

TAXI HOTSPOT PREDICTION MODEL USING PYTHON 3

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering Degree in Computer Science and Engineering

By

PARIKSHITH PRABAKAR (Reg. No. 38290061)

RATHISH.E (Reg. No. 38290069)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING**

**SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600119, TAMILNADU**

MARCH 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by

NAAC



(Established under Section 3 of UGC Act, 1956)

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI- 600119

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **PARIKSHITH PRABAKAR (Reg. No. 38290061)** and **RATHISH.E (Reg. No. 38290069)** who carried out the project entitled "**TAXI HOTSPOT PREDICTION MODEL**" under our supervision from November 2020 to March 2021.

Internal Guide

Dr. J . ALBERT MAYAN M.E., Ph.D.,

Head of the Department

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I, **PARIKSHITH PRABAKAR (Reg. No. 38290061)** hereby declare that the Project Report entitled "**TAXI HOTSPOT PREDICTION MODEL USING PYTHON 3**" **done by me under the guidance of Dr. J.ALBERT MAYAN, M.E.,Ph.D.**, is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. SASIKALA, M.E., Ph.D.**, Dean, School of Computing and **Dr. S. VIGNESHWARI, M.E., Ph.D.**, and **Dr. L. LAKSHMANAN, M.E., Ph.D.**, Head of the Department, Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. J. ALBERT MAYAN, M.E., Ph.D.**, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Computer Science and Engineering who were helpful in many ways For the Completion of the project.

ABSTRACT

We live in the age of fast-growing technology where nowadays people find it easier to book transportation such as cabs on an application to move from Point A to Point B. Be it 5km or 50km, applications such Ola and Uber seem to be the Holy Grail for those in need of transportation. For anyone a cab is just a click away. But do we truly understand how the applications and services work? Do we truly know the availability of cabs in and around? Are the cab rates truly justified? These doubts can be cleared by simply visualizing the scenarios using the TAXI HOTSPOT PREDICTION MODEL.

This model will aid people by collecting raw data from Uber / Ola datasets and visualizing the data in the form of pie charts and graphs. People will easily be able to understand the changes in rates due to demand, the availability and unavailability of cabs. This will make it easier for people to always be prepared, either for booking a cab and preparing other means of transportation.

All one has to do is to access the TAXI HOTSPOT PREDICTION MODEL_in the Internet or Google Maps and choose to visualize from 3 available categories;

1. The trip status Pie Chart
2. The pickup point Hotspots
3. Total Count of All trip Status in 5 time slots

TABLE OF CONTENTS

Chapter No.	TITLE	Page No.
--------------------	--------------	-----------------

ABSTRACT

1	INTRODUCTION - OVERVIEW OF PROJECT	
----------	--	--

2	AIM AND SCOPE 2.1 Aim of the Project 2.2 Scope and Objective 2.3 System Requirements 2.4 Overview of the project - Problem Statement - Advantages - Modules - Risks / Drawbacks 2.5 Software Used - Jupyter Notebook	
----------	---	--

3	METHODS 3.1 Design Methodology	
----------	--	--

- Existing System
- Proposed System

3.2 Module Description

3.3 Architecture

4 RESULTS

5 CONCLUSIONS

SOURCE CODE

CHAPTER_1

INTRODUCTION

OVERVIEW OF PROJECT

In the 21st century, technology is evolving at an unprecedented rate. The world is come to a state wherein people no longer just merely walk or cycle from Point A to Point B, rather most of us nowadays make us of the facility of being able to use applications such as Uber and Ola to book a means of transportation. Uber and Ola have garnered a handful of customers over the years, but how much do we know actually know about how these services actually work?

Uber and Ola may give the facility to book cabs, but since they don't cover a certain perimeter pertaining to the pickup location, we can never be sure of being able to book a cab with a short waiting time. Not to mention the cancellation fees we may be charged if we would rather not wait for a long period of time. These apps also tend to attract heavy user traffic, especially at peak times of the day, which tend to increase the rates of cabs when supply is unable to meet demands.

