

New York City TLC Project Data Summary

Project Overview

The NYC Taxi & Limousine Commission has contracted with Automatidata to build a regression model that predicts taxi cab fares. In this data summary, the Automatidata Data team inspects the data provided by the NYC Taxi and Limousine Commission to determine the key variables for EDA.

Key Insights

- This dataset includes variables that should be helpful for building prediction model on taxi cab ride fares.
- The identified unusual values are trips that are a short distance but have high charges associated with them, as shown in the total_amount variable.

Details

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

The total_amount variable indicates the necessity of further analyzing outlier variables.

Next Steps

1. Conduct exploratory data analysis (EDA).
2. Perform any data cleaning and data analysis steps to understand unusual variables (e.g., outliers).
3. Use descriptive statistics to learn more about the data.
4. Create and run a regression model.
5. Test the model on test data.