

**[https://www.canva.com/design/DAGkaqZD3AM/s1cGrBdHROWtn9LoEv6fHQ/edit?
utm_content=DAGkaqZD3AM&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton](https://www.canva.com/design/DAGkaqZD3AM/s1cGrBdHROWtn9LoEv6fHQ/edit?utm_content=DAGkaqZD3AM&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton)**

Maestría en Ciencia de Datos

MODELO DE PREDICCIÓN DEL CONSUMO DE DIÉSEL EN UNIDADES DE TRANSPORTE PÚBLICO

Realizado por:
Est. André Ramírez

Índice de CONTENIDOS



01. Introducción

02. Entendimiento de Datos

03. Preparación de Datos

04. Modelado

05. Resultados

06. Plan de Implementación

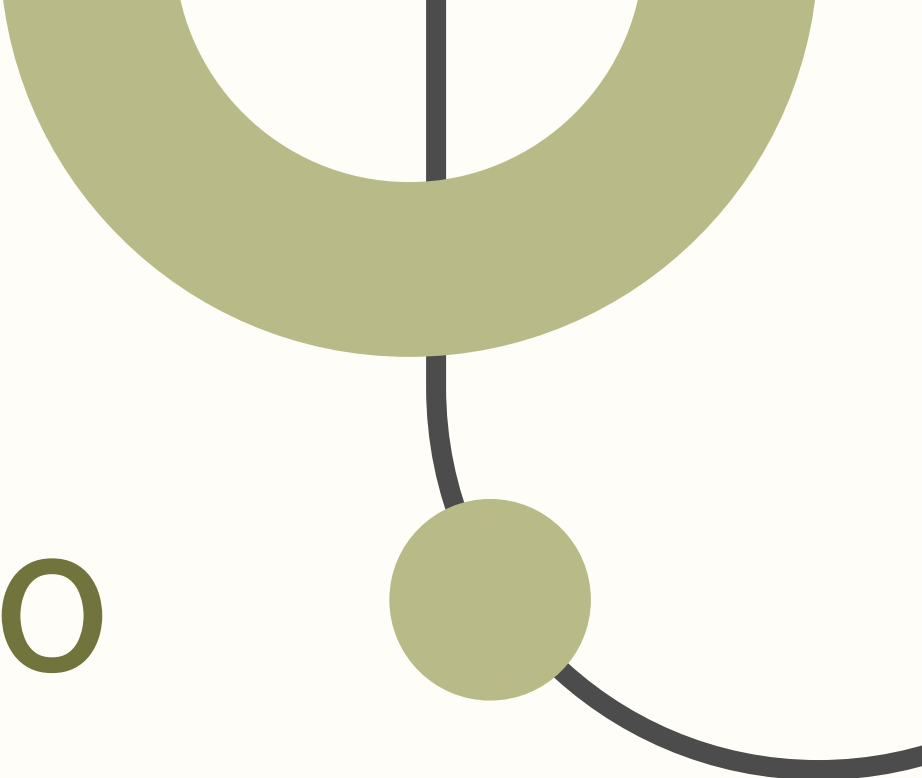
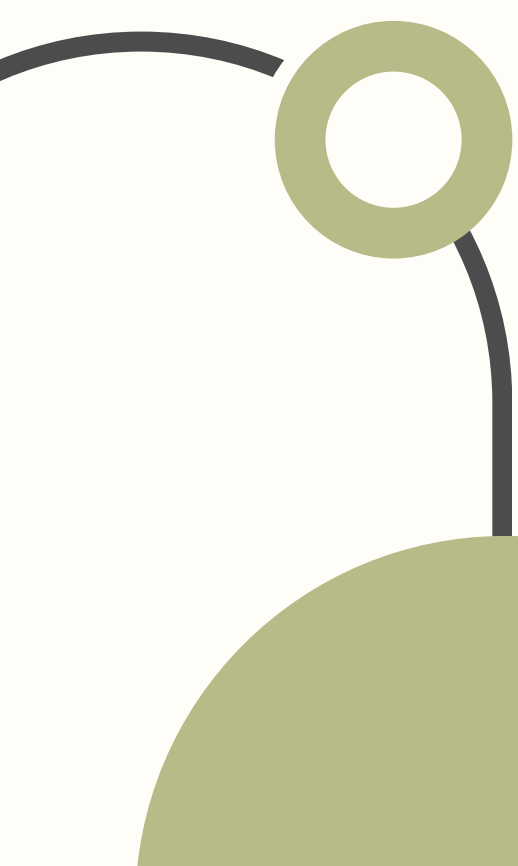
07. Conclusiones

