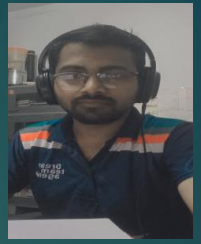
Capstone Project-2

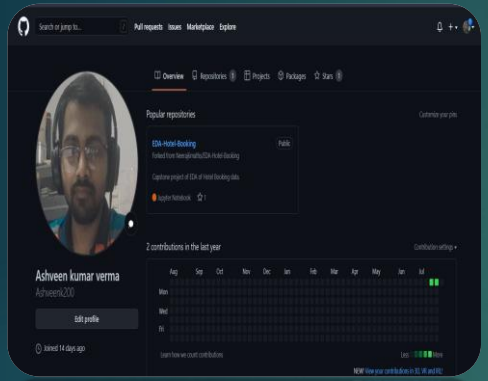
Demand Prediction for Public Transport

by

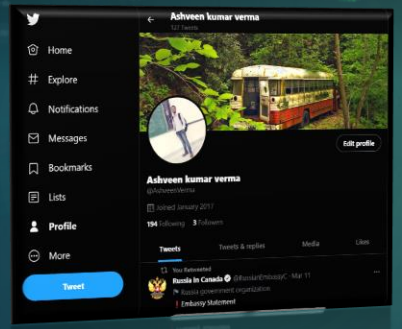
Ashveen Kumar Verma
(Self)

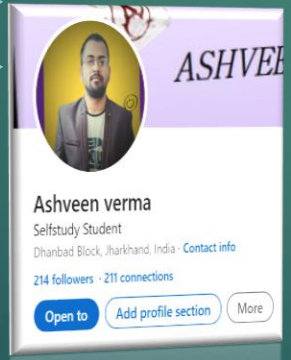
ASHVEEN KUMAR VERMA





- Live in Dhanbad Jharkhand
- Bsc graduate from pk roy memorial collage Dhanbad in 2020.
- Enrolled in DS program.
- This is my 2nd capstone project
- Linkedin- Ashveen verma
- Github- Ashveen kumar verma
- Twitter- Ashveen kumar verma





Project: Traffic Jam: Predicting People's Movement into Nairobi

AI

Problem Description

- 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.

Work Flow:-

- 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

- After collecting data it's important to understand your data . So we had 51645 rows and 10 columns. So lets understand this 10 columns.

ride_id

seat_number

payment_method

payment_receipt

travel_date

travel_time

travel_from

travel_to

car_type

max_capacity

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
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.

Data Description

Nairobi Transport Data.csv (zipped) is the dataset of tickets purchased from Mobiticket for the 14 routes from "up country" into Nairobi between 17 October 2017 and 20 April 2018. This dataset includes the variables: ride_id, seat_number, payment_method, payment_receipt, travel_date, travel_time, travel_from, travel_to, car_type, max_capacity.

Data is available for Nairobi through June 2018. Uber Movement provided historic hourly travel time between any two points in Nairobi. Any tables that are extracted from the Uber Movement platform can be used in your model.

Variables description:

ride_id: unique ID of a vehicle on a specific route on a specific day and time.

seat_number: seat assigned to ticket

payment_method: method used by customer to purchase ticket from Mobiticket (cash or Mpesa)

payment_receipt: unique id number for ticket purchased from Mobiticket

travel_date: date of ride departure. (MM/DD/YYYY)

travel_time: scheduled departure time of ride. Rides generally depart on time. (hh:mm)

travel_from: town from which ride originated

travel_to: destination of ride. All rides are to Nairobi.

car_type: vehicle type (shuttle or bus)

max_capacity: number of seats on the vehicle

Data cleaning and Manupulation :-

```
df.head()
```

	ride_id	seat_number	payment_method	payment_receipt	travel_date	travel_time	travel_from	travel_to	car_type	max_capacity
0	1442	15A	Mpesa	UZUEHCBUSO	17-10-17	7:15	Migori	Nairobi	Bus	49
1	5437	14A	Mpesa	TIHLBUSGTE	19-11-17	7:12	Migori	Nairobi	Bus	49
2	5710	8B	Mpesa	EQX8Q5G19O	26-11-17	7:05	Keroka	Nairobi	Bus	49
3	5777	19A	Mpesa	SGP18CL0ME	27-11-17	7:10	Homa Bay	Nairobi	Bus	49
4	5778	11A	Mpesa	BM97HFRGL9	27-11-17	7:12	Migori	Nairobi	Bus	49

