UBER

Data Analysis of Uber

Vaishali Kadam

© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Problem Statement

- Uber has been a major source of travel for people living in urban areas. Some people don't have their vehicles while some don't drive their vehicles intentionally because of their busy schedule. So different kinds of people are using the services of Uber and other taxi services.
- Uber is an app-based transportation network and taxi company. In its Airport rides in a particular city, many of its users facethe problem of cancellation by the driver or non-availability of cars.
- These very issues impact the business of Uber and it loses out on its revenue
- Pre Analysis: The problem of less number of pickups from the airport has two aspects to it. Uber is an aggregator of **demand and supply**. Demand-side is the riders who use the app to book cabs and supply side is the cab drivers registered with Uber. To form our solution, we'll be focusing on both the demand and supply side of the business activity.
- The objective is to first explore hidden or previously unknown information by applying exploratory data analytics on the dataset and to know the effect of each field on price with every other field of the dataset.
- By analyzing Uber trips, we can draw many patterns like which day has the highest and the lowest trips or the busiest hour for Uber and many other patterns. The dataset I'm using here is based on Uber trips from Specific Areas, a area with a very complex transportation system with a large residential community.
- You can do so much more with this dataset rather than just analyzing it. But for now, in the project, I have done the Uber Trips analysis using Python.

© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Proposed Solution

Push notification when the user lands:

- Uber app can geo-locate the user on the landing runways of the airports and send them push notifications to book their Uber
- An average total time from flight landing to the user coming out of the airport can be guess-estimated and the cab driver can be allocated to the user
- The solution potentially decreases the overall waiting time with an increase in algorithmic complexity at Uber's technology layer

Pre-book After Drop off:

- Users can be given an option to pre-book their return after a successful drop-off
- Uber can better pre-plan the pickup based on their return flight data
- Uber can also track the flight status through PNR details
- The solution will decrease the expected waiting time for a user

Uber Parking Spots (with geotagging on the Uber app):

- Uber can plan Uber Parking spots where all Uber cabs will be parked. Users can be educated about this via the app itself. Through the app, both the user and the cab know a common spot to meet and it reduces inconvenience

Enhanced Locating Assistance:

- Users can have personalized locating assistance through the app. If the user and the cab are moving in a different direction than expected, the app can send a push notification. The app can also suggest a common spot to the user where the cab can arrive at the same time

© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Descriptive Analysis and loading of data

First we have to load all the Library which we are needed and then load the Data through csv file…

[1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df=pd.read_csv(r'C:\Users\LENOVO\OneDrive\Desktop\New folder (2)\Uber Request Data.csv'
df
```

	Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp
0	619	Airport	1.0	Trip Completed	11-07-2016 11:51	11-07-2016 13:00
1	867	Airport	1.0	Trip Completed	11-07-2016 17:57	11-07-2016 18:47
2	1807	City	1.0	Trip Completed	12-07-2016 09:17	12-07-2016 09:58
3	2532	Airport	1.0	Trip Completed	12-07-2016 21:08	12-07-2016 22:03
4	3112	City	1.0	Trip Completed	13-07-2016 08:33	13-07-2016 09:25
...	...	...	...	...	...	...
6740	6745	City	NaN	No Cars Available	15-07-2016 23:49	NaN

Driver id,Drop timestamp seems to contain null values.

Hence any further analysis involving Driver id,Drop timestamp the NaN values has to be considered.

Checking if dataset contains any null

```
nan_values = df.isna()
nan_columns = nan_values.any()

columns_with_nan = df.columns[nan_columns].tolist()
print(columns_with_nan)
```

```
['Driver id', 'Drop timestamp']
```

```
df.isna()
nan_values.any()
```

```
Request id          False
Pickup point        False
Driver id            True
Status              False
Request timestamp   False
Drop timestamp       True
dtype: bool
```

