# New York City TLC Project Preliminary Data Summary

**Executive summary report**

Commission Prepared by **Automatidata**

## OVERVIEW

The NYC Taxi & Limousine Commission partnered with Automatidata to develop a regression model that predicts taxi fares. In this phase, the Automatidata data team conducted a preliminary review of the provided dataset to understand key variables and ensure the data is suitable for generating clear and meaningful insights.

## PROJECT STATUS

- Identified outliers in the dataset.
- Selected key variables for predictive modeling, such as total fare and trip distance.
- Explored potential interactions between selected variables.
- Assessed the most useful components of the dataset for generating meaningful insights.
- Prepared the foundation for future exploratory analysis, visualizations, and modeling.

## NEXT STEPS

1. Conduct a full exploratory data analysis.
2. Perform data cleaning and analysis to understand unusual variables (e.g., outliers).
3. Use descriptive statistics to gain deeper insights into the dataset.
4. Build and run a regression model.

## KEY INSIGHTS

- This dataset includes variables that should be helpful for building prediction model(s) on taxi cab ride fares.
- The identified outliers are short-distance trips with unusually high charges, as reflected in the `total_amount` variable. Reference screenshots illustrate these cases:

***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 indicates the necessity of further analyzing outlier variables.*