mobiticket

People's Movement into Nairobi

Group 5 - Data Tuners.

The Team

- **Biko Mwangi - Data mining.**
- **Joy Muli - Data mining and slides.**
- **Pruno Juma - Data mining.**
- **Vicky Mumo - Tableau and slides.**

What is MobiTicket?

- **Mobiticket is a service that enables you to book, pay and get bus tickets using your phone.**

- **Ticket reservations and payments can be done using MPESA checkout or using mobile wallet and it directly connects you to partner bus operators.**

Why Nairobi?

- **According to CNBC Africa Nairobi is one of the most heavily congested cities in Africa.**
- **With people moving in and out of Nairobi frequently, there is bound to be traffic congestion.**

Focus

- We want to analyse the movement of people into Nairobi.

- The goal is to have a model that can predict the number of passengers we can expect based on the time, the ride, and the location of origin.

Data Involved.

- The datasets used were provided by Zindi.
- There are 14 routes in the dataset. All of the routes end in Nairobi and originate in towns to the North-West of Nairobi towards Lake Victoria.
- Some of the important columns include; ride_id, travel_date, travel_time, travel_from, travel_to, car_type, max_capacity.

Technologies Used.

- **Python Programming - Used for manipulation of the data in the data mining process.**
- **Google Colabs - Used as an environment for perfoming the various data mining tasks.**
- **Tableau - Used to create interactive dashboards.**
- **Canva Slides - Used to create the presentation slides.**

Experimental Design

- Loading and Previewing Data
- Data Cleaning
- Exploratory Data Analysis
- Exploration of different models
- Decision of a model and its tuning
- Conclusions
- Recommendations

