

“La comida sabe mejor cuando llega a tiempo”

Analizando Tiempos de Entrega

GUILLERMINA LAZZARI



¿QUÉ ES ?

- Empresa de Tecnología Finlandesa
- Entrega de Alimentos y Mercancías

Startup --> Presente en
23 países y **180** ciudades



Las empresas alimentarias se vieron muy afectadas por la pandemia

2 de 5

Restaurantes cerraron o redujeron sus operaciones.

Los pedidos de comida a domicilio en línea aumentaron en un mes en

142%

La Pandemia COVID-19
Aceleró la revolución digital y disparó muchas de las transformaciones

DATASET



**18706 Filas
14 columnas**

“Wolt Flow of Orders - Autumn 2020”
Entrega de alimentos realizados a través
de la aplicación

- Tipos de alimentos pedidos y artículos
- Ubicaciones de los clientes y restaurantes
- Horas en que se realizaron y entregaron los pedidos
- Clima
- Velocidad del Mensajero

GLOSARIO

Field Details



VARIABLES

TIMESTAMP - (Creación de la orden)

ACTUAL_DELIVERY_MINUTES - (Minutos reales de entrega)

ESTIMATED_DELIVERY_MINUTES - (Minutos de entrega estimados)

DIF_ACTUAL_ESTIMATED_DEL_MIN (Diferencia entre Minutos REALES de entrega - Minutos de entrega ESTIMADOS)

CLOUD_COVERAGE - (Cobertura de nubes)

TEMPERATURE - (Temperatura)

WIND_SPEED - (Velocidad del viento)

PRECIPITATION - (Precipitación)

ITEM_COUNT - (Recuento de elementos)

USER_LAT - (Latitud del usuario) - (cliente)

USER_LONG - (Longitud del usuario) - (cliente)

VENUE_LAT - (Latitud del lugar/establecimiento)

VENUE_LONG - (Longitud del lugar/establecimiento)

HIPÓTESIS

Para poder concretar a tiempo un pedido de un cliente es importante analizar muchas variables

- Existen días donde la temperaturas pueden ser muy bajas o altas, los repartidores (couriers) pueden experimentar condiciones más desafiantes que podrían afectar la eficiencia de las entregas.
- Analizar los pedidos, ya que la preparación del paquete tambien lleva su tiempo.
- Importante analizar el contexto, entre ellos los dias de la semana y los fin de semanas ante las demandas de pedidos.



OBJETIVO

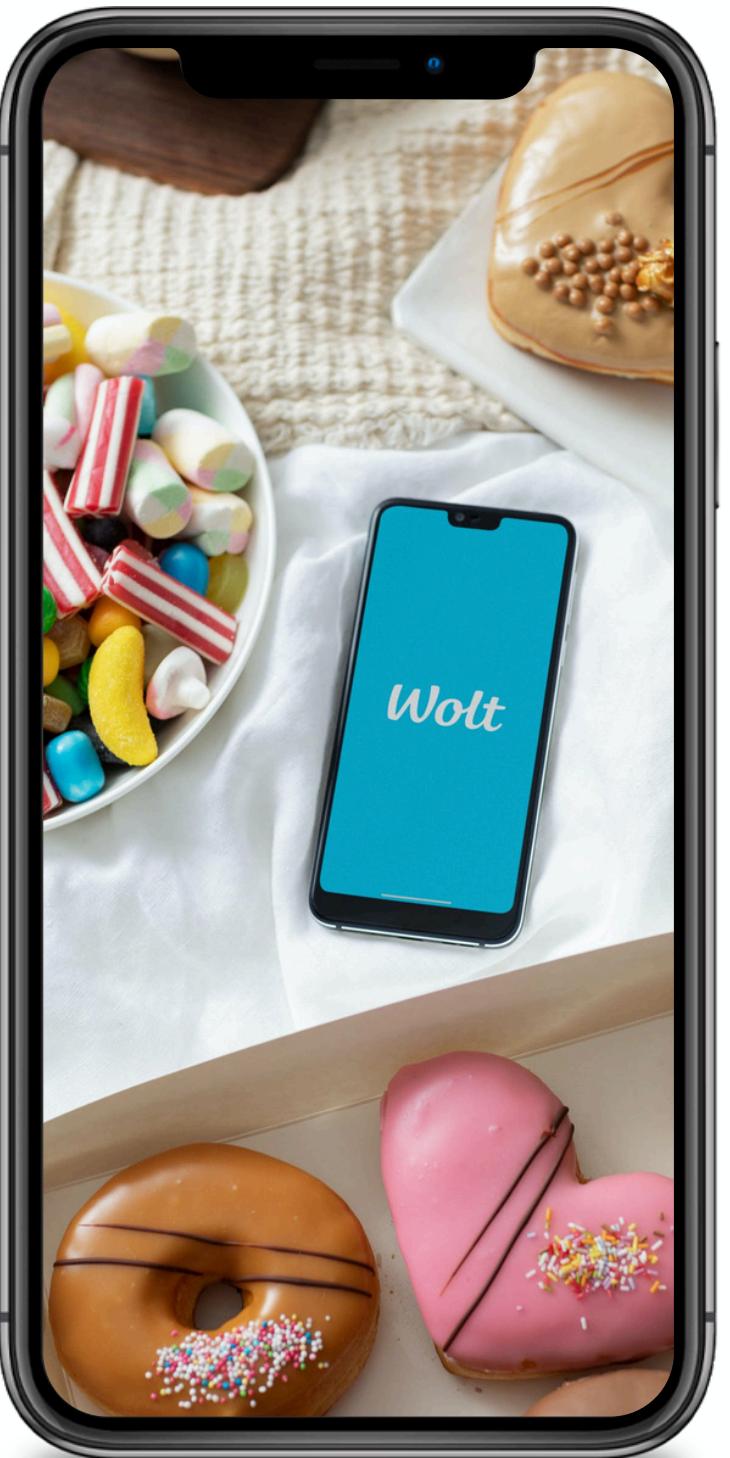
Predecir el tiempo de entrega

Analizar otras variables

- Número de pedidos
- Ordenes, los días más concurrentes o solicitados
- Las condiciones climáticas: ¿Afectan los tiempos de entrega?



