0	non-null float64
35	Estado Transacción	13894	non-null object
36	Tipo Tarjeta	4328	non-null object
37	Consumo En	13894	non-null object
38	Id Terminal	13894	non-null float64
39	Placa	13894	non-null object
40	Conductor	13894	non-null object
41	Observaciones	3782	non-null object
42	Parámetro Personalizado1	0	non-null float64
43	Parámetro Personalizado2	0	non-null float64
44	Parámetro Personalizado3	0	non-null float64
45	Parámetro Personalizado4	0	non-null float64
46	Parámetro Personalizado5	0	non-null float64
47	Id Transacción		

Identificación y activos

Variable	Tipo de dato	Categoría semántica	Definición operativa, contexto y utilidad
Id	Int64	Llave Primaria (Técnica)	Identificador secuencial único. Su utilidad es puramente informática para la trazabilidad del registro en caso de aclaraciones técnicas.
Num tarjeta	Object	Identificador de medio de pago	Código numérico asociado al plástico o tag NFC utilizado para la transacción. En la gestión de flotas, este es el vínculo financiero crítico. Una auditoría de este campo permite detectar "tarjetas fantasma" o activos dados de baja que siguen generando cargos. Es vital que este número tenga una relación 1:1 estricta con el Id Vehículo.
Id vehículo descripción vehículo	Object	Identificador de activo (Flota)	Identificador interno de la unidad. Aquí se observa una nomenclatura interna corta. Es crucial para cruzar información con sistemas de Telemetría y Mantenimiento. La consistencia en este campo es fundamental.
Placa	Object	Identificador legal	La matrícula oficial del vehículo (SPF - Servicio Público Federal). Es el dato maestro para el cumplimiento normativo ante la SICT. Permite cruzar multas y validar que el vehículo que carga corresponde legalmente a la flota.
Conductor	Object	Identificador de personal	Nombre del operador responsable. El análisis de esta variable permite crear perfiles de desempeño por conductor, identificando quiénes tienen los mejores rendimientos y quiénes presentan hábitos de carga sospechosos.

Geográficas y de red de suministro

Variable	Tipo de Dato	Categoría Semántica	Definición operativa, contexto y utilidad
Grupo Región	Object	Macro-Localización	Agrupación geográfica de alto nivel (ej. "CENTRO", "SURESTE"). Vital para el análisis de costos, ya que el precio del diésel varía significativamente entre regiones debido a costos logísticos de distribución de PEMEX y diferencias fiscales en zonas fronterizas.
Región	Object	Micro-Localización	Subdivisión operativa. Permite identificar corredores logísticos específicos. Un análisis detallado aquí puede revelar ineficiencias de ruta o desvíos no autorizados.
Id Terminal Estado Transacción	Int64 / Object	Punto de Venta (POS)	Identifica la estación de servicio específica. El análisis de frecuencia por terminal permite negociar descuentos por volumen corporativo. El "Estado" valida si la transacción fue exitosa, rechazando intentos fallidos que no deben sumar al costo.
TAR	Object (Null)	Origen de Suministro (Infraestructura)	Siglas de "Terminal de Almacenamiento y Reparto". Representa el nodo logístico de PEMEX del cual proviene el combustible. ⁵ Su correcta captura permitiría trazar problemas de calidad de combustible hasta su origen. Su ausencia total en el dataset es una pérdida significativa de trazabilidad.

Métricas y financieras

Variable	Tipo de Dato	Categoría Semántica	Definición operativa, contexto y utilidad
Cantidad Mercancía	Float64	Volumen (Input Energético)	Litros despachados. Es la variable independiente más crítica. Cualquier valor que exceda la capacidad nominal del tanque (variable Capacidad de Tanque) debe ser marcado automáticamente como fraude potencial o uso de tanques no autorizados.
Importe Transacción	Float64	Costo Total (Output Financiero)	Monto monetario total. Debe analizarse en conjunto con el volumen para derivar el precio unitario implícito y detectar anomalías en la facturación.
Precio Unitario Merc	Float64	Driver de Costo Variable	Precio por litro al momento de la carga. Su monitoreo es esencial para validar que se estén respetando los precios de contrato o para identificar estaciones con precios depredadores por encima del mercado.
Monto IVA / % IVA	Float64	Fiscalidad	Componente impositivo. Fundamental para el flujo de caja y la recuperación de impuestos. Errores aquí generan contingencias fiscales.
Certificación Cualli	Object	Garantía de Calidad	Indicador binario o categórico que señala si la estación cuenta con la certificación de calidad de PEMEX ("Cualli"). Cargar en estaciones certificadas reduce el riesgo de recibir "litros incompletos" o combustible adulterado que dañe los inyectores. ⁶

Telemetría indirecta

Variable	Tipo de Dato	Categoría Semántica	Definición operativa, contexto y utilidad
Km Transacción	Float64	Odometría Actual	Lectura del odómetro ingresada en la terminal punto de venta. Es la variable más crítica y más vulnerable del dataset. De su precisión depende todo el cálculo de rendimiento. Se observa una alta incidencia de errores humanos (dedazos, ceros).
Km Ant Transacción	Float64	Odometría Referencia	Lectura previa registrada. El sistema la utiliza para calcular el recorrido. Si este campo es nulo (como sucede en el 70% de los casos 2), el cálculo de eficiencia es imposible.
Rendimiento Real	Float64	KPI Calculado (Sistema)	Resultado de la fórmula (km actual - km anterior)/litros. Debido a la mala calidad de los inputs de kilometraje, este campo suele contener "basura" estadística que requiere una limpieza profunda antes de ser analizado.

3. Inspección técnica y de calidad

Antes de realizar cualquier inferencia estadística, es necesario validar la integridad de los datos. La inspección técnica realizada revela tanto fortalezas como vulnerabilidades operativas.

3.1 Telemetría

Valores nulos:

El análisis de la telemetría muestra un conjunto de datos limpio. La consulta de valores nulos arroja un resultado de cero para todas las columnas. Esto puede implicar que el dataset crudo (donde la pérdida de datos es común debido a fallos) ya pasó por un proceso ETL (Extract, Transform, Load) previo, donde los datos faltantes fueron imputados o las filas incompletas eliminadas. Esto simplifica el análisis pero introduce el riesgo de sesgo: no estamos viendo los momentos en que la flota estuvo "invisible".

Duplicados:

Se detectaron solo 9 en el dataframe. Aunque 9 registros representan una fracción baja (<0.13%), su eliminación es necesaria para evitar la inflación artificial de métricas acumulativas como la distancia total.

3.2 Peajes

Brecha Tag-Vehículo:

El hallazgo más crítico en la fase de inspección proviene de la comparación de cardinalidad en el dataset de Peajes:

- **Tags Únicos:** 403
- **Unidades Económicas Únicas (Unidades):** 377

Existe un excedente de 26 Tags activos en comparación con los vehículos registrados. Esto constituye una brecha con implicaciones a analizar ya que se puede tratar de dispositivos que están generando costos pero que no están asignados oficialmente a una unidad activa en este periodo. Por otro lado, puede indicar que los tags están siendo rotados entre vehículos (swapping) sin actualizar el sistema de registro.

Valores nulos:

Se identificaron 5,791 valores nulos en las columnas *Fecha Aplicacion* y *Hora Aplicacion*. Sin embargo, las columnas Fecha y Hora (del cruce real) están completas. Generalmente, este atributo se refiere al momento en que la institución bancaria procesa el cobro. Los nulos aquí indican transacciones en tránsito bancario. Para fines de análisis operativo (¿cuándo pasó el camión?), las columnas principales son suficientes, por lo que estos nulos no invalidan el estudio.

3.3 Viajes

Valores nulos:

Los valores nulos en el dataset no siempre son aleatorios, puede depender del estado del proceso, siendo crucial distinguir entre "información pendiente" (proceso en curso) e "información perdida" (error de captura). Los hallazgos principales se explican a continuación:

- **Peso Descarga (92% nulos):** La falta de este dato impide el cálculo sistemático de mermas o diferencias de inventario.
- **Dolly (86% nulos):** Al contrastar este dato con los registros de "Remolque 2" , se revela una inconsistencia de activos: miles de viajes en configuración "Full" no tienen asignado el Dolly correspondiente en el sistema.
- **Liquidación (33% nulos):** Representa un pasivo operativo latente. Un tercio de los viajes carece de registro de pago al operador, sugiriendo retrasos administrativos o pasivos laborales ocultos.
- **Fecha Salida (24% nulos):** Indica un error de ejecución sistémica. Si el viaje aparece como "Terminado", la ausencia de fechas de cierre implica que el monitoreo no cerró el ciclo en tiempo real.
- **Factura (3.4% nulos):** Aunque es un porcentaje bajo, constituye un riesgo directo de ingresos (fuga de flujo de caja) y un riesgo de incobrabilidad por antigüedad.

Para cada atributo, se realizó la imputación de nulos dependiendo de la mejor medida derivada del contexto.

Duplicados:

En el análisis se detectó un duplicado crítico: un folio asociado al cliente "NUEVA WAL MART DE MEXICO" (ruta TEYA/WM CANCÚN), aparece dos veces con atributos idénticos.

Aunque estadísticamente un solo caso parece insignificante frente a 108k registros, operativamente representa un riesgo de calidad severo. Un viaje duplicado puede derivar en doble facturación, causando rechazos administrativos. Se procedió a la eliminación del duplicado, manteniendo la primera ocurrencia para sanear la base.

Outliers:

Se identificó una anomalía que compromete la integridad de los datos temporales: la existencia de viajes futuros. Específicamente, un registro presentó una fecha de llegada en el año 2133, atribuible a un error de captura manual.

El impacto analítico de este único valor es desproporcionado, ya que distorsiona severamente las medias aritméticas para el "Tiempo de Tránsito" o "Ciclo de Viaje", anulando su utilidad estadística. Como acción correctiva, se implementó una máscara lógica para filtrar y aceptar solo fechas dentro del rango operativo razonable (2024-2025), segregando los registros fuera de rango como errores de captura.

3.4 Combustible

Al igual que los otros datasets, el análisis de este revela métricas de valores faltantes (*nulls*) que no son aleatorias, sino estructurales.