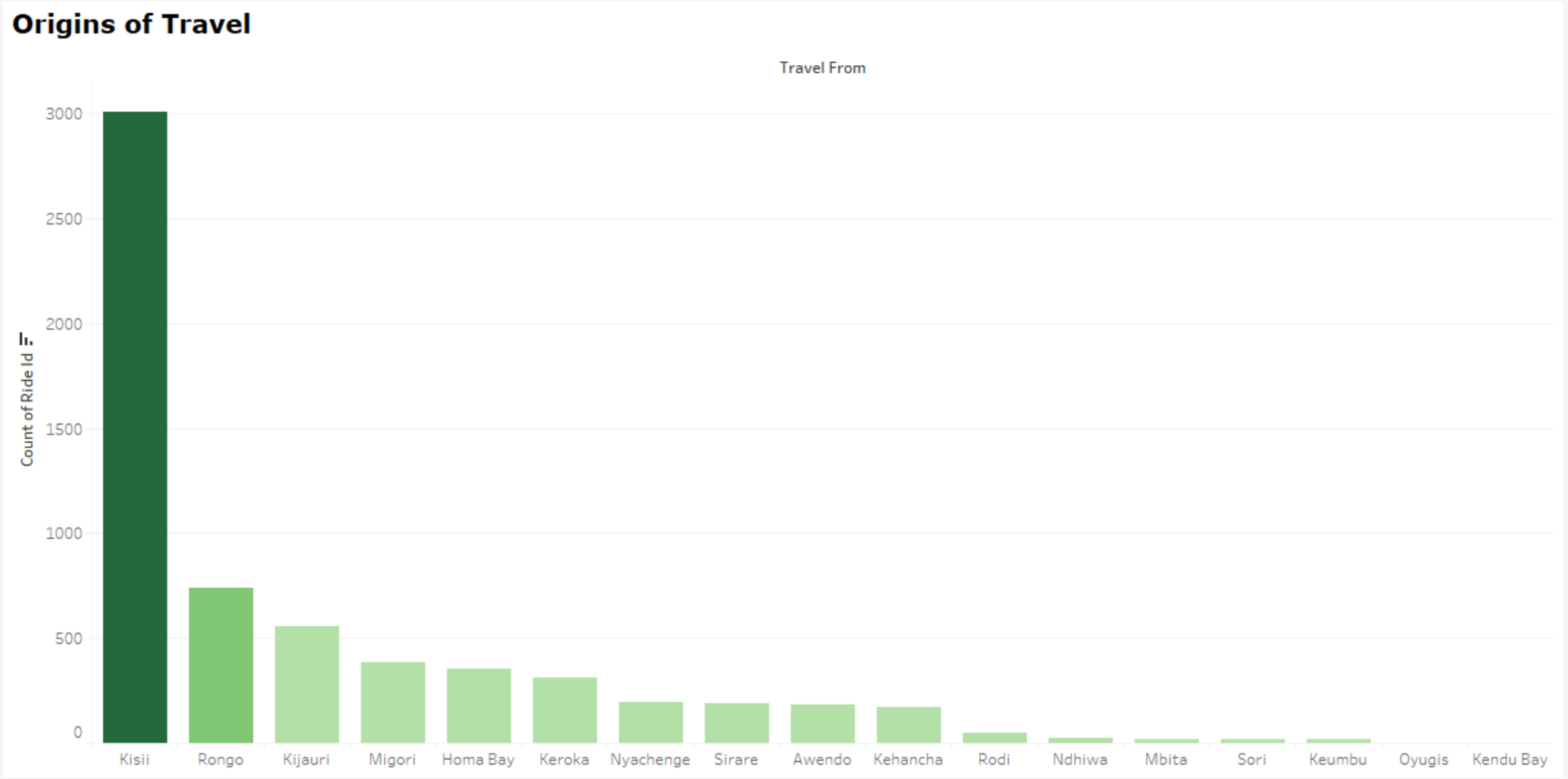
Point to note*

- **In our data cleansing, We created a new column ‘number_of_tickets’ that showed the number of