PASOS CLAVES



PREPROCESAMIENTO

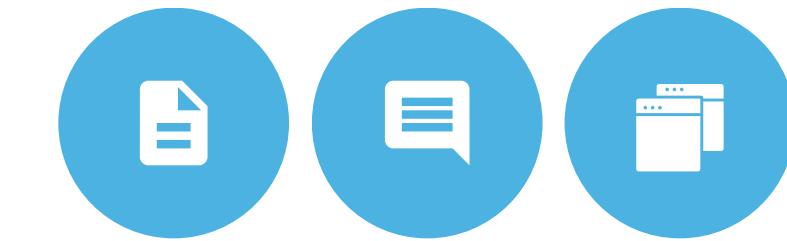
- Planteo de hipótesis previas
- Análisis Univariado/BI/TRI
- Identificación de valores atípicos
- Remoción de valores atípicos
- Imputación de valores faltantes, etc
- Encoding de variables categóricas
- Análisis Bivariado y correlación
- Split Train/Test

MODELADO

- Selección del Modelo
- CrossValidation
- Reducción de la dimensionalidad
- Validación del Modelo Final
- Se evalúa el rendimiento del modelo



VARIABLE OBJETIVO (TARGET)



[ACTUAL_DELIVERY_MINUTES](#)



MINUTOS REALES DE ENTREGA

Representa la cantidad de minutos que realmente tomó entregar el pedido desde el momento en que se realizó.

DISTRIBUCIÓN

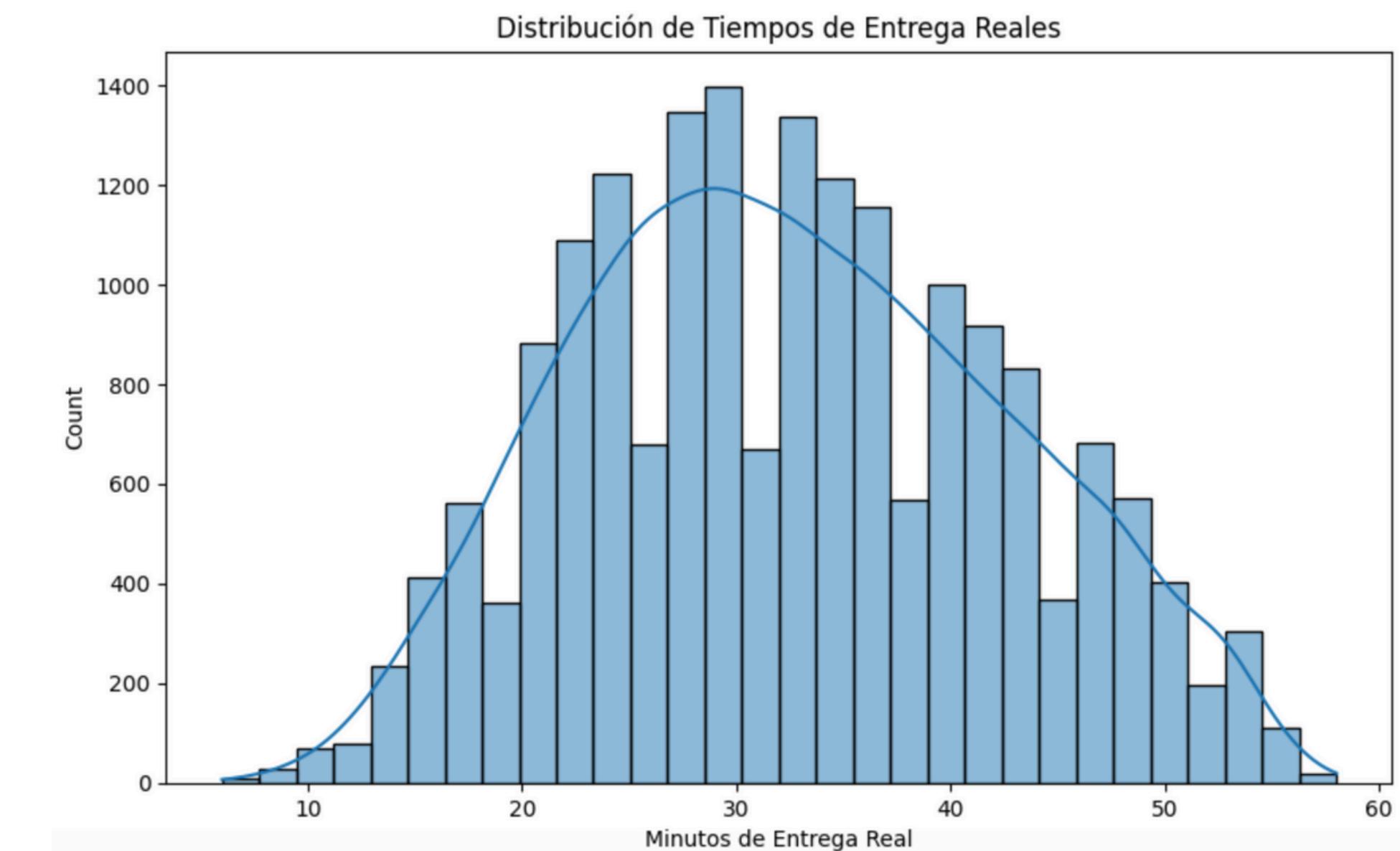
count	18706.00
mean	32.61
std	10.02
min	6.00
25%	25.00
50%	32.00
75%	40.00
max	58.00

ANÁLISIS

Distribución del tiempo de entrega
ACTUAL_DELIVERY_MINUTES

OBSERVACIÓN

Los tiempos de entrega de los pedidos comúnmente estan entre los 22' y los 38' minutos aprox.



