# Capstone Project Submission

**Team Member's Name, Email, and Contribution:**

1 Sandesh Gaikwad
Email:- sandeshg3.14159@gmail.com
Contribution: - Everything(individual Project)

**Please paste the GitHub Repo link.**

https://github.com/GaikwadSandesh/NYC_taxi_prediction

**Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches, and your conclusions. (200-400 words)**

**Problem:** To build a model that predicts the total ride duration of taxi trips in New York City. The primary dataset is one released by the NYC Taxi and Limousine Commission, which includes pickup time, geo-coordinates, number of passengers, and several other variables

**Approach:** First Handel all null values. Remove irrelevant columns. Remove anomaly. use feature engineering to distance between two locations. use box cox method to get a gaussian distribution of the skewed variable. fit linear models and tree-based models.

**Conclusions / Recommendations:**

- Linear, Lasso, and Ridge regressions are giving similar results.
- In the Decision Tree regressor, the results are slightly improved.
- Out of all tried models, Random Xgboost Regressor is giving the best result.

**Future challenges:**

- Distance and time of day are not sufficient to predict the time of ride it requires a thorough understanding of traffic and that the actual path of the ride is important.
- Deep Neural Networks can improve this performance.