tickets sold per ride_id.**

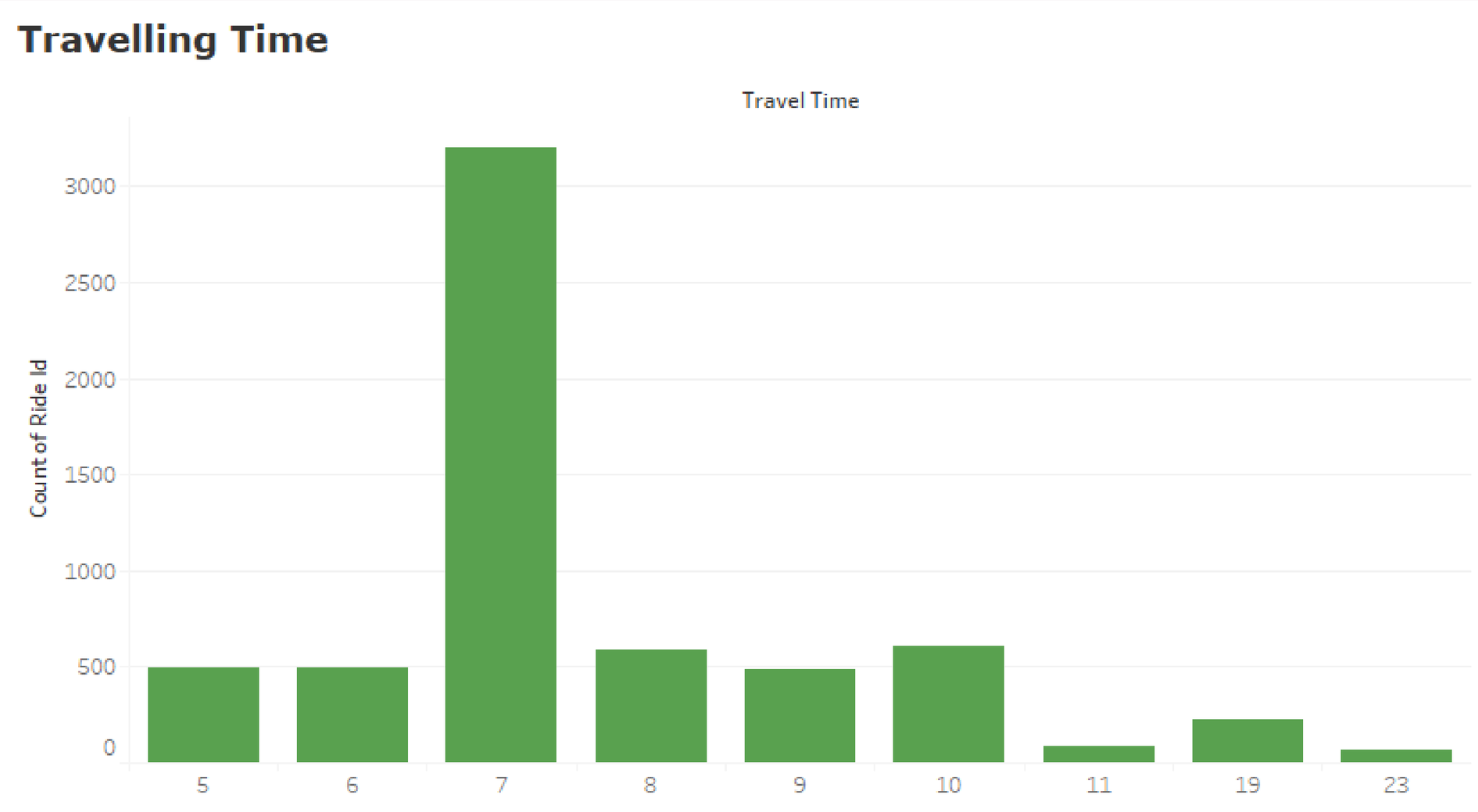
- **We used it as our target variable to predict people’s movement in Nairobi.**

