**New York City TLC Project Preliminary Data Summary**

**Executive summary report**

**Commission Prepared by Noor Ul Ain Zahid**

OVERVIEW :

The New York City Taxi & Limousine Commission (TLC) hired Automatidata to develop a model that can predict taxi fares. As part of this project, I, **Noor Ul Ain Zahid**, carried out a preliminary inspection of the data provided by the TLC. In this step, I reviewed and understood the key variables such as trip distance, passenger count, and fare amounts. I also ensured that the dataset was clean and reliable, making it suitable for producing clear and meaningful insights in the later stages of analysis.

PROJECT STATUS :

* Explored the 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 depict a taxi cab 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, visualizations, and models.

KEY INSIGHTS :

* This dataset includes variables that should be helpful for building prediction model(s) on taxi cab 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:

***Total\_amount variable***

 

NEXT STEP :

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.