

# Mejora del Desempeño de un servicio de Flotillas Terrestres con Inteligencia de Negocios y datos Telemáticos

Forza Transportation Services, Inc.

Presentan:

- Andrés Julián López Hurtado - A01793899
- Nathalia Milena Prada Hernandez - A01793999
- Víctor Alejandro Regueira Romero - A01794404

Asesor: Dr. Horacio Martínez Alfaro

Patrocinador(es): Mtro. Fernando Sebastián Sánchez Cardona

## Problemática

Forza Transportation es una empresa de transporte especializada en el servicio de **Full Truck Load (FTL)**



## Objetivo: Mejora del Desempeño del servicio

# 1

Aprovechar los datos generados por el sensor GeoTab para **identificar los factores clave** que afectan el rendimiento del combustible.

# 2

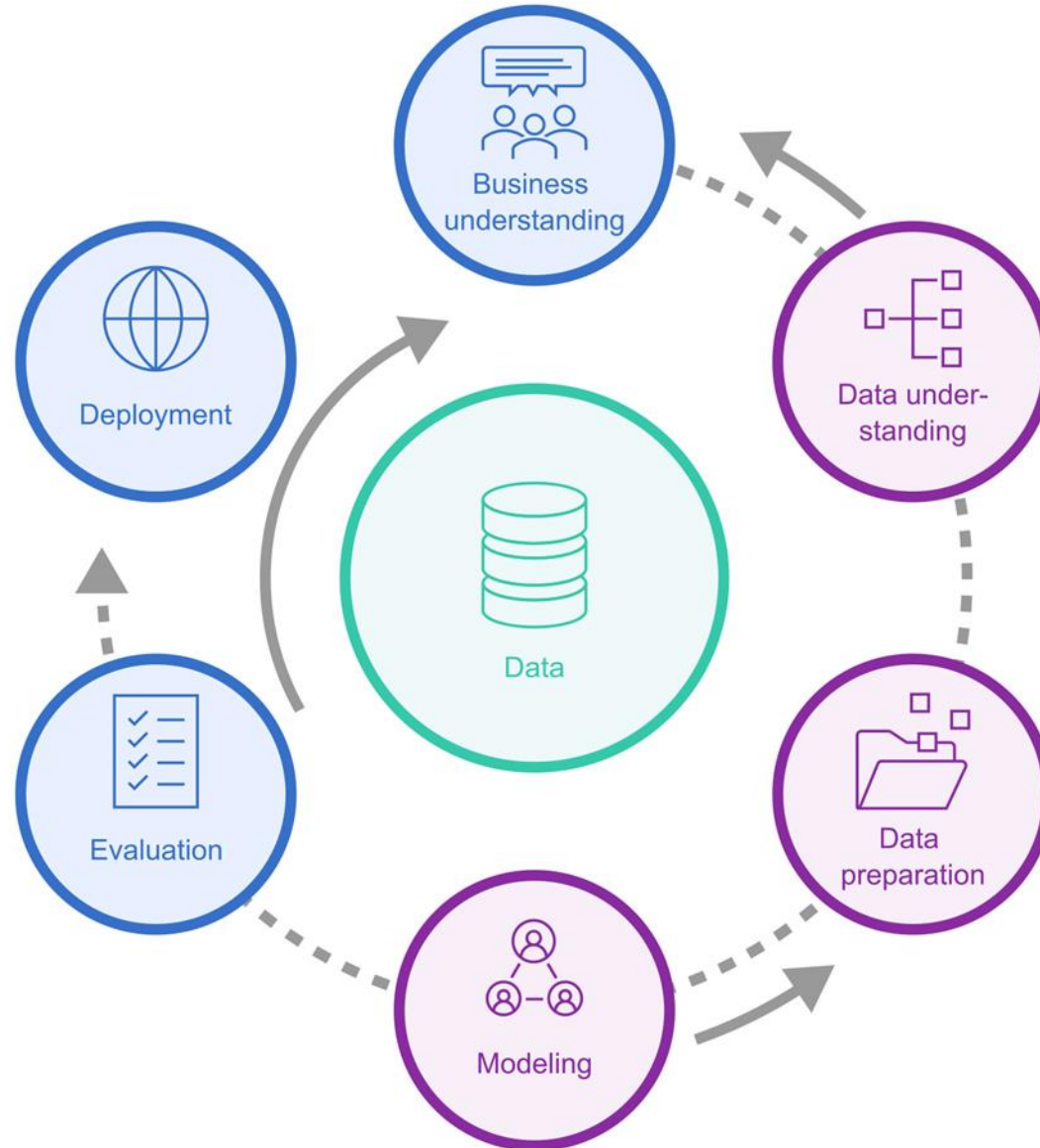
Generar **pronósticos de consumo** que permitan anticipar las necesidades de cada vehículo y optimizar la planificación operativa para reducir el consumo total.

# 3

Mejorar la agilidad y precisión en la gestión de la operación mediante una **visualización de datos y documentación de procesos** claras y accesibles.

