

FACULTAD  
DE CIENCIAS  
ECONÓMICAS



FAMAF  
Facultad de Matemática,  
Astronomía, Física y  
Computación



Universidad  
Nacional  
de Córdoba

# Diplomatura Universitaria en Ciencia de Datos, Inteligencia Artificial y sus aplicaciones en Economía y Negocios

## Estimación de días de entrega de una plataforma e-commerce

Autoría: **GRUPO 8**

- Alejandro Mezio
- Diego Nicolás Zallio
- Ignacio Cordoba
- Sebastián Lezama
- Federico Churquina



Tutor: Ignacio Frachetti  
2025

# ¿Qué es Olist?

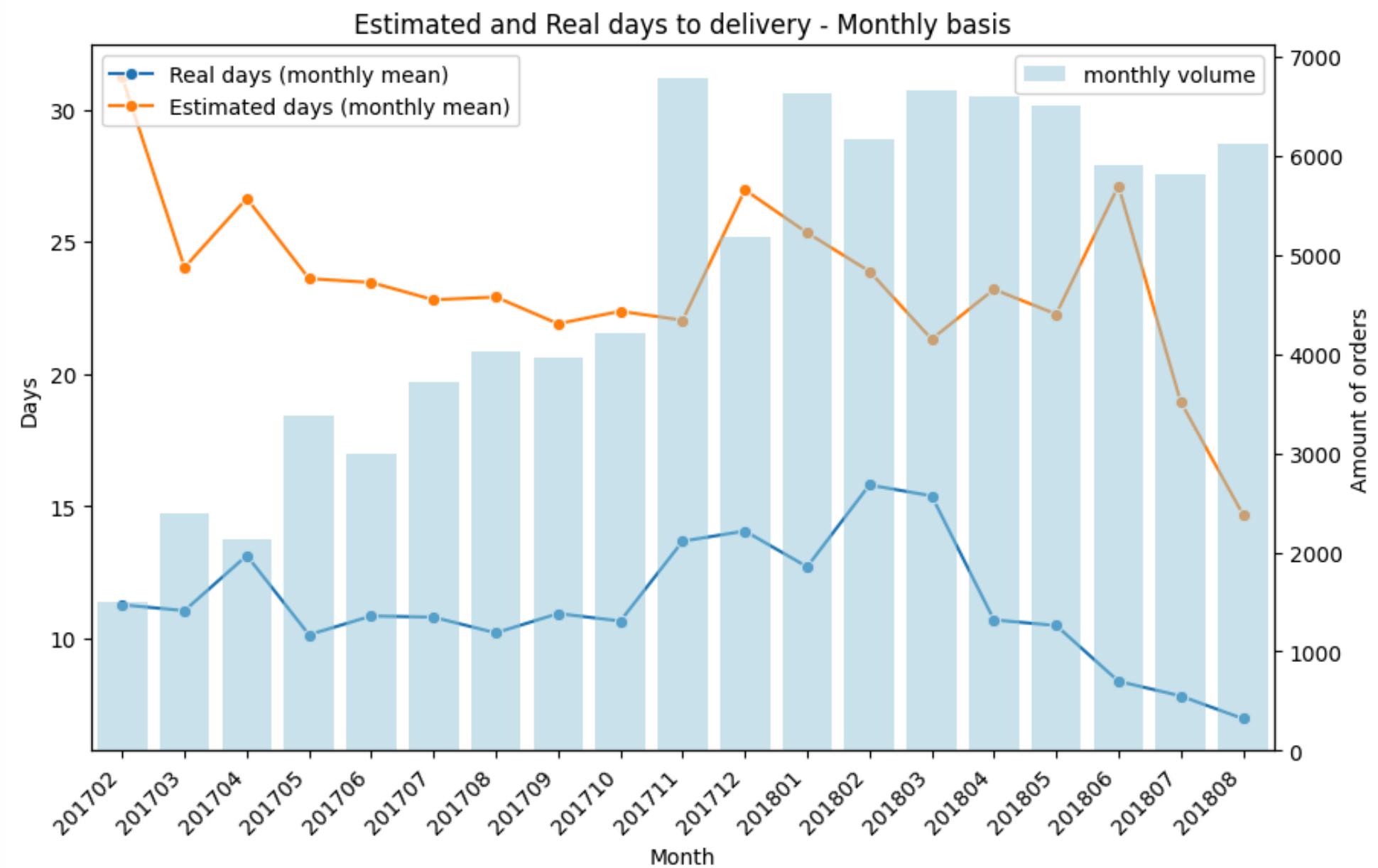
**Olist**, fundada en 2015 en Curitiba (estado de Paraná), Brasil, comenzó como una plataforma de marketplace enfocada en facilitar que comerciantes menos grandes pudieran acceder a los grandes marketplaces brasileños.

En los años más recientes (2023-2024) ha ampliado sus servicios hacia logística y capital (financiamiento para comerciantes), ofreciendo un ecosistema completo para los comercios permitiéndoles vender en múltiples canales, gestionar logística, finanzas y operaciones de venta de forma integrada.



# Oportunidad de mejora

Prestando atención a las variables temporales relacionadas con el proceso de entrega, detectamos una oportunidad de mejora: en muchos casos, los plazos de entrega estimados por la plataforma resultaban significativamente mayores que los tiempos de entrega reales.