After Performing our initial steps of loading and cleaning of the data, we went on to perform exploratory data analysis and we had the following results;

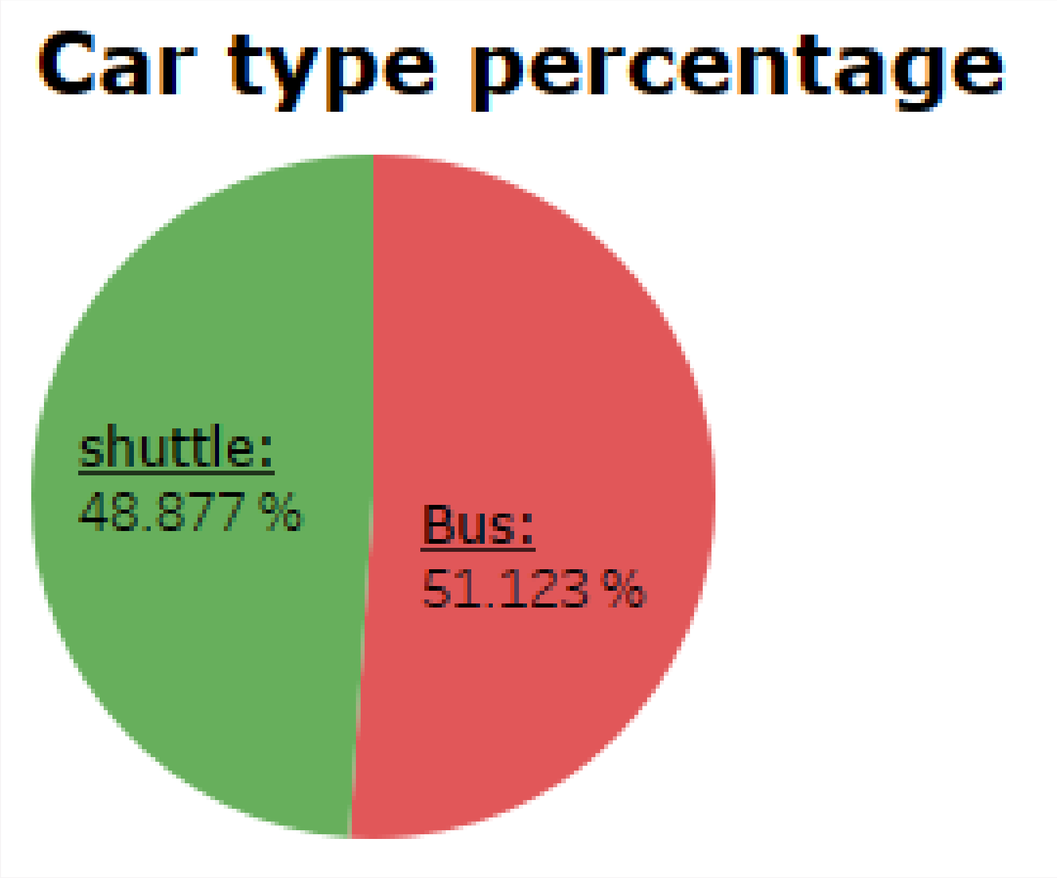