# Solución propuesta

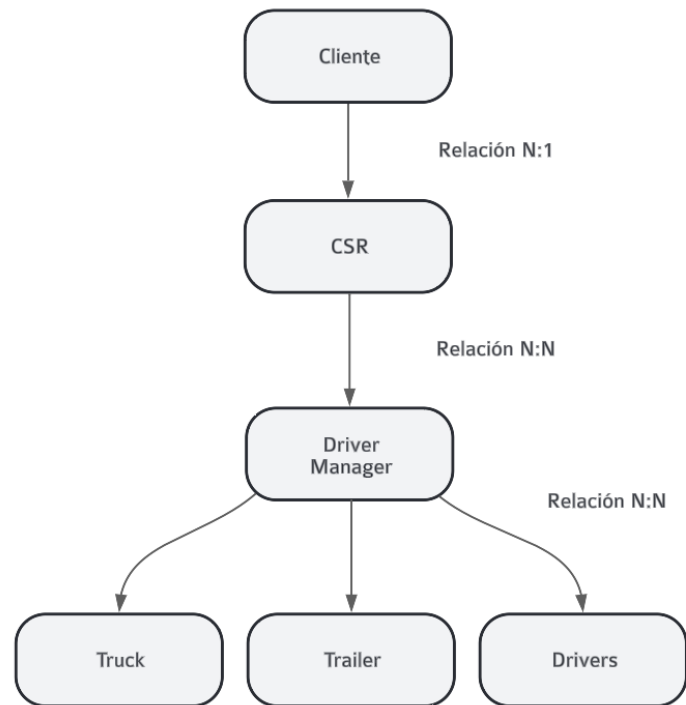
Metodología: CRISP-ML(Q)



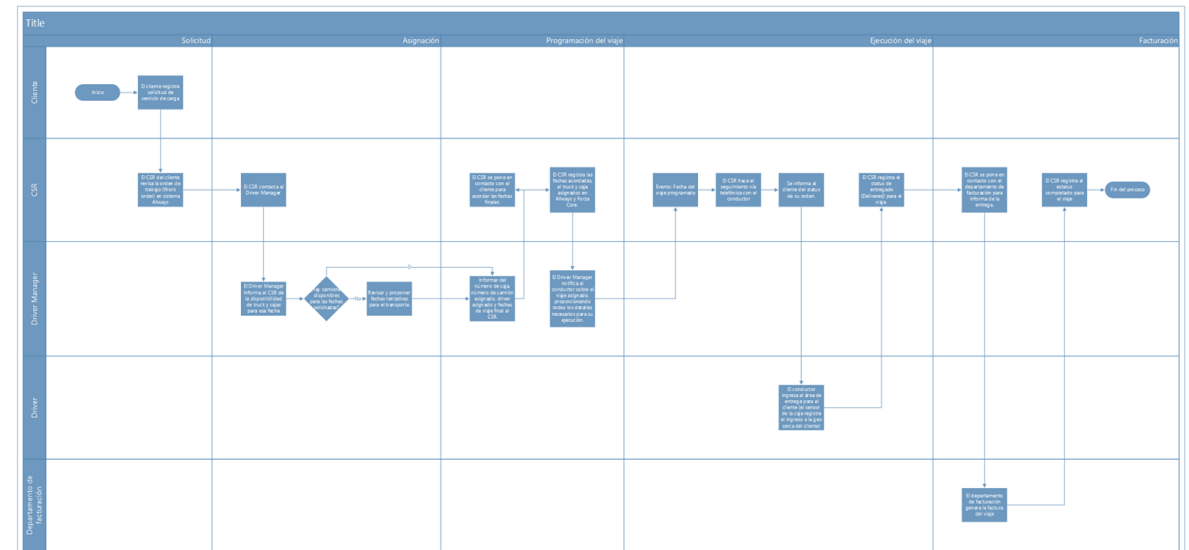


## Solución propuesta: Entendimiento del negocio

### Identificación de entidades de negocio



### Documentación del proceso



Data under-  
standing

## Solución propuesta: Entendimiento de los datos

### Tablas

| Safety                    |           |
|---------------------------|-----------|
| name                      | string pk |
| SpeedLimit1Count          | float     |
| TotalDrivingDuration      | float     |
| AverageSpeed              | float     |
| StopCount                 | float     |
| IdlingDuration            | float     |
| AfterHoursTripCount       | float     |
| TotalDistance             | float     |
| TotalStopsCount           | float     |
| SpeedingCount             | float     |
| SpeedingDistance          | float     |
| ExcessiveSpeedingCount    | float     |
| ExcessiveSpeedingDistance | float     |
| HardAccelerationCount     | float     |
| HardAccelerationDistance  | float     |
| HarshBrakingCount         | float     |
| HarshBrakingDistance      | float     |
| HarshCorneringCount       | float     |
| HarshCorneringDistance    | float     |
| MaxSpeed                  | float     |

| consumption             |           |
|-------------------------|-----------|
| date                    | date pk   |
| name                    | string pk |
| total_distance_used_sum | float     |
| total_fuel_used_sum     | float     |
| consumption             | float     |
| performance             | float     |

| Idling               |           |
|----------------------|-----------|
| date                 | date pk   |
| name                 | string pk |
| distance             | float     |
| idlingDuration       | string    |
| idlingPercent        | float     |
| idlingDuration_hours | float     |
| CurrentFuelPrice     | float     |
| IdlingCost           | float     |