```
Driver id,Drop timestamp seems to contain null values.
Hence any further analysis involving Driver id,Drop timestam
```

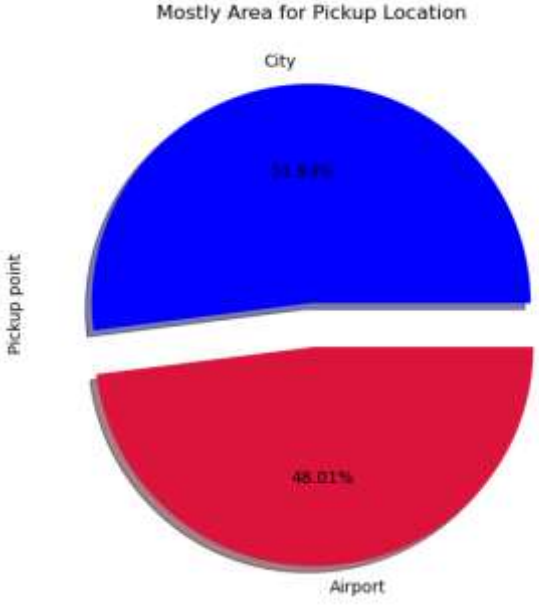
© All rights reserved by Fireblaze Technologies Pvt. Ltd.

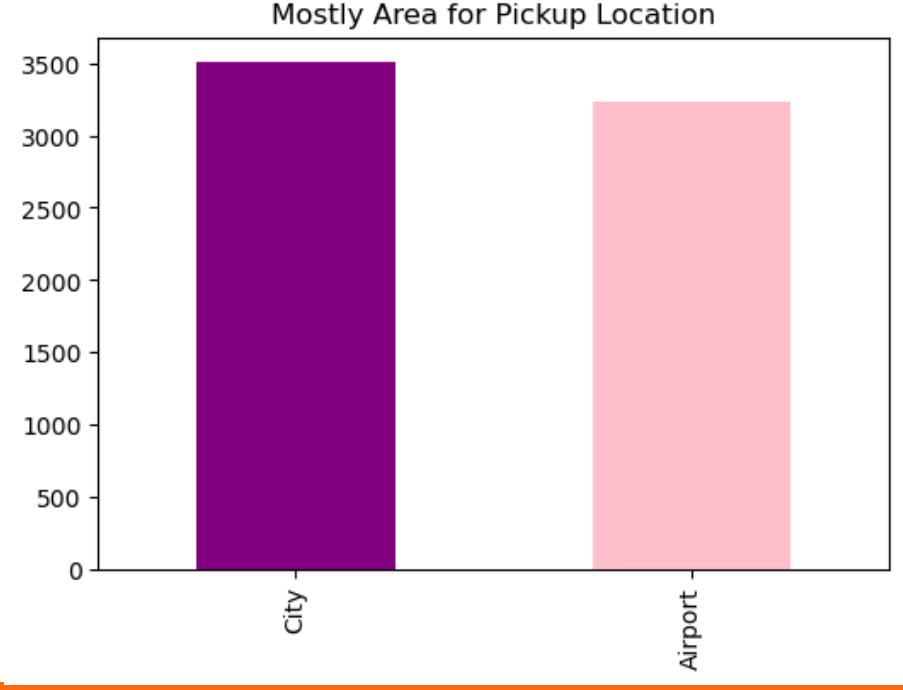
Brief Analysis of Data Visualization of Uber data

In this section, we will try to understand and compare all columns.

Let's start with checking the unique values in dataset of the columns with object datatype.

Get the data from the Mostly pickup areas...

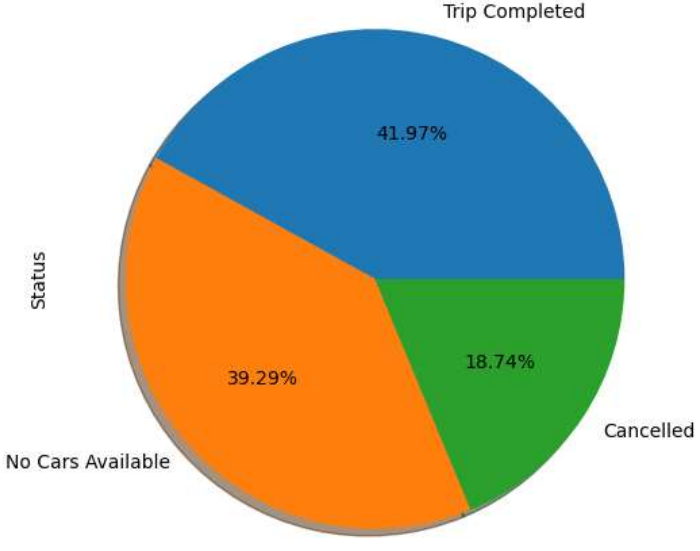


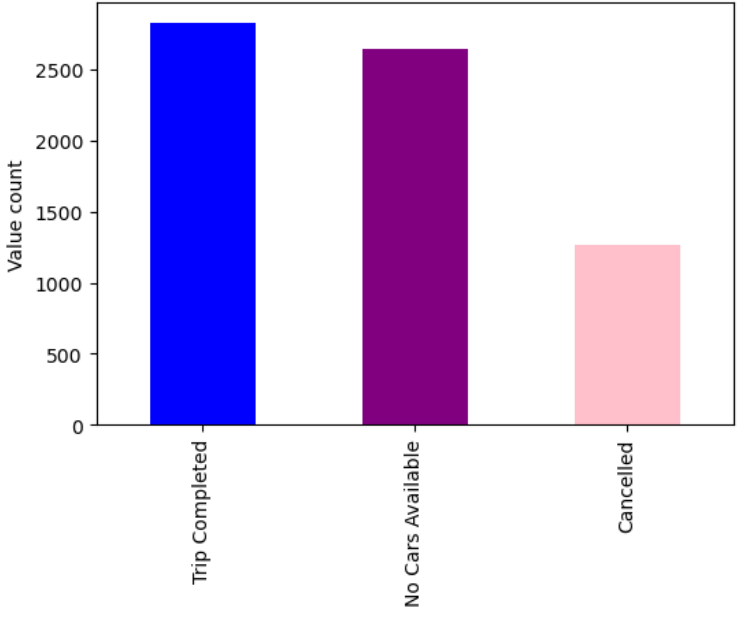


© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Descriptive Analysis

The survey seems the Trip Completed , Cancelled and No Cars Available...in mostly pickup area