Number of tickets purchased per town



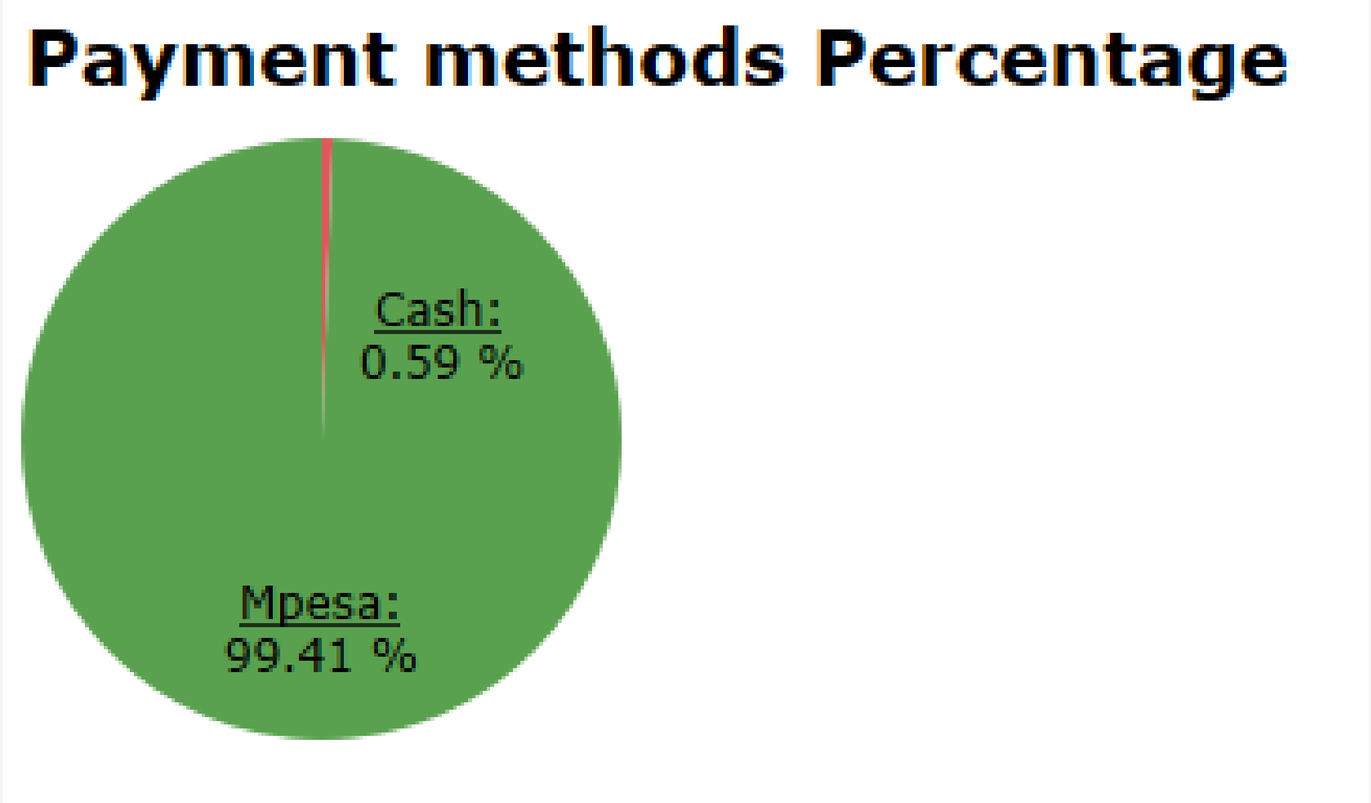
Travelling Time Distribution



Car Type used



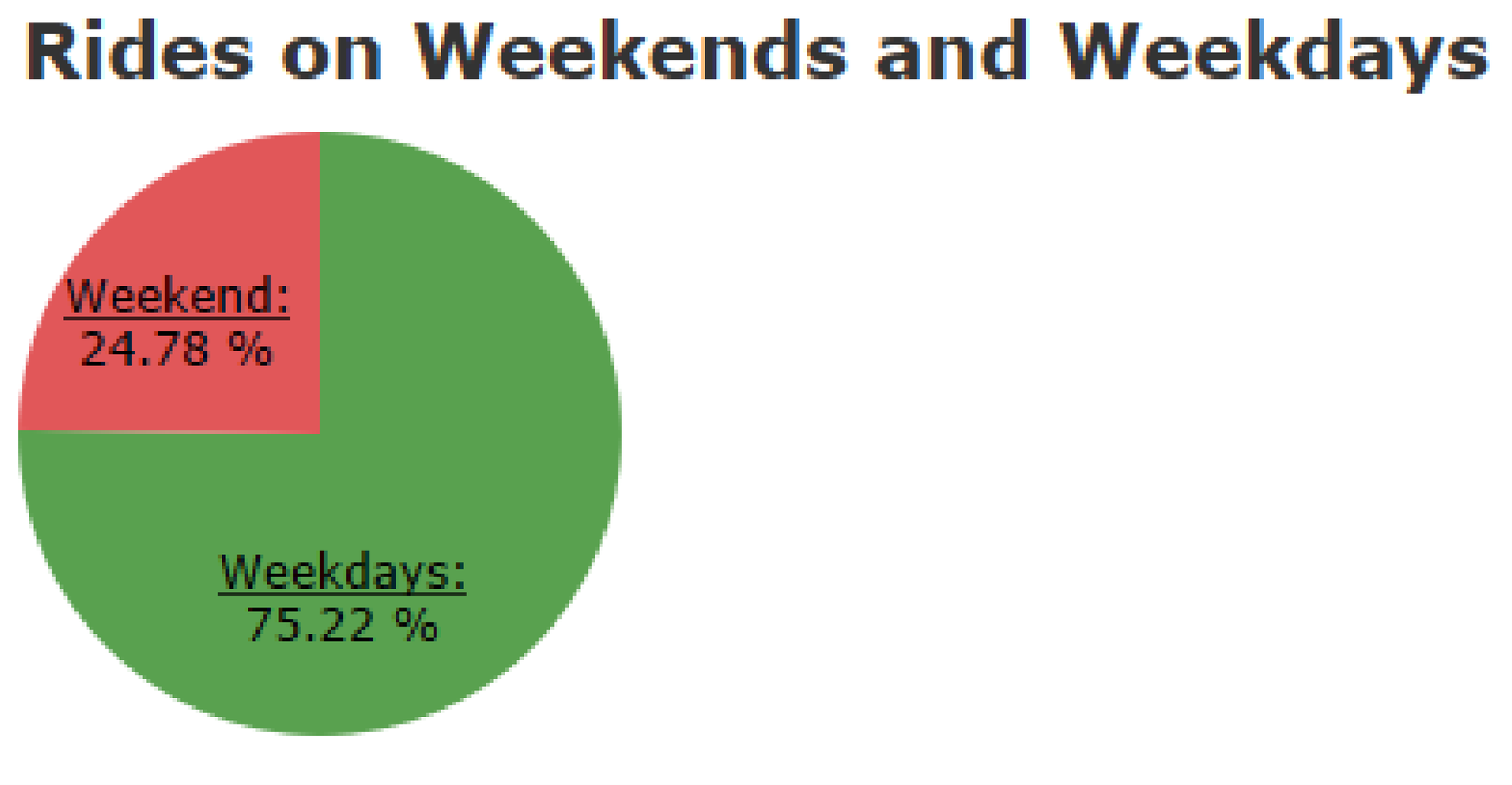