# Tipo

CSV

# Registros

177,653

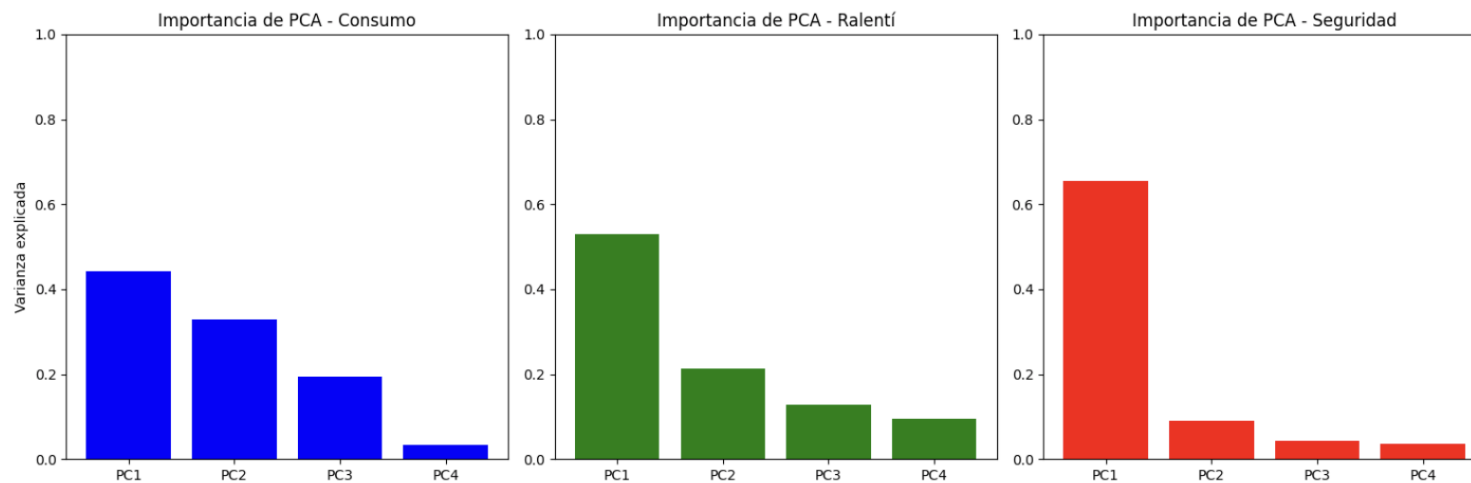
# Fecha

1 Julio – 31 Agosto

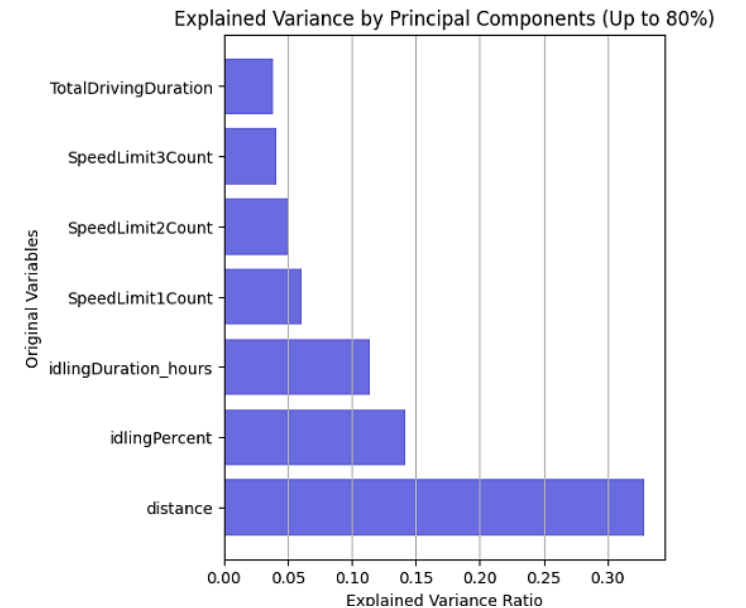


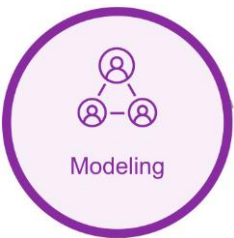
## Solución propuesta: Preparación de los datos

