

Kainat Khalid

I22-2242

CS-Z

◆ Report Structure

Your report should include the following sections:

1 □ Introduction

- Briefly describe the dataset and problem statement.
- Mention the **NYC Taxi Trip Duration Dataset** and its source.

2 □ Dataset Selection & EDA Summary

- Explain why you chose this dataset.
- Describe key insights from **Exploratory Data Analysis (EDA)**:
 - Missing values
 - Outliers
 - Correlations
 - Visualizations

3 □ Data Preprocessing & Feature Engineering

- Describe how missing values, categorical variables, and feature scaling were handled.
- Mention any new features created (e.g., **trip speed**).

4 □ Model Selection & Training

- Discuss different models tested:
 - **Linear Regression**
 - **Decision Tree**
 - **Random Forest**
 - **Gradient Boosting**
- Explain how the dataset was split (Train/Test).

5 □ Model Fine-Tuning & Evaluation

- Show performance metrics: **RMSE, MAE, R² Score**.
- Explain **hyperparameter tuning** (GridSearchCV, RandomizedSearchCV).
- Include **Cross-Validation Results**.

6 **Final Model & Conclusion**

- Discuss which model performed best and why.
- Highlight key findings and limitations.