La columna TAR (*Terminal de Almacenamiento y Reparto*) presenta un **100% de valores nulos** en todos los períodos. La TAR es el nodo logístico clave (ej. TAR 602) para rastrear el origen y la calidad de los lotes de combustible despachados.¹³ La ausencia total de este dato genera una pérdida de trazabilidad. Ante fallas mecánicas, el equipo de mantenimiento está ciego para correlacionar estos eventos con lotes de combustible contaminado, impidiendo detectar causas raíz externas.

Un hallazgo que se puede destacar de igual manera es la gran cantidad de nulos presentes, aproximadamente un 70%, que existen en ciertas variables de telemetría (más relacionadas con el

combustible): *Km Ant Transacción, Recorrido, Rendimiento Vehículo y Capacidad de Tanque*.

Esto nos puede decir que en 7 de cada 10 cargas, el sistema no tiene una referencia histórica de la unidad. Esto puede llegar a ocurrir debido a que la captura del kilometraje en la carga anterior no fue la adecuada, lo que interrumpe la cadena necesaria para calcular información importante.

- **Consecuencia:** El KPI de "Rendimiento del Combustible (Km/L)" no se puede calcular para la mayoría de los registros. Utilizar los promedios de únicamente el 30% que si es calculable introduce mucho sesgo a los análisis. Esto produce que solo se analice a quienes cumplen el el procedimiento en su totalidad, lo que puede llevar a posiblemente ignorar unidades con mal desempeño.

Campos administrativos y de control

- **Ticket de bomba (100% nulos):** No tener los números de ticket de manera digital complica de gran manera las auditorías. Es de gran importancia tener alguna referencia para no caer por "cargas fantasma"
- **Observaciones (72.8% nulos):** Este nivel de campos nulos se podría decir que es aceptable.

Consistencia de identificadores y gestión de activos

La integridad de la información maestra es sumamente importante para el funcionamiento en conjunto con los sistemas de Peajes y Telemetría.

- **Nomenclatura no estandarizada:** Se pudieron detectar identificadores no estandarizados, es importante poder adaptar la nomenclatura para que sea homogénea y no pueda llevar a futuros problemas la mala nomenclatura.
- **Acción que se recomienda:** Es crucial poder mudarse hacia una nomenclatura que sea a prueba de errores, que no haya duplicados, que sea sencillo de identificar y que no resulte confusa para futuros análisis

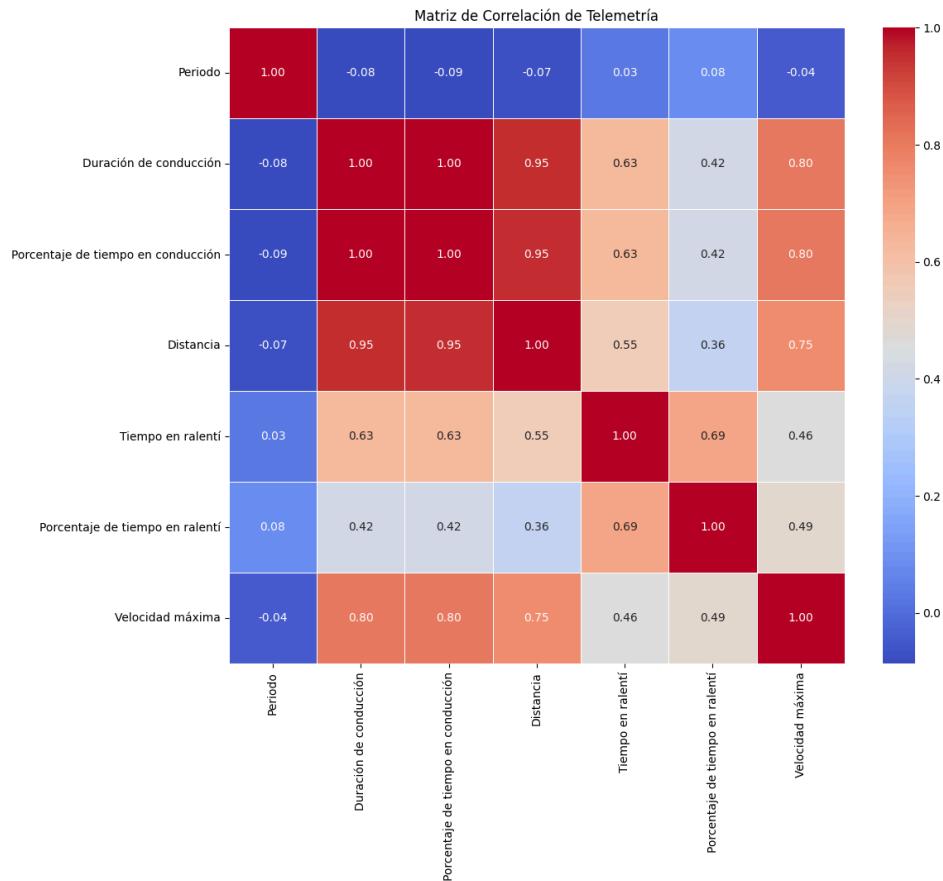
Transacciones Atípicas

El análisis de distribución reveló dos diferentes tipos de transacciones atípicas; El primero siendo dirigido a cuando el importe es igual a cero, lo cual corresponde a cuando las transacciones fueron rechazadas por algún motivo, cabe resaltar que es importante poder filtrar estos registros para que no sesgue el promedio de los costos. Por otro lado, cuando el kilometraje es igual a cero, se puede atribuir a que el operador no hizo el registro adecuadamente de datos.

4. Análisis exploratorio de datos (EDA): Telemetría

El análisis de las variables físicas de la flota, en este caso telemetría, nos permite entender el comportamiento operativo de la empresa.

4.1 Correlaciones

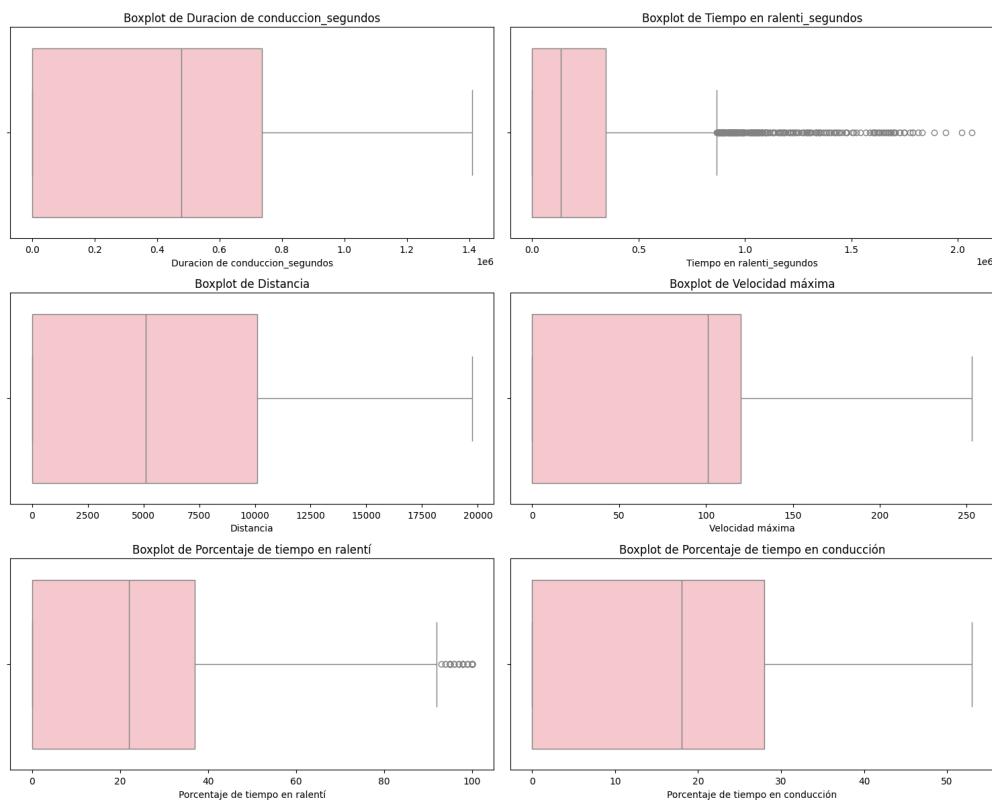


La matriz de correlación revela la estructura lógica de la operación.

- **Duración – Porcentaje de tiempo en conducción ($r = 1.00$):** La correlación perfecta indica que el porcentaje de tiempo en conducción es una transformación de la duración. Es decir, que ambas variables contienen la misma información operativa. Su interdependencia implica que solo una debe utilizarse en modelos analíticos para evitar multicolinealidad.
- **Distancia - Duración ($r = 0.95$):** La correlación de Pearson de 0.95 entre Distancia y Duración de conducción tienen sentido. Sin embargo, el 5% de varianza no explicada es causada por la variabilidad en la velocidad media, influenciada por el tráfico, la topografía y la carga.

- **Velocidad máxima – Duración ($r = 0.80$):** Esta alta correlación sugiere que los viajes más largos exponen al operador a tramos donde se alcanzan velocidades más altas. Sin embargo, la velocidad máxima no se mantiene constante debido a otras variables como el comportamiento del conductor o restricciones de seguridad.
- **Velocidad máxima – Distancia ($r = 0.75$):** La correlación fuerte indica que quienes recorren mayores distancias tienden a alcanzar velocidades más elevadas. Esto es coherente con rutas de larga distancia donde existe menor interrupción y el vehículo puede mantener velocidades más altas durante intervalos prolongados.
- **Ralentí - Duración:** Existe una correlación positiva moderada ($r = 0.63$) entre la duración del viaje y el tiempo en ralentí. Esto es consistente con la logística de larga distancia: viajes más largos implican más paradas obligatorias, semáforos y colas de peaje. Sin embargo, aquellos conductores que logran altas distancias con bajo incremento de ralentí representan el estándar de eficiencia a replicar.
- **Velocidad máxima – Porcentaje de tiempo en ralentí ($r = 0.49$):** Aunque la correlación es moderada, sugiere que quienes pasan más tiempo en ralentí tienden a no alcanzar velocidades altas, probablemente debido a operación en zonas urbanas, tráfico denso o rutas con detenciones frecuentes.