Payment Method Analysis



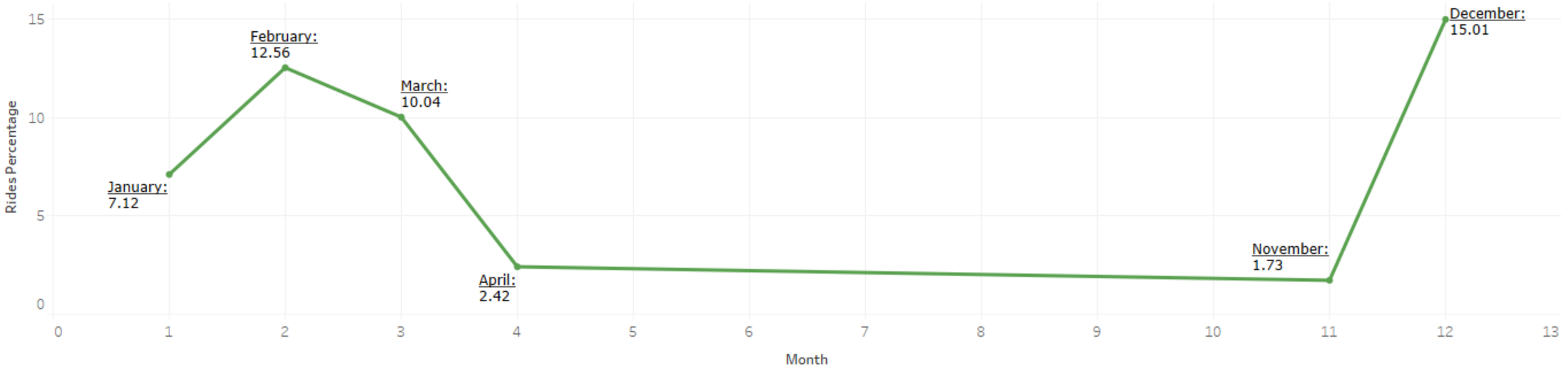
Rides behaviour Through Time

The following graphs show the behaviour of rides through different spaces of time;