The Taxi Hotspot Prediction Model_will enable people to understand and view the various zones in which the taxi count is higher / lower. Once the taxi dataset is loaded into the model, the code can be used to read the dataset file and visualize the input data in the form of pie charts and graphs. The code we used will be run in the Jupyter Notebook (Python 3)

The Data visualization will cover status of Trip counts over 5 periods of time in a day, the pickup point hotspots, the zones covered and a trip status pie chart.

The Model and its various divisions will be able to view as an extension of Google Maps and on the Internet.

Chapter 2

Aim and Scope

2.1 – Aim of the Project

Analysing the Taxi Hotspot zones and the effect of user application traffic on the rates of available taxis is the main aim of this Prediction Model. This model will be made available as an extension in the internet and Google maps. This will enable people to be prepared for any possible scenario, be it preparing to shed a little extra money for transportation to preparing a another means of transport.

2.2 – Scope and Objective

The purpose of this model is to determine the accuracy of being able to pinpoint the geographical zones where taxis are the most and least found.

To identify how the availability of taxis in said geographical zones and the unprecedented increase in rates due to user traffic are co-related.

To identify how people might understand the graphs and statistics produced by the prediction model.

To identify how one plots his/her next move based on the information gathered from the prediction model.

2.3 – System Requirements

- i) Hardware Requirements:
 - RAM: 4 – 8 GB
 - ROM: 64 GB and above
 - Processor: Intel i3 and above

- ii) Software Requirements
 - Windows / MAC
 - Visual code
 - Anaconda 3
 - Jupyter Notebook

2.4 – Overview of the project

. The existing problems:

1. Lack of geographical accuracy (geotag)
2. Unprecedented increase in rates due to untimely high demand
3. Availability of more conditional on time rather than location
4. App interactivity affecting the demand and supply chain

. To overcome these drawbacks, Taxi Hotspot prediction model concept has been approached.

Tackling the problem:

. The Taxi demand and supply chain keeps changing every day.

. But as demand increases for taxis, the supply does not necessarily increase, rather the fare increases the match the hire count of taxis.

. Due to this the increase in fare seems to be in a way linked to the app interactivity and refresh rate.

. This is why we feel the best way to tackle this would be to reduce the traffic garnered by the app and provide an alternative to take note of the taxi availability and hotspots.

CHAPTER 3

Methods

3.1 – Proposed System:

1.Input:

. The input of this project is Taxi hotspot prediction dataset, namely a csv file.

2.Output:

. The output of this project is to display the predicted taxi hotspots – with regard to the Zone, Hour Slot and Hire Count

3.Method:

