Capstone Project

**Transport Demand prediction**

by

**Ghanshyam Menariya**

--------------------------------------------\*\*\*------------------------------------------------------------

**Ghanshyam Menariya**

**G-Mail:** [ghanshyammenariya50@gmail.com](mailto:ghanshyammenariya50@gmail.com)

Contact no : 7689965076

**GitHub Link~~**

**Ghanshyam Menariya-** https://github.com/ganny55/play-store-review.git

# Problem Statement

# 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 Victo

**Transport Demand Prediction**

# 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.

### T**he 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:**

1. **Kawangware: the first stop in the outskirts of Nairobi**
2. **Westlands**
3. **Afya Centre: the main bus terminal where most passengers disembark**

**All of points are maped**

### 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.

### Uber Movement traffic data can be accessed [here](https://movement.uber.com). 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 Preparation

Pre-processing is important into transitioning raw data into a more desirable format. Undergoing the preprocessing process can help with completeness and compellability. For instance, you'll see if certain values were recorded or not. Also, you'll see how trustable the info is. It could also help with finding how consistent the values are. We need preprocessing because most real-world data are dirty. Data can be noisy i.e. the data can contain outliers or simply errors generally. Data can also be incomplete i.e. there can be some missing values.

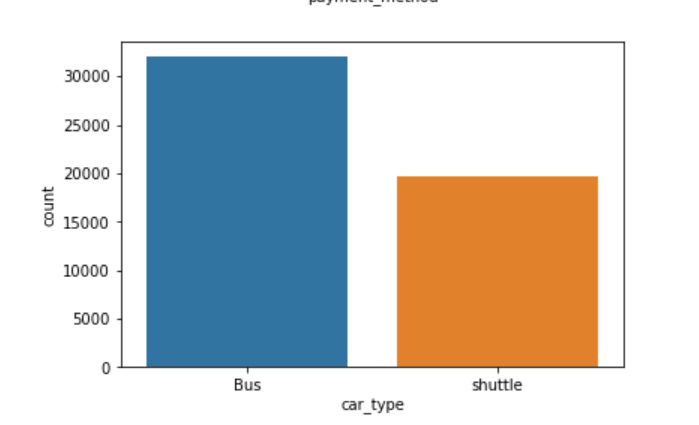
The available data is raw and unusable for Exploratory data analysis, so before we do anything with the data we will have to explore and clean it to prepare it for data analysis

# EXPLORATORY DATA ANALYSIS

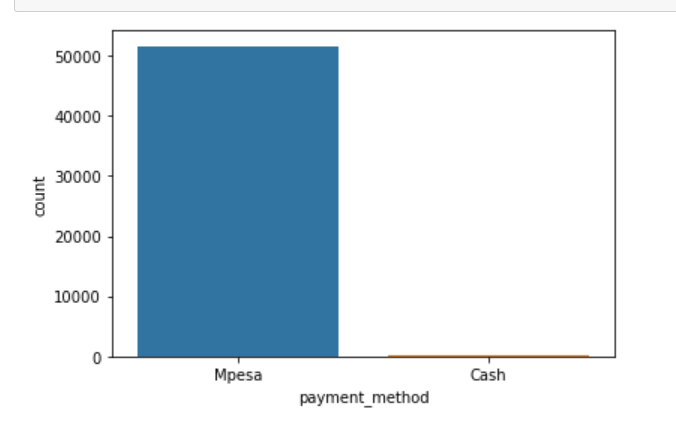
Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset.

EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better. In this article, we will understand EDA with the help of an example dataset. We will use **Python** language (**Pandas** library) for this purpose.