### PCA



### PCA Tabla unida





## Solución propuesta: Desarrollo del Modelo

**Tendencia a la  
baja**

Inicios de Julio

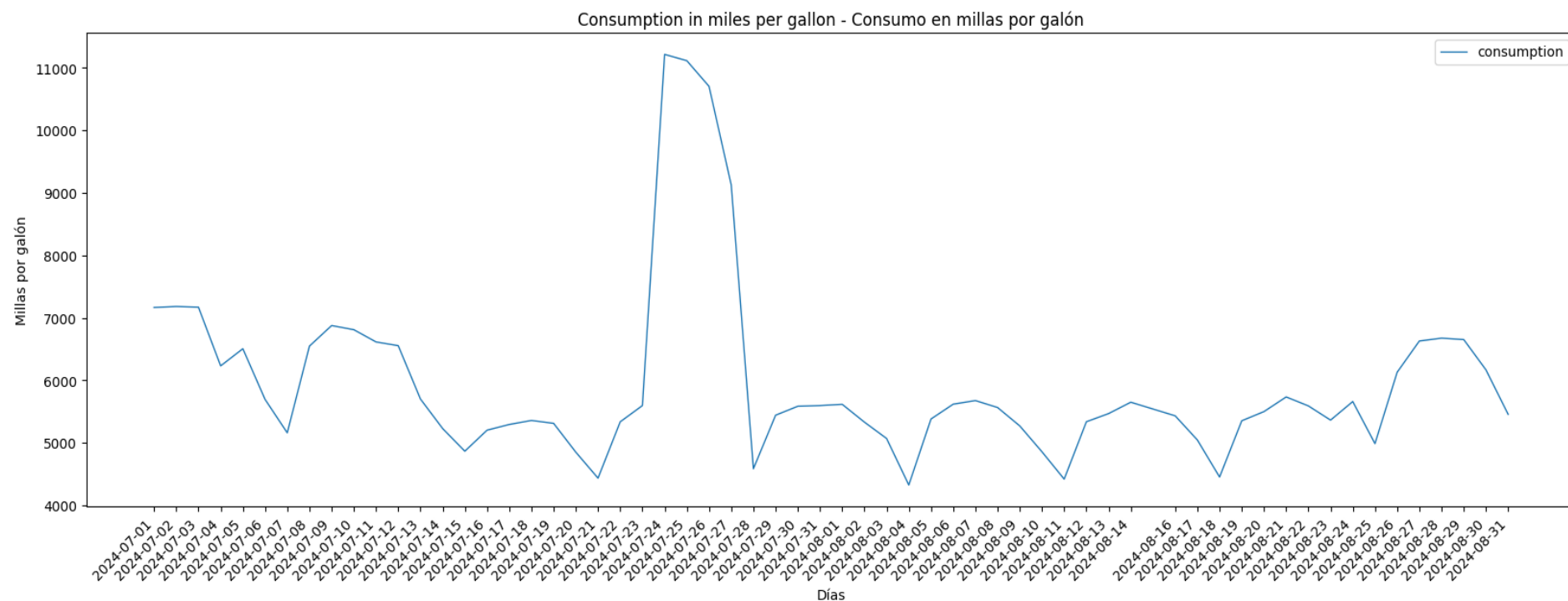
**Pico de consumo**

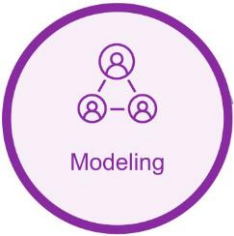
24 y 25 de Julio

**Consumo  
estable**

Agosto

### Histórico





Solución propuesta: Desarrollo del Modelo

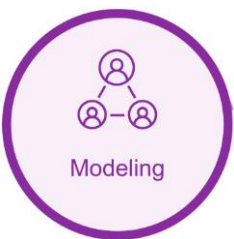
Tabla consumos

|            | consumption | day | dayofweek | month | quarter | year | dayofyear |
|------------|-------------|-----|-----------|-------|---------|------|-----------|
| date       |             |     |           |       |         |      |           |
| 2024-07-01 | 7166.415676 | 1   | 0         | 7     | 3       | 2024 | 183       |
| 2024-07-02 | 7182.232302 | 2   | 1         | 7     | 3       | 2024 | 184       |
| 2024-07-03 | 7171.074568 | 3   | 2         | 7     | 3       | 2024 | 185       |
| 2024-07-04 | 6231.688432 | 4   | 3         | 7     | 3       | 2024 | 186       |
| 2024-07-05 | 6506.612911 | 5   | 4         | 7     | 3       | 2024 | 187       |
| ...        | ...         | ... | ...       | ...   | ...     | ...  | ...       |
| 2024-08-27 | 6630.083851 | 27  | 1         | 8     | 3       | 2024 | 240       |
| 2024-08-28 | 6675.791509 | 28  | 2         | 8     | 3       | 2024 | 241       |
| 2024-08-29 | 6653.025786 | 29  | 3         | 8     | 3       | 2024 | 242       |
| 2024-08-30 | 6168.360344 | 30  | 4         | 8     | 3       | 2024 | 243       |
| 2024-08-31 | 5456.706931 | 31  | 5         | 8     | 3       | 2024 | 244       |

Hiperparámetros

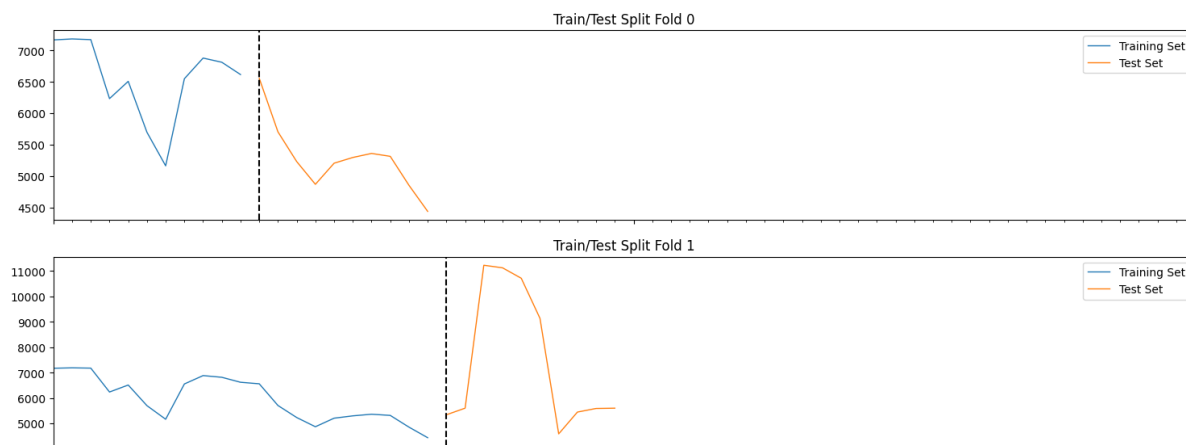
| Característica         | Trial 32                                     | Trial 97                                     |
|------------------------|--|--|
| learning_rate          | 0.1841                                       | 0.1846                                       |
| max_depth              | 8  | 10   |
| subsample              | 0.9188                                       | 0.9040                                       |
| colsample_bytree       | 0.8703                                       | 0.8813                                       |
| reg_alpha              | 0.000231                                     | 0.000132                                     |
| reg_lambda             | 1.6468                                       | 1.7343                                       |
| min_child_weight       | 3  | 3  |
| gamma                  | 0.2373                                       | 0.2048                                       |
| n_estimators           | 343  | 355  |
| Puntajes por partición | -788.75, -1034.10, -1861.16, -198.95, -68.17 | -789.24, -1036.60, -1898.17, -215.01, -77.43 |
| Puntaje promedio       | -790.22                                      | -803.29                                      |
| MAE después de Optuna  | 1390.40                                      | 1406.49                                      |