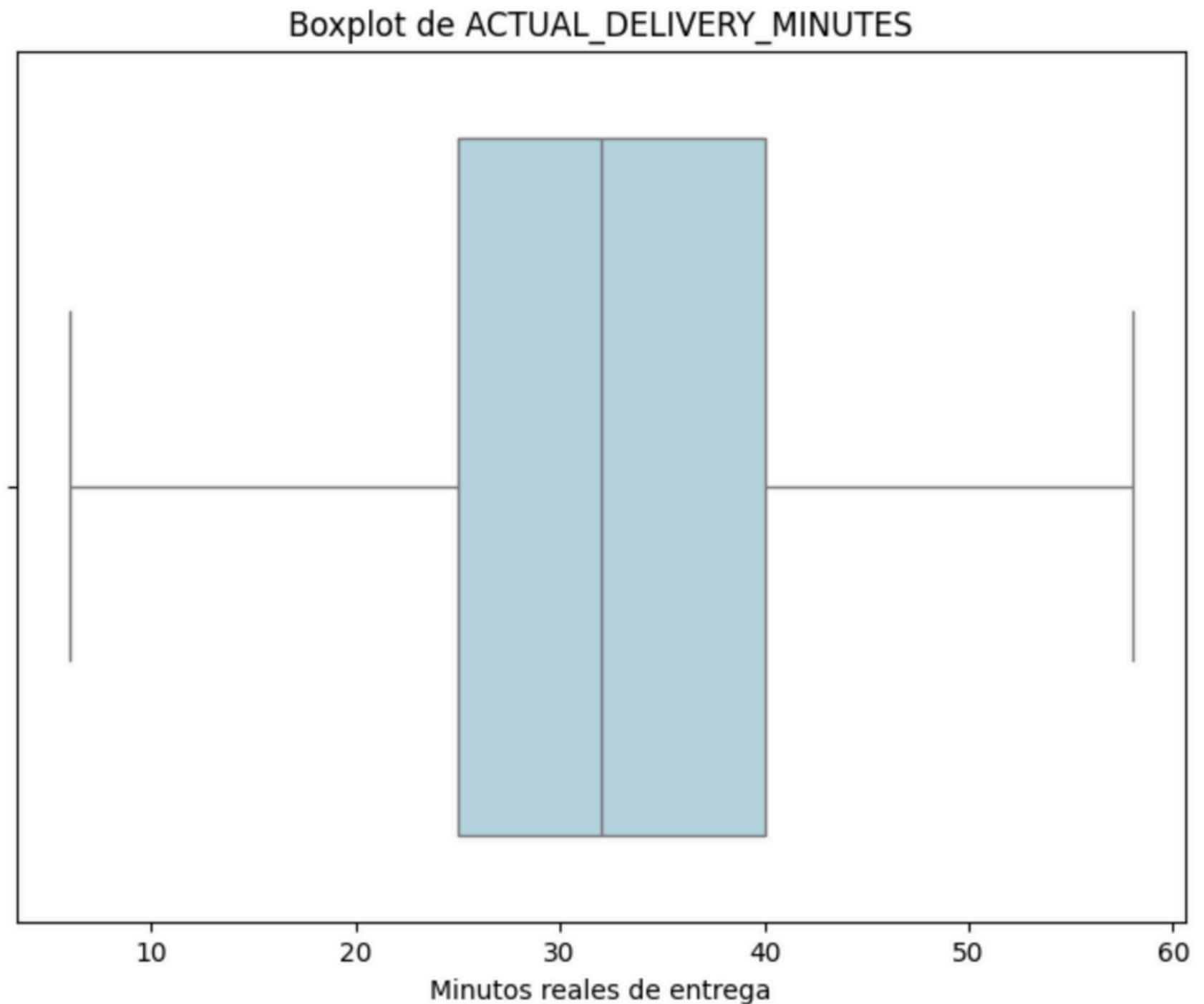
HISTROGRAMA

ANÁLISIS

Distribución del tiempo de entrega
ACTUAL_DELIVERY_MINUTES

OBSERVACIÓN

Aquí podemos observar mejor que en el histograma que los minutos reales de entrega van desde los 25' a los 40' min aprox