Above table displayed that the data of Trip Completed is more than the No Cars Availble and Cancelled.

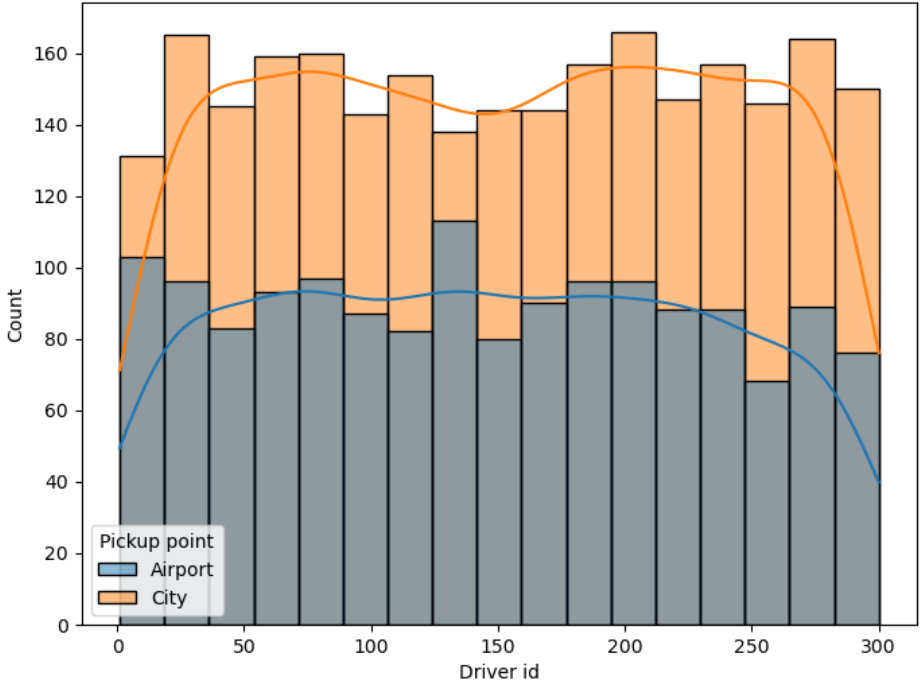
So, we can predict that the Uber is growing day by day...And people give good reviews...

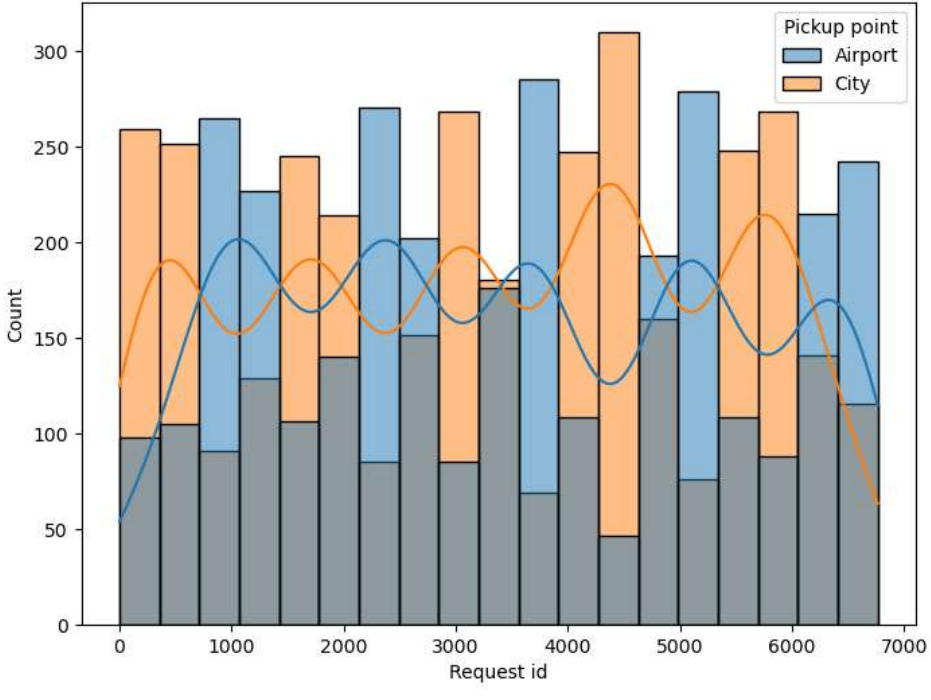
© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Descriptive Analysis

Let’s do the same for Driver id column and Pickup id,

here we will be using the Driver id column and Pickup id which we have extracted above wrto Pick point.





© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Descriptive Analysis

For further analysis we would need data in hourly and daily basis.

Splitting date-time columns into separate date and time columns.