Introducción

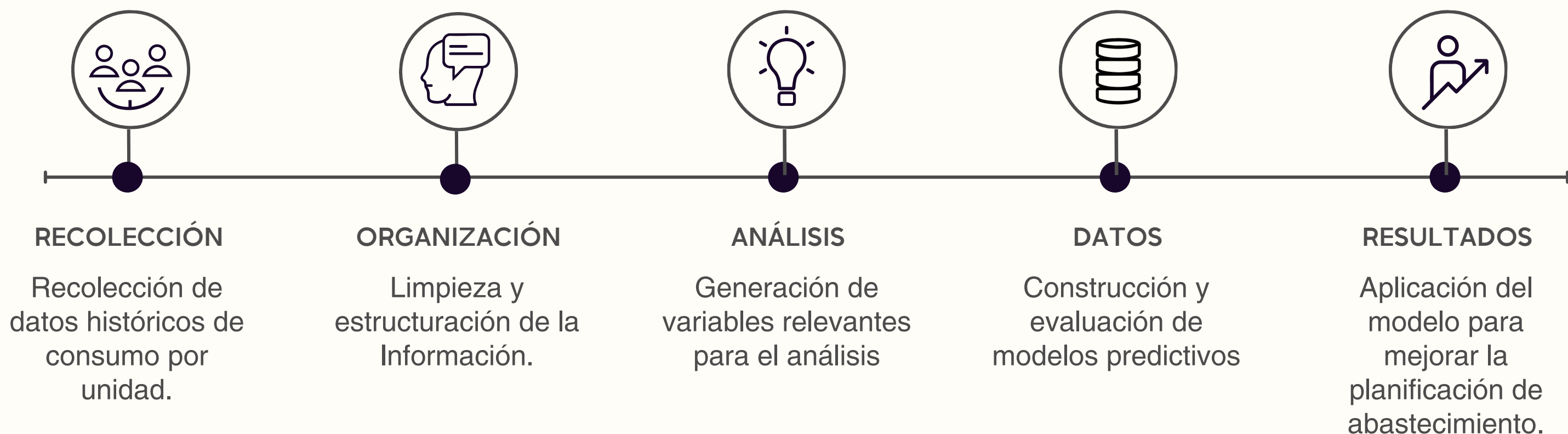
En el Transporte Público de Quito, una buena gestión del diésel es clave para que las unidades sigan operando sin interrupciones. El problema es que actualmente la planificación del consumo no se basa en datos históricos ni en predicciones. Esto puede causar retrasos, desperdicio de recursos y descoordinación en el mantenimiento de los buses y los surtidores. Por eso, el problema principal es la falta de un modelo que permita anticipar el consumo de diésel por unidad o por flota.

Oportunidad de Impacto

Una planificación eficiente del consumo de diésel tiene un impacto directo en la operación del transporte público de Quito. Actualmente, la planificación se hace sin considerar datos históricos o pronósticos, lo que causa ineficiencias operativas y logística.



Oportunidad de Impacto



Entendimiento de Datos

Partimos de un dataset con información histórica de consumo de diésel. Para garantizar su utilidad, realizamos:

- Limpieza de datos faltantes y registros inconsistentes.
- Estandarización de formatos de fecha y hora.
- Identificación de tendencias y estacionalidades.
- Análisis visual para entender los patrones de consumo en diferentes unidades y horarios.