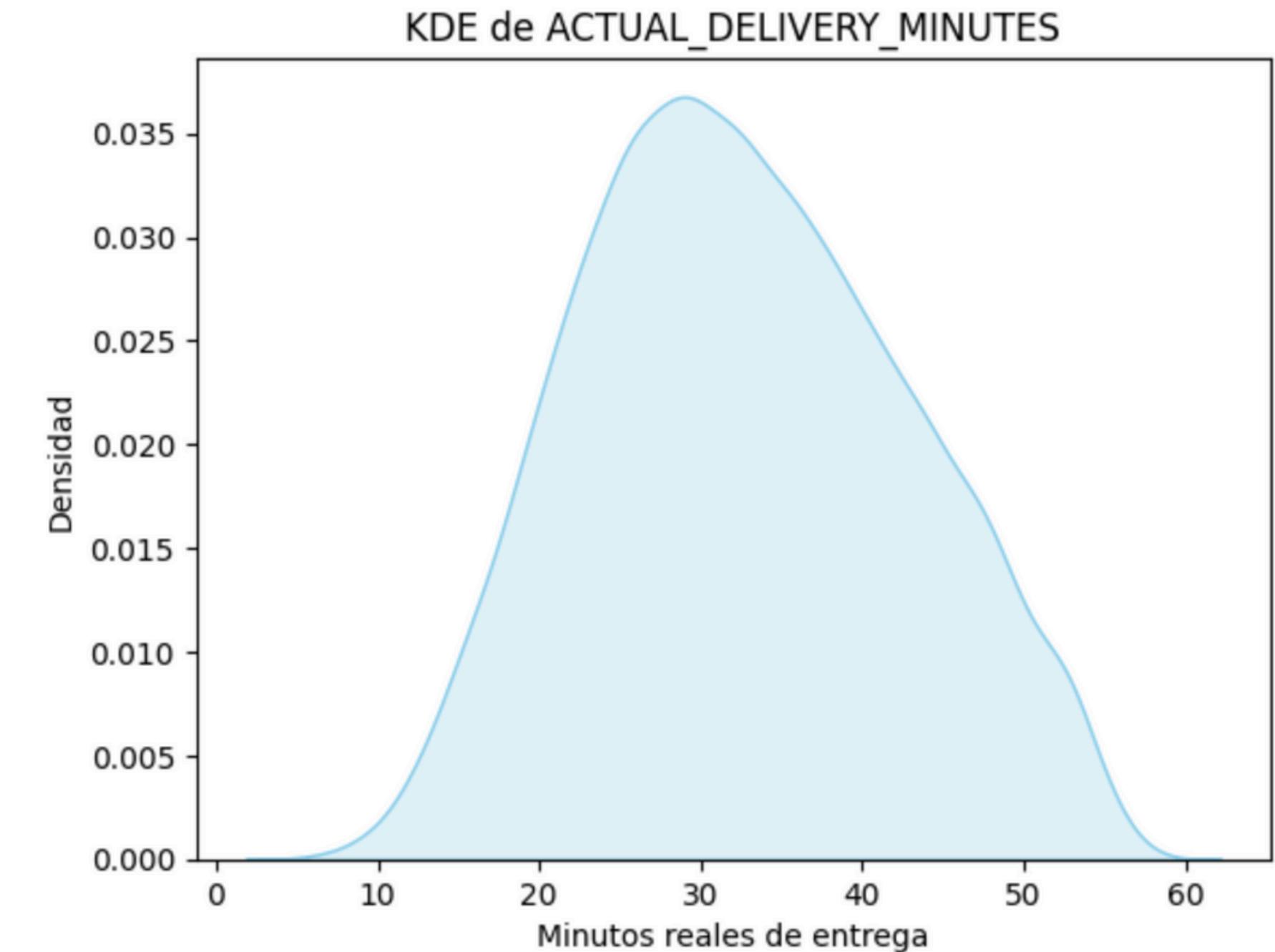
BOXPLOT (Diagrama de Caja)

ANÁLISIS

Distribución del tiempo de entrega
ACTUAL_DELIVERY_MINUTES

OBSERVACIÓN

Este gráfico nos permite una visión más continua de la distribución de datos.
Hasta aquí hemos realizado un análisis univariado



KDE - DENSIDAD del kernel
(Kernel Density Estimate)