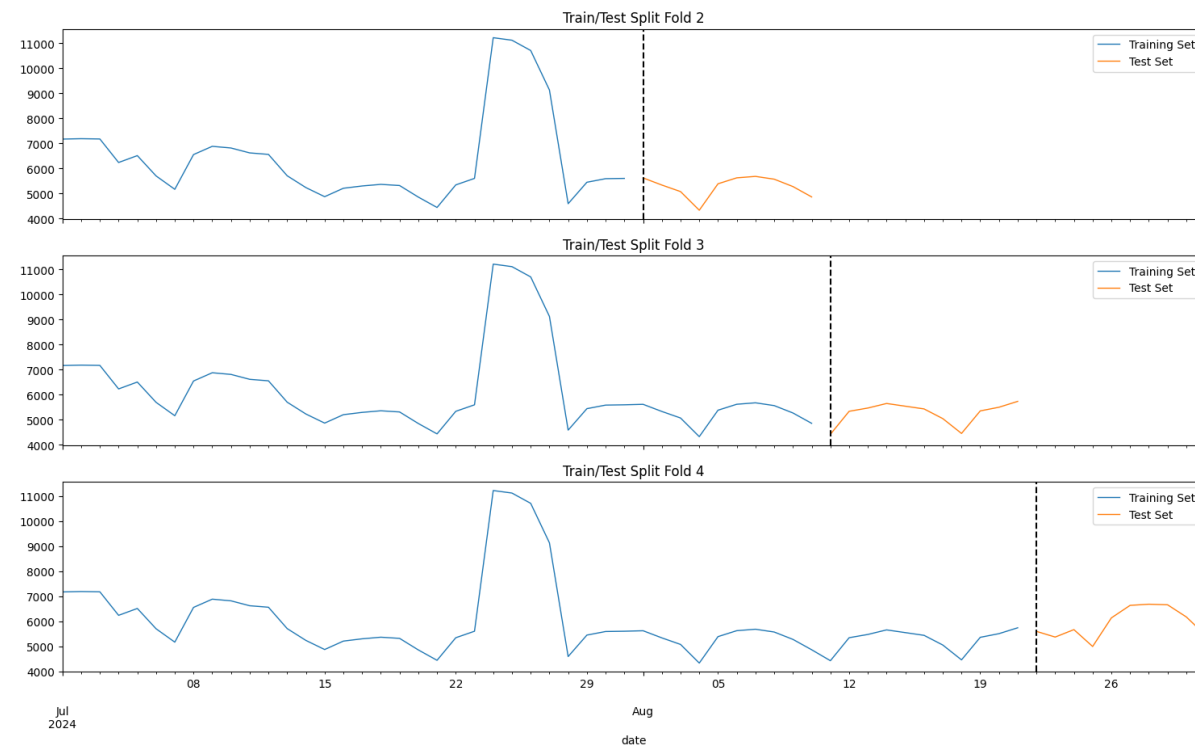


## Solución propuesta: Desarrollo del Modelo

### Validación cruzada (parte 1)



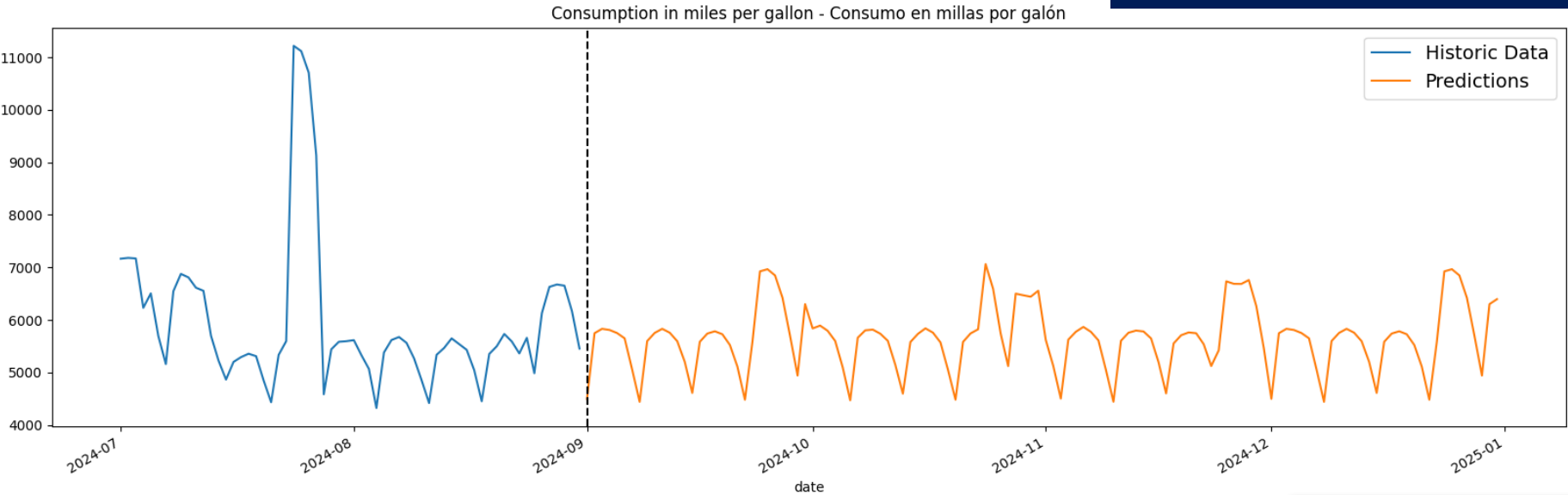
### Validación cruzada (parte 2)





# Solución propuesta: Evaluación

## Pronóstico



| Característica         | Trial 32                                     |
|------------------------|--|
| learning_rate          | 0.1841                                       |
| max_depth              | 8  |
| subsample              | 0.9188                                       |
| colsample_bytree       | 0.8703                                       |
| reg_alpha              | 0.000231                                     |
