

# New York City TLC Project Preliminary Data Summary

## Executive Summary Report

Prepared by Nawaz Quraishi

### Project Overview

This project conducted an initial exploration of the NYC Taxi and Limousine Commission dataset to familiarize the team with the key data variables. Additionally, it aimed to verify the relevance of the provided information for further analysis and the derivation of meaningful insights.

### Details

### Key Insights

- The dataset comprises of numeric data types including floats and integers, as well as strings.
- The dataset does not contain null values.
- Further investigation is required to examine some questionable values in the dataset.
- VeriFone Inc., provided most number of rides among the vendors.
- Within the dataset there are four distinct payment methods.
- Credit cards emerged as the most favored method of payment.
- Excluding taxes and surcharges, the average fare amount is \$13.02.
- For constructing predictive models, the variables trip\_distance and fare\_amount appear to be the most useful.

- Python, with a primary focus on Pandas used to investigate the dataset.
- The dataset contained 22,699 rows and 18 variables.
- The average fare calculated includes negative values. Once the negative values are rectified, the average fare will yield a higher number.
- As the scope of the project did not involve data cleaning or Exploratory Data Analysis (EDA), the dataset was unaltered and the essential groundwork for subsequent analysis and exploration has been accomplished.

### Next Steps

- Investigate the negative values in the dataset.
- Clean the data.
- Perform a comprehensive Exploratory Data Analysis.
- Feature Engineer variables.
- Build a predictive model.
- Create visualizations to communicate findings.