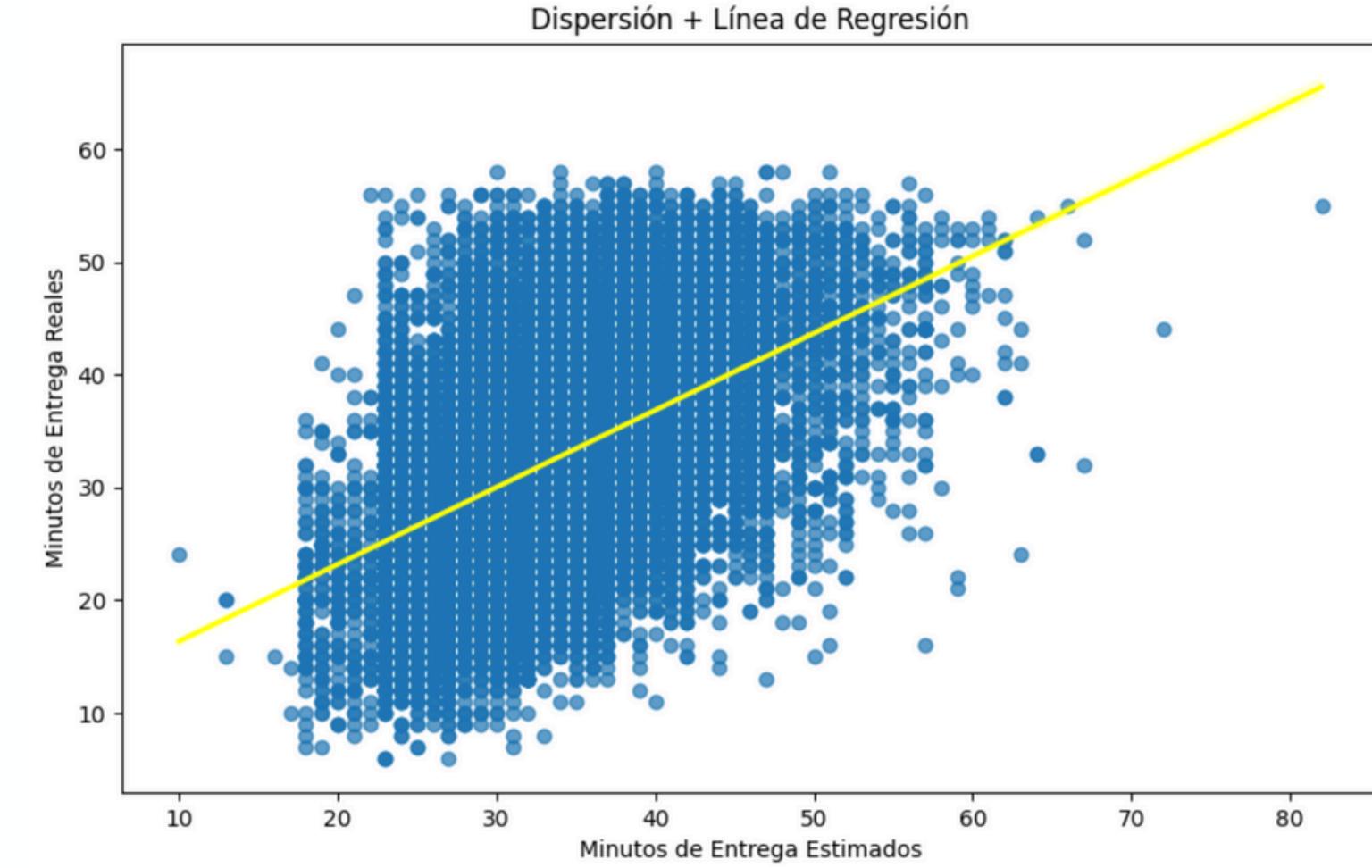
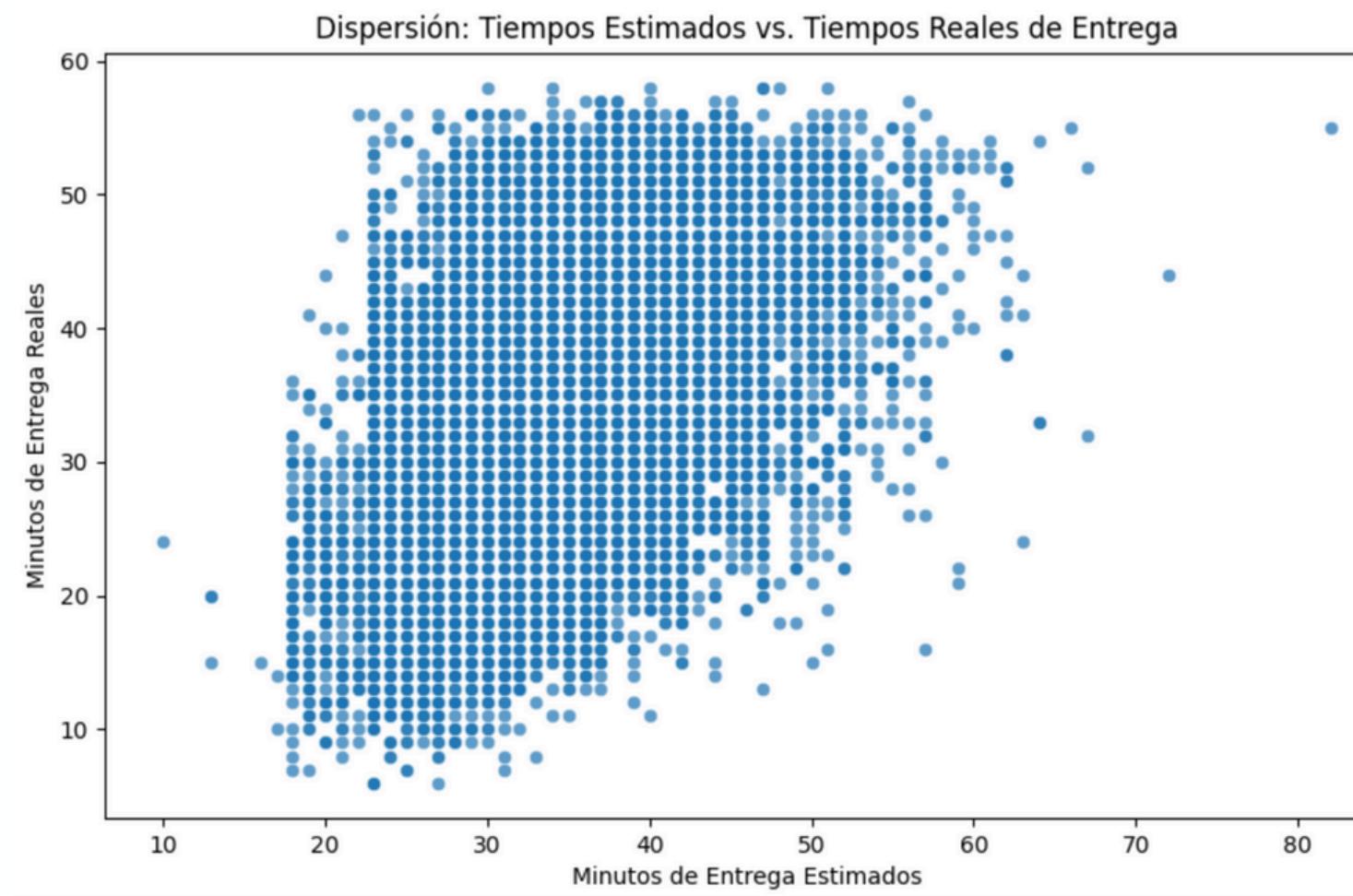
ANÁLISIS BIVARIADO

VARIABLES

ACTUAL_DELIVERY_MINUTES
ESTIMATED_DELIVERY_MINUTES

GRÁFICO DE DISPERSIÓN

Los gráficos de dispersión visualizan la relación entre dos variables numéricas de forma que una variable se muestra en el eje x y la otra, en el eje y.



