

TLC Fare Prediction Project EDA Summary

Executive Summary Report for Automatidata

Overview

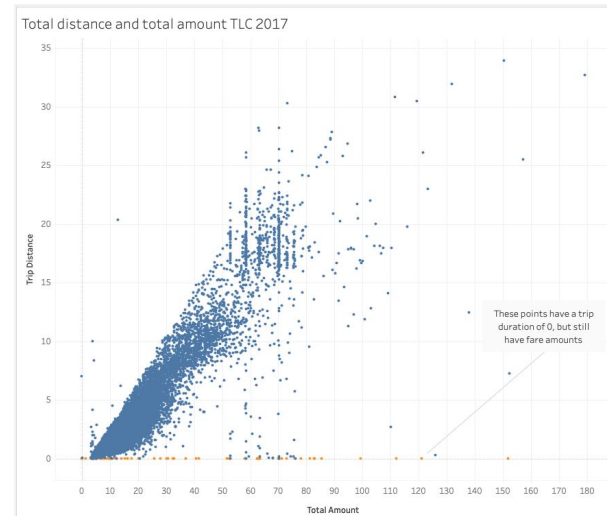
The New York City TLC data is ready for exploratory data analysis (EDA) in Python. Clean, validate, and create a visualization for the taxi commission data.

Objective

- Load data, explore, and extract the New York City TLC data with Python
- Use custom functions to organize the information within the New York City TLC dataset
- Build a dataframe for the New York City TLC project

Results

- Identified outliers using describe(), unique(), box plots, etc.
- Filtered out outliers with condition statements, for example, `df[total_amount] < 180` to remove rows with total_amount values more than 180.
- Dropped least applicable columns from dataset.
- Created data visualizations for specific columns, such as histograms, bar charts, scatter plots, etc.
 - Scatter plot of Total Amount vs Trip Distance on Tableau (right).



Next Steps

- Review the data visualizations.