# Objetivo

**Desarrollar un modelo de machine learning capaz  
de predecir con mayor precisión el tiempo  
estimado de entrega de un pedido**

# Análisis Exploratorio y Limpieza de Dataset

01

- Depuración de variables;
- Eliminación de duplicados
- Detección valores faltantes

02

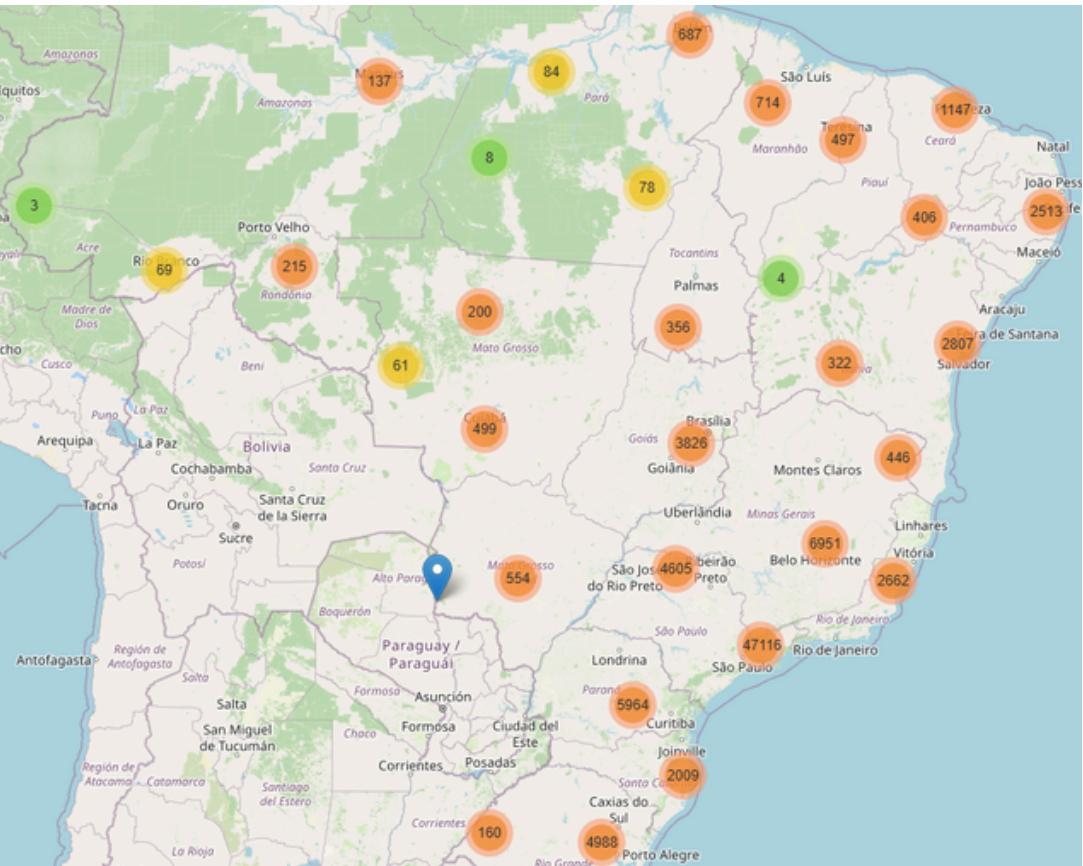
- Creación de variables:
- Peso volumetrico
  - Distacia vendedor-comprador
  - Days\_to\_delivered