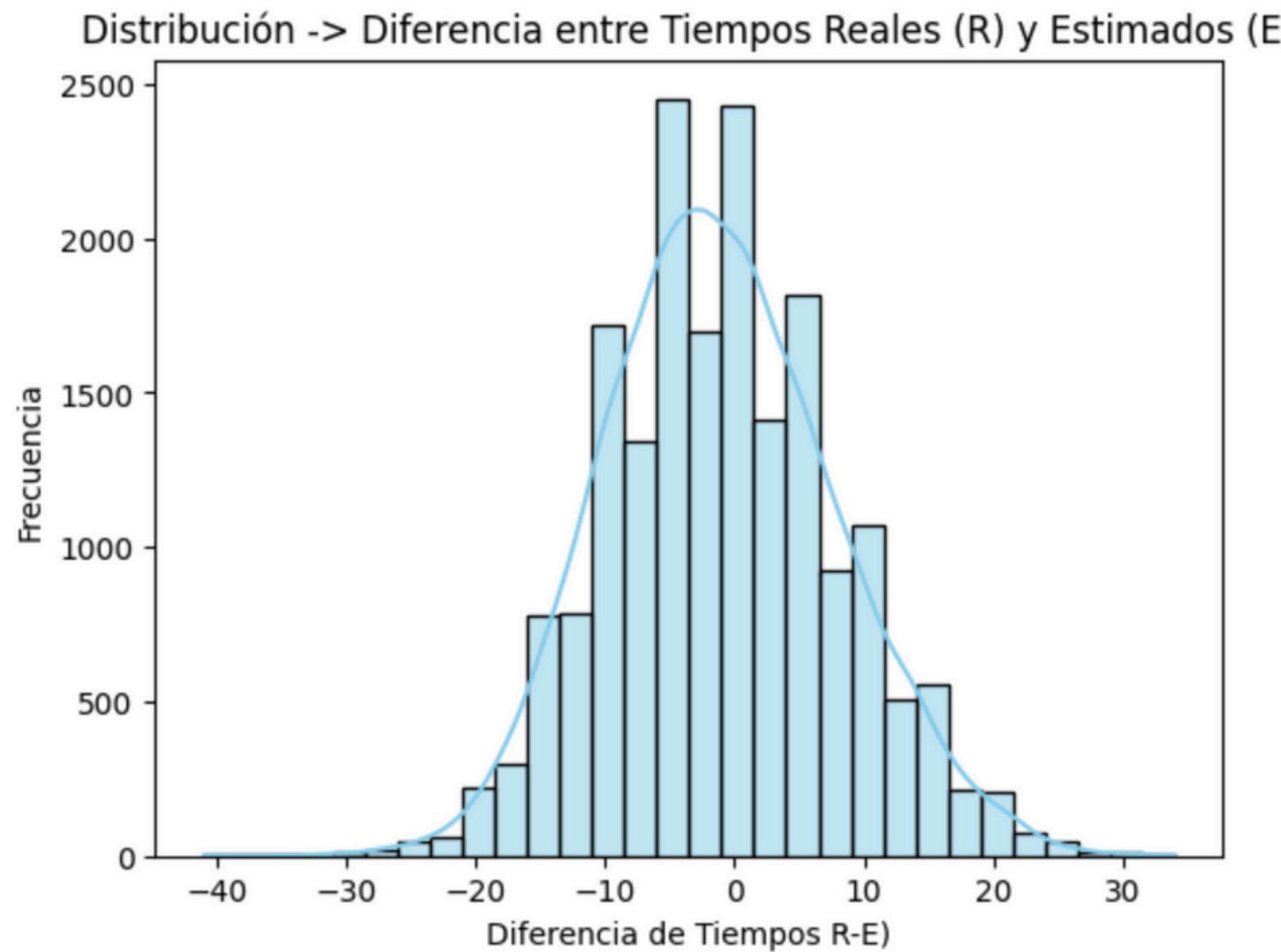
Se puede observar que existe una correlación positiva (tal vez un tanto moderada) Me indica que a medida que los tiempos estimados de entrega AUMENTAN también tienden a AUMENTAR los tiempos reales de entrega y al revés también.

ANÁLISIS BIVARIADO

VARIABLES

KDE (Kernel Density Estimate)

GRAFICAMOS LAS DIFERENCIAS DE TIEMPO



DIFERENCIA_TIEMPO = df1['ACTUAL_DELIVERY_MINUTES'] - df1['ESTIMATED_DELIVERY_MINUTES']

Centro de la Distribución

- El pico del histograma se encuentra alrededor de cero
- Las estimaciones son precisas, el tiempo real de entrega se acerca al tiempo estimado en la mayoría de los casos

Cola Positiva

- En algunos casos los tiempos reales de entrega fueron mayores que las estimaciones
- En ciertas situaciones el servicio tiende a tardar más de lo estimado
- Podría deberse a: condiciones del tráfico, situaciones imprevistas o una demanda inusualmente alta en ciertos momentos

Cola Negativa

- En algunos casos los tiempos reales fueron menores que las estimaciones.
- En ciertos casos el servicio tiende a ser más rápido de lo estimado.
- Podría deberse a eficiencias en el proceso de entrega, menor demanda en ciertos momentos o una planificación efectiva