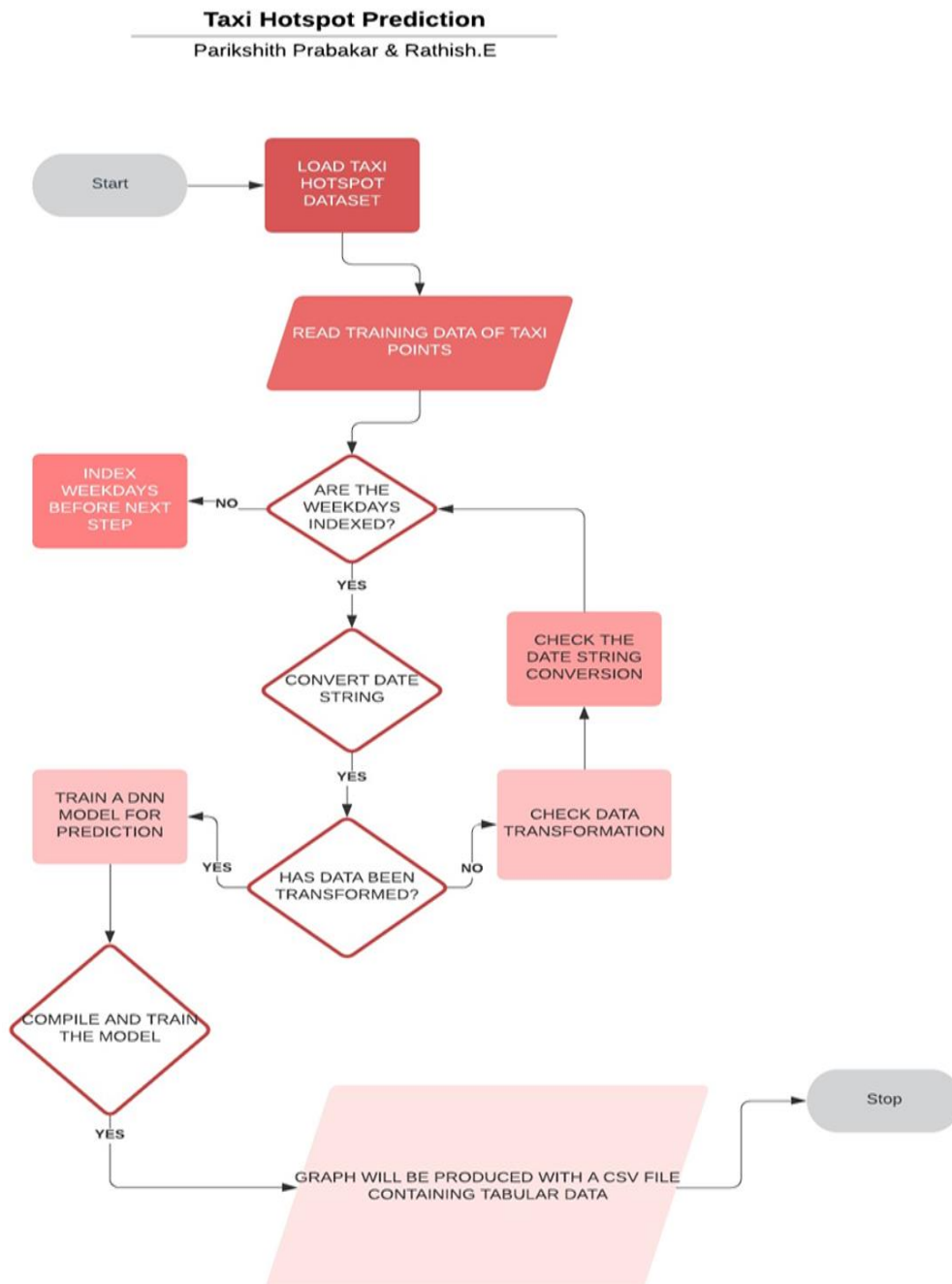
. By using the Jupyter Notebook, we can read the dataset and come to 3 conclusions:

1. The location of training zones
2. The zone with the higher taxi count
3. How app interactivity affects demand and supply

3.2 – Modules used:

- a. Taxi Hotspot Dataset Preparation and Preprocessing through Google Colab IDE (i.e preprocessing of the datasets to remove null values)
- b. Data transformation and conversion into 5 different time slots / zones(for efficiency and accuracy)
- c. Hotspot zone identification and classification (i.e predicting the hire count and hour slot in each zone)

3.3 – Architecture:



Architecture of the work flow of the Taxi Hotspot Prediction Model

CHAPTER 4

Results

4.1 :

