**NYC Taxi & Limousine Commission | Preliminary Data Analysis**

Prepared by: Kaisaer John

**OVERVIEW**

A regression model is being developed to forecast taxicab fares for the NYC Taxi & Limousine Commission. In this phase of the project, a preliminary examination of the data provided by the NYC Taxi and Limousine Commission has been conducted to convey essential descriptions of data variables and confirm the suitability of the information for deriving clear and meaningful insights.

**PROJECT STATUS**

**KEY INSIGHTS**

* Explored dataset to find any unusual values.

* Considered which variables are most useful to build predictive models (in this case: ‘total\_amount’ and ‘trip\_distance, which work together to depuct a taxicab ride).
* Considered potential interactions between the two chosen variables.
* Examined which components of the provided data will provide relevant insights.
* Built the groundwork for future exploratory data analysis, visualization, and models.
* This dataset includes variables that should be helpful for building prediction model(s) on taxicab ride fares.
* The identified unusual values are trips that are a short distance but have high charges associated with them, as shown in the ‘total\_amount’ variable. Reference screenshots:

A screenshot of a cell phone

Description automatically generated

**NEXT STEPS**

1. Conduct a complete exploratory data analysis.
2. Perform any data cleaning and data analysis steps to understand unusual variables (e.g. outliers).
3. Use descriptive statistics to learn more about the data.
4. Create and run a regression model.