| reg_lambda             | 1.6468                                       |
| min_child_weight       | 3  |
| gamma                  | 0.2373                                       |
| n_estimators           | 343  |
| Puntajes por partición | -788.75, -1034.10, -1861.16, -198.95, -68.17 |
| Puntaje promedio       | -790.22                                      |
| MAE después de Optuna  | 1390.40                                      |

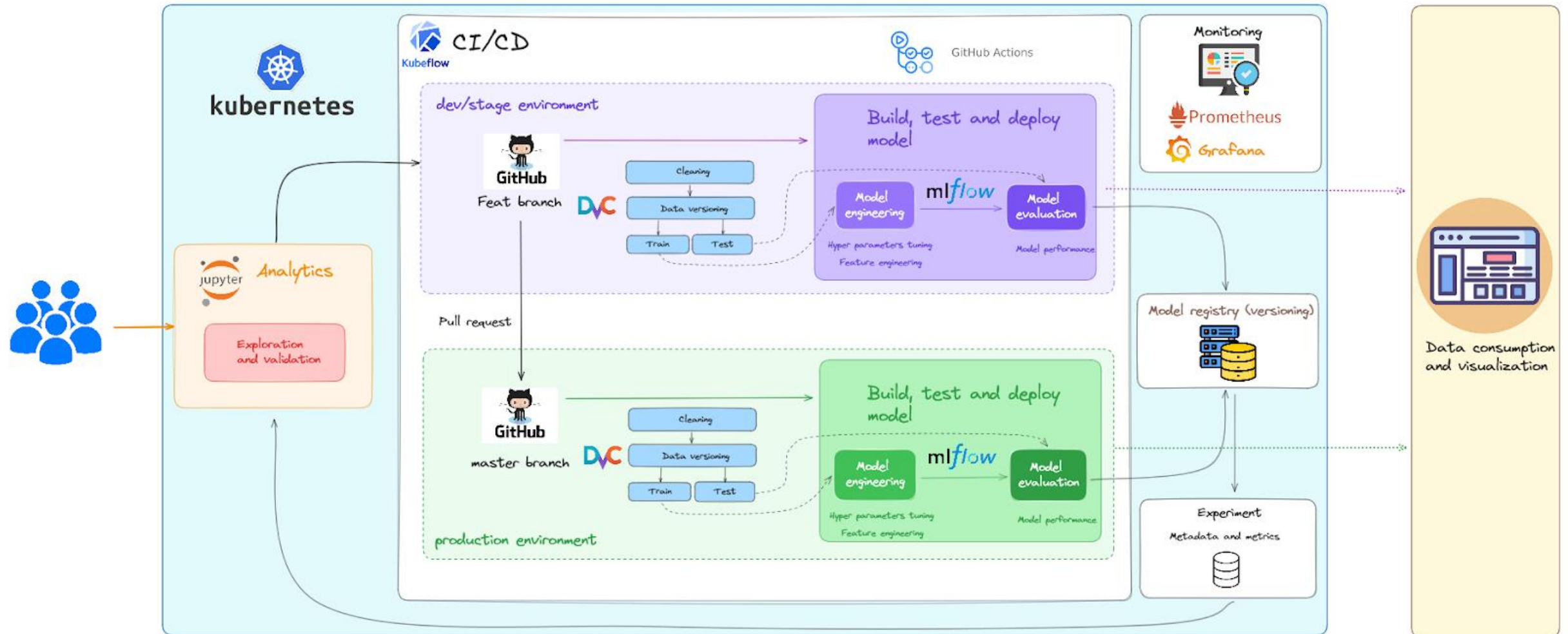
```
[0] validation_0-rmse:1236.08345
[100] validation_0-rmse:18.21589
[200] validation_0-rmse:3.28521
[300] validation_0-rmse:0.81316
[342] validation_0-rmse:0.71778
MAE después de ajustar el modelo: 0.48
```



