

Resumen de datos preliminares del proyecto TLC de la ciudad de Nueva York

Informe de resumen ejecutivo
Comisión Elaborada por Automatidata

OVERVIEW

La Comisión de Taxis y Limusinas de la Ciudad de Nueva York ha contratado a Automatidata para construir un modelo de regresión que prediga las tarifas de los taxis. En esta parte del proyecto, el equipo de datos de Automatidata realizó una inspección preliminar de los datos suministrados por la Comisión de Taxis y Limusinas de la Ciudad de Nueva York para informar al equipo de las descripciones de las variables de datos clave y garantizar que la información proporcionada sea adecuada para generar información clara y significativa.

PROJECT STATUS

- Se ha explorado el conjunto de datos para encontrar valores inusuales.
- Se consideró qué variables son más útiles para construir modelos predictivos (en este caso: `total_amount` y `trip_distance`, que trabajan juntos para representar un viaje en taxi).
- Se consideraron las posibles interacciones entre las dos variables elegidas.
- Examinó qué componentes de los datos proporcionados proporcionarán información relevante.
- Sentó las bases para futuros análisis exploratorios de datos, visualizaciones y modelos.

NEXT STEPS

1. Realizar un análisis exploratorio completo de los datos.
2. Realice cualquier paso de limpieza y análisis de datos para comprender variables inusuales (por ejemplo, valores atípicos).
3. Utilice estadísticas descriptivas para obtener más información sobre los datos.
4. Cree y ejecute un modelo de regresión.

KEY INSIGHTS

- Este conjunto de datos incluye variables que deberían ser útiles para crear modelos de predicción sobre las tarifas de los viajes en taxi.
- Los valores inusuales identificados son viajes que son de corta distancia pero que tienen cargas altas asociadas, como se muestra en la variable `total_amount`. Capturas de pantalla de referencia:

Total_amount variable

trip_distance	fare_amount
2.60	999.99
0.00	450.00
33.92	200.01
0.00	175.00
0.00	200.00
32.72	107.00
25.50	140.00
7.30	152.00
0.00	120.00
33.96	150.00

[Alt-text] La variable `total_amount` indica la necesidad de seguir analizando las variables atípicas.