

New York City TLC Project Preliminary Data Summary

Executive summary report

Project Overview

The aim of this project was to conduct an initial examination of data from the NYC Taxi and Limousine Commission. The purpose was to familiarize the team with the essential descriptions of data variables and confirm that the data is appropriate for producing insights that are concise and significant.

Details

Key Insights

- The dataset contains variable that should be useful in building predictive model(s) on taxicab rides
- Some trips were unusual:
- High fare and low distance

| fare_amount | trip_distance |
|-------------|---------------|
| 999.99 | 2.60 |
| 450.00 | 0.00 |
| 200.01 | 33.92 |
| 175.00 | 0.00 |
| 200.00 | 0.00 |
| 107.00 | 32.72 |

- Trips with negative total fare

| | |
|---------|------|
| -4.80 | 0.24 |
| -5.30 | 0.64 |
| -5.30 | 0.40 |
| -5.80 | 0.46 |
| -5.80 | 0.70 |
| -120.30 | 0.17 |

- Analyzed the dataset to identify any atypical variables.
- Evaluated which variables would be most valuable in developing predictive models, with a focus on total_amount and trip_distance.
- Explored potential correlations or interactions between the selected variables.
- Investigated the relevant components of the data to extract meaningful insights.
- Established a foundation for forthcoming exploratory data analysis, visualizations, and modeling activities.

Next Steps

- 1) Carry out a thorough exploratory analysis of the dataset, including data cleaning and analysis to identify any anomalies such as outliers.
- 2) Utilize descriptive statistics to gain deeper insights into the data.
- 3) Develop and implement a regression model to further understand the relationships within the dataset.