

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

Team Member's Name

Email id

Syed Aquib

syedaquib153@gmail.com

Ganesh Subramanian

gsubramanian5@gmail.com

Mayur Chulbhare

chulbharemayur@gmail.com

Abhishek Kirar

abhishekkirar27@gmail.com

Contribution:

Syed Aquib :

- Data Wrangling
 - Loading and Pre-processing
 - Structuring data
 - Enriching data
- Data Mining
- Data Analysis
- Visualizations
 - Bar graphs and Distribution Graph
- Machine Learning -- Modelling and Predicting using Algorithms
 - Linear Regression , Lasso and Ridge Regression , XGBoost Regression
- Observation
- Summarization
- Conclusions
- Technical Document
- Power Point Presentation

Ganesh Subramanian :

- Data Wrangling
 - Loading and Pre-processing
 - Structuring data
 - Enriching data
- Data Mining
- Data Analysis
- Visualizations
 - Bar graphs and Distribution plot
- Machine Learning – Modelling and predicting using Algorithms
 - Linear Regression , Lasso and Ridge Regression , XGBoost Regression
- Observation
- Summarization
- Conclusions
- Technical Document
- Power Point Presentation

Mayur Chulbhare :

- Data Wrangling
 - Loading and Pre-processing
 - Structuring data

- Enriching data
- Data Mining
- Data Analysis
- Visualizations
 - Bar graphs and Distribution plot
- Machine Learning – Modelling and predicting using Algorithms
 - Linear Regression , Lasso and Ridge Regression , XGBoost Regression
- Observation
- Summarization
- Conclusions
- Technical Document
- Power Point Presentation

Abhishek Kirar :

- Data Wrangling
 - Loading and Pre-processing
 - Structuring data
 - Enriching data
- Data Mining
- Data Analysis
- Visualizations
 - Bar graphs and Distribution plot
- Machine Learning – Modelling and predicting Algorithms
 - Linear Regression , Lasso and Ridge Regression , XGBoost Regression
- Observation
- Summarization
- Conclusions
- Technical Document
- Power Point Presentation

Please paste the GitHub Repo link.

Github Link:-

<https://github.com/syedaquib153/Nairobi-Transport-Demand-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)

Project Name: Nairobi-Transport-Demand-Prediction

Problem Statement:

Nairobi-Transport-Demand-Prediction provided to us is in unformatted manner, uneven data, and duplicate data and also some data columns in it is irrelevant, since it's a piled-up data coming from various different countries. For doing the analysis on the data the data needs to be in correct format and well organized form. The main objective of the analysis is to obtain the meaning full insights and facts from the given datasets, modelling and predicting the model using Machine Learning Algorithms.

Approach:

As a team we read the data present in the file and gone through the details in each and every column. The data set was huge in which some of the data was not required for the analysis so the data was cleaned by dropping some unwanted columns and created new data frame, with the columns we required for the analysis. The first problem we faced was the name of the columns which was not proper and the nan values present in the data. We renamed the columns by using dictionary format and replaced all the nan values to zero in int dtype and unknown in object dtype by using replace syntax. Each and every column were compared to gain the insights about the data. We worked individually to gain some insights by doing the exploratory data analysis using python. Cleaning the dataset, analysing the data and visualizing the data by plotting the data into different graph and charts so that the trend and relationship between the various indicators can be understand easily, Modelling and Predicting the model using Machine learning algorithms which model is best to predictor.

Conclusion:

We used different type of regression algorithms to train our model like, Linear Regression, Regularized linear regression (Ridge and Lasso), XGboost regressor and Also we tuned the parameters of XGboost regressor and also found the important features for training the model. Out of them XG boost with tuned hyperparameters gave the best result.

This resulting model can be used by Mobiticket and bus operators to anticipate customer demand for certain rides, to manage resources and vehicles more efficiently, to offer promotions and sell other services more effectively, such as microinsurance, or even improve customer service by being able to send alerts and other useful information to customers. We used different type of regression algorithms to train our model like, out of them XG Boost with tuned hyperparameters gave the best result.