4.2 Boxplots



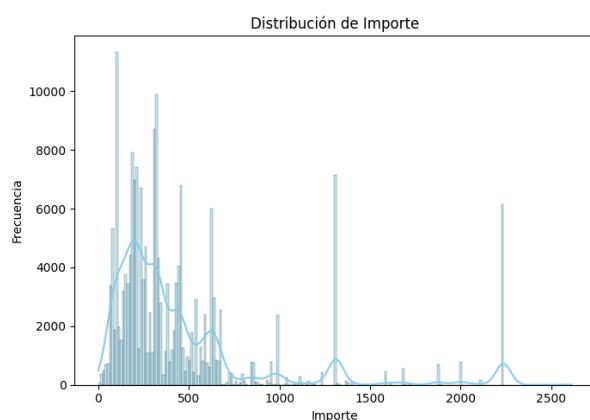
El análisis de diagramas de caja (Boxplots) permite identificar comportamientos anómalos.

- **Duración de conducción:** La mayoría de los trayectos tienen duraciones relativamente cortas. Existen viajes extraordinariamente largos que se presentan como extremos operativos, probablemente rutas especiales, errores de registro o vehículos que permanecieron en "conducción" por un sensor que no se desactivó. La amplitud entre Q1 y Q3 indica una alta variabilidad en la duración de los trayectos.
- **Tiempo de ralentí:** El boxplot muestra que el 75% de los viajes tiene tiempos de ralentí relativamente bajos. Sin embargo, hay una cola extremadamente larga con muchos puntos fuera del rango esperado, algunos incluso por encima de 2,000,000 segundos (23 días). Esto sugiere que puede haber sensores que registran el ralentí incluso cuando el vehículo no está realmente al ralentí, o vehículos encendidos durante largas jornadas sin desplazamiento (pueden reflejar ineficiencia, malos hábitos, o errores técnicos).
- **Velocidad Máxima:** Se espera una distribución con un límite superior rígido (techo) debido a los gobernadores de velocidad (típicamente a 95 km/h o 110 km/h en México). Cualquier valor atípico por encima de este límite representa una violación de seguridad grave (manipulación del gobernador o descenso en neutral "gravity assist").
- **Distancia recorrida:** La mayor parte de los viajes ocurren entre 0 y 10,000 km (75%). Además, se observa una cola larga que sugiere rutas demasiado largas. No hay tantos outliers puntuales como en ralentí, lo que indica que la distancia está más controlada.

5. Análisis Exploratorio de Datos (EDA): Peajes

El análisis del dataset de Peajes revela una realidad financiera multimodal que también debe ser analizada.

5.1 Histogramas



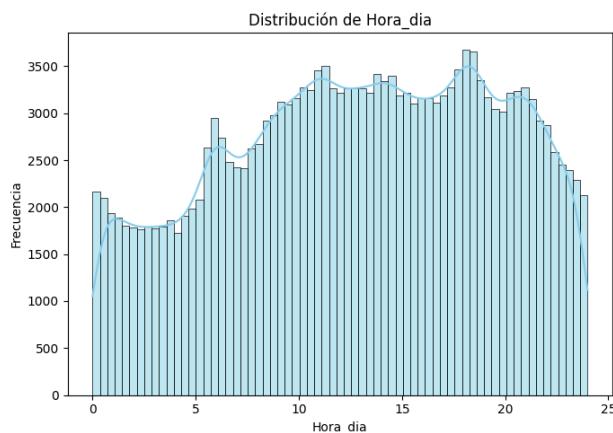
IMPORTE

El histograma de Importe no sigue una distribución normal, sino que es claramente multimodal, presentando múltiples picos.

- **Moda 1:** Hay una alta densidad de transacciones por debajo de \$200 MXN. Esto representa la operación de "última milla" y tramos locales.
- **Moda 2:** Transacciones en el rango medio (\$200 - \$600 MXN). La gran densidad a la izquierda indica que la operación principal de Bepensa consiste en tramos cortos o medianos.
- **Moda 3:** Picos distintivos alrededor de \$1,300 y \$2,200 MXN. Estos valores corresponden a tarifas de autopistas de alta especificación para vehículos pesados. La investigación contextual confirma que tarifas de este calibre son consistentes con tramos completos del **Arco Norte** y el **Circuito Exterior Mexiquense** para configuraciones de 7 a 9 ejes (T3-S2-R4).⁴

Este comportamiento multimodal implica que el uso de un "Costo Promedio de Peaje" (\$458 MXN) sería un error metodológico. Mezclar peajes urbanos de \$50 pesos con cruces de Arco Norte de \$1,000 pesos resulta en una media que no representa ninguna realidad operativa .Se deben segmentar los pronósticos en "Rutas Cortas" vs "Rutas Largas", porque el comportamiento de costo es totalmente diferente.

HORA



- **03:00 - 04:00 AM:** Se observa una caída significativa en la actividad durante la madrugada. Esto se alinea con las mejores prácticas de seguridad en el transporte de carga en México, donde se evita la circulación nocturna en corredores de alto riesgo para evitar el robo de mercancía.¹¹
- **Pico de 18:00 PM:** La actividad máxima se registra al final de la tarde. Operativamente,

esto implica que la flota está cruzando plazas de cobro en horas pico, lo que probablemente aumenta el tiempo de ralentí (colas en casetas) y el consumo de combustible.

5.2 Outliers

Al aplicar el método del Rango Intercuartílico (IQR) a los peajes se estableció un límite superior de \$1,062.50 MXN para considerar un valor como "normal". El análisis identificó 17,444 transacciones por encima de este límite. En estadística clásica, esto indicaría datos "sucios". Sin embargo, en el contexto logístico, un peaje de \$1,200 pesos no es un error, es el precio de mercado para un tráiler en una autopista con características diferentes. Es decir, estos 17,000 registros no son anomalías a eliminar; son un **Segmento de Negocio** distinto. Representan la operación de larga distancia en infraestructura de alta velocidad.

6. Análisis Exploratorio de Datos (EDA): Viajes

6.1 Dependencia de Walmart

El análisis de frecuencia de la variable Nombre Cliente indica una dependencia extrema de un solo socio comercial. El cliente "NUEVA WAL MART DE MEXICO" acumula **59,899 registros**, lo que representa aproximadamente el **55% del volumen total** de viajes registrados en el periodo.

Esta concentración tiene implicaciones estratégicas profundas que van más allá del volumen.

Al representar más de la mitad de la operación, los estándares operativos de Bepensa son: los estándares de Walmart. Esto significa que toda la flota debe operar bajo los requisitos de **OTIF (On-Time In-Full)** de Walmart.

Walmart exige a sus proveedores y transportistas un cumplimiento de OTIF del **98%**.⁹ El incumplimiento de esta métrica no solo afecta la relación comercial, sino que conlleva multas financieras que pueden ascender al **3% del Costo de los Bienes Vendidos (COGS)**.¹⁰

6.2 Operación:

Centralización operativa y rol de Mérida

El análisis de la variable **Sucursal** revela una clara concentración de la operación en el sureste de México. La sucursal "**MÉRIDA/YUC001**", con **31,686 registros**, equivale aproximadamente al **29%** del total de movimientos registrados. Esta magnitud posiciona a Mérida como el nodo dominante dentro de la red logística de Bepensa, funcionando en la práctica como el centro principal.

Desde una perspectiva logística, este comportamiento es consistente con la geografía de la región peninsular. Mérida actúa como punto articulador entre Yucatán, Campeche y Quintana Roo; sus rutas permiten conectividad eficiente hacia los principales centros consumidores y hacia los puertos estratégicos de Progreso y Campeche. Esta centralidad disminuye distancias promedio, reduce tiempos de respuesta y favorece la consolidación de servicios.

Implicaciones para el mantenimiento y gestión de activos:

La concentración de la flota en un solo nodo trae consigo ventajas operativas. La cercanía de las unidades facilita la centralización de actividades de mantenimiento, la consolidación de refacciones y la gestión unificada de talleres. Esto permite: Reducción de costos por inventario de refacciones al evitar duplicación entre sucursales. Mayor uniformidad en el servicio técnico y cumplimiento de estándares. Optimización del tiempo de inmovilización de las unidades, al tener mayor disponibilidad de técnicos y recursos. Sin embargo, esta centralización también vuelve al sistema vulnerable: cualquier saturación, indisponibilidad o falla en el taller de Mérida afectaría de forma desproporcionada la continuidad operativa.

10 rutas principales:

El análisis del Top 10 de Rutas revela una marcada concentración de la actividad logística en corredores de "circuito cerrado".. Destaca, en primer lugar, el ecosistema logístico de **Chalco** (Valle de México), que al sumar sus interconexiones con la Bodega Remota y la Fábrica de Jabón La Corona, domina el volumen total de operaciones. Esto sugiere que Chalco funciona como el gran *hub* de reabastecimiento para la zona centro.

RUTA	CONTEO DE REGISTROS
WM CEDIS CHALCO/WM BODEGA REMOTA CHALCO	3813
WM BODEGA REMOTA CHALCO/WM CEDIS CHALCO	3620
BB PONIENTE/BB PACABTUN	2344
BB PACABTUN/BB PONIENTE	2343
FABRICA DE JABON LA CORONA/WM CEDIS CHALCO	1539
WM CEDIS VILLAHERMOSA SECOS/PENSION SALINAS CRUZ	1465
BB PLAYA DEL CARMEN/BB CANCUN PLANTA	1443
PENSION SALINAS CRUZ/WM CEDIS VILLAHERMOSA SECOS	1412
BB CANCUN PLANTA/BB PLAYA DEL CARMEN	1406
WM CEDIS CHALCO/FABRICA DE JABON LA CORONA	1202

6.3 Comportamientos Atípicos en la Flota:

El inventario reporta 348 tractocamiones únicos activos. Entre ellos, el activo "T541" destaca por registrar 2,128 viajes en un periodo de 10 meses, lo que equivale a un promedio de 7.1 viajes diarios.

Análisis:

Un rendimiento operativo de 7 viajes por día no es físicamente viable para un tractocamión dedicado a rutas de media o larga distancia. A partir de este hallazgo se plantean tres hipótesis operativas:

1. Unidad de Patio :

Un vehículo dedicado exclusivamente a movimientos internos en un CEDIS o planta puede realizar múltiples traslados cortos durante un mismo día.

2. Operación Shuttle o Transferencias Locales:

Rutas repetitivas y extremadamente cortas (como planta a CEDIS dentro de la misma zona urbana) generan un volumen elevado de viajes diarios.