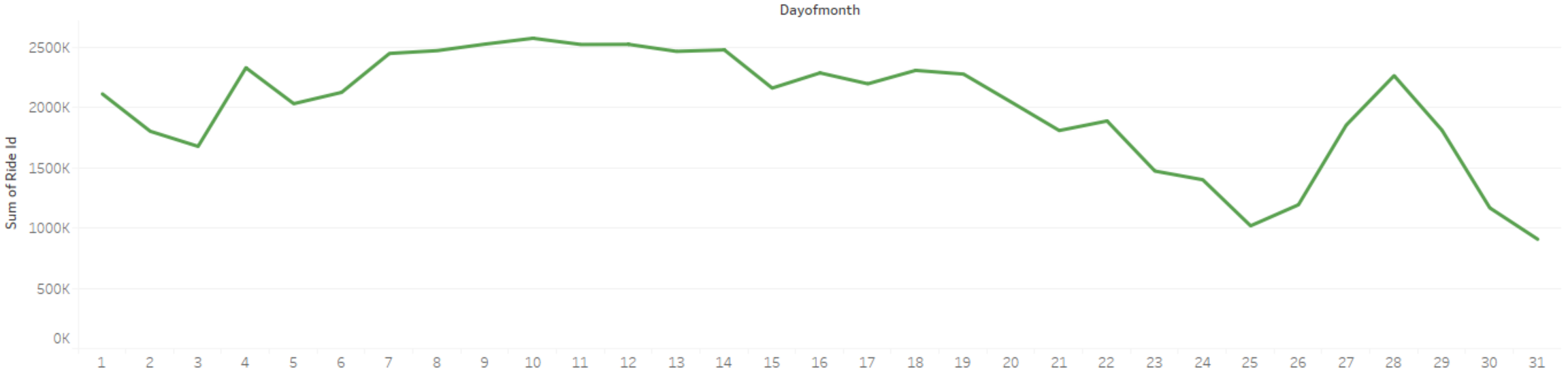
Rides distribution through the week



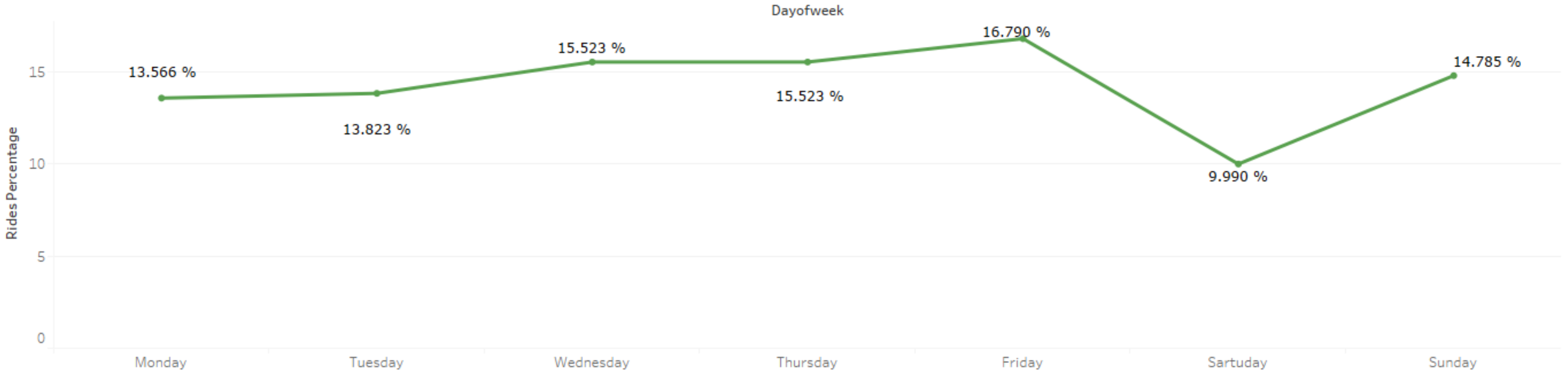
Rides on Months of the year.



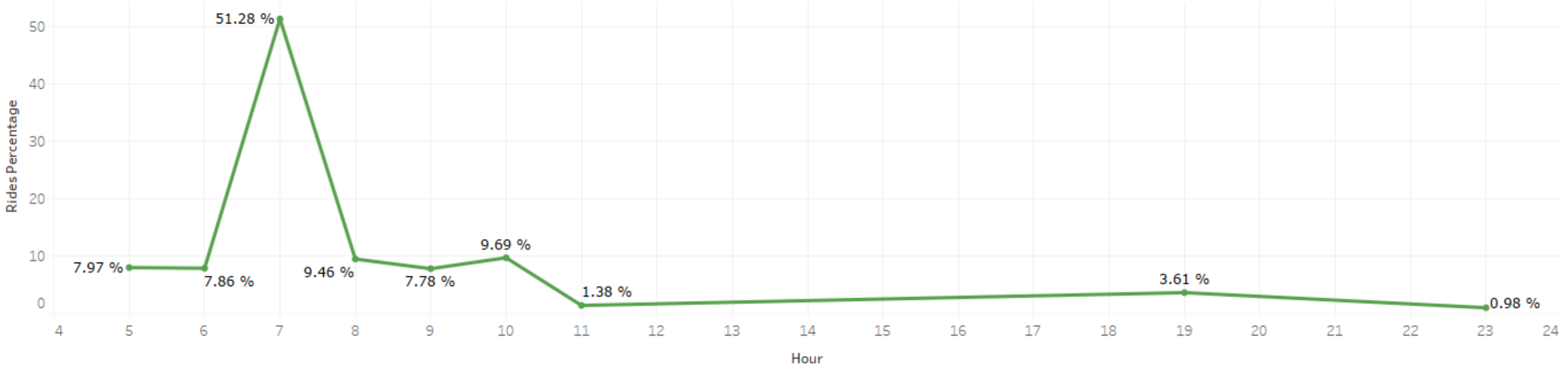
Rides through the month.



Rides on Days of the week.



Rides through the Day.



Modelling

- **We performed several models and we got the following corresponding Mean Absolute Errors ;**

