New York City TLC Project Preliminary Data Summary

Executive summary report
Commission Prepared by **Automatidata**

OVERVIEW

Automatidata has been contracted by the New York City TLC to build a regression model that can forecast taxi fares based on the distance of the trip. Currently, a preliminary examination of the data has been conducted to provide descriptions of important variables and guarantee that valuable insights can be derived.

PROJECT STATUS

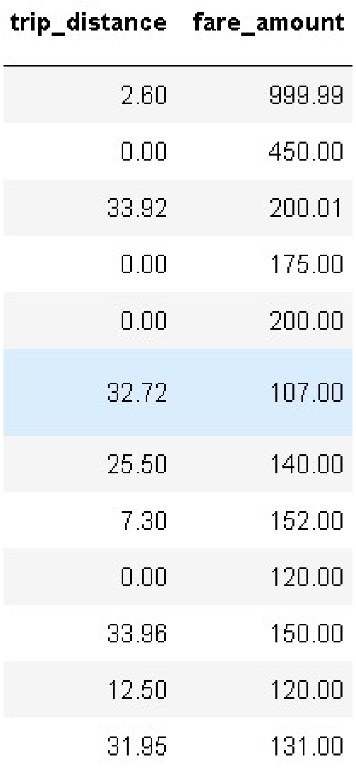
- The dataset has been analyzed to identify any anomalies.
- Trip distance and total amount for a taxicab ride were identified as the most useful variables for developing a predictive model.
- Paved the way for further exploration opportunities, including EDA, visualizations, and modeling.
- Potential correlations between the two selected variables have been explored.

NEXT STEPS

- Begin by performing a thorough exploratory analysis of the data.
- Clean and reformat the data as required.
- Identify any outliers and decide on the appropriate action to take, such as removing them or further investigation.
- Utilize descriptive statistics, including summaries and tables, to gain a better understanding of the data, particularly the necessary variables.
- Lastly, develop and test a regression model.

KEY INSIGHTS

- Key variables in the dataset were identified for creating a regression model to anticipate taxicab ride fares (trip_distance and fare_amount.)
- Unusual values were found, such as low trip_distance with high fare_amounts.
- Instances of trip_distance = 0 with fare amounts, negative values, and extremely high maximum values were also present in the data.



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
12.50	120.00
31.95	131.00