New York TLC Project Preliminary Data Summary

Executive Summary Report

OVERVIEW

New York City Taxi and Limousine Commission (New York City TLC), has hired the Automatidata team to create a regression model to predict the taxi cab fares. We execute a preliminary inspection of the data provided by New York City TLC. Thus, we would like to inform some key aspects that we found respecting to the variables that may affect the fare, so we can check if the insights are useful and proceed to the following steps.

PROJECT STATUS

- We performed a preliminary inspection of the data.
- We conducted a study of the two variables that will be useful to create the regression model.
- We examined another variables to see if it affects our study, such as payment type, the tip amount and the vendor ID.
- This work was useful because we gain insights on future complete data analysis and the creation of data visualizations.

NEXT STEPS

- Clean the dataset for future complete exploratory data analysis.
- Use descriptive analysis to learn more about the data,
- Create the regression model.

KEY INSIGHTS

- There some some unusual values for the fare amount and total amount, showing negative values.
- There is one value that shows that the longest trip doesn't correspond the maximum fare.
- Besides this unusual value, the other data shows a strong correlation.
- We also performed a relation on what is the preferred payment, being the credit card.
- We calculated the average tip amount for each passenger count with credit card.

trip_distance	fare_amount
33.96	150.00
33.92	200.01
32.72	107.00
31.95	131.00
30.83	80.00
30.50	90.50
30.33	52.00
28.23	52.00
28.20	52.00
27.97	52.00

Table 1. Trip distance vs. fare amount. It is important to note that there are an unusual value,