03

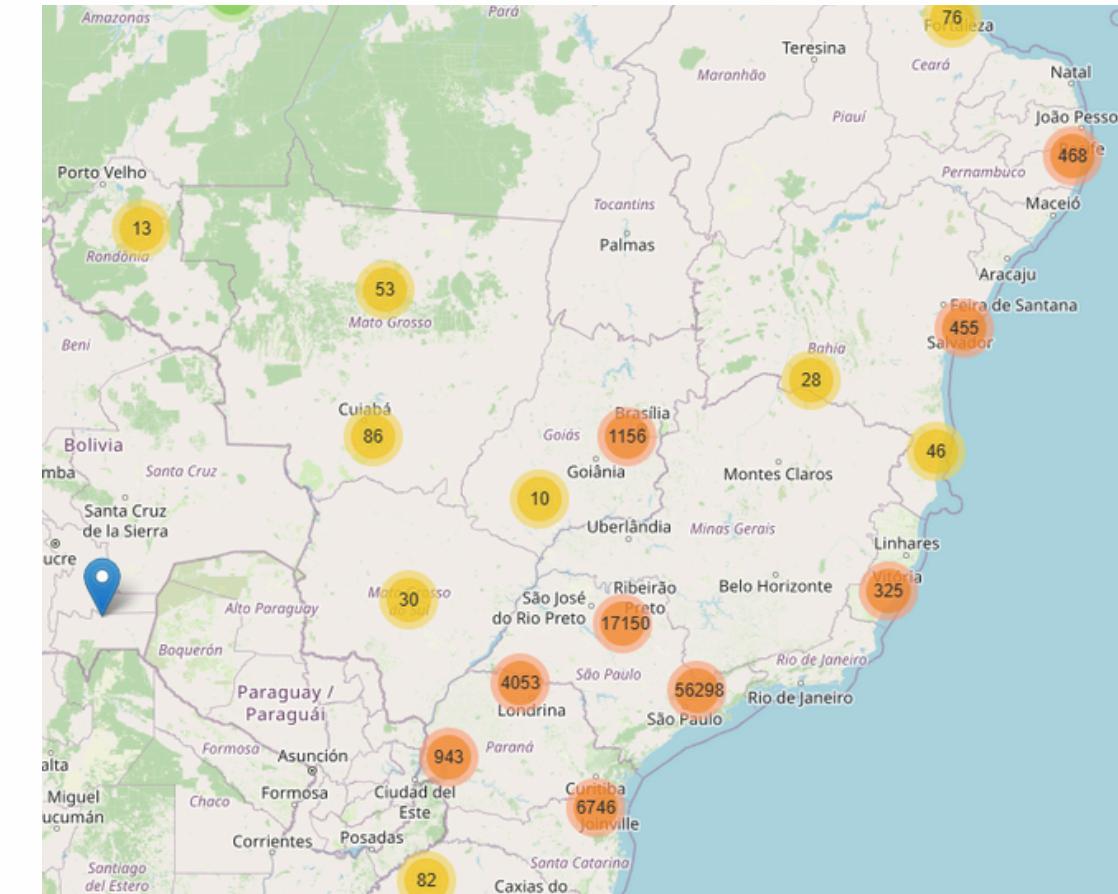
- Creación de variable categoricas:
- "rutas"
  - "product\_venta"
  - "sales\_same\_state"

# Dataset y Análisis Gráfico

## Compradores



## Vendedores



**El 70% de las órdenes salen desde vendedores que se encuentran en el estado de Sao Paulo y el 42% de las órdenes llegan a compradores en Sao Paulo.**

# Definición de variables

## Variable objetivo

**"days\_to\_delivered"**

Número de días transcurridos entre la fecha de compra (order\_purchase\_date) y la fecha de entrega al cliente (order\_delivered\_customer\_date)

## Features

- distance\_km
- price
- freight\_value
- product\_chosen\_weight
- sales\_same\_state (binaria)
- product\_venta
- rutas

# Entrenamiento y Test

Realizamos una división aleatoria train/test (80% y 20%) con random\_state fijo para reproducibilidad.

Para evaluación estable utilizamos validación cruzada (subrutina GridSearchCV) para la optimización de hiperparámetros.



# Modelos utilizados

0- Regresión Lineal Simple

1- Lasso Regression

2- Ridge Regression

3- Regresión Polinómica con Ridge

4- Árbol de Decisión

5- Random Forest

# Elección de métrica



## MAE

Corresponde a la **mediana** de la distribución

Tiene las **mismas unidades que la variable objetivo**, por lo que es **más interpretable**

