# Exploratory Data Analysis of New York City TLC Data

## Executive Summary Report

**Commision prepared by Automatidata**

The NYC Taxi & Limousine Commission has partnered with Automatidata to develop a regression model that can accurately predict the fares of taxi cab rides. Before modeling can take place, the data needs to be analyzed, explored, cleaned, and structured.

## Key Insights

**The Problem:** During the Exploratory Data Analysis of the dataset, certain issues were discovered. Several rows of data exist in which the trip distance value is missing, yet there is a corresponding total amount value. These data points have been identified as anomalies and should be removed from the dataset to ensure that models and ride fare predictions are not affected by skewed data.

---

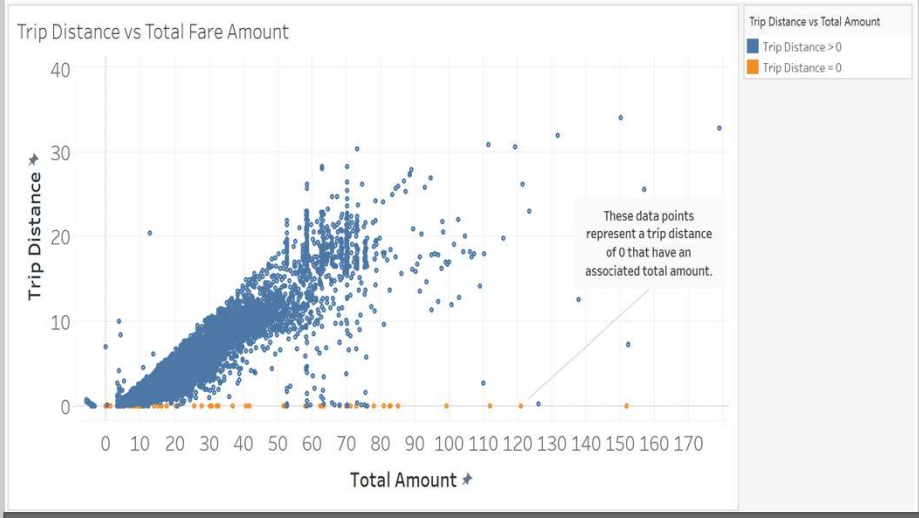
**The Solution:** Due to the nature of these data points, it is recommended to remove these outliers with a trip distance of 0 from the dataset.

---

**Important Considerations:** Come up with additional strategies to handle additional outliers, such as data points with low trip distance but high costs or high trip distance and low costs.

## Details

During the exploratory data analysis, it was found that the two key variables for predicting the fare amount are Trip Distance and Total Amount. The Scatter Plot demonstrates the relationship between these variables, indicating a relatively linear connection.




## Next Steps

- Investigate any anomalous data points that could affect fare amount predictions and conduct further analysis to determine which key variables have the strongest correlation with trip fares.
- Consider which variables are the most relevant for regression models and statistical inference.