Entendimiento de Datos

FECHA	DISPENSADOR	HORA	VEHICULO	FLOTA	GALONES	KILOMETRAJE
2020-01-01	1	10:35:00	S024	B12M	53,568	411475,8
2020-01-01	1	12:12:00	S029	B12M	54,197	479353,4
2020-01-01	1	12:20:00	S011	B12M	56,556	463696,8
2020-01-01	1	13:10:00	S074	B12M	65,614	451279,5
2020-01-01	1	13:30:00	S015	B12M	42,879	418781,5

Entendimiento de Datos

```
raw_diesel.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 228209 entries, 0 to 228208  
Data columns (total 7 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   FECHA           228209 non-null object  
1   DISPENSADOR     228209 non-null int64  
2   HORA            228209 non-null object  
3   VEHICULO        228207 non-null object  
4   FLOTA           193574 non-null object  
5   GALONES         228209 non-null object  
6   KILOMETRAJE     227988 non-null object  
dtypes: int64(1), object(6)  
memory usage: 12.2+ MB
```

```
raw_diesel.isna().sum()
```

✓ 0.0s

```
FECHA           0  
DISPENSADOR     0  
HORA            0  
VEHICULO        2  
FLOTA           34635  
GALONES         0  
KILOMETRAJE     221  
dtype: int64
```

Entendimiento de Datos

```
print('El dato duplicado es el siguiente: ')\nraw_diesel[raw_diesel.duplicated()]
```

El dato duplicado es el siguiente:

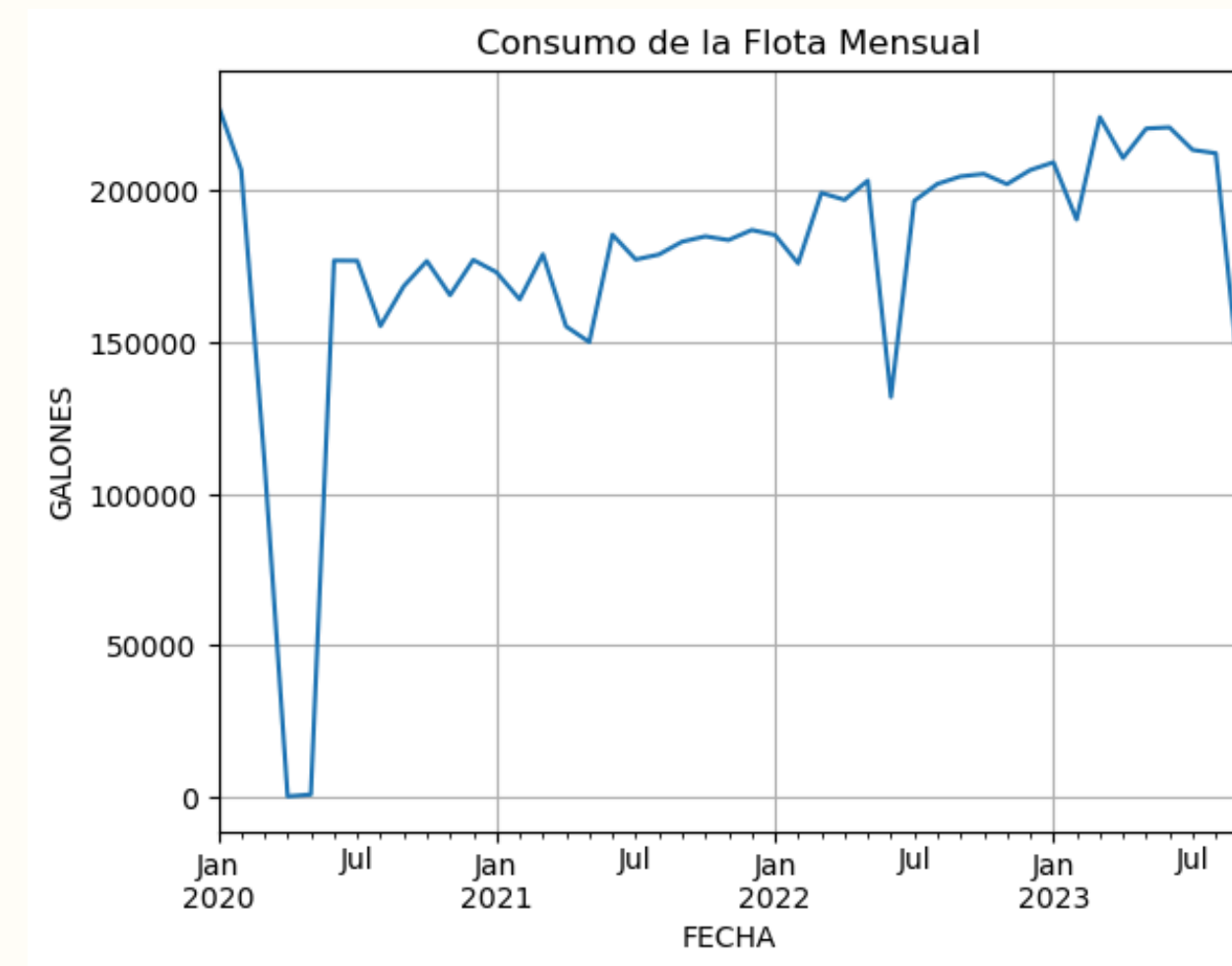
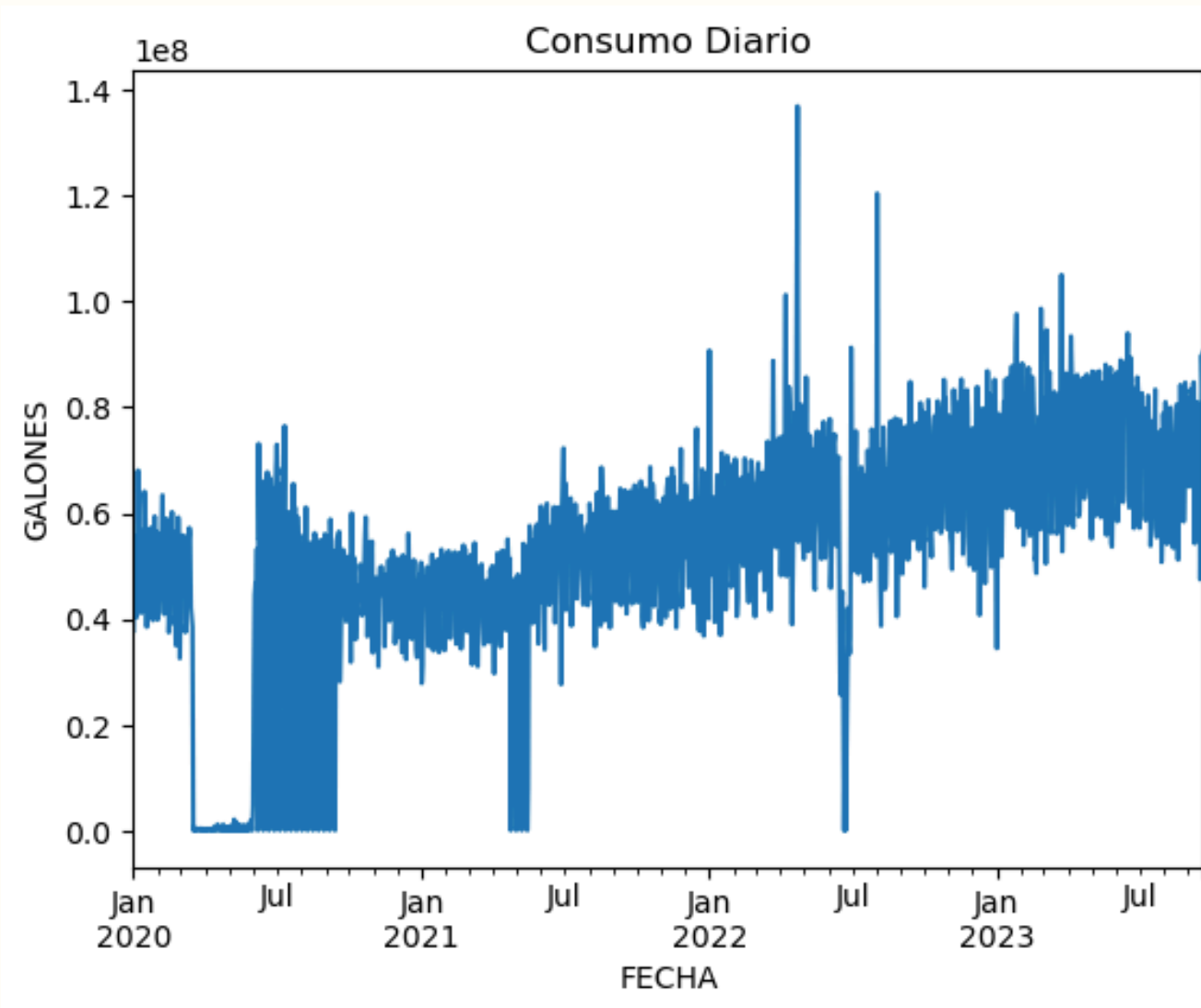
	FECHA	DISPENSADOR	HORA	VEHICULO	FLOTA	GALONES	KILOMETRAJE
23196	2020-08-03	4	10:15:00	T008	TROLEBUS	2,831	63703