3. Error en la Asignación de Datos:

Es posible que T541 funcione como un “comodín” dentro del sistema, utilizado para registrar viajes cuando la unidad real no está especificada. Esto sugiere un problema crítico: la pérdida de trazabilidad real en más de 2,000 viajes.

Riesgo:

Si se confirma la tercera hipótesis, existen implicaciones graves:

- Inconsistencia en el cálculo de costos por unidad.
- Imposibilidad de correlacionar gastos de combustible, peajes y mantenimiento con la unidad correcta.
- Subestimación o sobreasignación del desgaste mecánico.
- Riesgo de incumplimiento en auditorías internas y externas.

6.4 Vehículos y Dollys

El análisis de las columnas Remolque 1, Dolly y Remolque 2 muestra discrepancias profundas entre la operación registrada y la operación físicamente posible.

Hallazgos clave

- **Remolque 1:** 99,679 registros, casi todos los viajes involucran un remolque.
- **Remolque 2:** 42,909 registros, operación significativa de Full (doble remolque).
- **Dolly:** 14,485 registros, severamente inferior a lo esperado.

Para un viaje en configuración Full, el Dolly es un componente indispensable. El hecho de que existan aproximadamente 28,500 viajes con Remolque 2 pero sin Dolly registrado implica:

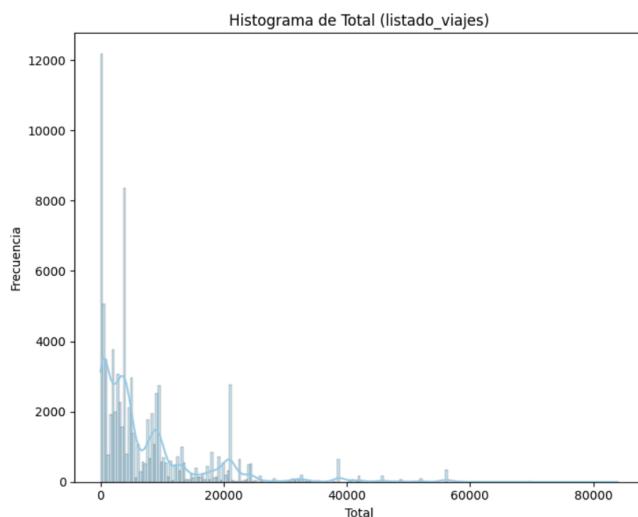
- Pérdida del 66% de trazabilidad de los Dollys.
- Riesgo elevado de robo, pérdida o daño no rastreado.
- Impacto negativo en mantenimiento preventivo (rellenos, rótulas, acoples).
- Inconsistencias en el cumplimiento de normativas federales y estatales.

El control de Dollys es uno de los puntos más críticos en cualquier operación Full debido a su valor, susceptibilidad al desgaste y alto riesgo de sustracción.

6.5 FINANCIEROS

Distribución sesgada del Ingreso

El análisis del 'Total' muestra una distribución altamente asimétrica. La mayoría de los viajes se encuentran en rangos muy bajos de ingreso (0–\$5,000 MXN), mientras que existen valores atípicos (outliers) que alcanzan hasta \$84,000 MXN.



Interpretación de la media y mediana

- **Media:** \$5,444 MXN

- **Mediana:** \$2,591 MXN

La media es más del doble de la mediana, lo que confirma una estructura de ingresos bimodal:

1. **Operación local:**

La mayoría de viajes generan ingresos bajos, típicos de movimientos urbanos o regionales.

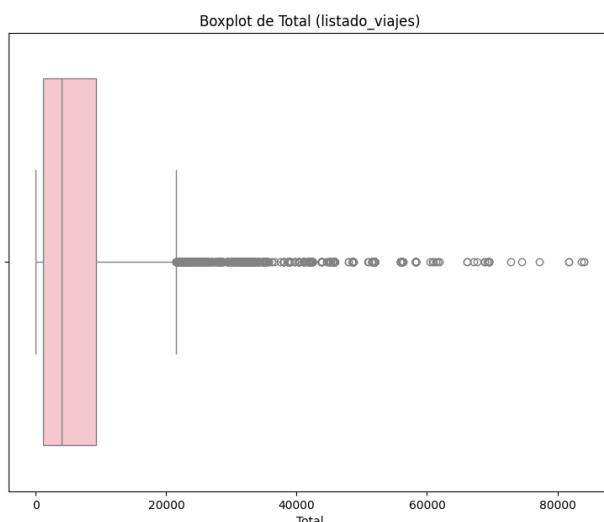
2. **Viajes de alto valor:**

Un pequeño porcentaje de viajes de larga distancia generan ingresos significativamente mayores y sostienen financieramente el promedio general.

Riesgo de gestión basada en promedios

Utilizar el ticket promedio como referencia puede llevar a:

- Sobreestimación del ingreso diario real.
- Proyecciones incorrectas de flujo de efectivo.
- Evaluaciones erróneas de rentabilidad por ruta.

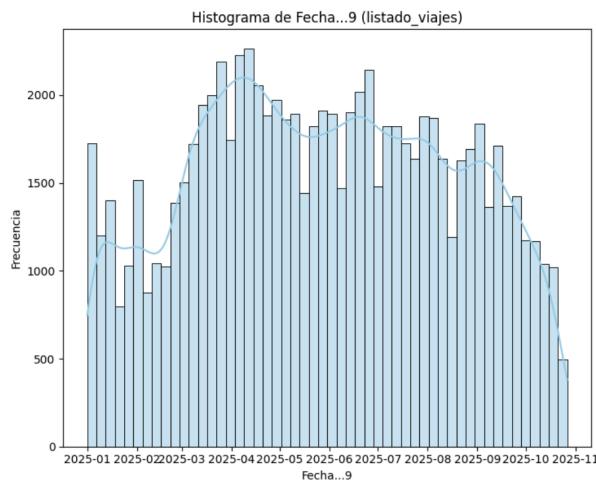


El boxplot del 'Total' presenta un cuerpo central estrecho y una extensa cadena de outliers a la derecha. Esto implica:

- Alta concentración de viajes de bajo monto.
- Importante dependencia financiera en un grupo minoritario de viajes de alto valor.

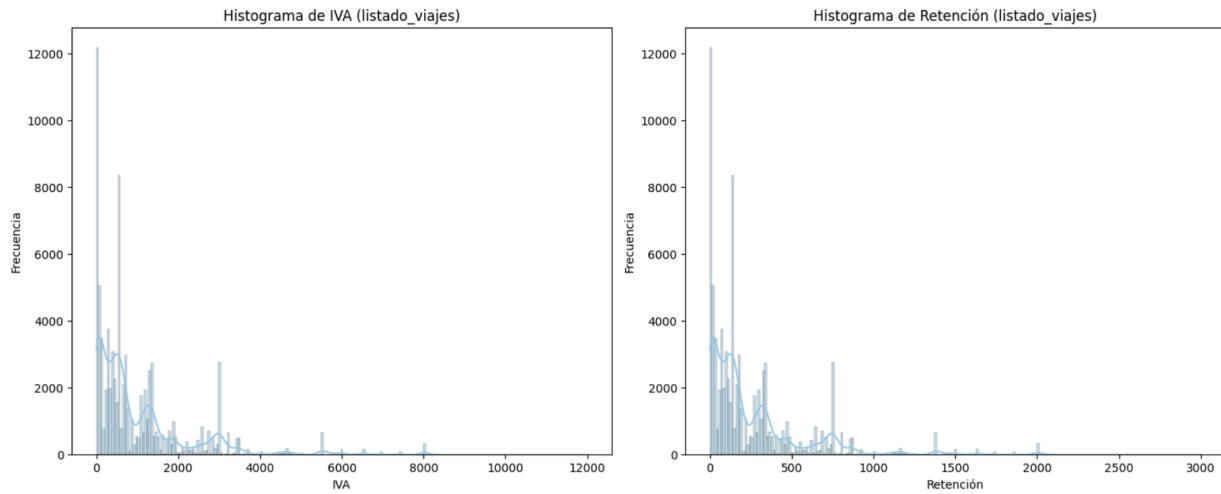
Desde la perspectiva de negocio, estos outliers deben ser gestionados con gran precisión debido a su impacto desproporcionado sobre la rentabilidad.

Comportamiento de Viajes a través del tiempo:



El histograma muestra que los pedidos tienen una estacionalidad clara a lo largo del año. De enero a abril el volumen sube de forma constante, alcanzando su punto más alto en marzo-abril. De mayo a agosto la actividad se mantiene estable, sin picos fuertes. A partir de septiembre el volumen disminuye progresivamente hasta noviembre. En resumen, el negocio tiene un inicio de año muy activo, una fase media estable y un cierre de año con menor demanda

Columna retención



La retención del 4% (LIVA Art. 1-A, fracción II, inciso c) debe ser calculada como:

$$\text{Retención Teórica} = \text{SubTotal} \times 0.04$$

La existencia de esta columna permite auditar:

- La integridad del cálculo fiscal.
- La correcta configuración del sistema de facturación.
- La consistencia entre SubTotal, IVA y Total.

Desviaciones sistemáticas podrían indicar fallos significativos con potencial sancionatorio del SAT.

Liquidación:

La columna **Liquidación** contiene **35,993 valores nulos**, es decir:

Un 33% de los viajes no están liquidados.

Impacto:

- Imposibilidad de conocer el costo real por viaje.
- Riesgo de anticipo sin comprobación.
- Pérdida de control sobre gastos de operador y viáticos.
- Distorsión de KPIs de rentabilidad y margen por ruta.

En términos empresariales, esto constituye una **deuda técnica administrativa** severa.

OUTLIERS

Se calculó la diferencia entre *Fecha Salida* y *Fecha Llegada*, limpiando valores imposibles. Se obtuvo:

- **Promedio:** 9h 36m
- **Mediana:** 2h 22m
- **IQR:** 6h 16m
- **$Q3 + 1.5 \times IQR \approx 16.5$ horas** (límite superior razonable)

Conclusión:

La mediana confirma que la mayoría de los viajes son de **operación local**.