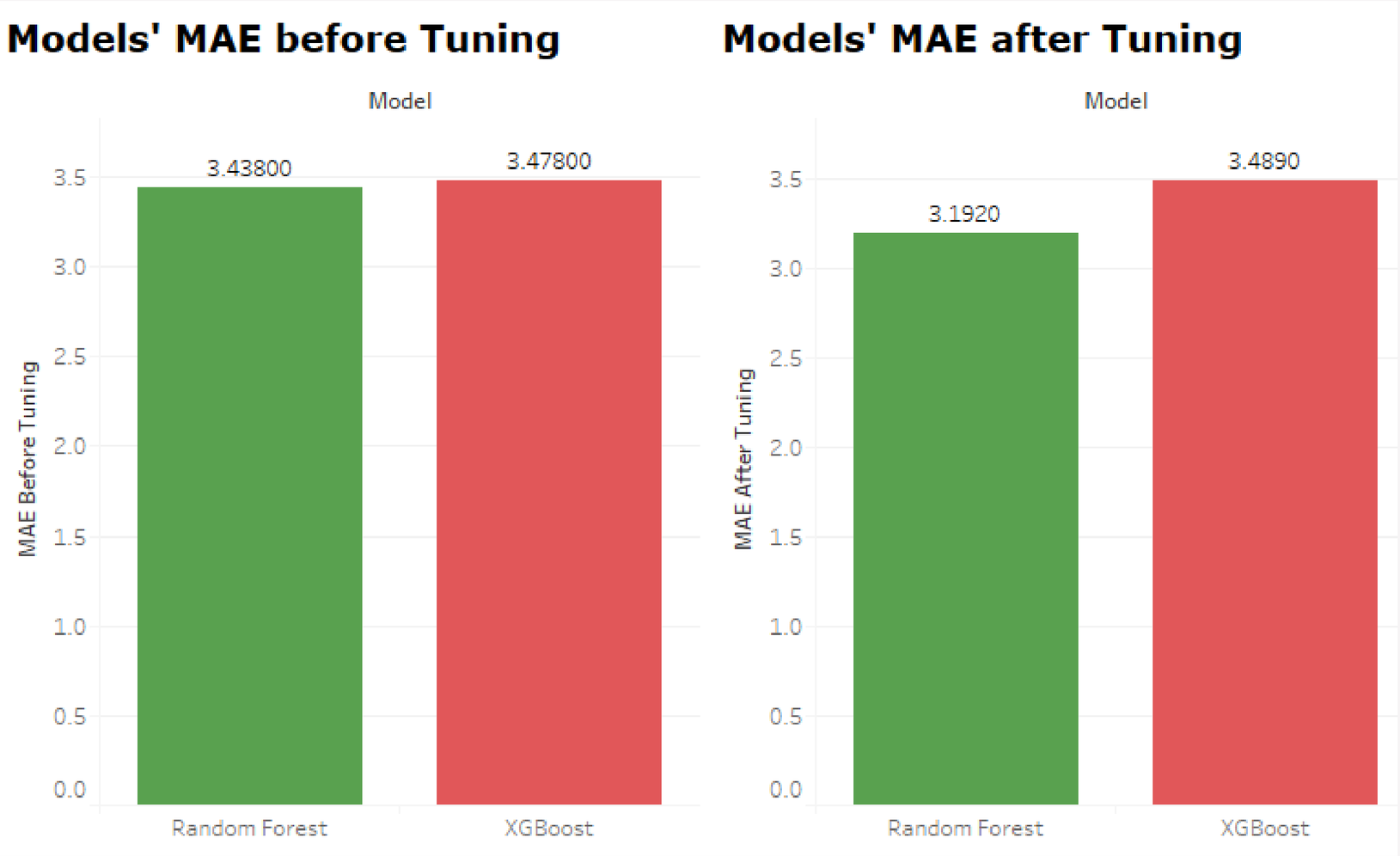
- **Linear Regression (Baseline Model) - 5.33**
- **XGBoost - 3.478**
- **Random Forest Regressor - 3.438**

Hyperparameter Tuning

- **Hyperparameter tuning was performed on XGBoost and Random Forest and only Random Forest's marginal absolute error improved**

- **XGBoost - 3.489**
- **Random Forest Regressor - 3.192**

The graph shows the models' perfomance before and after the tuning was done.



Conclusions

- **The top four important features : Origin, Month, Day of the week and hour**
- **The tuned Random Forest with the MAE of 3.192 was great for use.**

Challenges

- **Hyper-Parameter Tuning.**

Future Plans

- **Source for more data that shows arrival times in order to get a picture of travel times within different time periods.**

Recommendations

- **The optimal model can be used by Mobiticket and bus operators to anticipate customer demand for certain rides.**

- **This will help manage resources and vehicles more efficiently, as well as offering promotions and providing other services.**

- **For areas like Kisii, Mobiticket should advice the partner bus operators to avail more buses and few shuttles to areas like Sori and Oyugis**

Any questions?