Initial Data Insights for NYC TLC Project

Executive Summary: Commission Report of Automatidata

OVERVIEW

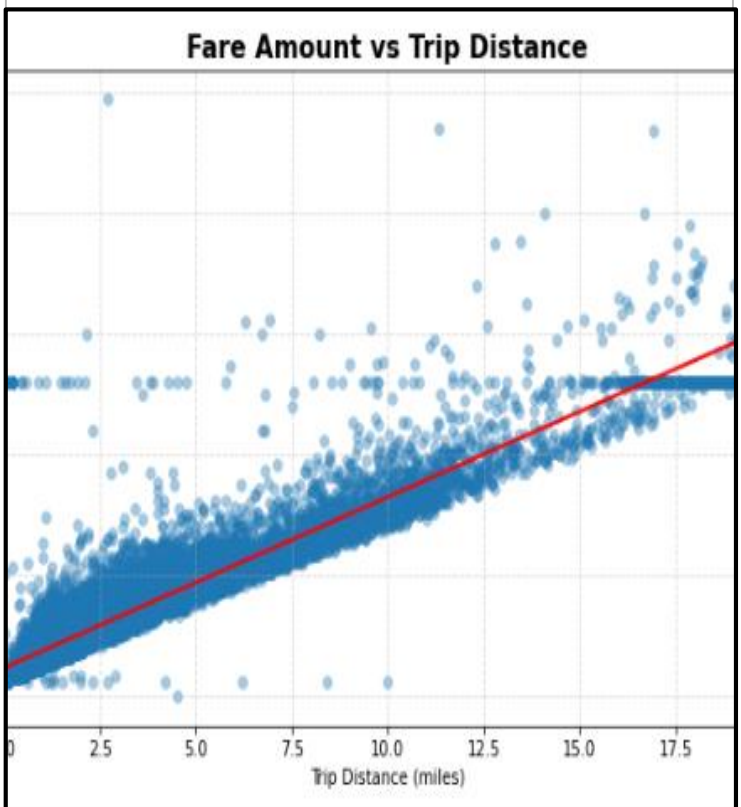
The NYC Taxi & Limousine Commission has contracted with Automatidata to build a regression model that predicts taxi fares before each ride using distance, time of day, and other relevant factors. In this stage of the project, the Automatidata data team conducted a preliminary inspection of the detailed taxi and rideshare trip data provided by the NYC TLC to review key variable descriptions, assess data quality, and ensure the dataset is suitable for generating accurate and meaningful insights.

PROJECT STATUS

- Explored dataset to find unusual values in columns.
- Examined features that can provide relevant information.
- The mentioned three features are considered to possible interaction.
- Built groundwork for future data analysis, data visualization, and model development.

KEY INSIGHTS

- The two most helpful variables for building a model to predict tips or fare amounts for NYC TLC are trip_distance, fare_amount, and payment_type**.**
- No null values are present in any column but .there is a need of data cleaning.
- The maximum recorded trip distance is 33.96 miles.



NEXT STEPS

- Perform Exploratory Data Analysis.
- Conduct data cleaning to handle outliers, remove irregular values, and anomalies
- Use descriptive statistics for better understanding of data.
- Develop and execute regression model.