Capstone Project Submission

Instructions:

i) Please fill in all the required information.
ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

sibani choudhury
Mail id choudhurysibani120@gmail.com
Contribution

- Preview Data
- Check total number of entries and column types
- Check the null values
- Plot distribution of numeric data
- Plot distribution of categorical data
- Remove the outliers
- Correlation through heatmap
- Building the model
  Linear regression
  XgBOOST
  Decission tree
  Gradient boosting

Model interpretation
Conclusion

Rasik Jain
Mail id rasik6627265@gmail.com
Contribution

- Data Cleaning
- EDA
- Feature engineering and Feature selection
- distribution check for dependent and independent features numeric data
- Outlier detection and elimination
- Correlation
- PPT and Team Colab Building Contribution
- Building and evaluating the model Linear regression
  Adaboost
  GBM
  Linear Regression
  XGBoost
  LightGBM
- Conclusion

Sangamesh chandankera
Mail id sangameshchandan@gmail.com
Contribution

- Loading the data
- Data Wrangling / Explore the data
- Check the null values
- Descriptive statistics of the dataset

- Exploratory data analysis(EDA)
- Feature Enginering and data visualization
- Categorical values – One hot encoding
- PPT and Team Colab Building Contribution
- Building the model
  - Linear regression
  - Decision tree
  - Adaboost
  - Gradient boosting
  - XgBOOST
- Conclusion

Please paste the GitHub Repo link.

Github Link:- https://github.com/Sangameshchandan/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)

A taxi company faces a common problem of efficiently assigning the cabs to passengers so that the service is smooth and hassle free. One of main issue is determining the duration of the current trip so it can predict when the cab will be free for the next trip.

Our first step is to prepare dataset for our machine learning models. After loading the dataset we performed Exploratory Data Analysis by comparing our target variable that is trip duration with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables. We will do certain steps like dropping unnecessary columns and do the one hot encoding for the required columns.

Once after exploring the data, started to check out the null values present in the each column of the dataset. After which went for the visualization part to get the insights of each variable.

After data handling we fit our Machine learning models like Linear regression,Xgboost, Lightgbm to our data .And with the help of ML metrices like r2 score, RMSE we decide that which machine learning model is the best fit for our dataset.

We are mostly concerned with the information of pick up latitude and longitude and drop off latitude and longitude, to get the distance of the trip .
LightGBM will be the best model to predict the trip duration for a particular taxi.
.