

Estimación Predictiva de Tarifas de Taxi en Nueva York (TLC)

Modelo de regresión para estimar tarifas antes del inicio del viaje

Descripción general del proyecto

Automatidata colabora con la Comisión de Taxis y Limusinas de Nueva York (TLC) para desarrollar un modelo analítico que permita estimar la tarifa de un viaje de taxi antes de que ocurra. Utilizando datos históricos de viajes, el proyecto aplica análisis exploratorio, preparación de datos y un modelo de regresión lineal múltiple con el fin de identificar los factores clave que influyen en el costo del viaje y apoyar la toma de decisiones operativas y financieras.

Key Insights

- La distancia promedio del viaje es el principal factor que explica la tarifa, seguida por la duración promedio.
- Existen tarifas reguladas (flat fares) que no dependen de distancia ni tiempo, lo que genera patrones discretos en los datos.
- El modelo final explica **entre 84 % y 87 %** de la variabilidad de la variabilidad de la tarifa, con desempeño consistente entre datos de entrenamiento y prueba.
- Variables temporales (día, mes, hora punta) y operativas tienen un impacto secundario en el monto final.
- La limpieza de valores atípicos y la eliminación de variables no disponibles antes del viaje mejoraron significativamente la capacidad de generalización del modelo.

Details

- Se trabajó con una muestra representativa de viajes de taxi de NYC (2017), previamente validada y depurada.
- Se aplicó un modelo de regresión lineal múltiple, asegurando el cumplimiento de supuestos básicos y evitando fuga de información.
- El error promedio del modelo es de aproximadamente 2 USD, con un error típico cercano a 4 USD, adecuado para estimaciones previas al viaje.
- El modelo es interpretable, permitiendo explicar claramente el efecto de cada variable sobre la tarifa estimada.

Next Steps

- Separar explícitamente los viajes con tarifas fijas de aquellos con tarifas variables para mejorar la precisión.
- Evaluar modelos alternativos o segmentados (por tipo de viaje o zona).
- Integrar el modelo como una herramienta de estimación y comunicación, no como tarifa exacta.
- Actualizar el modelo con datos más recientes y validar su desempeño bajo cambios regulatorios..