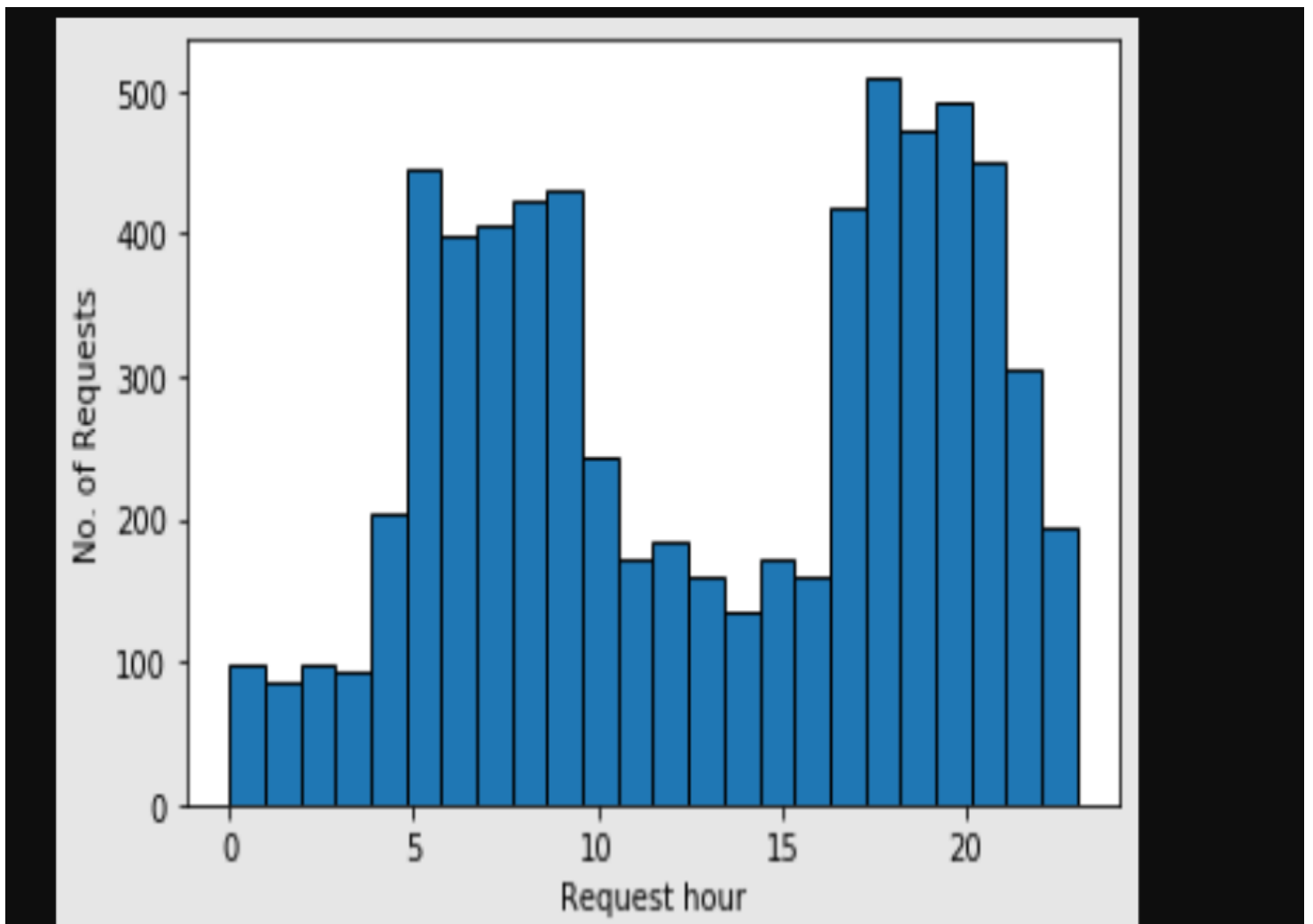


Fig: Request Hour VS Number of Requests

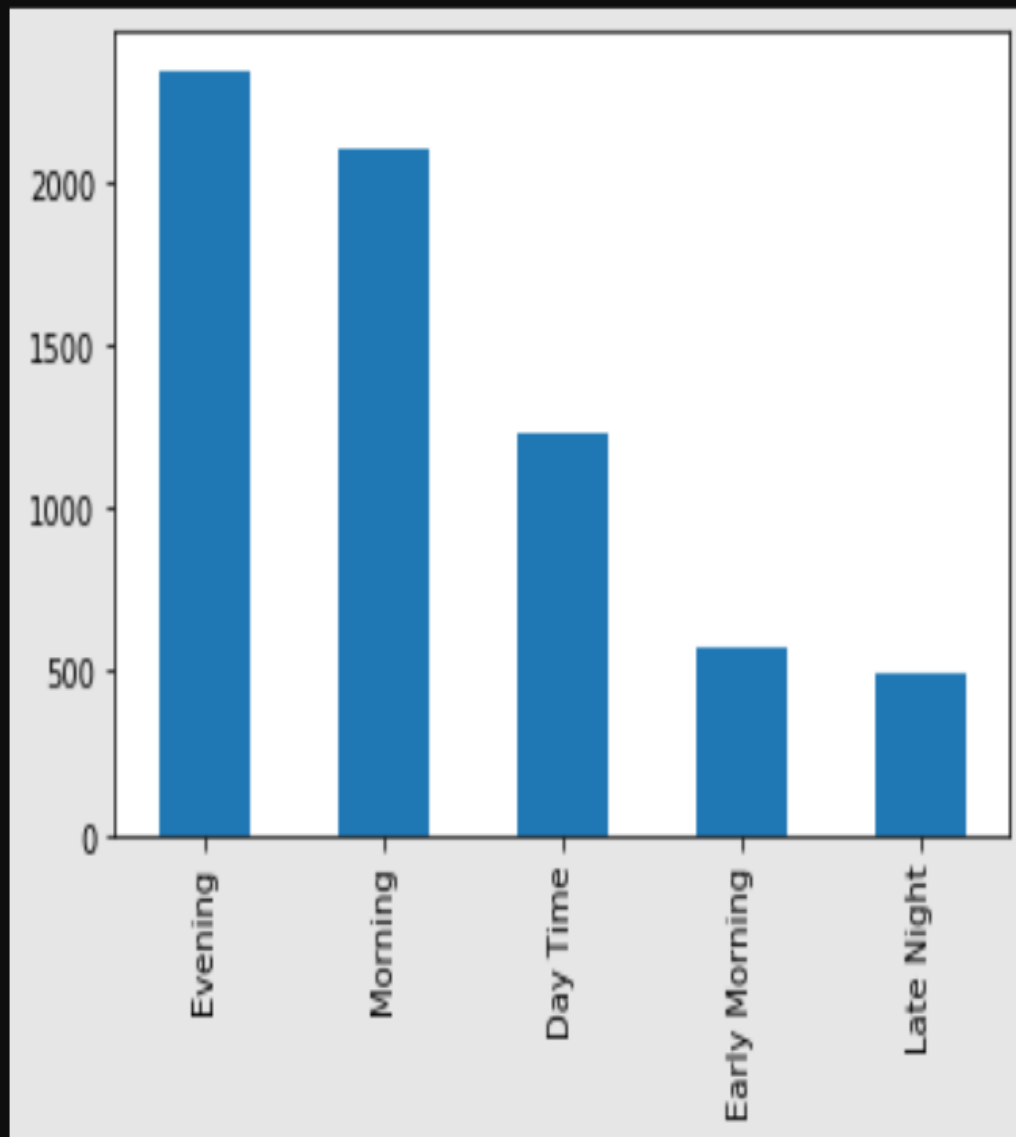


Fig: Number of Trips / Booking per slot

4.3 –

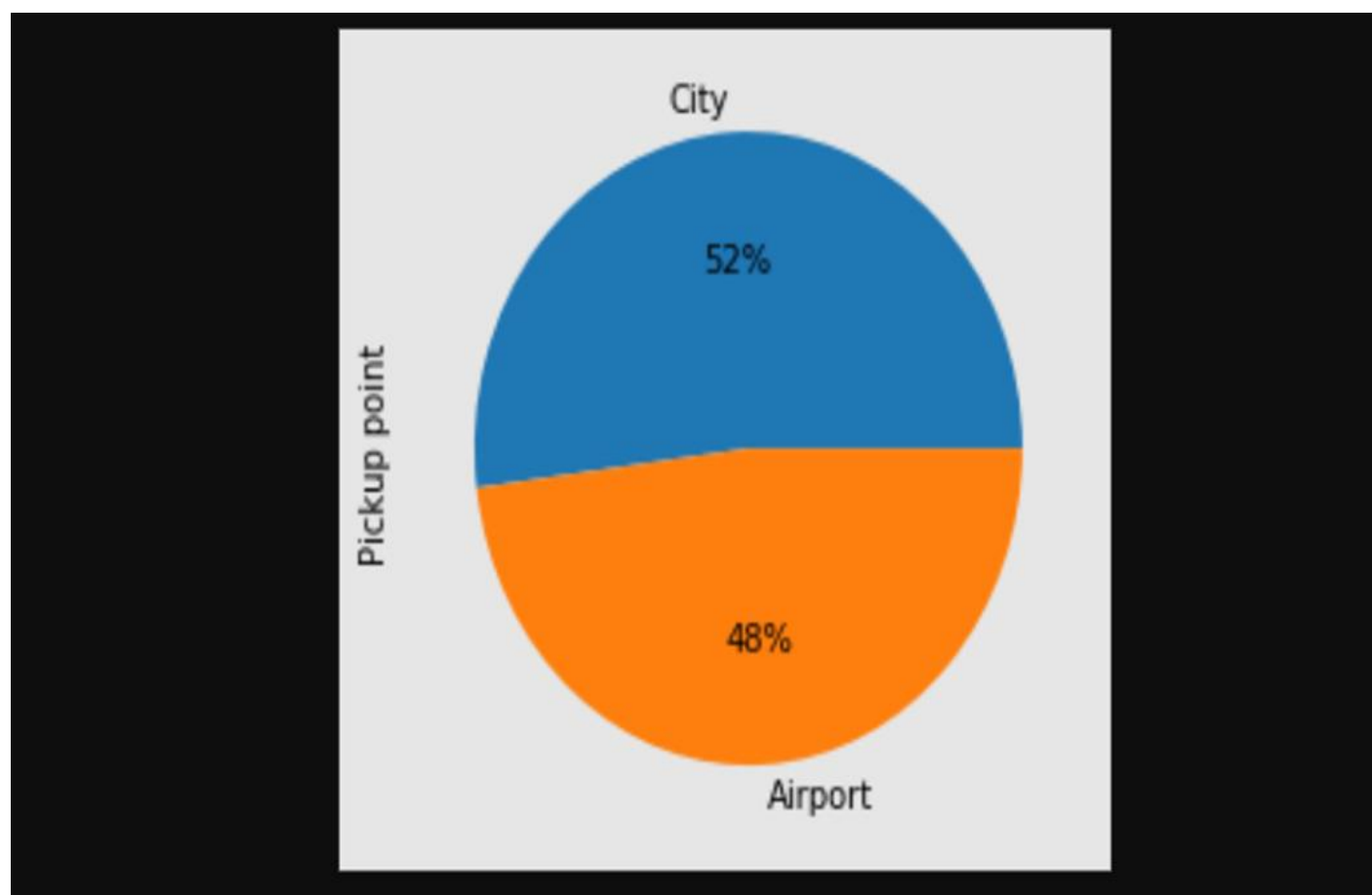


Fig: Pickup Points

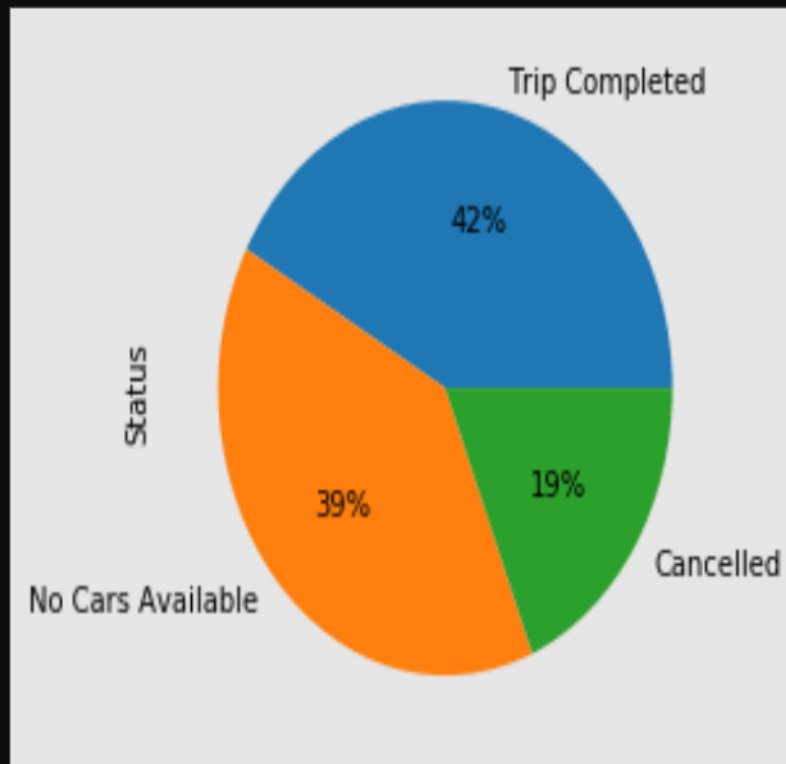


Fig: Trip Status

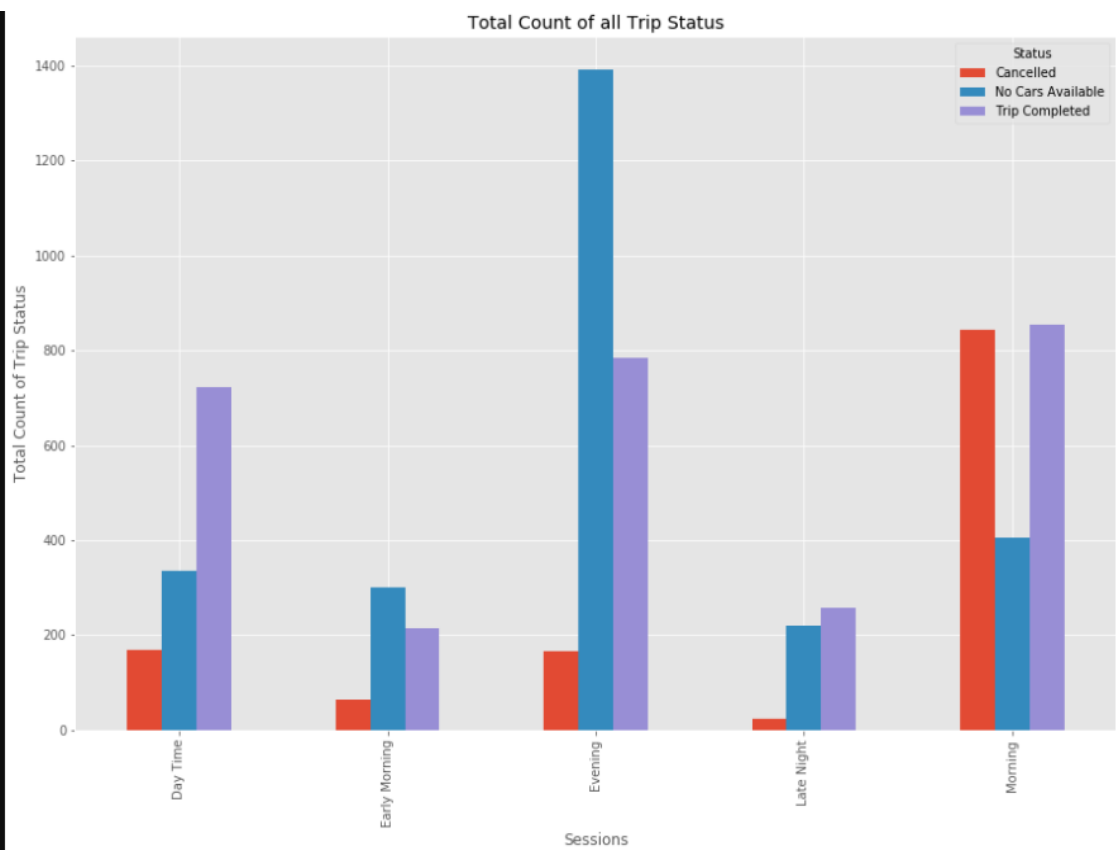


Fig: Total Count of All Trip Status



Fig: The Correlation Matrix

Chapter 5

Conclusion

. To overcome the drawbacks in taxi hotspot prediction, I propose to use the Jupyter Notebook in the modified prediction process.

. Project Outcome:

