

# Taxi Company Project Preliminary Data Summary (Milestone 2)

Executive summary report  
Prepared by PhD Aleksandar Osmanli

## OVERVIEW

The taxi company has asked to build a regression model that predicts taxi cab fares. In this part of the project, it was performed a preliminary inspection of the data supplied by the taxi company in order to inspect the key data variable descriptions, and ensure the information provided is suitable for generating clear and meaningful insights.

## PROJECT STATUS

- 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 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.

## 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.

## 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 fare\_amount variable. Reference screenshots:

### *fare\_amount variable*

trip_distance	fare_amount
2.60	999.99
0.00	450.00
33.92	200.01
0.00	175.00
0.00	200.00
32.72	107.00
25.50	140.00
7.30	152.00
0.00	120.00
33.96	150.00

[Alt-text] The fare\_amount variable indicates the necessity of further analyzing outlier variables.