New York TLC Project Preliminary Data Summary

Executive summary report
Prepared by **Automatidata**

OVERVIEW

The New York Taxi and Limousine Commission seeks to develop a regression model so as to help their riders estimate taxi fares before the ride. In this part of the project, the Automatidata data team performed a preliminary inspection of the data supplied by the NYC Taxi and Limousine Commission.

PROJECT STATUS

- Project Proposal had been approved.
- The Data was provided by TLC Data team.

The tasks were:
- To build a Data-Frame for the TLC dataset
- To examine data type of each column
- To gather descriptive statistics
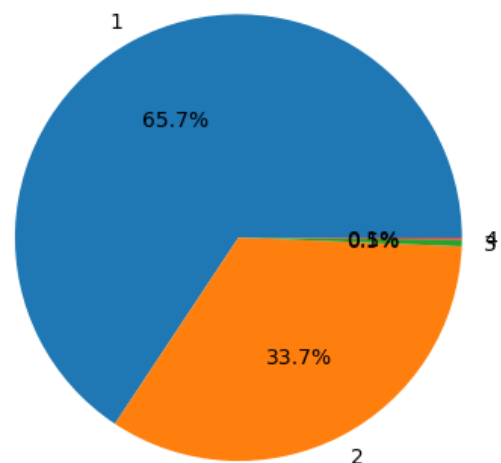- To build groundwork for future EDA, visualizations, and models.

NEXT STEPS

- Conduct Exploratory Data Analysis.
- Clean Data for Analysis.
- Build a regression model.

KEY INSIGHTS

- The data contains errors.
- Distance is 0 for some rides.
- Total amount is negative for some rides.
- Payment type distribution can be seen in the visualization.
- Average tip for credit card payments is $2.61
- Average Total amount is similar for both vendors ~$15.64.
- Tips are larger for passenger count greater than 6.

Payment type distribution

1 — 65.7%
2 — 33.7%
0.5%

Total Count of payment types: 5161674