ANÁLISIS MULTIVARIADO

OBSERVACIÓN

USER_LAT y VENUE_LAT

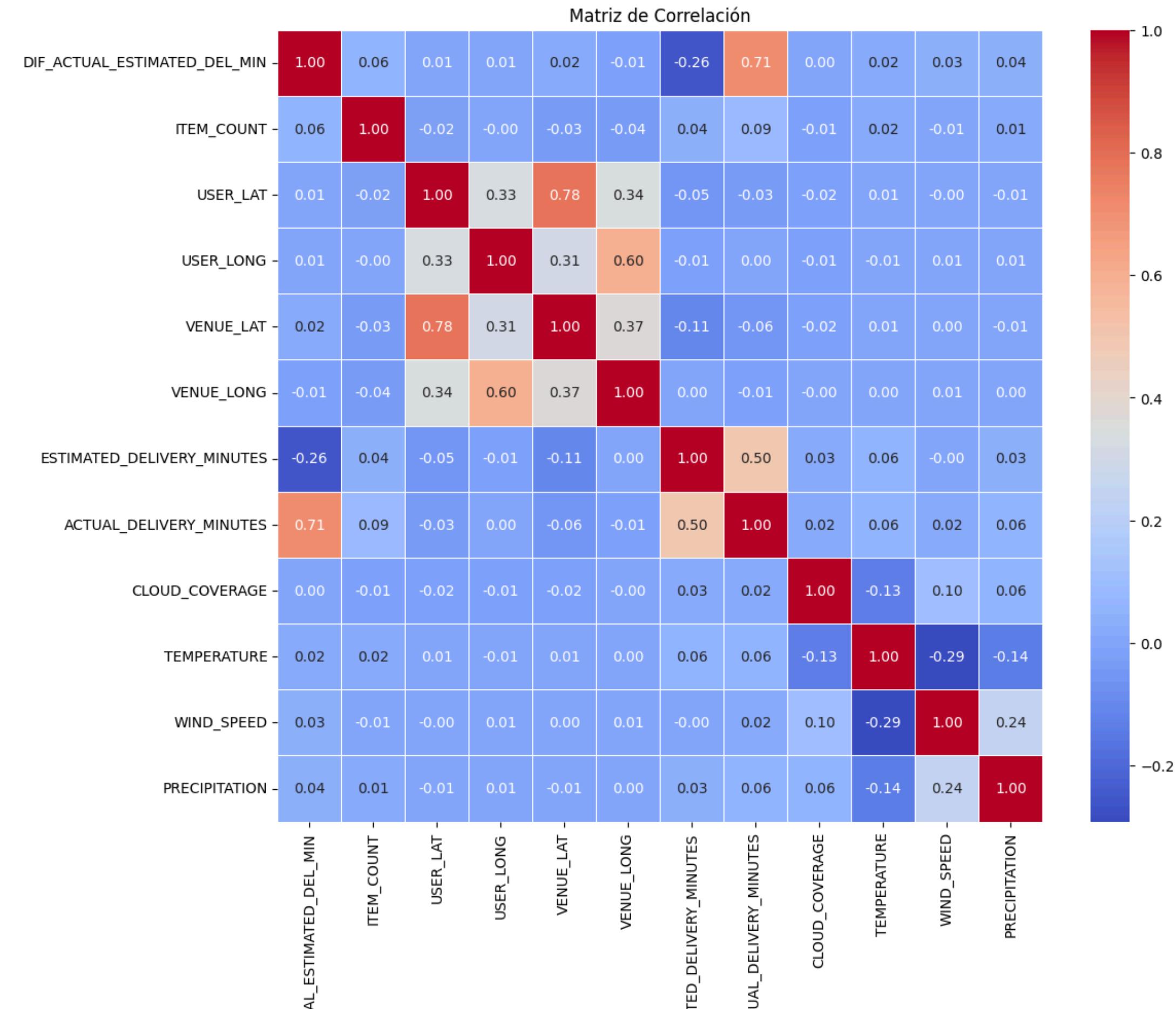
Correlación positiva. Esto sugiere que, en el conjunto de datos, las latitudes de los usuarios y los lugares (venues) tienden a aumentar juntas.

0.78

WIND_SPEED y TEMPERATURE

Correlación negativa débil. Esto significa que, en general, a medida que la velocidad del viento aumenta, la temperatura tiende a disminuir, pero esta relación no es muy fuerte.