Outlier crítico: viaje de 212 días

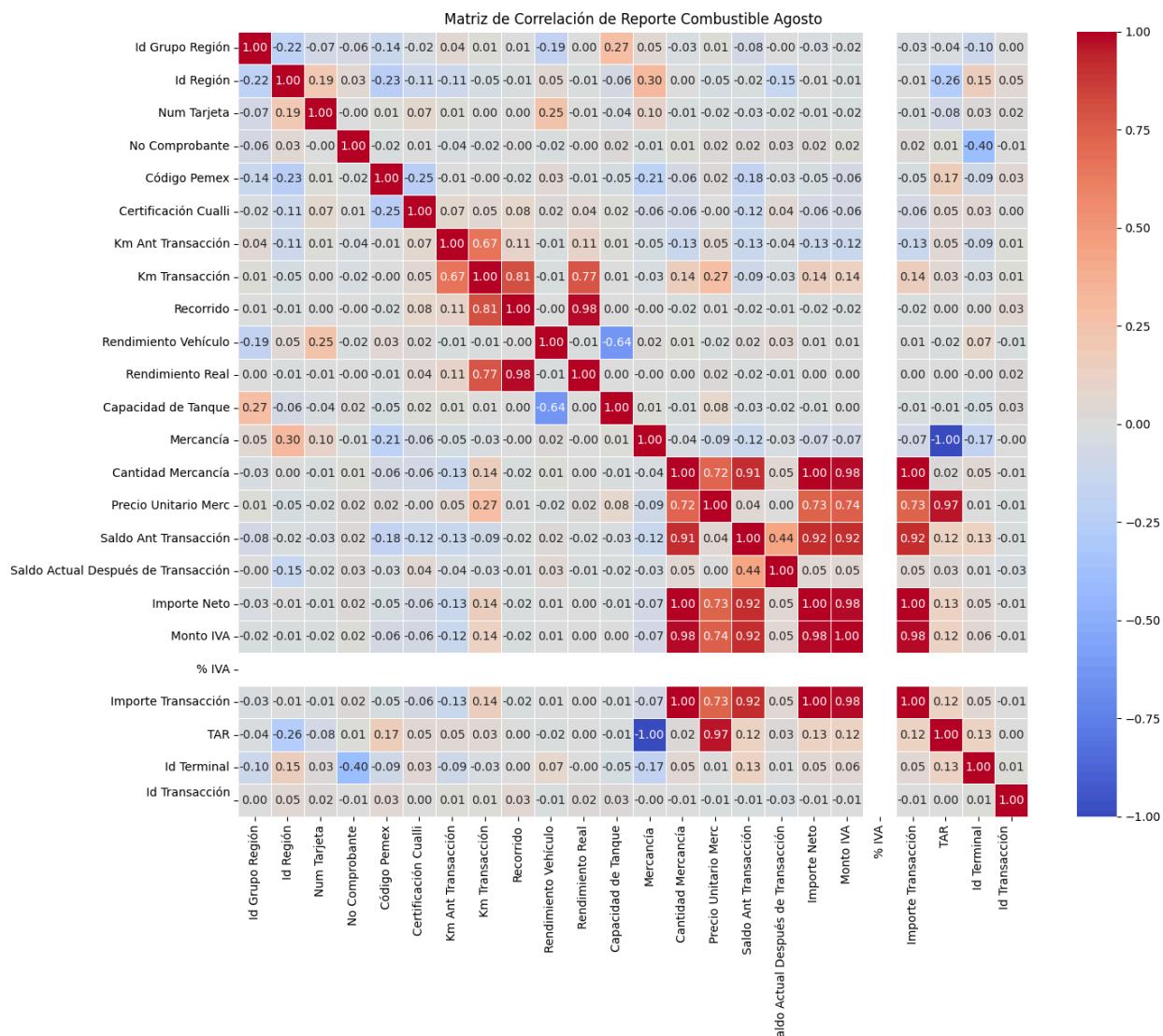
Un registro de 212 días es evidencia de:

- Falta de cierre administrativo.
- Captura manual deficiente.
- Deterioro grave del KPI “tiempo de tránsito”.

Estos registros deben excluirse del cálculo operacional, pero incluirse como evidencia de fallas administrativas.

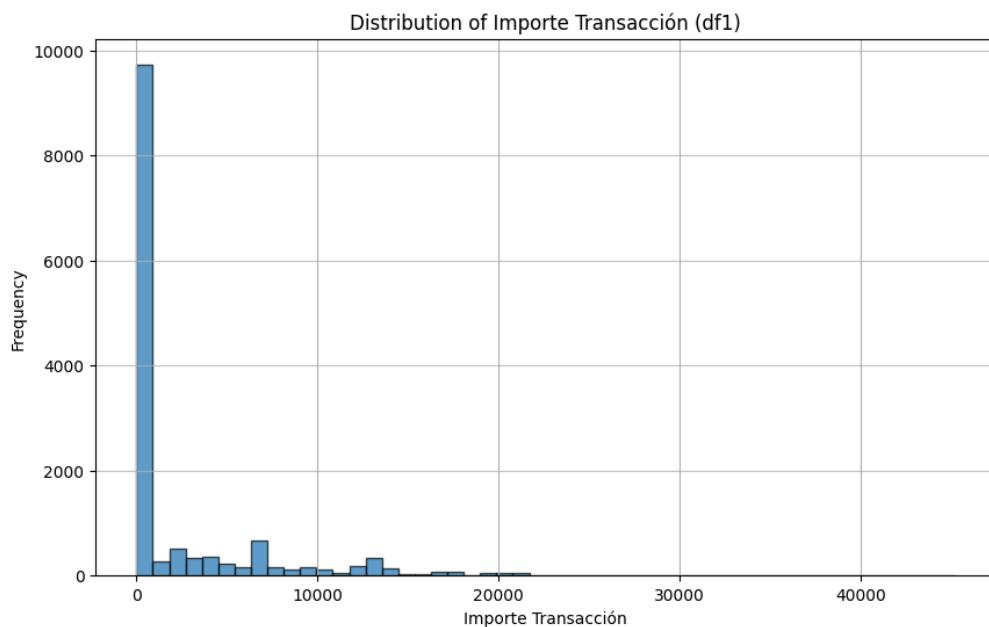
7. Análisis Exploratorio de Datos (EDA): Combustible

7.1 Mapa de calor



Se implementó un mapa de calor para poder determinar cuáles eran las variables que más relación tenían entre sí, y con ello poder enfocarnos en las que consideramos más relevantes para el estudio de combustible.

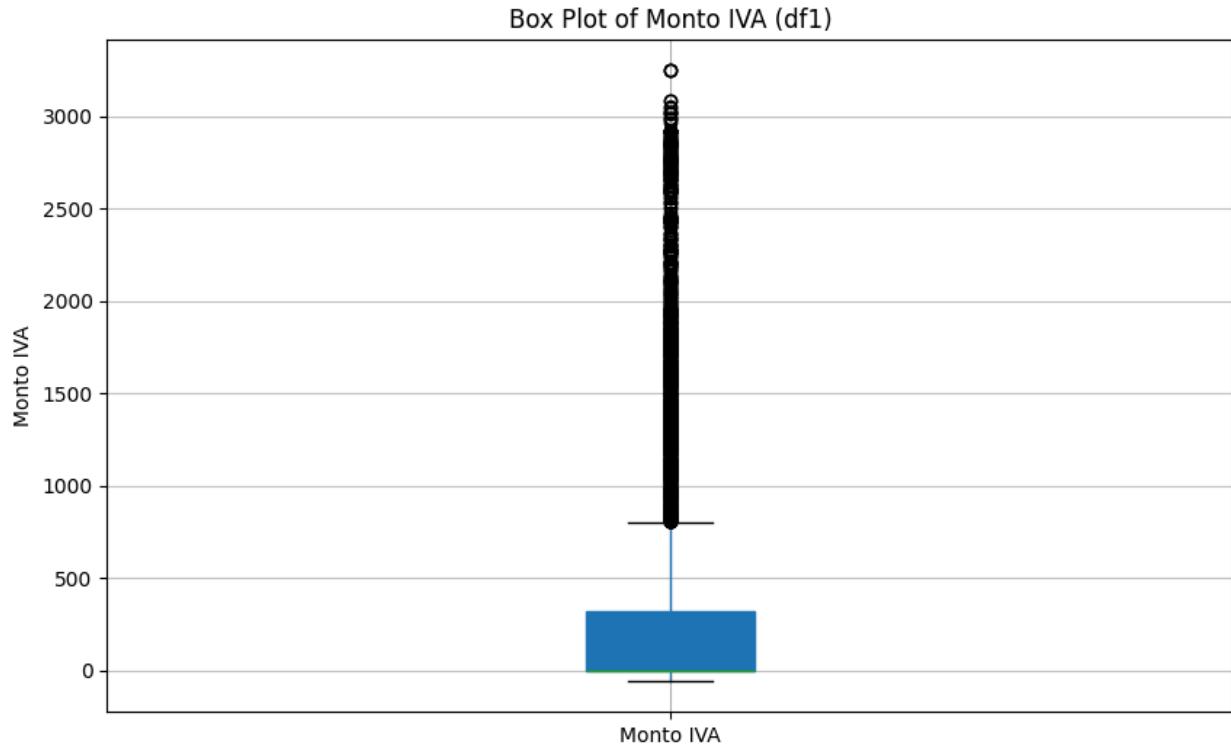
7.2 Importe Transacción



El histograma de Importe Transacción ² muestra una distribución asimétrica positiva (sesgada a la derecha), lo cual es indicador clave de datos económicos ya que generalmente se repiten más los importes de menor cantidad, no obstante revela comportamientos operativos específicos. Existe una concentración significativa de transacciones en rangos de importe bajos y medios. Lo cual podría sugerir una gran cantidad de viajes en vacío, con poco llenado o un error.

