



Evaluación Formativa

Taller de métodos cuantitativos FMSA315

Primer Trimestre 2025

1 Instrucciones

- Cuenta con 120 minutos para realizar la actividad.
- Debe cargar un archivo PDF (incluir nombre y RUT en el documento), cualquier otro tipo de documento no será recibido y será calificado con nota mínima, sin apelación.
- Crear un breve informe con la solución de cada pregunta, comentando los principales resultados.
- No se asignarán puntos a respuestas sin justificación.
- Cualquier situación o problema en el transcurso de la evaluación debe ser informado de manera inmediata a su docente . Cualquier justificación una vez terminada la evaluación no será considerada.

2 Contextualización

El dataset taxis (de la librería seaborn) incluye registros de viajes en taxis en un período de tiempo específico, que cubren detalles esenciales como las fechas de recogida y entrega, número de pasajeros, distancia recorrida, tarifa, propinas, peajes, pago total, color del taxi y método de pago. Se incluyen estadísticas detalladas sobre la duración de los viajes, las distancias, las tarifas y los pagos.

Las variables del dataset son las siguientes:

1. pickup: fecha y hora de recogida.
2. dropoff: fecha y hora del término de viaje
3. passengers: Número de pasajeros
4. distance: Distancia recorrida (en millas).
5. fare: Tarifa del viaje.
6. tip: Propina extra.
7. tolls: impuestos.
8. total: Precio final del viaje, incluyendo tarifa, propina e impuestos.
9. color: Color del taxi.
10. payment: Medio de pago (credit card, cash).
11. pickup borough: Zona de recogida (Manhattan, Bronx, Queens, Brooklyn)
12. dropoff borough: Zona de llegada

En base a estos datos se pide responder a las siguientes preguntas.

3 Problemas

1. Carga la base de datos indicando que tipo de datos tiene cada variable, además del total de observaciones de cada una de ellas.
2. Representar la variable 'total' en un histograma, junto con el valor medio y la mediana de este. Luego determinar lo siguiente:
 - (a) ¿La distribución parece simétrica o sesgada? ¿Hacia qué valores se está sesgando?
 - (b) ¿La media y la mediana son similares? ¿Qué nos dice esto sobre la distribución?
3. Crea un gráfico de barras que muestre la distancia recorrida promedio por zona de recogida
 - (a) ¿Cuál es la zona de recogida, en donde se recorre una mayor distancia hasta llegar a destino?
 - (b) ¿Hay alguna zona de recogida con una gran desviación en la distancia total recorrida, según las medidas descriptivas (como la desviación estándar)?
4. Se desea visualizar la distribución del precio total del viaje (variable total) mediante una gráfica de cajas comparativas (boxplot):
 - ¿Cuál es la mediana de la variable 'total' en cada zona de recogida?
 - ¿Existen valores atípicos en cada categoría?
 - ¿Qué zona de recogida tiene la mayor variabilidad en precios totales pagados?
5. Determinar la probabilidad de que un taxi cuyo punto de recogida es en Queens, lleve más de 2 pasajeros y sea de color amarillo.