Trata todos los errores de manera **lineal**

## MSE

Corresponde a la **media** de esa distribución

Tiene **unidades al cuadrado**, lo que lo hace menos interpretable

Penaliza los **errores grandes de forma más severa**

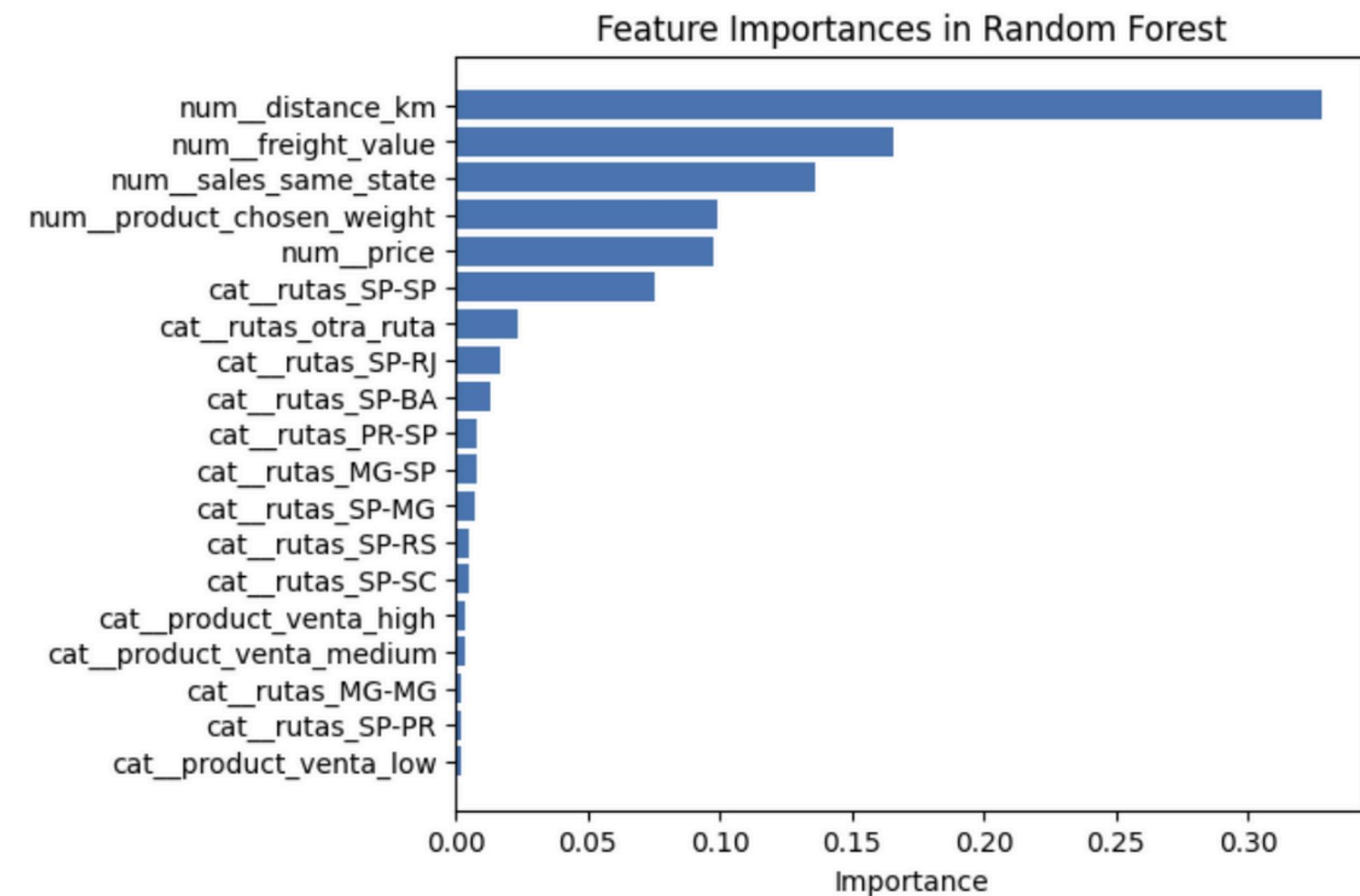
# Elección del mejor modelo

Tras comparar el valor de