New York City TLC Project Preliminary Data Summary

Executive summary report

OVERVIEW

The NYC Taxi & Limousine Commission has asked Automatidata to build a regression model that predicts taxi cab fares. In this part of the project, the Automatidata data team performed an exploratory analysis of the data supplied by the NYC Taxi and Limousine Commission in order identify key variables, and ensure the data provided is suitable for generating clear and meaningful insights.

PROJECT STATUS

- Explored dataset to find outlier or unusual values.
- Identified variables most useful to build predictive models (total_amount and trip_distance, which are key components of a taxi cab ride).
- Considered potential interactions between the two chosen variables.
- Built the groundwork for future exploratory data analysis, visualizations, and models.

NEXT STEPS

1. Conduct a complete exploratory data analysis.
2. Perform any data cleaning and data analysis steps to understand unusual variables (e.g., outliers).
3. Use descriptive statistics to summarize the data.
4. Create and run a regression model.

KEY INSIGHTS

- This dataset includes variables that should be helpful for building a 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. Reference screenshots:

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