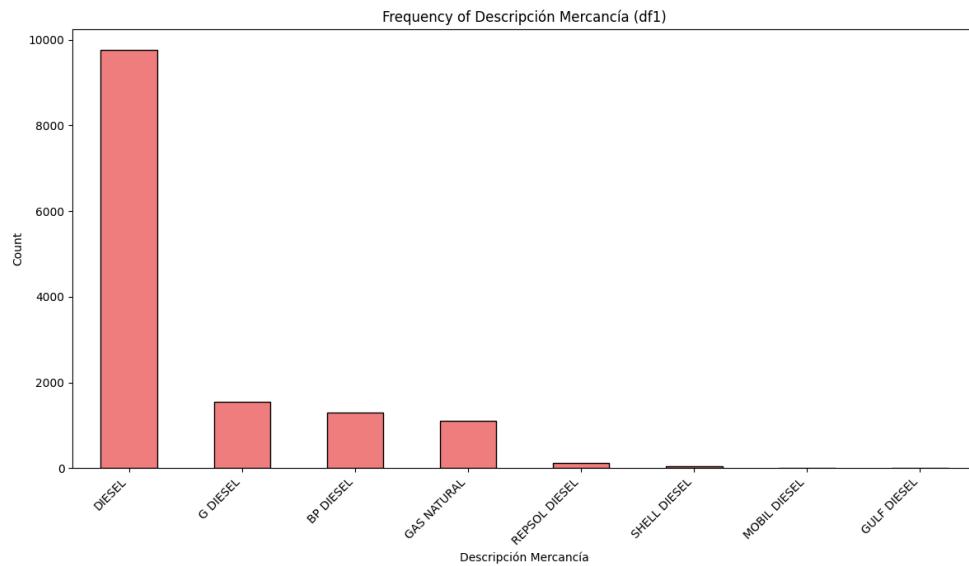
También se observan outliers en el extremo superior de la distribución de importes. Lo cual sugiere que son menos frecuentes los viajes en los que las unidades van con carga completa (Tracto Full).

7.3 Monto IVA



Usamos el IVA como proxy del monto total, ya que son directamente proporcionales. Los outliers en IVA son outliers en Importe Total.

La gráfica de caja del componente fiscal (IVA) confirma la existencia de valores extremos considerables. Mientras el 50% de las transacciones se mantiene en un rango estándar y bajo, los outliers alcanzan valores superiores a 3,000 MXN sólo de IVA (aprox. \$18,750 MXN de carga total).



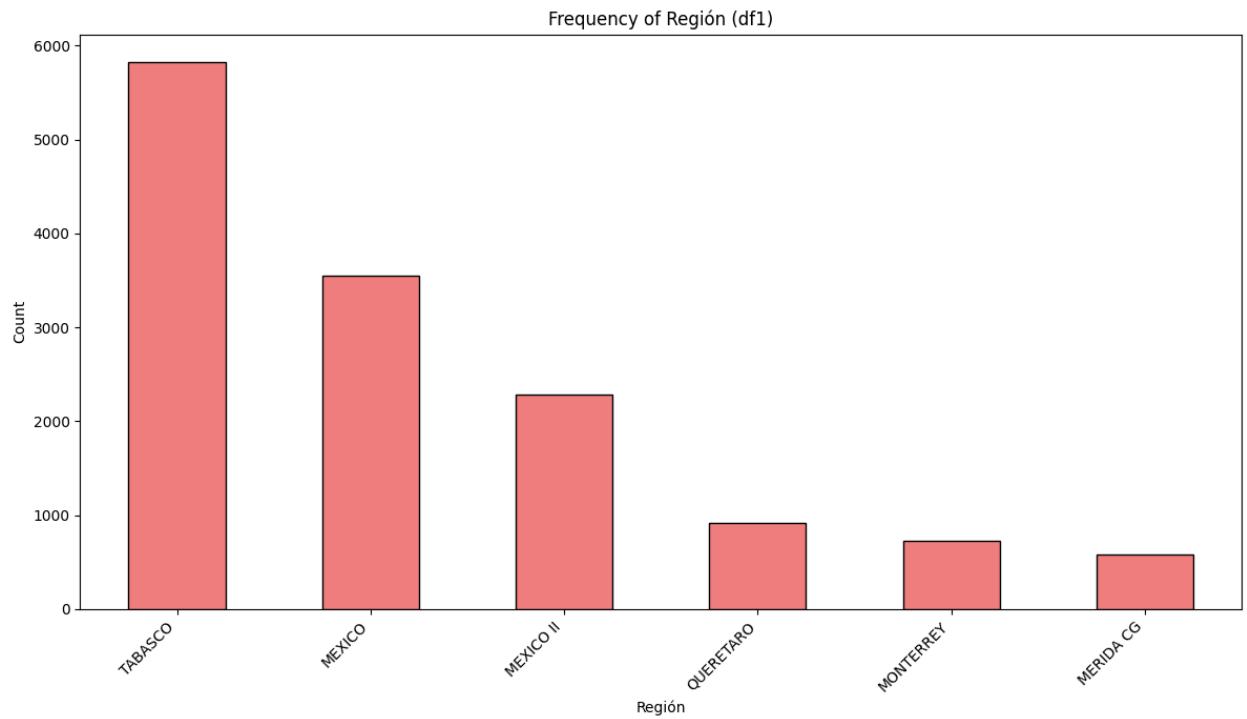
El análisis de frecuencia por tipo de producto válida que la flota es casi exclusivamente diésel.

Por lo que también se considera que el precio del combustible en México no es uniforme. Factores como la logística de distribución de PEMEX, impuestos estatales y la competencia local crean zonas de precios diferenciados. El análisis por Grupo Región² revela la huella económica de la flota.

La inmensa mayoría de las transacciones se agrupan bajo la región "CENTRO". Esto es consistente con la estructura logística nacional, donde el eje Ciudad de México - Querétaro - Puebla concentra la mayor actividad de distribución.

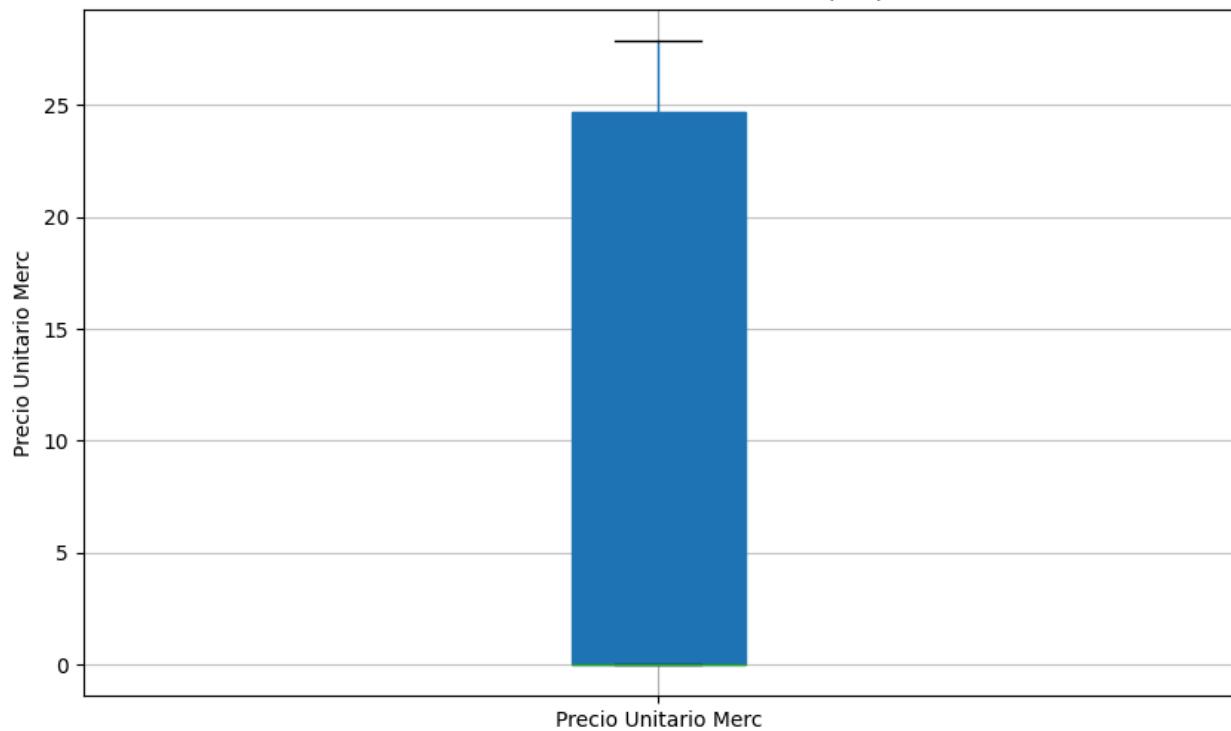
La visualización destaca a TABASCO y MÉXICO como los nodos dominantes de consumo. Esta concentración geográfica no sólo dicta la logística, sino la exposición al riesgo. La altura desproporcionada de e

7.4 Región

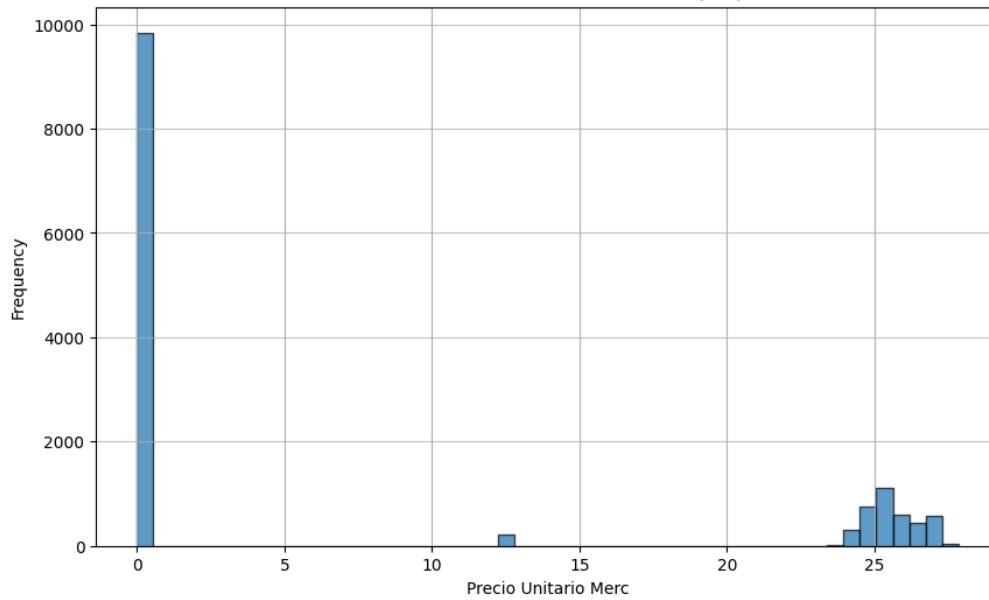


Estas barras indican que cualquier interrupción en el suministro o aumento de inseguridad en estas dos zonas específicas podría paralizar una porción significativa de la operación total. Por lo que sería buena opción buscar alternativas de rutas para reducir la exposición a siniestros.