```
df['Request timestamp']=pd.to_datetime(df['Request timestamp'])
```

```
df.dtypes
```

```
Request id                    int64
Pickup point                 object
Driver id                   float64
Status                       object
Request timestamp    datetime64[ns]
Drop timestamp               object
dtype: object
```

```
df['Request timestamp_day']=df['Request timestamp'].dt.day
```

```
df.head(1)
```

Request id	Pickup point	Driver id	Status	Request timestamp	Drop timestamp	Request timestamp_day

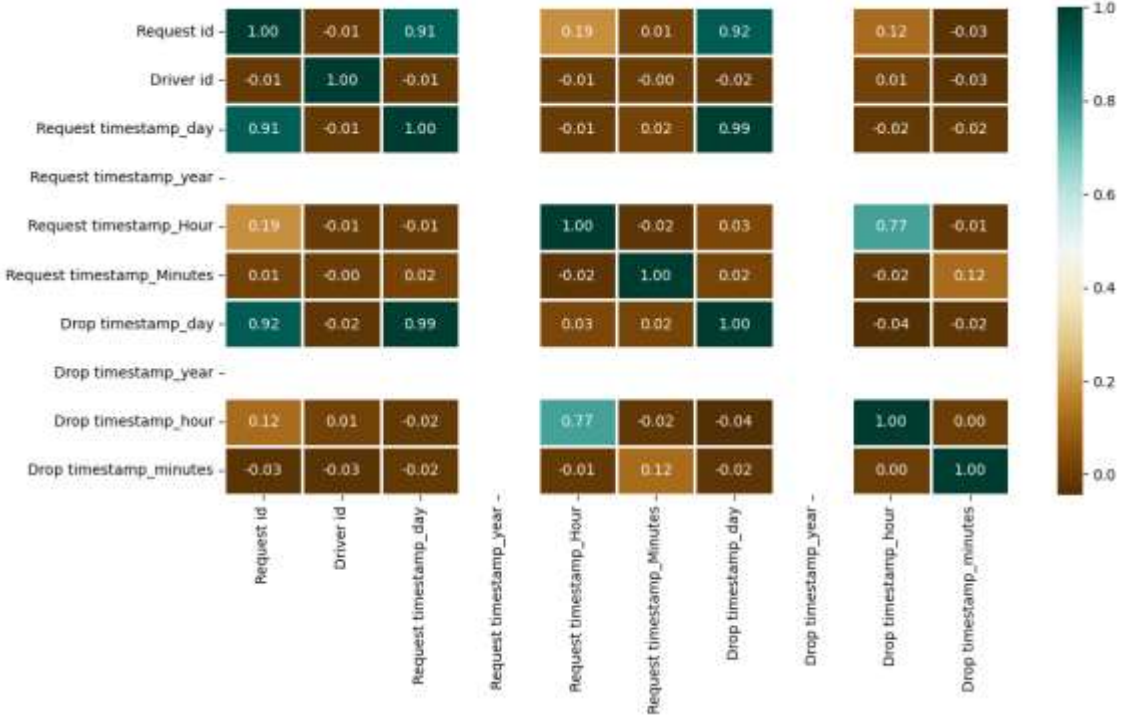
```
df.dtypes
```

