

Capstone Project Submission

Instructions:

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

Team Member's Name, Email and Contribution:

1. Aniket Suresh Satpute

Email-id: - aniketsatpute000@gmail.com

❖ **Contribution-**

- Values count for payment mode, car type, maximum capacity
- Finding out target variable and removing constant feature.
- Comparing day of month v/s number of tickets.
- Exploration of hour v/s number of tickets.
- Encoding categorical variables.
- Implementing and evaluating model:
- XG boost and random forest.

2. Kaiwalya Dashrath Zankar

Email-id: - kaiwalyazankar@gmail.com

❖ **Contribution -**

- Checking the presence of NAN values by using the tool missing value matrix.
- Visualization of the towns from which these routes originate.
- Explorations of travel from v/s number of tickets.
- Feature engineering for adding some variables.
- Finding difference between next and previous buses.
- Training linear regression and gradient boosting regression.

Please paste the GitHub Repo link-

GitHub Link: - [Aniket-Satpute/Transport-Demand-Prediction. \(github.com\)](https://github.com/Aniket-Satpute/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)

Nairobi is one of the most heavily congested cities in Africa. Each day thousands of Kenyans make the trip into Nairobi from towns such as Kisii, Keroka, and beyond for work, business, or to visit friends and family. The journey can be long, and the final approach into the city can impact the length of the trip significantly depending on traffic. How do traffic patterns influence people's decisions to come into the city by bus and which bus to take? Does knowing the traffic patterns in Nairobi help anticipate the demand for particular routes at particular times.

The aim of the competition is to create a predictive model using traffic data provided from Uber Movement and historic bus ticket sales data from Mobiticket to predict the number of tickets that will be sold for buses into Nairobi from cities in "up country" Kenya.

In this project, we are analyzing the various aspects with different use cases which covers many aspects of airbnb listings. It helps in not only understanding the meaningful relationships between attributes but it also allows us to do our own research and come-up with our findings.

The data set contains 51,645 observations and 10 columns

To understand data integrated details, we find out done null values and unique values from which and we came to know there no null values

From the EDA point of view kisii is the top place from where the number of rights originate and also, we need to remove the constant feature that is tribal to and through that we found out our target variable as number of tickets.

then we have then some future engineering and that the observes the most of the tickets were sold at 7 AM and 8 PM and that seems true because in the morning most of the people goes to the work and office and also there is no ride between 12:00 pm to 5:30 pm.

While training the modern we used different ML algorithms amongst all XG boost with they won't ever parameters gives us the best result and the speed having most important features amongst all of them.

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 micro-insurance, or even improve customer service by being able to send alerts and other useful information to customers.

Drive Link:- [Transport Demand Prediction - Google Drive](#)