

Análisis exploratorio de datos de TLC de la ciudad de Nueva York

Informe de resumen ejecutivo
Comisión Elaborada por Automatidata

Project 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 viajes en taxi. En esta parte del proyecto, los datos deben analizarse, explorarse, limpiarse y estructurarse antes de cualquier modelado.

Key Insights

El problema: Después de ejecutar un análisis exploratorio inicial de datos (EDA, por sus siglas en inglés) en una muestra de los datos proporcionados por la TLC de la ciudad de Nueva York, está claro que algunos de los datos resultarán un obstáculo para la predicción precisa de las tarifas de los viajes. Es decir, viajes que tienen un costo total ingresado, pero una distancia total de "0". En este punto, nuestro análisis indica que se trata de anomalías o valores atípicos que deben tenerse en cuenta en el algoritmo o eliminarse por completo.

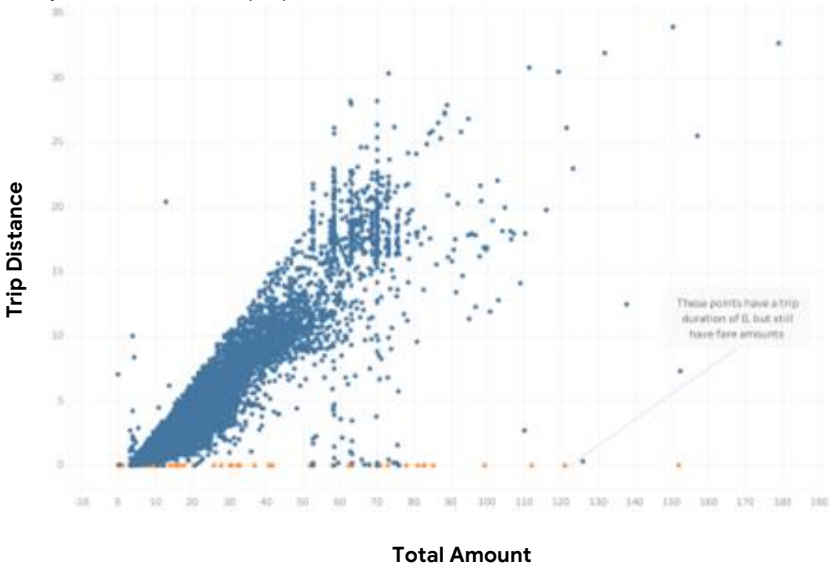
Solución propuesta: Después del análisis, recomendamos eliminar los valores atípicos con un registro de distancia total de 0.

Claves del éxito

- Asegurar con New York City TLC que la muestra proporcionada sea un reflejo exacto de sus datos en su conjunto.
- Planifique el manejo de otros valores atípicos, como la baja distancia de viaje junto con los altos costos.

Details

Como resultado del análisis exploratorio de datos realizado, el equipo de datos de Automatidata consideró la distancia del viaje y la cantidad total como variables clave para representar un viaje en taxi. El diagrama de dispersión proporcionado muestra la relación entre las dos variables. Este diagrama de dispersión se creó en Tableau para mejorar la visualización proporcionada.



Texto alternativo: Gráfico que muestra los datos de TLC de la ciudad de Nueva York que trazan variables para la distancia total y la cantidad total.

Next Steps

- Determinar cualquier punto de datos inusual que pueda plantear un problema para análisis futuros en la predicción de tarifas de viaje..
 - Por ejemplo, ubicaciones que tienen duraciones más largas.
- Determine las variables que tienen el mayor impacto en las tarifas de los viajes.
- Filtre hacia abajo para tener en cuenta las variables más relevantes para ejecutar la regresión, el análisis estadístico y el ajuste de parámetros.