Capstone Project Submission

Team Member's Name, Email and Contribution:

Name : Jayalaxmi Mekap
Email id : youmailjaya.inbox@gmail.com

Contribution:
- Introduction to Data
- Data cleaning
- Correlation
- Understanding the data
- Payment method across each rider
- Car or bus type across each rider
- Max capacity in car and bus
- Max rider travel from
- Rider spending time in bus or car
- Number of tickets Vs rider
- Travel from vs number of tickets
- Day of the month Vs number of tickets
- Number of tickets vs hours
- Merging of data frames on apps
- Speed vs number of tickets
- Encodings categorical features
- Training the model
- Linear regression
- Implementing lasso regression
- Ridge regression
- Training gradient boosting regressor
- Training XGboost
- Random forest regressor
- Grid search cv on random forest
- Grid search cv on XGboost algorithm

Please paste the GitHub Repo link.

Gitthub Link :

https://github.com/Jaya9522/Demand_Prediction_for_Public_Transport_Capstone_Project
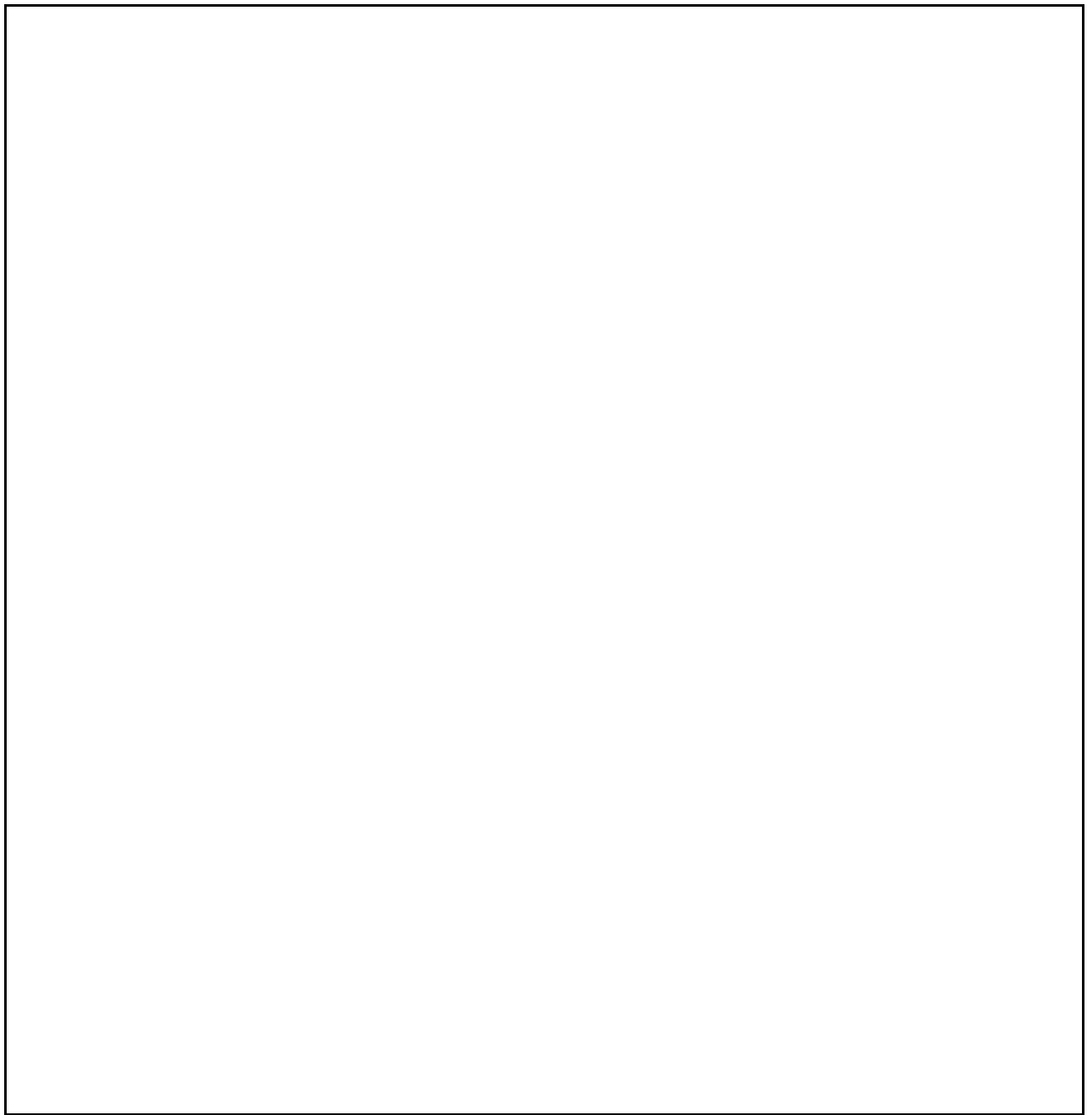
Drive Link :
https://drive.google.com/drive/folders/1Fhr_gAbhknkbwPhjuWJU4QsJrVgVXidl?usp=sharing

Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions.

(200-400 words)

- This challenge asks you to build a model that predicts the number of seats that Mobiticket can expect to sell for each ride, i.e. for a specific route on a specific date and time. There are 14 routes in this dataset. All of the routes end in Nairobi and originate in towns to the North-West of Nairobi towards Lake Victoria.

- The towns from which these routes originate are: Awendo Homa Bay Kehancha Kendu Bay Keroka Keumbu Kijauri Kisii Mbita Migori Ndhiwa Nyachenge Oyugis Rodi Rongo Sirare Sori

- The routes from these 14 origins to the first stop in the outskirts of Nairobi takes approximately 8 to 9 hours from time of departure. From the first stop in the outskirts of Nairobi into the main bus terminal, where most passengers get off, in Central Business District, takes another 2 to 3 hours depending on traffic.

- The three stops that all these routes make in Nairobi (in order) are: Kawangware: the first stop in the outskirts of Nairobi Westlands Afya Centre: the main bus terminal where most passengers disembark All of these points are mapped here.Passengers of these bus (or shuttle) rides are affected by Nairobi traffic not only during their ride into the city, but from there they must continue their journey to their final destination in Nairobi wherever that may be. Traffic can act as a deterrent for those who have the option to avoid buses that arrive in Nairobi during peak traffic hours. On the other hand, traffic may be an indication for people's movement patterns, reflecting business hours, cultural events, political events, and holidays.

- There are 51645 observations in the dataset. And has 10 columns.There are total **61** unique seats in this dataset
- travelers have used 2 types of payment method and most of the people have used **Mpesa** to pay for their ticket.

- The record of 149 days out of 2 year is present in this dataset.

- There are 2 different types of car and most of them are bus.There are two type of payment methods people have used to buy the tickets.
- There are two type of cars Bus and shuttle and the maximum capacity of the bus is 49 while shuttle can contain 11 travelers.

- We can see that there is the gap between 5 to 11 in the day of the month. We can assume that there is official holyday of public transport between these days. we can also say that the number of tickets in all the days of month are same.
-
- We used diffent type of regression algorithms to train our model like, Linear Regression, Regularized linear regression (Ridge and Lasso), GBM,Random Forest Regressor, XGboost regresssor. and Also we tuned the parameters of Random forest regressor and XGboost regressor and also found the important features for training the model. Out of them XGboost with tuned hyperparameters gave the best result.