la métrica de evaluación (MAE), el **Random Forest** mostró el mejor equilibrio entre precisión y generalización.

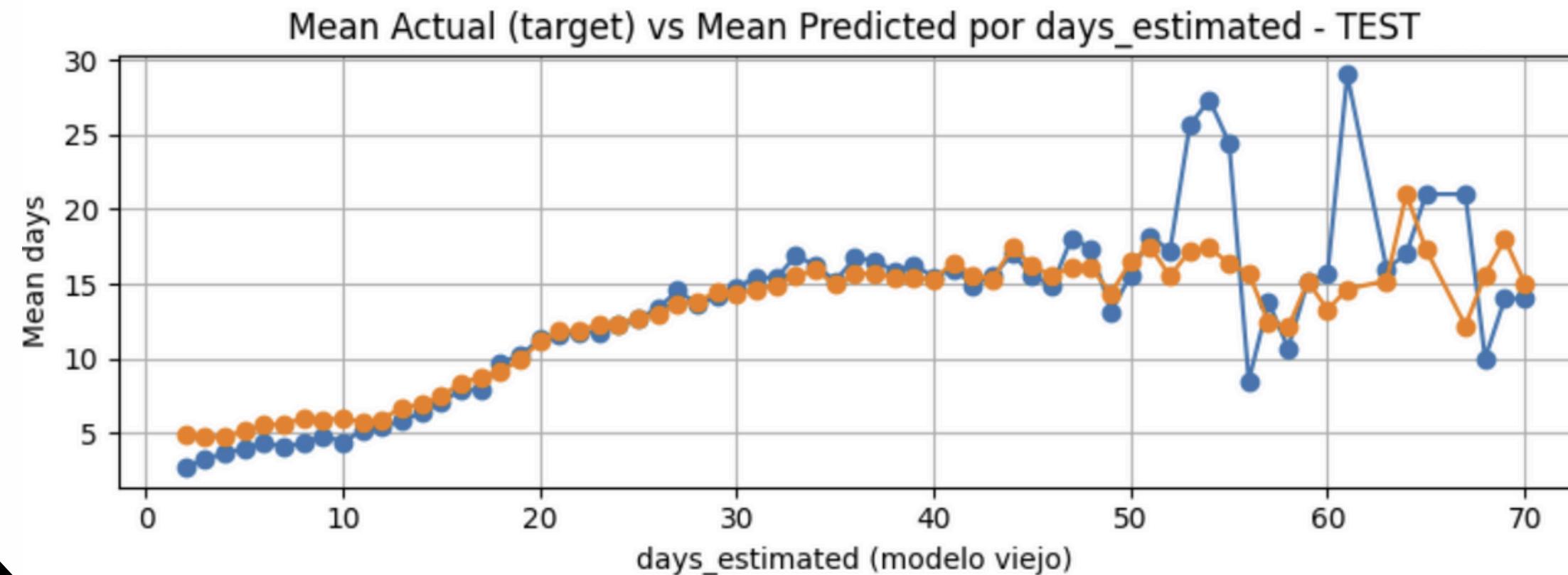
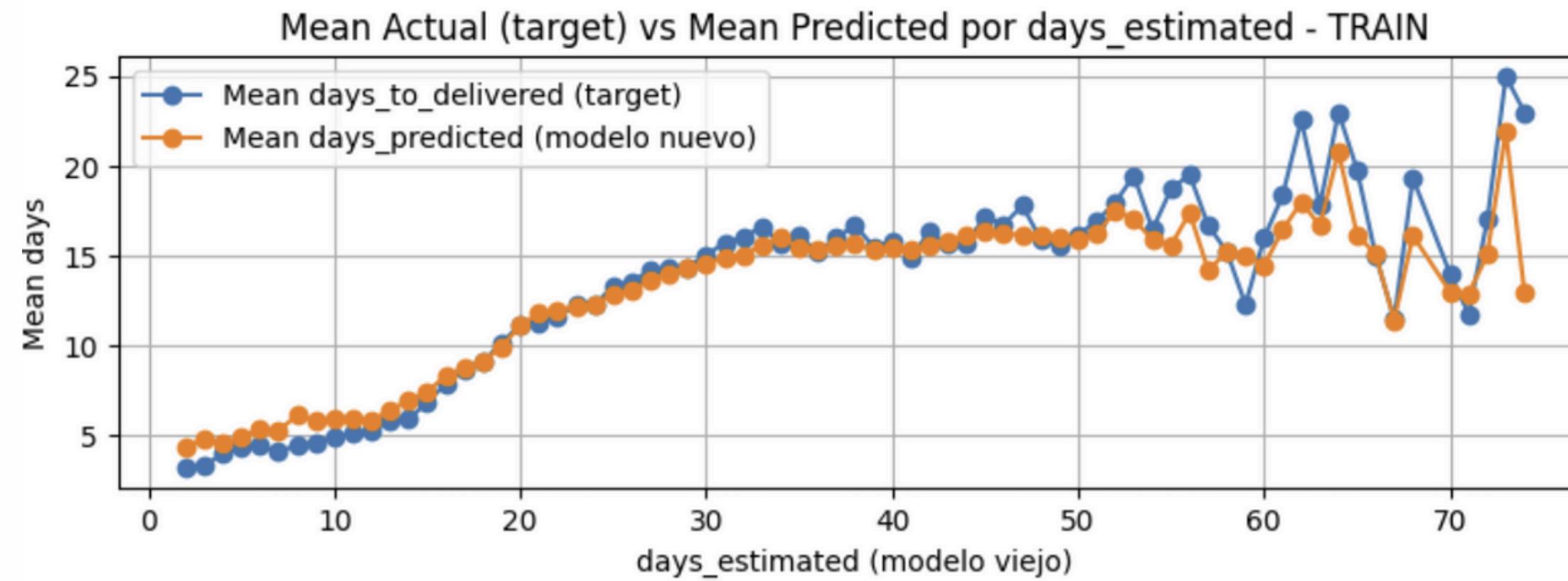
Metrica/Modelo	linreg_base	lasso	ridge	poly	tree	forest
MAE	4,84	4,84	4,84	4,80	4,76	4,67
RMSE	7,18	7,18	7,18	7,15	7,12	7,03

