

# New York City TLC Project - Preliminary Data Summary

## Executive Summary Report I

Commission Prepared by **Automatidata**

### Project Overview

The NYC Taxi & Limousine Commission has partnered with Automatidata to develop a regression model for predicting taxi fares. In this phase of the project, the Automatidata team carried out an initial review of the dataset provided by the NYC Taxi and Limousine Commission. This was done to better understand the key variables and confirm that the data is appropriate for deriving valuable insights.

### Key Insights

- This dataset includes key variables that should assist in constructing prediction models for taxi fares.
- The unusual values identified involve short trips with disproportionately high charges, as indicated in the variable for the total amount. Refer to the attached screenshot for more details.

trip_distance	fare_amount	total_amount
2.60	999.99	1200.29
0.00	450.00	450.30
33.92	200.01	258.21
0.00	175.00	233.74
0.00	200.00	211.80
32.72	107.00	179.06
25.50	140.00	157.06
7.30	152.00	152.30
0.00	120.00	151.82
33.96	150.00	150.30

The variable for the total amount highlights the need for a deeper analysis of outlier values.

### Details

- Inspected the dataset for any outliers or anomalies.
- Identified the most relevant variables for building predictive models, particularly the total amount and the trip distance, which are essential for representing a taxi ride.
- Considered potential relationships between the selected variables.
- Reviewed which parts of the dataset would yield meaningful insights.
- Laid the foundation for further exploratory data analysis, visualizations, and model-building.

### Next Steps

1. Conduct a thorough exploratory data analysis.
2. Carry out data cleaning and analysis to identify unusual values, such as outliers.
3. Use descriptive statistics to gain deeper insights into the dataset.
4. Develop and test a regression model.