Box Plot of Precio Unitario Merc (df1)



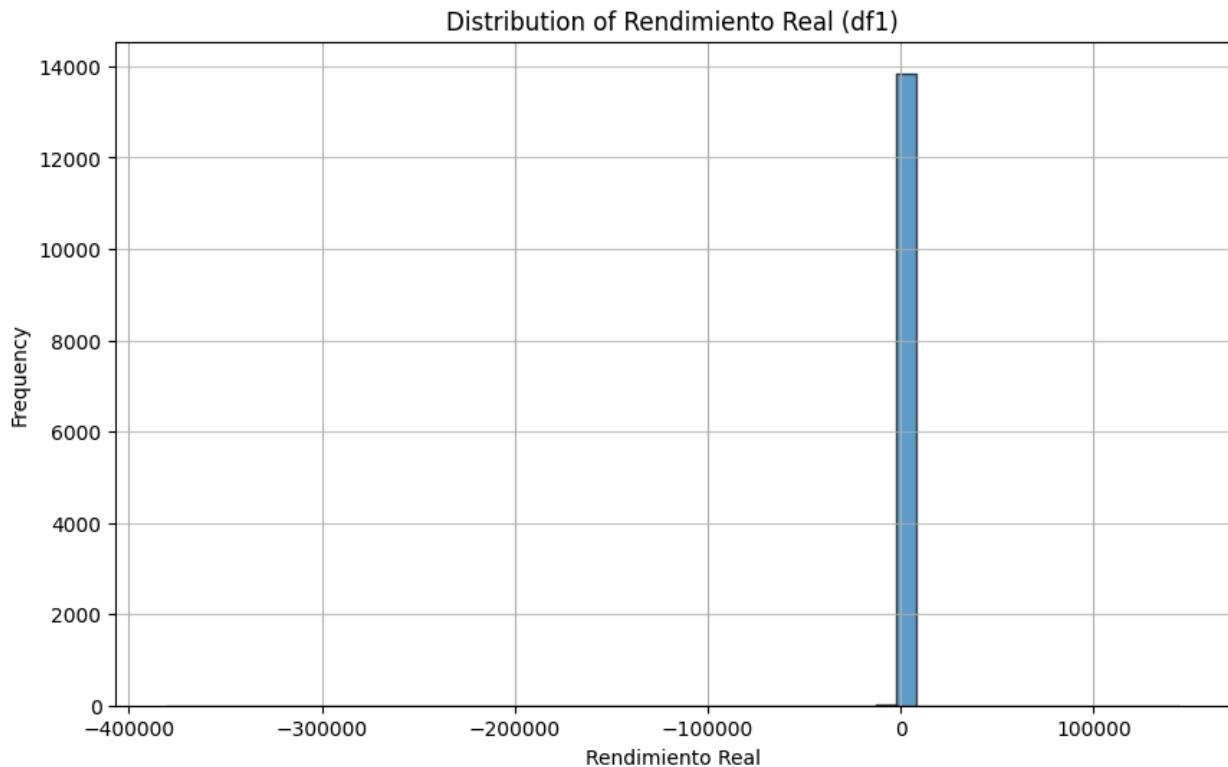
Distribution of Precio Unitario Merc (df1)

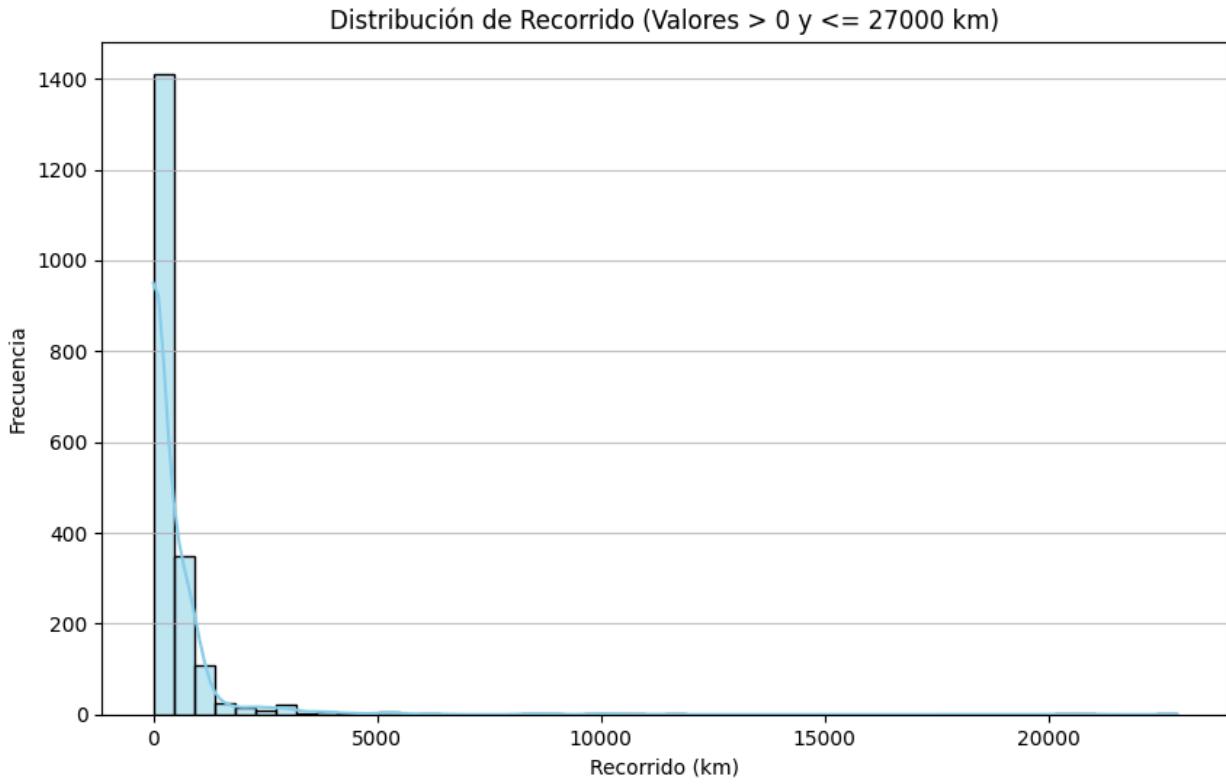


El histograma de precios revela un comportamiento concentrado en una regiones con precios muy definidos (cerca de los 25 MXN). Lo cual es esperable debido a que a nivel nacional el costo del diesel es de 26.27 MXN.

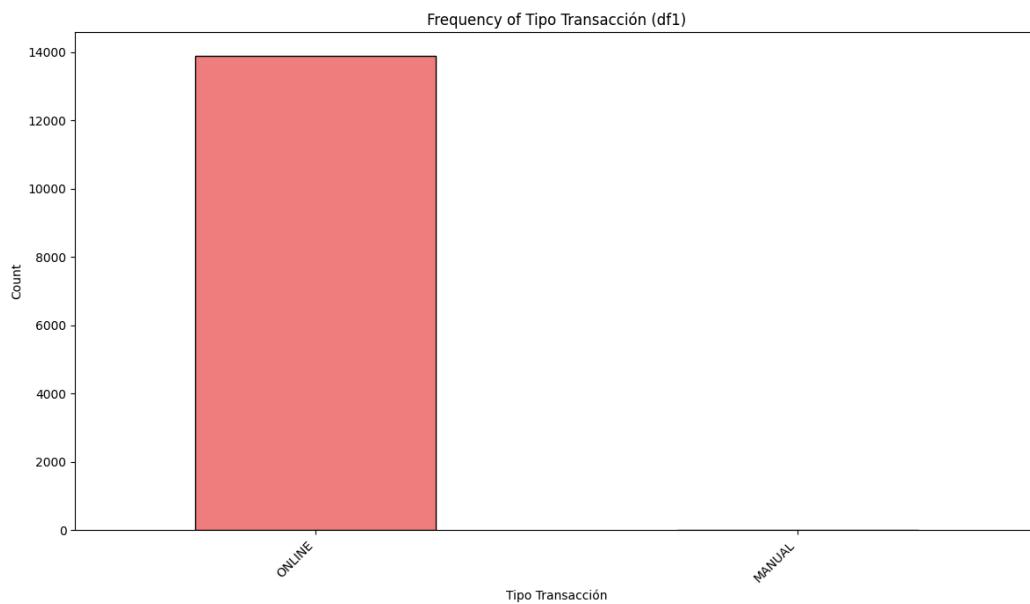
7.5 Rendimiento Real

En el apartado de Rendimiento real se observó que había muchas inconsistencias en la base de datos, debido a que habían valores negativos que superan por mucho el valor de rendimiento de vehículo cuyos valores oscilan entre (2.4 y 2.6) . Sin embargo al momento de limpiar la base se pudo observar que el rendimiento se encuentra en promedio en 1.8.