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51645 entries, 0 to 51644
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   ride_id          51645 non-null  int64 
 1   seat_number      51645 non-null  object
 2   payment_method   51645 non-null  object
 3   payment_receipt  51645 non-null  object
 4   travel_date      51645 non-null  object
 5   travel_time      51645 non-null  object
 6   travel_from      51645 non-null  object
 7   travel_to        51645 non-null  object
 8   car_type         51645 non-null  object
 9   max_capacity     51645 non-null  int64 
dtypes: int64(2), object(8)
memory usage: 3.9+ MB
```

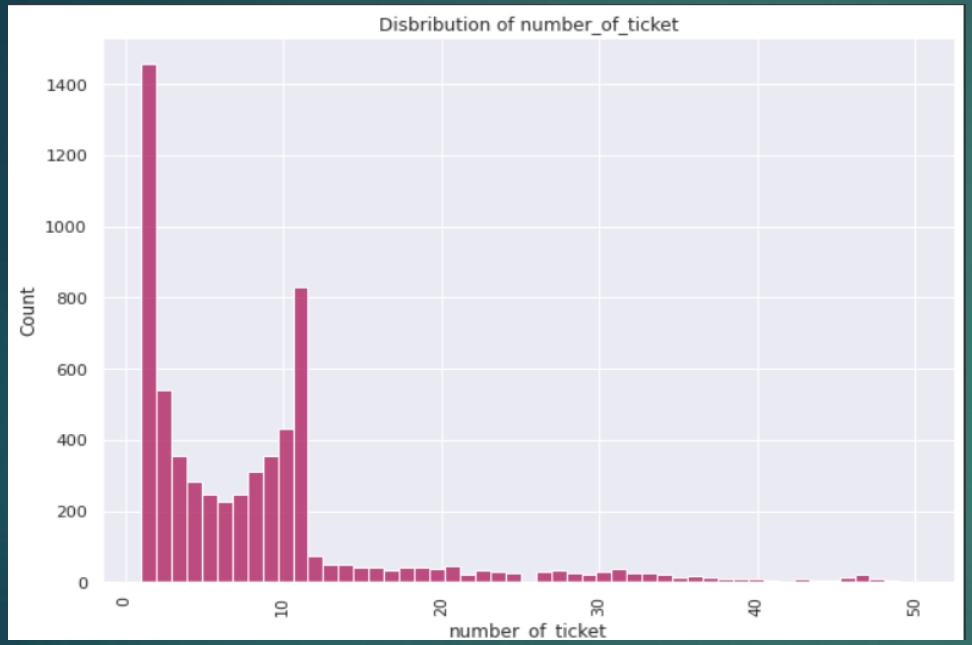
```
[ ] df.describe()
```

	ride_id	max_capacity
count	51645.000000	51645.000000
mean	10188.645793	34.534321
std	2211.295708	18.451193
min	1442.000000	11.000000
25%	8287.000000	11.000000
50%	10353.000000	49.000000
75%	12029.000000	49.000000
max	20117.000000	49.000000