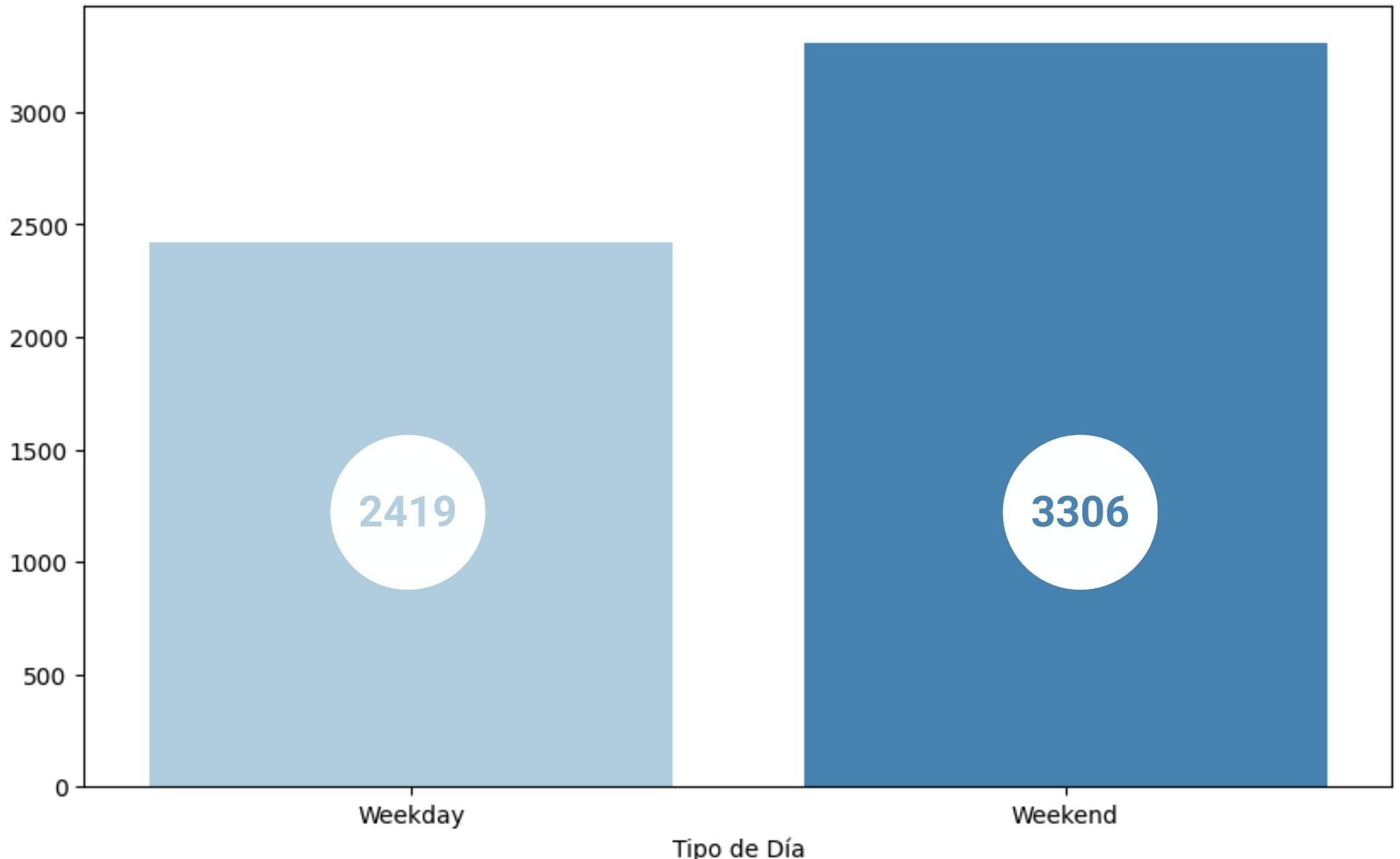
-0.29



MATRIZ DE CORRELACIÓN

ANÁLISIS DE LOS DÍAS

- El promedio de pedidos por día de fin de semana (3306) es mayor que el promedio de pedidos por día de semana (2419). Los datos indican que, en promedio, hay más pedidos realizados durante los fines de semana.
- Interpreto que puede deberse a diversos factores, como el aumento de la actividad social y de ocio durante el finde y puede llevar a un mayor número de pedidos.



MODELADO

MACHINE LEARNING

Aprendizaje Supervizado



REGRESIÓN LINEAL
RANDOM FOREST
XGBoost

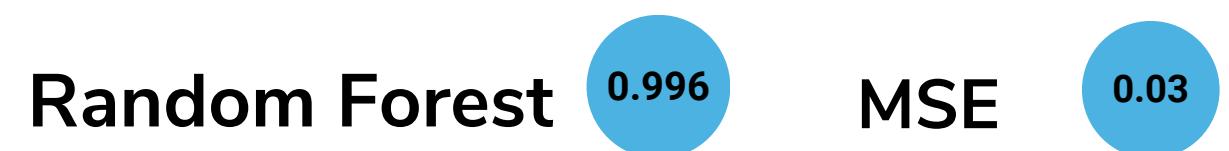
¿Por qué estos 3 modelos?

Ofrecen un rango de capacidades de modelado, desde la simplicidad y rapidez de la Regresión Lineal hasta la flexibilidad y potencia de los métodos basados en árboles como Random Forest y XGBoost.

- Predicción Numérica
- Variable Target de tipo Cuantitativa
- Feature Selection
- Foward Selection
- Optimización de hiperparámetros
- Crossvalidation

CONCLUSIONES

En este proyecto, analicé un conjunto de datos de Wolt con el objetivo de predecir el tiempo de entrega de los pedidos. Realicé el análisis exploratorio de datos (EDA) de manera ordenada, seleccioné las variables relevantes utilizando el método de Forward Selection, y entrené y evalué tres modelos de aprendizaje automático: Regresión Lineal, XGBoost y Random Forest. Los resultados indicaron que el modelo Random Forest tuvo el mejor rendimiento, destacándose por un alto coeficiente de determinación (R^2) y un bajo Error Cuadrático Medio (MSE) durante la validación cruzada (Cross-validation).



¿Se puede predecir el tiempo de entrega?

Se ha demostrado que se puede predecir el tiempo de entrega y poder así mejorar la eficiencia operativa, optimizar la planificación logística y mejorar la experiencia del cliente enfocandonos en dominar la logística de entregas.

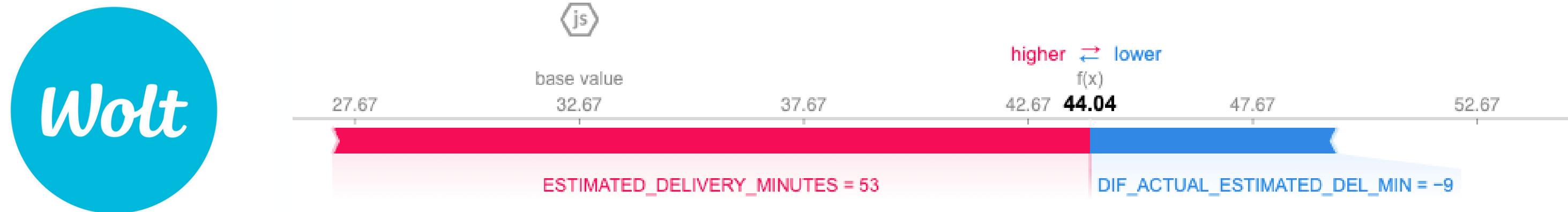
¿Cuáles son las características que mas influyen?

La cantidad de ítems en el pedido puede afectar significativamente el tiempo de preparación y, en consecuencia, el tiempo total de entrega. Por otro lado los patrones de demanda pueden variar significativamente entre días de semana y fines de semana, afectando el tiempo de entrega debido a diferentes volúmenes de pedidos y disponibilidad de recursos.



GRÁFICO DE FUERZA

SHAP: SHapley Additive exPlanations



- Base Value representa la predicción promedio sin considerar características específicas.
- $f(x)$ es la predicción ajustada basada en la muestra específica. En este caso, el modelo predice que el tiempo de entrega para un pedido con 4 ítems es de aproximadamente 44.04 minutos.
- Partes de Color muestran cómo cada característica influye en la predicción final. La parte roja agrega valor a la predicción y la parte azul resta valor a la predicción.