Capstone Project Submission

Team Member's Name, Email and Contribution:

Name : Puneet Subanji
Email : puneetsubhanji@gmail.com

- **Data understanding**
  Data Analysis
  Data pre processing and exploration

- **Feature Analysis**
  Univariate
  Bivariate and multivariate analysis

- **Feature engineering**
  Null value check
  Feature creation , etc

- **Data visualisation**

- **Algorithm implementation**
  Linear Regression
  Lasso Regression
  Ridge Regression
  Decision tree Regressor
  XGBoost Regressoin

- **Research Analytics**
  Technical documents

Please paste the GitHub Repo link.

Github Link:- https://github.com/PUNEETSUBHANJI/NYC_Taxi_Trip_Time_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

The task is to create a model that forecasts the entire ride time for taxi trips in New York City. Your primarydataset is one released by the NYC Taxi and Limousine Commission, which includes pickup time, geo- coordinates, number of passengers, and several other variables.

APPROACH

- In first step, imported the data set to carry out the descriptive analysis over the data set to understand the information of data available.
- Checked the supplied data set for missing values and instances of repeated values.
- Exploring all the variables of the data set (such as Vendor-Id, Pickup_point, Dropoff_point, Distance, Speed, Storage_frwd, etc) with respect to trip duration, to determine the factors and understand factors of trip duration.
- Explored the correlation using data visualization and several types of plots with Trip duration and different variables.
- Encoding of categorical columns and fitting the different models, we used below algorithms :-
    - Lasso Regression
    - Ridge Regression
    - Decision tree Regressor
    - XGBoost Regressor
- Then tuning into Hyperparameters and perfomance evaluation to identify best fit Model.

CONCLUSION
- Most of the trips durations took 10-30 mins to complete at a speed range of 10-20 km/h.
- Trip duration on Thursday is longest among all days.
- Highest amount of trips were taken by a single passenger
- Large group of people travelling together is rare compared to single passenger.
- The busiest and most popular time of day is the evening.
- From February, we can see trip duration rising every month. The reason for this progressive rise in trip duration over time may be due to seasonal factors like wind and rain.
- The trips taken by the two vendors do not differ significantly.
- XGbooster is the best performing model amongst all.