# New York City TLC Project Preliminary Data Summary

**Executive summary report**
Commission Prepared by **Automatidata**

## OVERVIEW

The purpose of this project was to conduct an initial examination of the data obtained from the NYC Taxi and Limousine Commission. The objective was to gain an understanding of the key data variable definitions and determine if the information available is appropriate for producing relevant and valuable insights.

## PROJECT STATUS

- Checked dataset for anomalies.
- Determined the most relevant variables for building predictive models, which were identified as total_amount and trip_distance.
- Analyzed potential variable interactions.
- Examined relevant data components for insights.
- Built 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

- This dataset includes variables that should be helpful for building prediction model(s) on taxi cab rides.
- The identified unusual values are trips that are a short distance but have high charges associated with them, as shown in the total_amount variable. See screenshot.

### *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 |