Deployment

## Solución propuesta: Despliegue

### Arquitectura MLoops (Open Source)



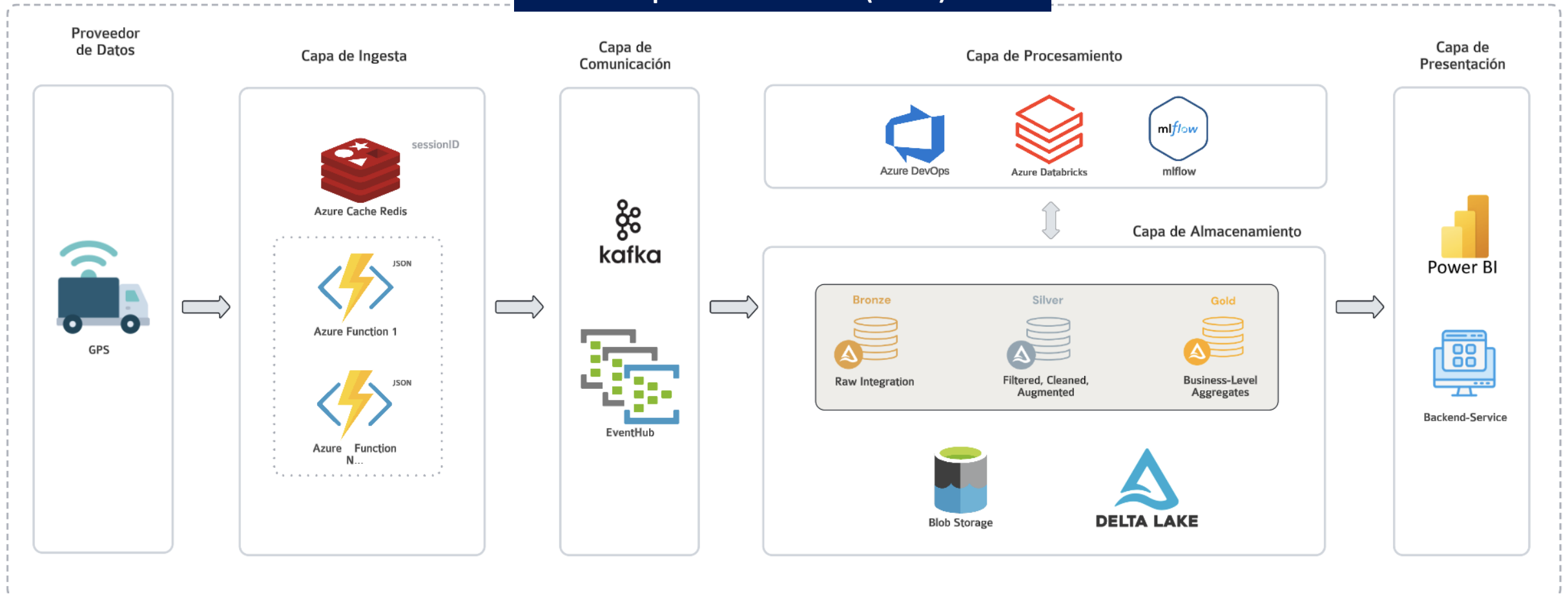


Deployment

## Solución propuesta: Despliegue

### Arquitectura General (Azure)

\$ 5,000 dls

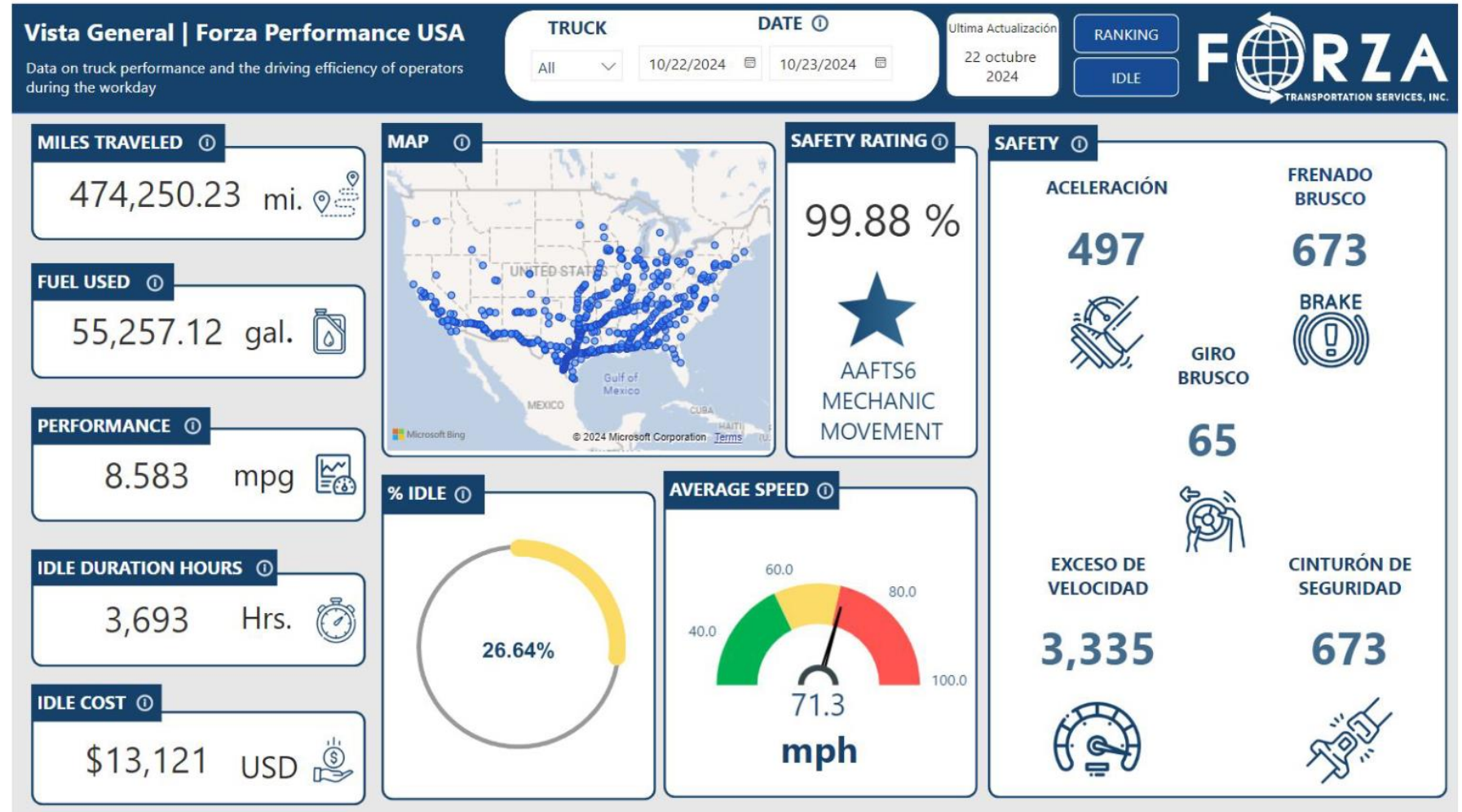


## Recomendaciones Comerciales Fundamentales

### Creación de dashboard

Visualizar la  
operación

Toma de decisiones  
basada en datos



## Beneficios y resultados esperados



Reducción de costos operativos.



Mejora de la rentabilidad del negocio.



Toma de decisiones informadas.



Competitividad en las operaciones



## Trabajos Futuros

- **Integración de datos externos:** Expandir el modelo de pronóstico incorporando datos adicionales, como precios dinámicos de combustible y condiciones climáticas, para mejorar la precisión de las predicciones.
- **Predicción a largo plazo con aprendizaje profundo:** Explorar el uso de algoritmos de aprendizaje profundo para mejorar la capacidad de hacer predicciones precisas a largo plazo.
- **Mantenimiento preventivo:** Considerar factores de mantenimiento preventivo basados en el consumo y uso de los vehículos para optimizar la eficiencia operativa.



# Hallazgos e ideas clave

## Ingeniería de Datos :

- El análisis mostró que las variables tienen una correlación débil con el consumo de combustible.
- El Análisis de Componentes Principales (PCA) identificó dos factores clave para la reducción del consumo de combustible:
  - Porcentaje de ralentí en el viaje.
  - Frecuencia de excesos de velocidad.

## Modelos de clasificación:

- Árbol de decisión: Precisión promedio de 0.82 y puntaje F1 de 0.78.