```
] df.shape
```

```
(51645, 10)
```

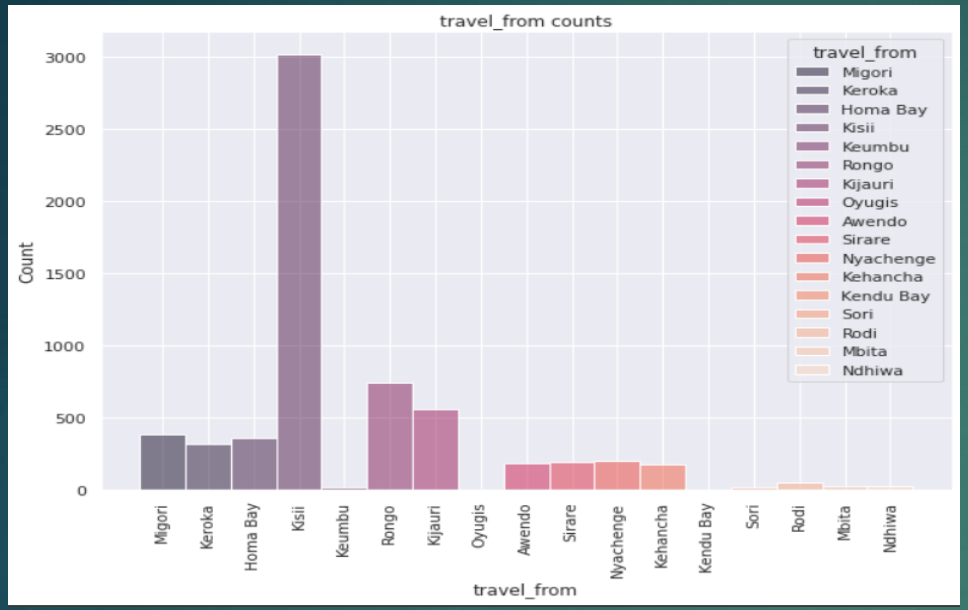
Exploratory Data Analysis(EDA)



Here we can se that the maximum ticket sold under 10

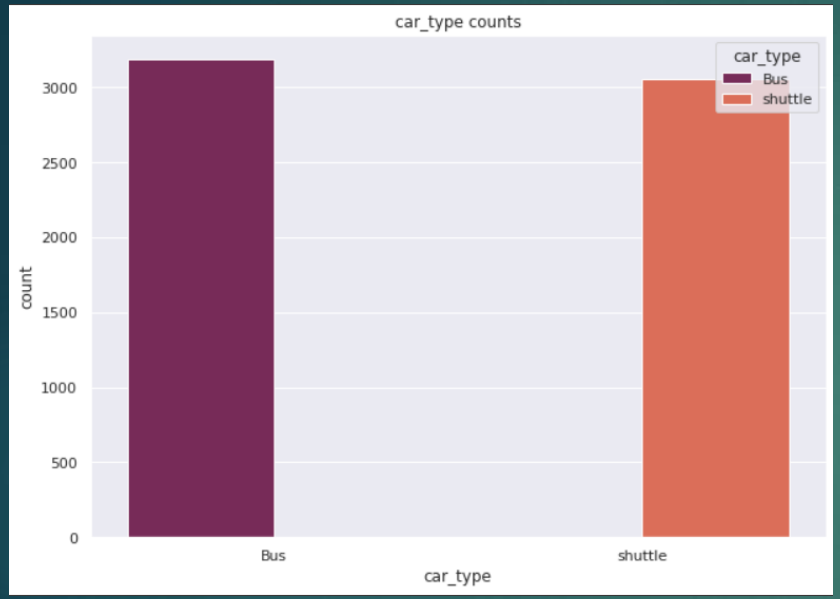
We also say from this graph that maximum person prefer to travel alone.

Exploratory Data Analysis(EDA)



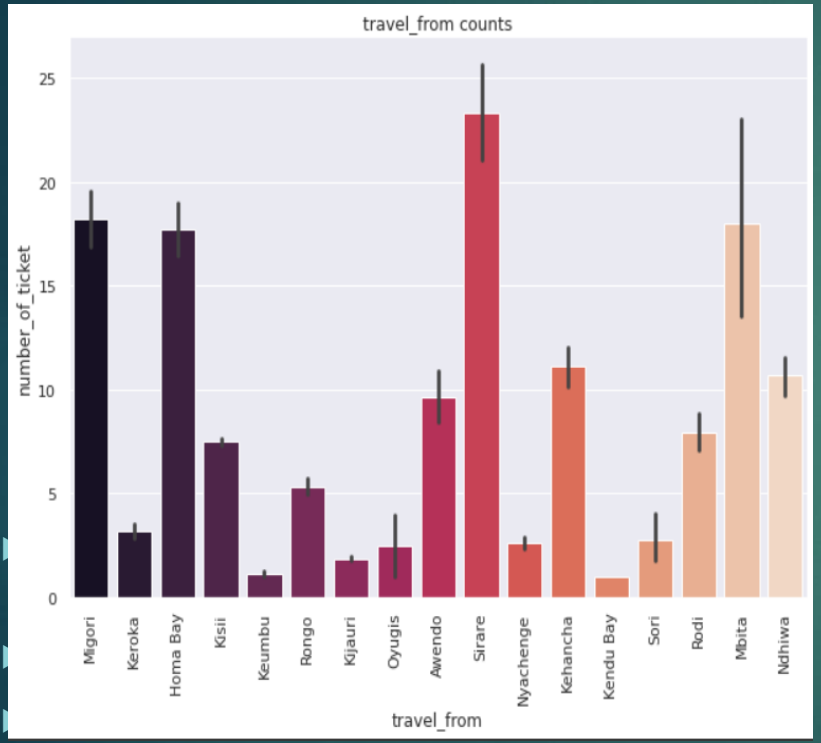
- Conclusion:-
- The graph shows kisii has maximum number of bookings and ndhiwa have lowest number of bookings for travel
- Here we also conclude that kisii may be a tourist hotspot of very religious place.

Exploratory Data Analysis(EDA)



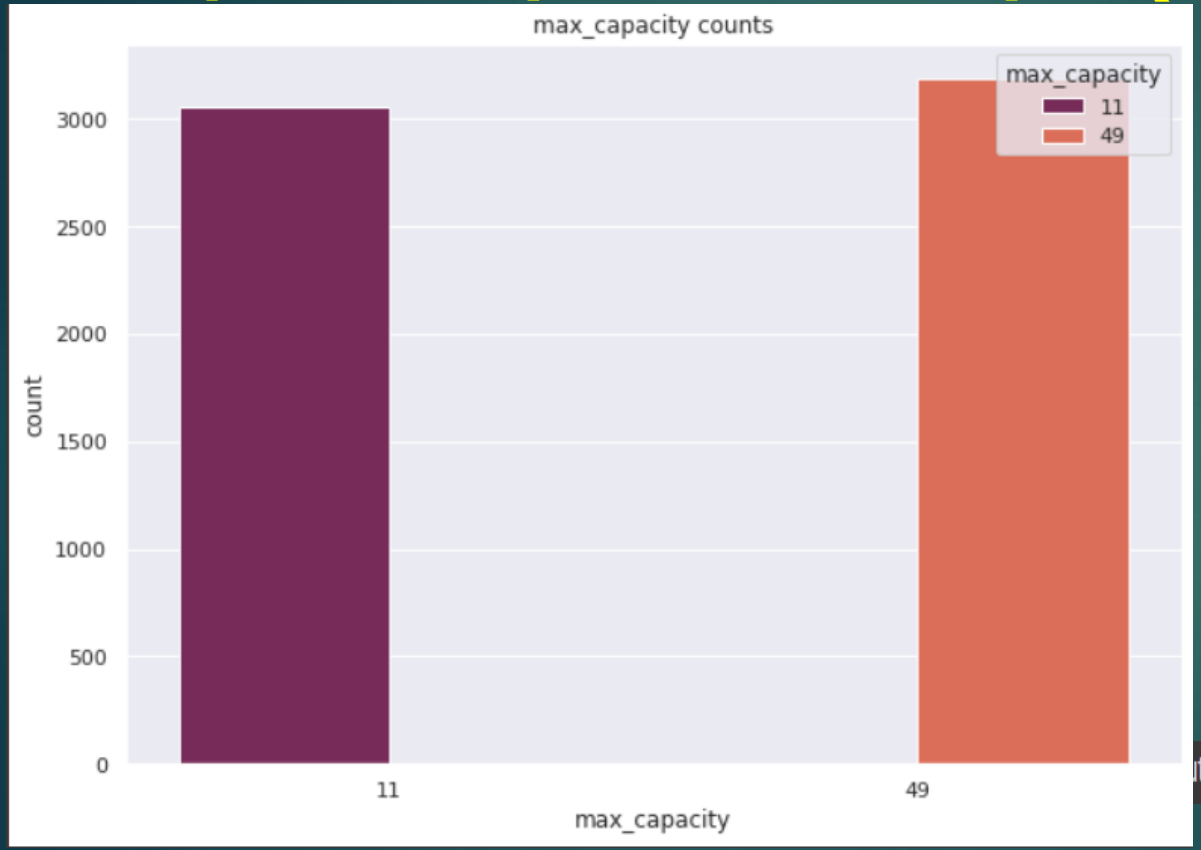
- Conclusion
- From graph we say that passengers travels from bus and shuttle rather than train.
- Buses value is quite higher than shuttle.

Exploratory Data Analysis(EDA)



Most of the people in sirare and keumbu and kendu bay has least travel_from_count.

Exploratory Data Analysis(EDA)



- Conclusion
- From graph we can say that the maximum capacity is of 11 and 49 is 3000 and 3500 approximately.

t for sudden visits other mediums are most preferred.

Model building and pridection :

AI