1. To chalk out the zones with a greater hire count
2. To provide a clear picture and understanding of app user traffic
3. To control, to an extent, control the demand and supply mismatch problem
4. To reduce the cancellation rate of taxis

SOURCE CODE

1)

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from mpl_toolkits.mplot3d import Axes3D
from sklearn.preprocessing import StandardScaler
```

2)

```
uber_data = pd.read_csv('Uber Request Data.csv')
print(uber_data.shape)
```

3)

```
print(uber_data.columns)
```

4)

```
uber_data.head()
```

5)

```
uber_data.info()
```

6)

```
# Converting Request_timestamp and drop_timestamp to uniform datetime format
```

```
uber_data["Request_timestamp"] = uber_data["Request_timestamp"].apply(lambda x:
pd.to_datetime(x))
```

```
uber_data["Drop_timestamp"] = uber_data["Drop_timestamp"].apply(lambda x:
pd.to_datetime(x))
```

7)

```
uber_data.head()
```

8)

```
#Check for null values
```

```
uber_data.isnull().sum()
```

9)

```
uber_data.Status.value_counts()  
10)
```

```
#Extract the hour from requested timestamp
```

```
uber_data["Request hour"] = uber_data["Request timestamp"].dt.hour  
uber_data.head(5)
```

```
11)  
plt.hist(uber_data["Request hour"],edgecolor='black',bins=24)  
plt.xlabel("Request hour")  
plt.ylabel("No. of Requests")  
plt.show()
```

```
12)  
# Divide the time of the day into five categories
```

```
def time_period(x):  
    if x < 5:  
        return "Early Morning"  
    elif 5 <= x < 10:  
        return "Morning"  
    elif 10 <= x < 17:  
        return "Day Time"  
    elif 17 <= x < 22:  
        return "Evening"  
    else:  
        return "Late Night"
```

```
13)  
uber_data["Time Slot"] = uber_data["Request hour"].apply(lambda x: time_period(x))  
uber_data.head()
```

```
14)  
uber_data["Time Slot"].value_counts().plot.bar()  
plt.show()
```

```
15)  
uber_data["Pickup point"].value_counts().plot.pie(autopct='%1.0f%%')  
plt.show()
```

```
16)  
uber_data["Status"].value_counts().plot.pie(autopct='%1.0f%%')  
plt.show()
```

```
17)
plt.style.use('ggplot')
uber_data.groupby(['Time Slot' , 'Status']).Status.count().unstack().plot.bar(legend = True,
figsize=(15,10))
plt.title('Total Count of all Trip Status')
plt.xlabel('Sessions')
plt.ylabel('Total Count of Trip Status')
plt.show()
```