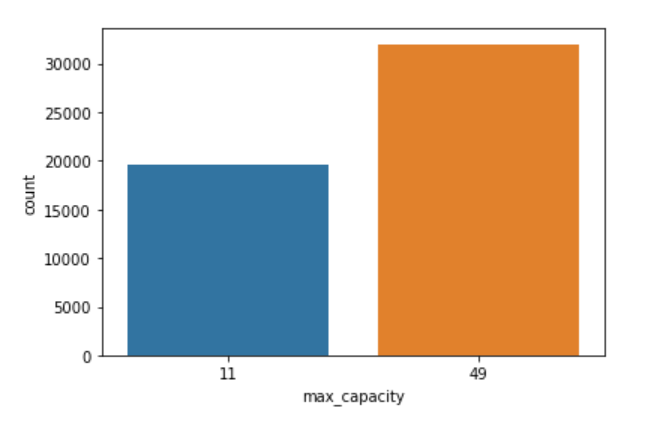
Count the Vehicle type



Count the payment method



Count the Capacity of Vehicle

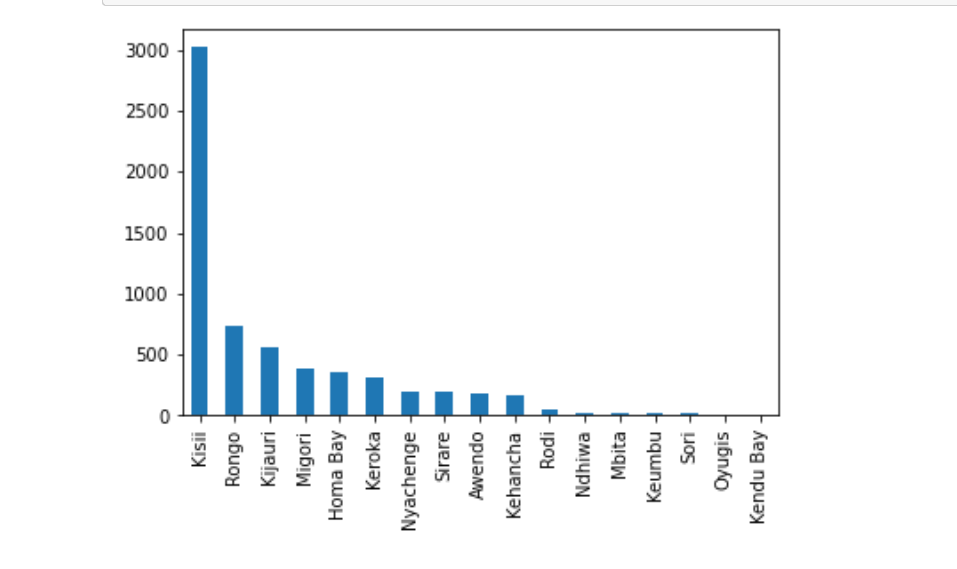


## Summary

There are two type of payment methods people have used to buy the tickets.

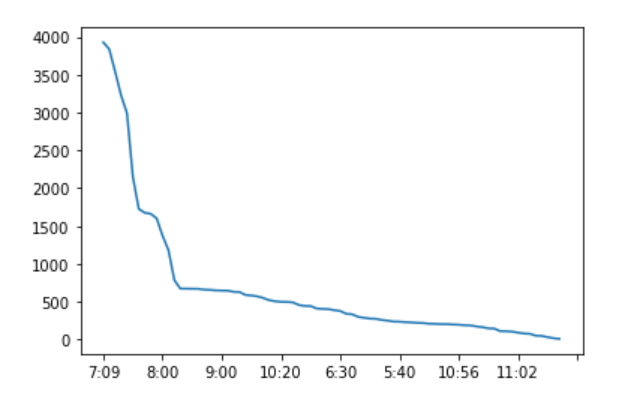
There are two type of cars Bus and shuttle and the maximum capacity of the bus is 49 while shuttle can contain 11 travellers

Count people travel location



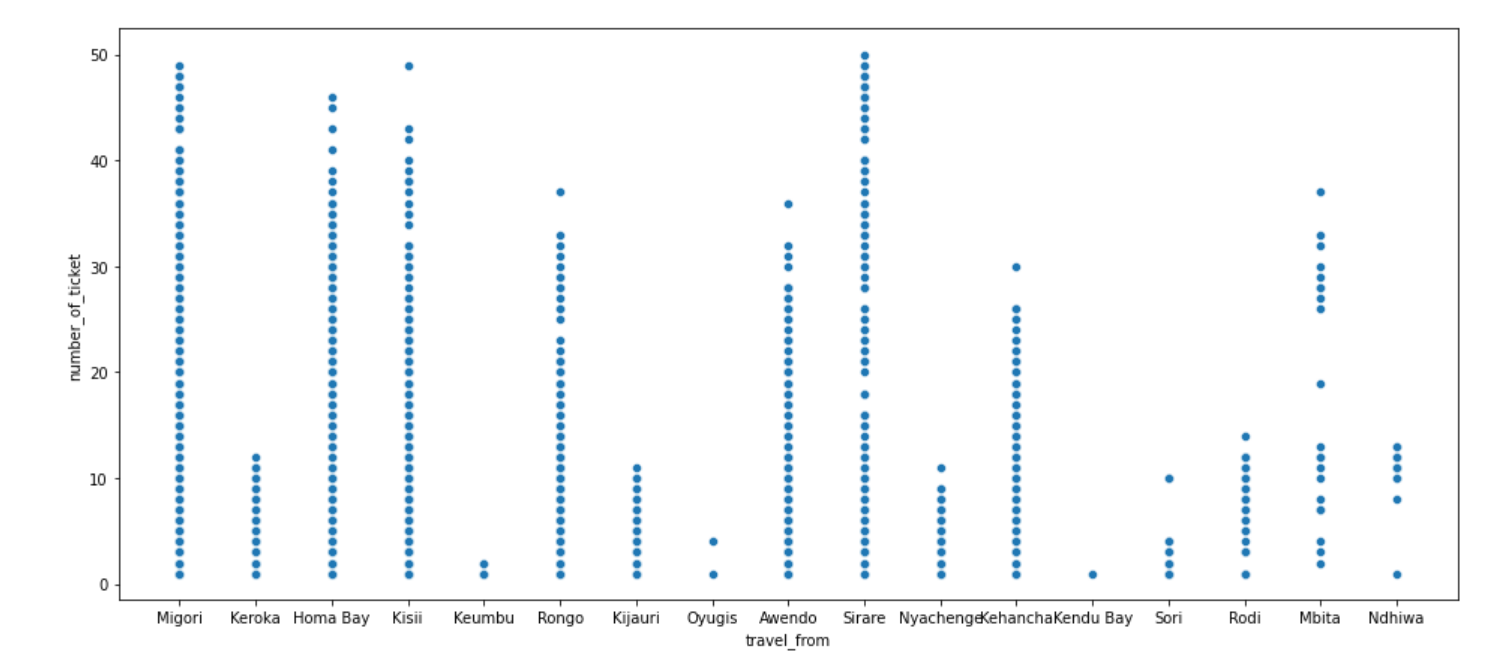
According to this bar plot we can see the most of people travel from kisii and rongo. but something 3000 people travel from kisii .

Count travellers travel time



By this chart we can see the most of people travel between 7’o clock to 8 o clock. We assume this is a office time of people .

Count the number of tickets on traveling location:



Migori , homa bay , srare ,kissi that’s the place where most number of ticket is sell .