```
[ ]  #separate table in x and y

     x= df[['max_capacity','number_of_ticket']]
     y= df['car_type']
```

```
     ##training module

     from sklearn.model_selection import train_test_split
     x_train, x_test,y_train,y_test = train_test_split(x,y, test_size = 0.2)
```

```
[ ]  from sklearn.svm import SVC
     model_svc = SVC()
     model_svc.fit(x_train,y_train)
```

```
     SVC()
```

```
[ ]  prediction1 = model_svc.predict(x_test)
     #predictiong the accuracy


     from sklearn.metrics import accuracy_score
     print (accuracy_score(y_test , prediction1)*100)
```

```
     100.0
```

```
[ ]  #model 1

     #logestic regression

     from sklearn.linear_model import LogisticRegression
     model_LR=LogisticRegression()
     model_LR.fit(x_train,y_train)
```

```
     LogisticRegression()
```

```
[ ]  #model 2

     #prediction the accuracy
     prediction2 = model_LR.predict(x_test)
     from sklearn.metrics import accuracy_score
     print(accuracy_score(y_test,prediction2)*100)
```

```
     100.0
```

Conclusion

In this project, we have used regression-based algorithms like linear regression and we found the important features for training the model. In the linear regression model Regressor algorithm gives the best results with an accuracy of around 100%.

AI