- XGBoost: Precisión de 0.85 y puntaje F1 de 0.81.
- Ambos modelos indicaron que la mayoría de las variables, excepto las no accionables (distancia recorrida y consumo total de combustible), carecían de alta significancia. Esto llevó a optar por un modelo de pronóstico como complemento para respaldar decisiones estratégicas y operativas.

## Modelo de Pronóstico (Forecasting):

- Se identificaron patrones estacionales en el consumo de combustible, lo que facilita la planificación de abastecimiento.
- El modelo mostró un error absoluto medio (MAE) de 0.48, lo que indica un buen ajuste y precisión en las proyecciones futuras del consumo.



## Bibliografía

- Mssaperla. (2024, 1 marzo). *What is the medallion lakehouse architecture? - Azure Databricks*. Microsoft Learn. <https://learn.microsoft.com/en-us/azure/databricks/lakehouse/medallion>
- Kumar Mukhiya, S., y Ahmed, U. (2020). *Hands-On Exploratory Data Analysis with Python*. Packt Publishing. <https://learning.oreilly.com/library/view/hands-on-exploratory-data/9781789537253/0957090f-fa4d-4145-95dd-6d3782e5c04d.xhtml>
- Visengeriyeva, L., Kammer, A., Bär, I., Kniesz, A., y Plöd, M. (2023). *CRISP-ML(Q). The ML Lifecycle Process*. MLOps. INNOQ. <https://ml-ops.org/content/crisp-ml>
- Geotab. (2022). *7 strategies to reduce fleet fuel costs: Quick guide for managing fuel economy and idling*. Recuperado de [https://www.geotab.com/CMS-GeneralFiles-production/NA/ebooks/7%20Strategies%20to%20Reduce%20Fleet%20Fuel%20Costs\\_Geotab%20%5BPublic%5D.pdf](https://www.geotab.com/CMS-GeneralFiles-production/NA/ebooks/7%20Strategies%20to%20Reduce%20Fleet%20Fuel%20Costs_Geotab%20%5BPublic%5D.pdf)