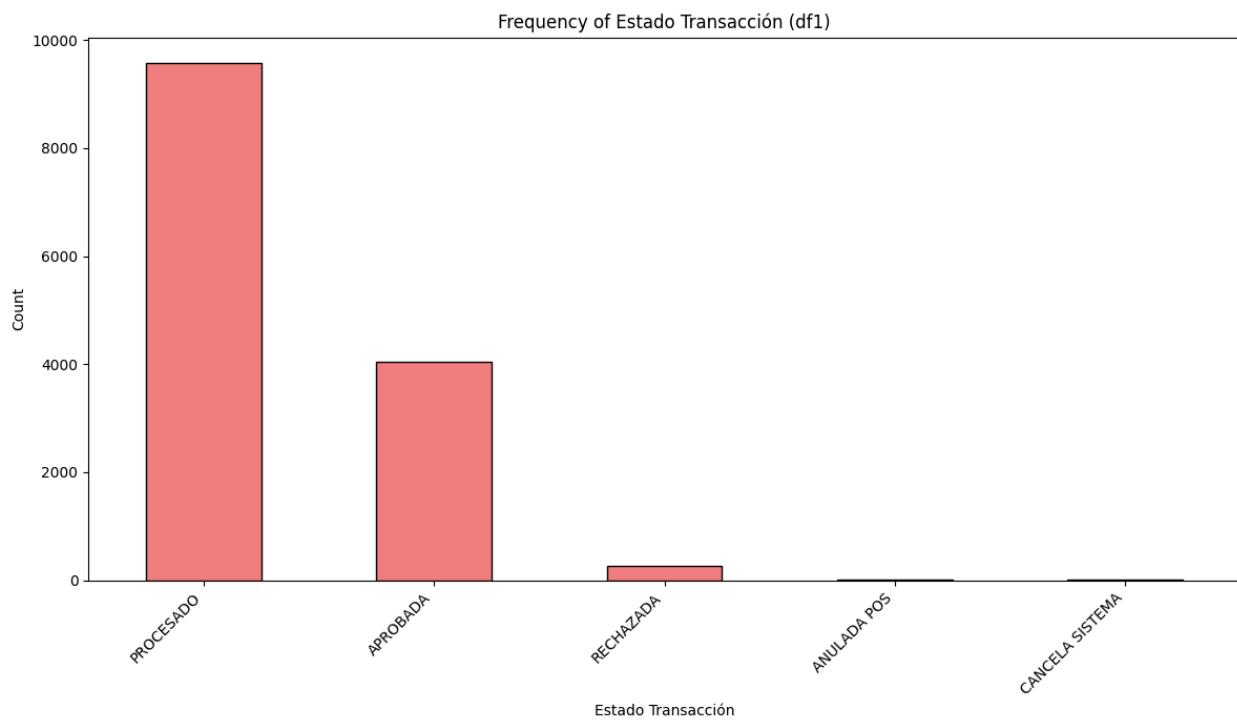
Al momento de estudiar la distribución de los kilómetros nos percatamos que habían demasiados valores atípicos. Por lo que después de tratar los datos logramos observar que la mayoría de los valores se encuentran entre los 1400 km. Por lo que inferimos que la mayoría de las unidades no hacen viajes largos.



La gráfica muestra una predominancia absoluta de transacciones "**ONLINE**".

Esto es una excelente señal de integridad y control de datos. Una transacción "Online" implica que la validación de saldo, NIP y reglas de negocio (litros máximos, tipo de combustible) se ejecutó correctamente.

7.6 Transacción



El análisis de los estados de la transacción revela la salud operativa de la flota en las estaciones de servicio. La inmensa mayoría de los intentos resultan exitosos. Esto indica que las tarjetas están correctamente configuradas y con fondos suficientes.

Por otro lado, la presencia de situaciones en las que fue rechazada muestra que todavía falta mejorar la situación de las transacciones.

6. KPIs

6.1 La Importancia de los KPIs

Los Indicadores Clave de Desempeño (KPIs) actúan como recursos de medición. Estos deben ser cuantificables, consistentes y alineados con los objetivos estratégicos de la organización.¹ En el análisis siguiente, no solo calcularemos estadísticas descriptivas, sino que propondremos una taxonomía de KPIs derivados directamente de la fusión de las tablas de datos, cubriendo dimensiones de tiempo, costo, productividad y calidad del servicio.

6.2 KPIs Propuestos

Basandonos en fuentes de consulta revisadas¹, se proponen los siguientes KPIs:

- **Costo de Infraestructura por Km (CPK):**

Importe (Peajes) / Distancia (Telemetría)

Objetivo: Minimizar (sin sacrificar velocidad)

Mide la eficiencia de la selección de ruta. Un CPK alto solo se justifica si la velocidad promedio aumenta proporcionalmente.

- **Integridad de Activos (Tag Ratio):**

Conteo tags / Conteo camiones

Objetivo: 1.0

Control de Fugas. Actualmente el ratio es 1.06 (403/377), indicando descontrol.

- **Ratio de Ralentí:**

Tiempo Ralentí / (Tiempo en conducción + Tiempo Ralentí)

Objetivo: < 10% (Estándar Industria)

Control de Desperdicio. Impacto directo en el consumo de combustible.

- **OTIF Proxy (Nivel de Servicio Estimado):**

(Número de viajes con estatus 'TERMINADO' y duración \leq Target / Número total de viajes) x100

Objetivo: Definido por ruta o ventana de cita del cliente.

Alinear la operación con el estándar del 98% exigido por el mercado retail.

- **Periodo de Financiamiento Operativo**

Fecha Vencimiento (Cobro)-Fecha Salida (Gasto Operativo)

Objetivo: Minimizar la brecha temporal entre la ejecución del gasto logístico (salida de ruta) y la exigibilidad del cobro (vencimiento de factura).

Los 3,693 registros sin Factura representan un ciclo infinito. Reducir los "Viajes sin Factura" es la palanca más rápida para mejorar el flujo de caja.

- **Rentabilidad por Kilómetro:**

SubTotal (Ingreso)/Distancia Real Recorrida (GPS)

Objetivo: Maximizar el retorno financiero por unidad de distancia recorrida, reduciendo la brecha entre los kilómetros facturados y los kilómetros realmente ejecutados (GPS).

Maximizar el ingreso por kilómetro es más eficiente que simplemente maximizar el volumen de viajes. Permite identificar rutas que, aunque generan ingresos altos, tienen costos de operación (distancia/tiempo) desproporcionados.

- **Costo por Kilómetro (CPK) - Combustible:**

Σ Importe_Combustible / Σ Km_Recorridos

Objetivo: Controlar el impacto financiero directo del movimiento.

Antes de saber cuánto ganar (rentabilidad), es necesario saber cuánto cuesta mover la unidad. Este KPI define el punto de equilibrio operativo.

- **Tasa de Calidad de Odometría (Calidad de datos):**

$$(\text{Transacciones_con_Km_Valido} / \text{Total_Transacciones}) \times 100$$

Objetivo: > 95%

Garantizar la fiabilidad de los datos para cálculos de rendimiento. Actualmente es crítica (~30%).

- **Índice de Anomalías de Carga**

$$(\text{Cargas_>_Capacidad_Tanque} + \text{Cargas_Frecuencia_Anormal}) / \text{Total_Cargas}$$

Objetivo: 0% (Tolerancia Cero)

Detectar y prevenir el robo de combustible y fraudes.

- **Ratio de Mantenimiento (Salud de Flota):**

$$\text{Eventos de Mantenimiento Correctivo} / \text{Eventos de Mantenimiento Preventivo}$$

Objetivo: Maximizar (> 1)

Mide la proactividad de la gestión de activos. Un ratio superior a 1 indica una cultura de prevención (ahorro); un ratio inferior a 1 indica que la operación vive en "modo bombero" (apagando fuegos), lo cual dispara los costos variables y el tiempo de inactividad no planificado.

- **Tasa de Ocio**

Fórmula:

$$\frac{\text{Horas Totales Disponibles} - \text{Horas Activas (Motor Encendido)}}{\text{Horas Totales Disponibles}} \times 100$$

Objetivo: Minimizar (Tendencia a 0%, idealmente <15% por tiempos de carga/descarga)

Representa el costo de oportunidad del activo. Un camión estacionado sigue generando costos fijos (seguro, depreciación, sueldo base) pero ingresó cero.

- **MAPE (Error Porcentual Absoluto Medio):**

$$\frac{1}{n} \sum |(\text{Demanda Real} - \text{Demanda Pronosticada}) / \text{Demanda Real}| \times 100$$

Objetivo: Minimizar

Mide la precisión de la planeación. Si el MAPE es alto, habrá un exceso de flota parada (costo de ociosidad) o falta de camiones para cubrir la demanda (venta perdida).

- **Coeficiente de Variación (CV) de la Demanda**

$$(\text{Desviación Estándar de Pedidos} / \text{Media de Pedidos}) \times 100$$

Objetivo: Clasificar (Bajo = Demanda Estable; Alto = Demanda Volátil)

Mide la volatilidad del cliente.

- **CV Bajo:** Permite asignar una flota dedicada fija (costo menor).
- **CV Alto:** Obliga a tener flota flexible o subcontratada (costo mayor). Ayuda a decidir qué tipo de contrato de transporte se necesita para cada ruta o cliente.

Obras citadas

1. KPIs en Logística. Principales y beneficios de utilizarlos. - Across Logistics, fecha de acceso: noviembre 19, 2025, <https://acrosslogistics.com/blog/kpis-en-logistica>
2. Real Time Data Integrations Best Practices - Microsoft Learn, fecha de acceso: noviembre 23, 2025,
<https://learn.microsoft.com/en-us/xandr/data-providers/best-practices-for-real-time-data-provider-integrations>
3. Cómo Funciona la Retención de IVA a Transportistas en 2024 - Grupo Altea MG, fecha de acceso: noviembre 23, 2025,
<https://www.grupoaltea.org/post/retenci%C3%B3n-de-iva-a-transportistas>
4. Tarifas - Autopista Arco Norte, fecha de acceso: noviembre 19, 2025,
<https://www.arconorte.com.mx/tarifas/>
5. Foros - Fiscal - ¿QUE SIGNIFICA TAR? - Página 1 - Fiscalia, fecha de acceso: noviembre 23, 2025,
<https://www.fiscalia.com/modules.php?name=Foros&sop=verTopico&idTema=522>
6. fecha de acceso: noviembre 23, 2025,
<https://www.factugas.com/nosotros.html#:~:text=PEMEX%20mediante%20su%20certificacion%20cualli,se%20realiza%20en%20dispensarios%20que>
7. Walmart Requires 98% OTIF Compliance from Suppliers - Blog, fecha de acceso: noviembre 23, 2025,
<https://blog.gettransport.com/walmart-requires-98-otif-compliance-from-suppliers/>
8. Walmart changes OTIF Requirements - Zipline Logistics, fecha de acceso: noviembre 23, 2025, <https://ziplinelogistics.com/blog/walmart-changes-otif-requirements/>
9. KPIs en la industria del transporte: ¿cuáles seguir? - AntsRoute, fecha de acceso: noviembre 23, 2025, <https://antsroute.com/es/solucion/kpi-transporte/>
10. Foros - Fiscal - ¿QUE SIGNIFICA TAR? - Página 1 - Fiscalia, fecha de acceso: noviembre 23, 2025,
<https://www.fiscalia.com/modules.php?name=Foros&sop=verTopico&idTema=522>