# Feature Importances

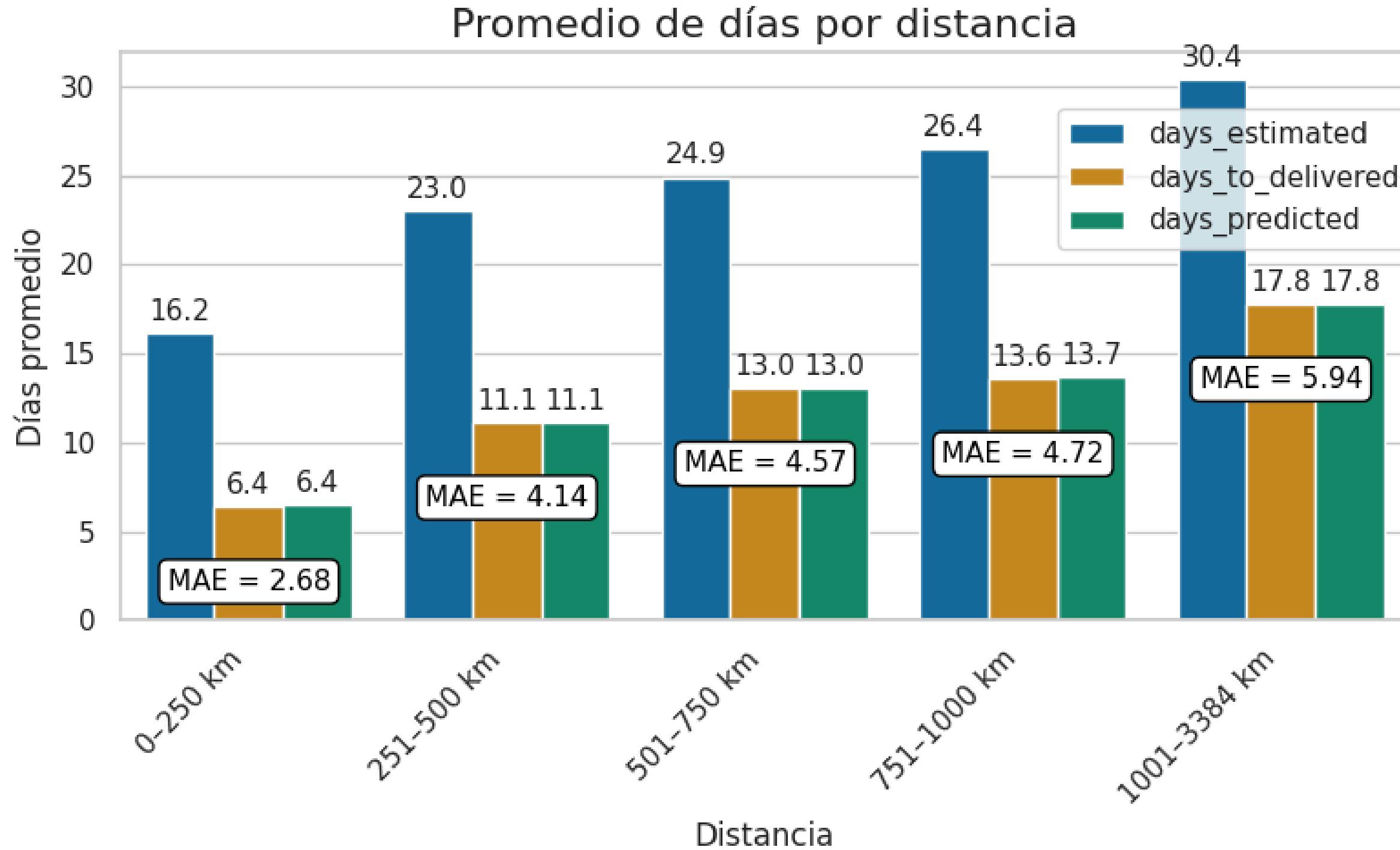
	Feature	Importance
3	num_distance_km	0.328252
1	num_freight_value	0.165548
4	num_sales_same_state	0.135751
2	num_product_chosen_weight	0.098646
0	num_price	0.097758
17	cat_rutas_SP-SP	0.075180
18	cat_rutas_otra_ruta	0.023405
14	cat_rutas_SP-RJ	0.016465
11	cat_rutas_SP-BA	0.012975
10	cat_rutas_PR-SP	0.008047
9	cat_rutas_MG-SP	0.007860
12	cat_rutas_SP-MG	0.007388
15	cat_rutas_SP-RS	0.004879
16	cat_rutas_SP-SC	0.004618
5	cat_product_venta_high	0.003613
7	cat_product_venta_medium	0.003609
8	cat_rutas_MG-MG	0.002188
13	cat_rutas_SP-PR	0.002088
6	cat_product_venta_low	0.001729



