

New York City TLC Project Preliminary Data Summary

Executive summary report

Commission Prepared by **Automatidata**

OVERVIEW

This project performed a preliminary inspection of NYC Taxi and Limousine Commission data to inform the team of key data variable descriptions, and ensure the information provided is suitable for generating clear and meaningful insights.

PROJECT STATUS

- Explored dataset to find any unusual values.
- Considered which variables are most useful to build predictive models (in this case: total_amount and trip_distance).
- Visualized the selected two variables to show potential interactions between them
- Examined which components of the provided data will provide relevant insights.
- Build 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 learn more about the data.
4. Create and run a regression model.

KEY INSIGHTS

- The Automatidata team assessed variables finding both missing and unusual data values. The dataset has unusual values which will cause future problems with predictive models.
- The identified unusual values are some trips that are a short distance but have high charges associated with them, as shown in the total_amount variable.
- This dataset includes variables that should be helpful for building prediction model(s) on taxi cab rides.

Total_amount variable

