

New York City TLC Project Preliminary Data Summary

Executive summary report
Commission Prepared by **Automatidata**

OVERVIEW

The New York City Taxi and Limousine Commission (TLC) is responsible for licensing and regulating taxi cabs in the city. The TLC partnered with Automatidata to create a regression model that predicts taxi fares before a trip. For this project, the Automatidata team explored the TLC’s trip data to examine its structure, identify potential issues, and confirm that key variables identified, and it is suitable for generating meaningful insights.

PROJECT STATUS

- Explored the dataset to find any outlier values.
- Identified which variables are most useful for building a predictive model (in this case: trip_distance and total_amount).
- Tested the relationship between trip_distance and total_amount.
- Looked at other variables that could provide helpful context.
- Prepared the data for deeper exploration, visualizations, testing, and modeling to answer questions and gain insights.

NEXT STEPS

- Perform a full exploratory data analysis (EDA).
- Clean the data and analyze it more to understand the outlier values.
- Use descriptive statistics to learn more about the dataset.
- Begin building and testing a regression model.

KEY INSIGHTS

- The dataset includes variables that can help build prediction models for estimating taxi ride fares.
- The identified outliers are trips with short distances but very high fares, as shown in the total_amount variable. Reference screenshots:

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

[Alt-text] The total_amount variable shows the need for more analysis of these outlier values.