Preparación de Datos

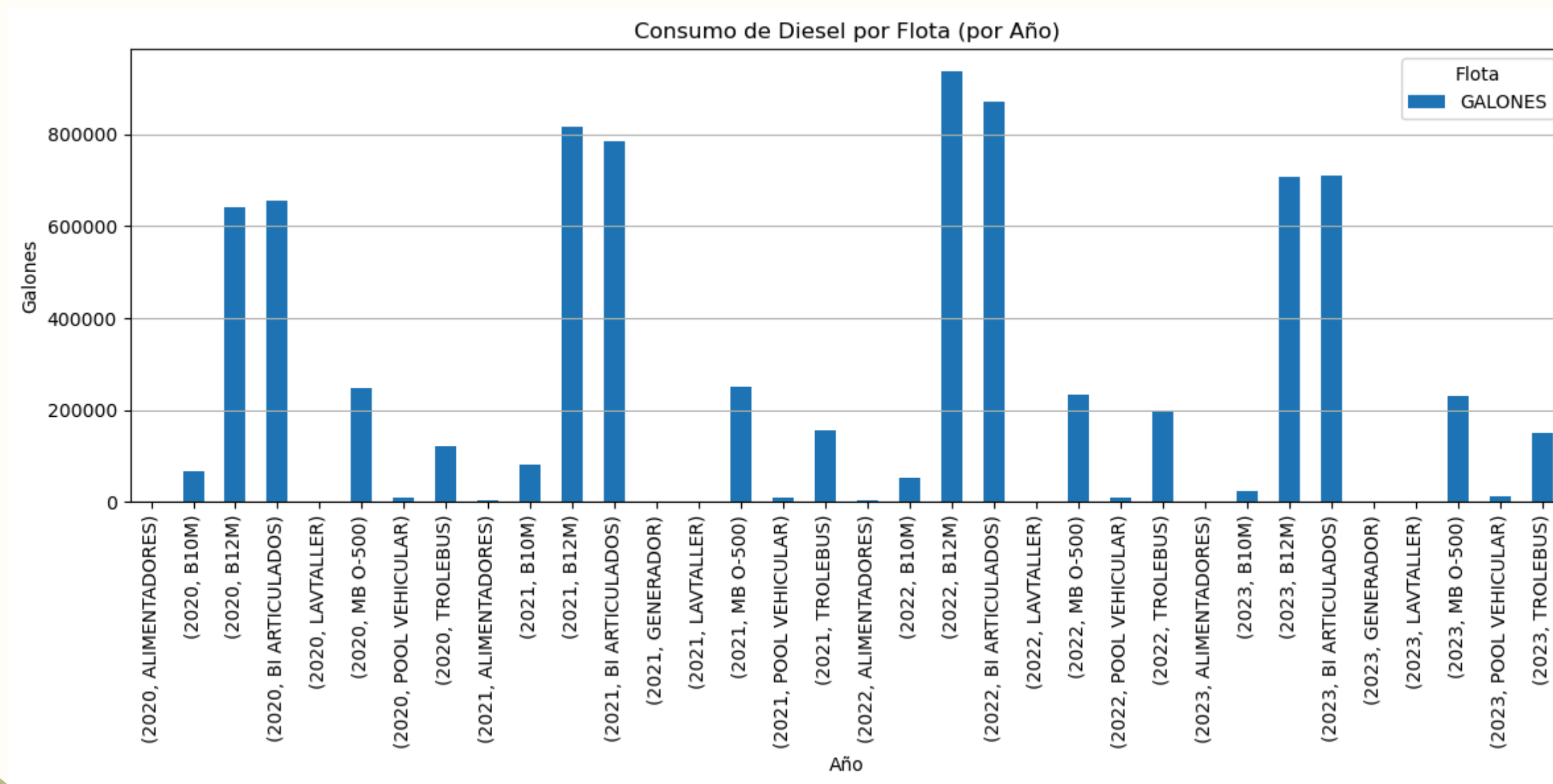
Transformamos los datos brutos en un conjunto estructurado y enriquecido mediante:

- **Variables temporales:** Año, mes, día, hora.
- **Variables de rendimiento:** Consumo por km recorrido.
- **Variables categóricas:** Identificador de unidad, tipo de día (laboral o fin de semana).
- **Otras transformaciones:** Creación de variables derivadas y codificación para uso en modelos.

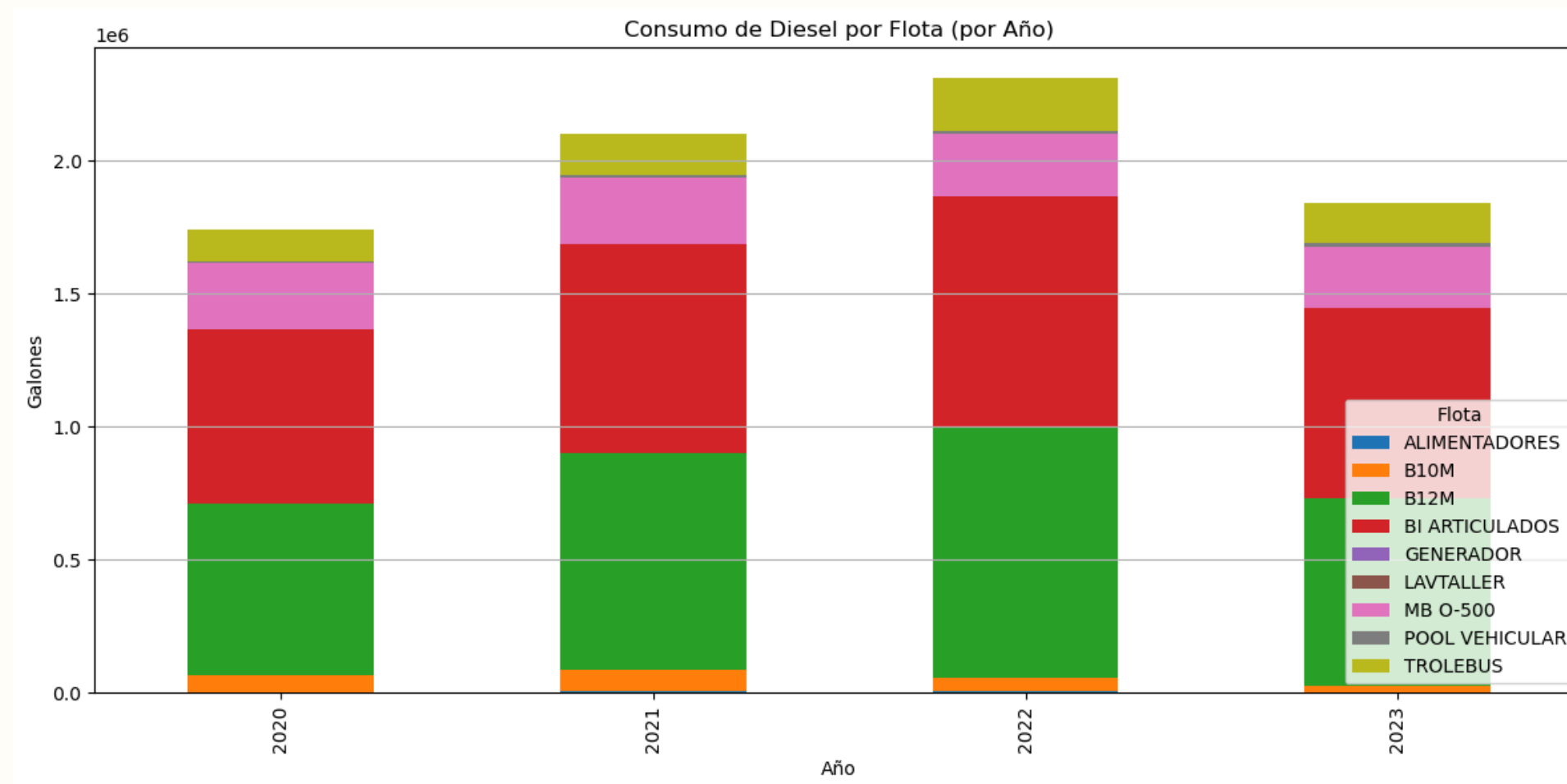
Preparación de Datos



Preparación de Datos



Preparación de Datos



El gráfico describe el consumo de diésel total durante el transcurso de 4 años, donde podemos apreciar que la flota trolebus es la que más diésel consume.

Modelado

Utilizamos distintos modelos de regresión para predecir el consumo de diésel a una hora específica, se utilizaron los siguientes modelos:

- Regresión Lineal.
- Árboles de Decisión.
- Random Forest
- Gradient Boosting Regressor (GBR).

Entrenamos, validamos y comparamos los modelos usando métricas RMSE, MAE y R2.

Resultados

Resultados de prueba:

	RMSE	MAE	R2
Regresion Lineal	1.800644e+05	1570.237159	0.999988
Decision Tree	1.759680e+07	93287.701677	0.884547
Random Forest	2.594157e+06	15750.818060	0.997491
Gradient Boosting	1.438646e+07	132966.421286	0.922830

Probamos cuatro modelos distintos y la Regresión Lineal demostró ser la más precisa, con una tasa de error mínima y una capacidad acertada de explicar el comportamiento del consumo de diésel. Esto nos permite anticipar la demanda con gran confianza, optimizando así el abastecimiento y la planificación operativa

Plan de Implementación

Este modelo podría integrarse en la operación diaria del sistema de transporte público mediante:

- Una herramienta de pronóstico diaria por unidad o grupo de unidades.
- Integración con los sistemas de planificación de abastecimiento.
- Alertas para identificar consumos fuera de lo esperado.
- Apoyo en la programación del mantenimiento preventivo de surtidores y unidades.

El sistema se puede automatizar para actualizarse con nuevos datos cada semana o mes.

Conclusiones

La Regresión Lineal tuvo el mejor rendimiento en todas las métricas:

- Bajo error absoluto (MAE): predice muy cercano al valor real.
- Excelente R2: explica casi el 100% de la variabilidad del consumo.
- Menor RMSE, lo que significa menor desviación en errores grandes.

Random Forest también es competitivo, con buen R2 y errores razonables, pero está por debajo en precisión frente a la regresión lineal

Decision Tree y Gradient Boosting presentaron errores significativamente más altos, lo que indica que no se adaptaron bien a la naturaleza de los datos.

Conclusiones

Implementar este modelo de pronóstico permitiría:

- **Reducir el riesgo de desabastecimiento de diésel.**
 - **Mejorar la eficiencia logística en estaciones de carga.**
 - **Optimizar el uso de recursos en mantenimiento y operación.**
 - **Tomar decisiones basadas en datos, alineadas con los objetivos operativos del sistema de transporte.**
-