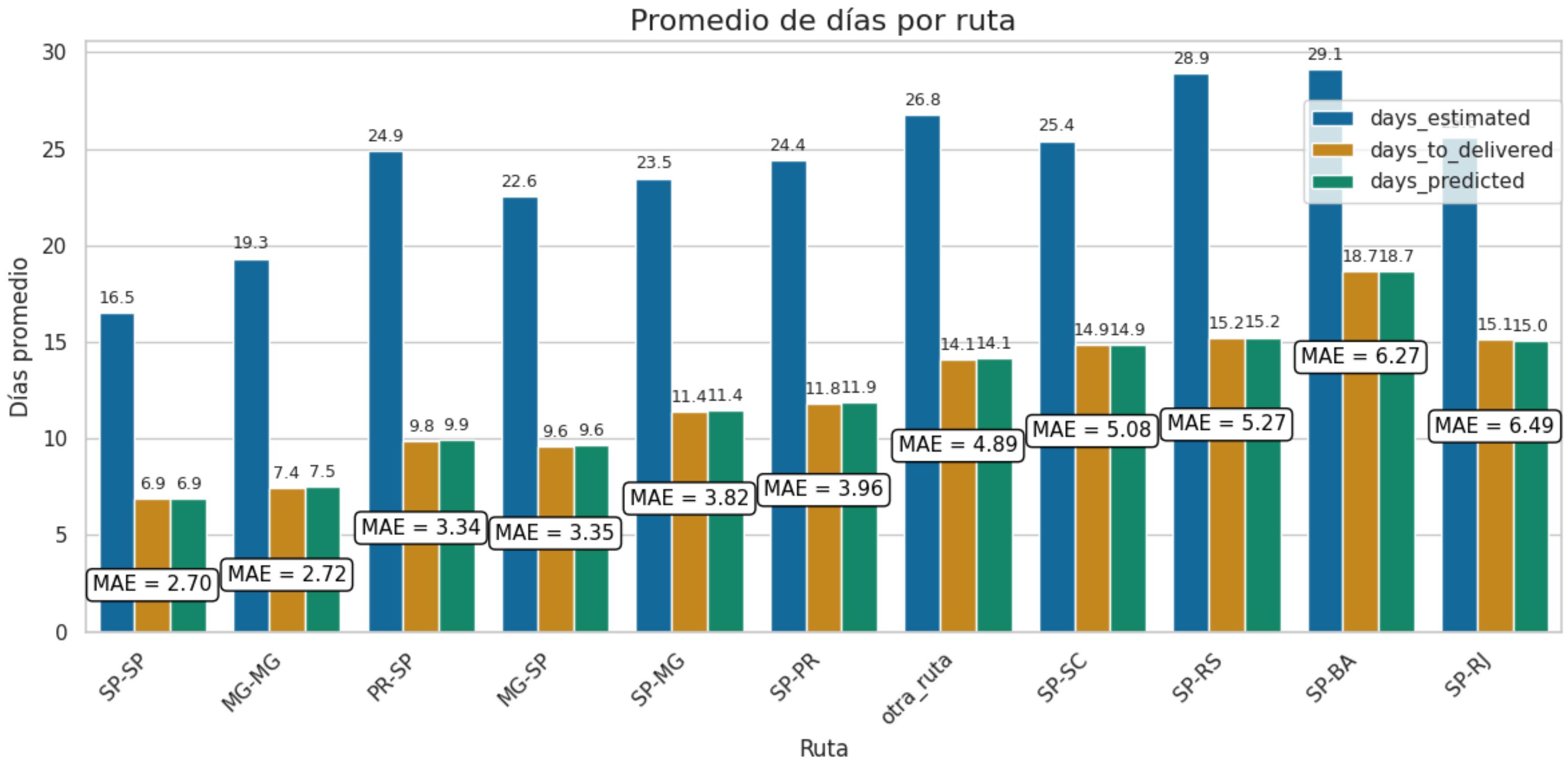
# Resultados Train vs Test



# Performance del modelo



# Performance del modelo



# Conclusión

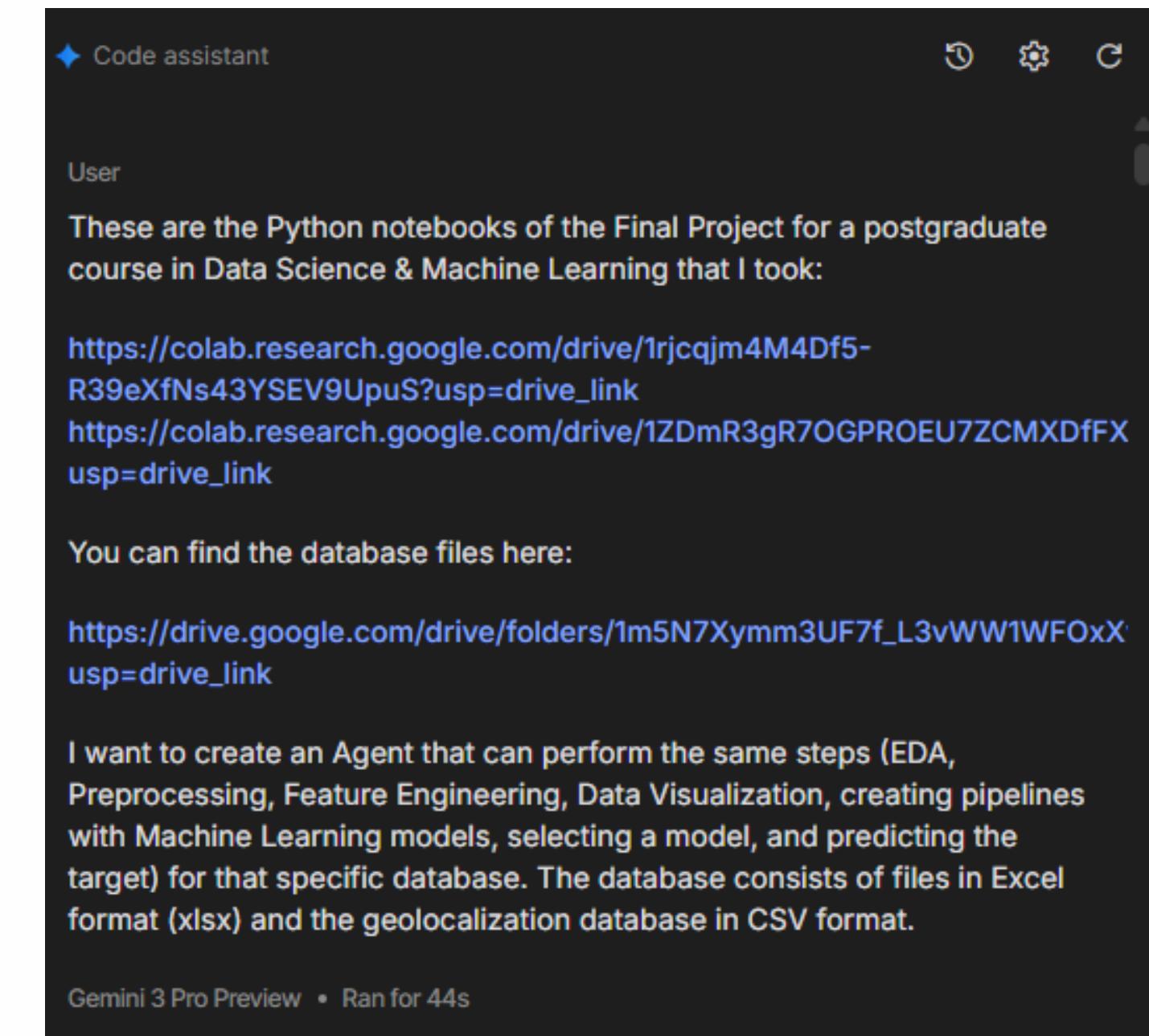
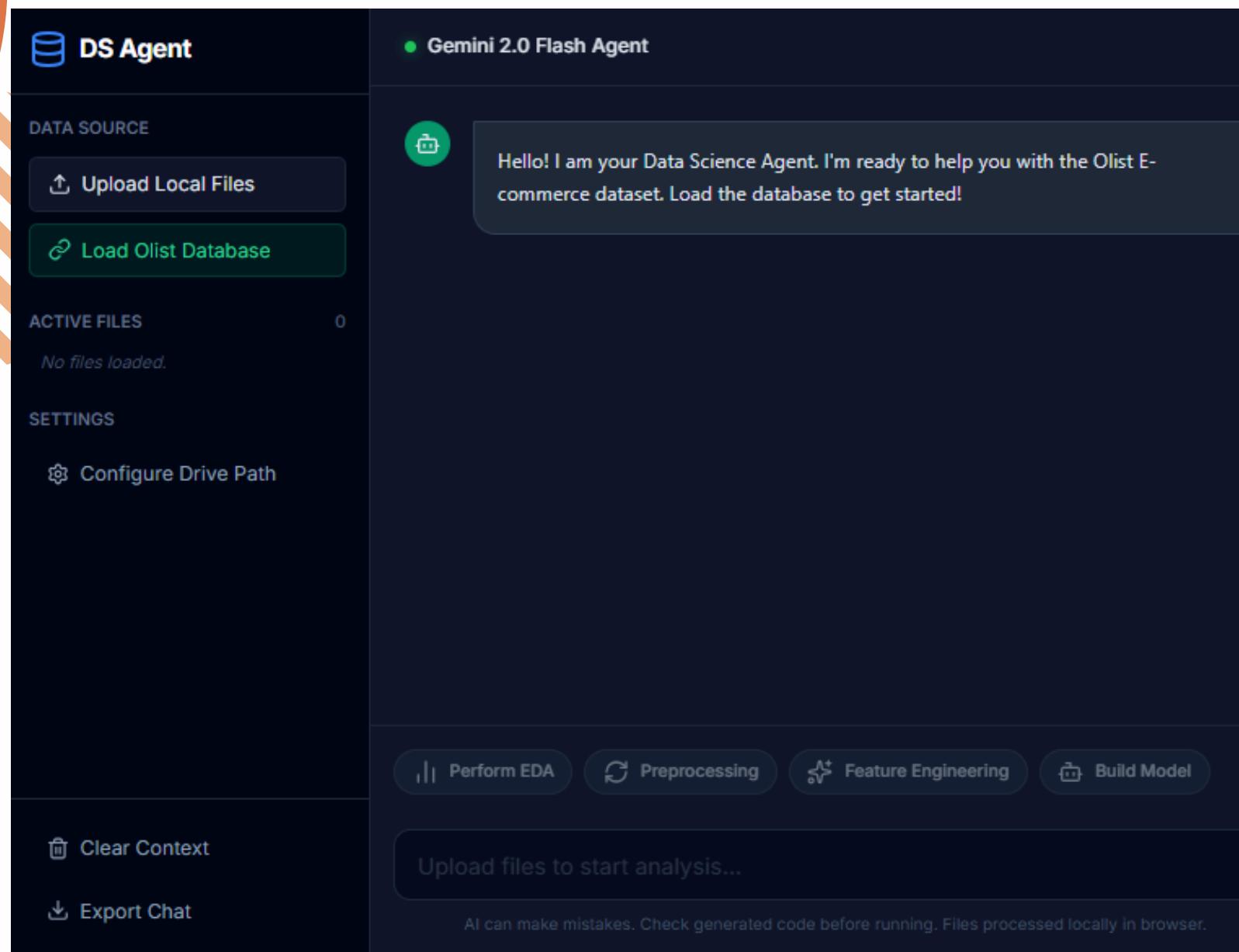
El análisis de la importancia de variables y las curvas de dependencia parcial muestra que la **precisión en la predicción de los días de entrega** se ve explicada principalmente por:

- La **distancia en kilómetros** entre comprador y vendedor.
- El hecho de que la compra y la venta se encuentren **dentro de un mismo estado**.
- El **valor del flete**, que actúa como proxy del peso, volumen y distancia.
- La **ruta comprador-vendedor** (aunque con menor incidencia, ya que su efecto se encuentra parcialmente representado en la variable de distancia y en la relación dentro/diferente estado).

# Mejoras a futuro

- Incorporar los datos de reviews: analizando si la calificación de ventas realizadas por compradores guarda relación con los tiempos de entrega.
- Crear una variable “SCORE” del vendedor: tomando como información los días de demora en despachar el producto.
- Acotar rango temporal de análisis: acotar el rango de tiempo de los datos para entrenamiento al año 2018 donde el volumen de ventas era mayor y estable, y donde la estimación de entrega original era más precisa.
- Readecuación de la selección de las variables explicativas, descartando las de menor incidencia en la predicción.
- Creación por IA de un agente que realice las mismas tareas que un científico de datos.

# AGENTE IA



**GRUPO 8**

**Muchas  
gracias!**