```
Request id                          int64
Pickup point                       object
Driver id                         float64
Status                             object
Request timestamp          datetime64[ns]
Drop timestamp             datetime64[ns]
Request timestamp_day               int64
Request timestamp_time             object
Request timestamp_weekday          object
Request timestamp_monthname        object
Request timestamp_year              int64
Request timestamp_Hour              int64
Request timestamp_Minutes           int64
Drop timestamp_day                float64
Drop timestamp_time                object
Drop timestamp_dayname             object
Drop timestamp_monthname           object
Drop timestamp_year               float64
Drop timestamp_hour               float64
Drop timestamp_minutes            float64
dtype: object
```

© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Descriptive Analysis

Here is the relation between the columns which contain numeric values....that we are added after...

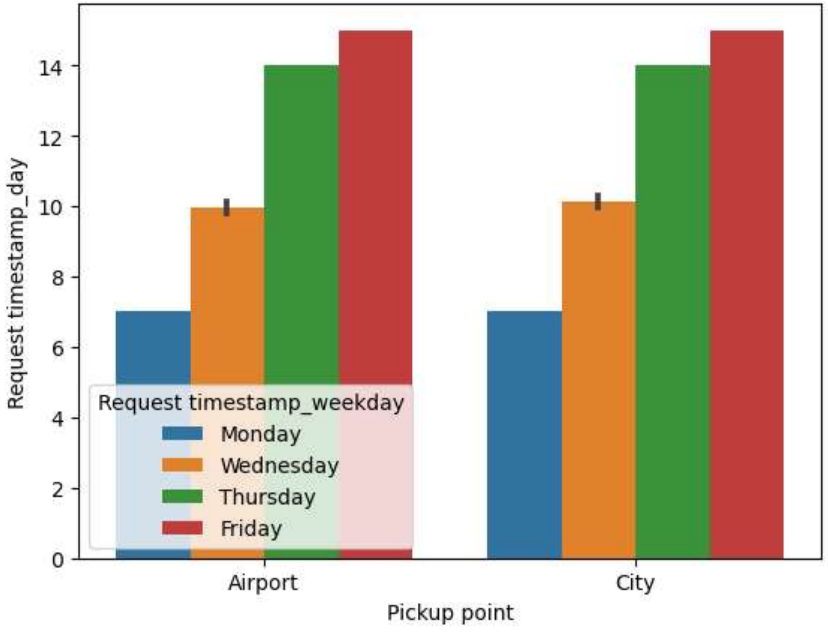


© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Descriptive Analysis

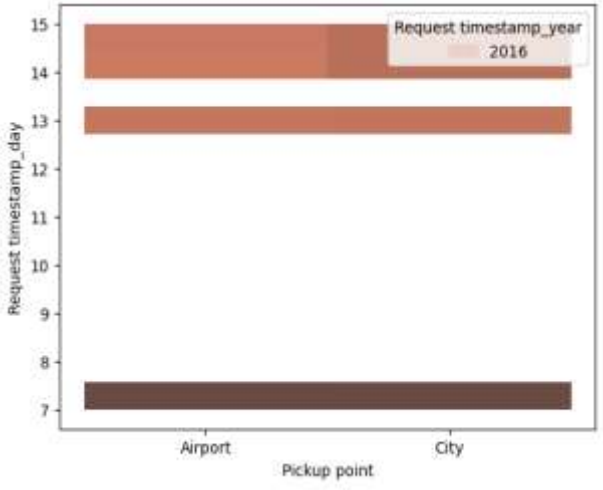
First up all we have to get deep knowledge about the data.....so distribute the data to get more clarification about it..

To get the idea about the Pickup time wrto the other data....

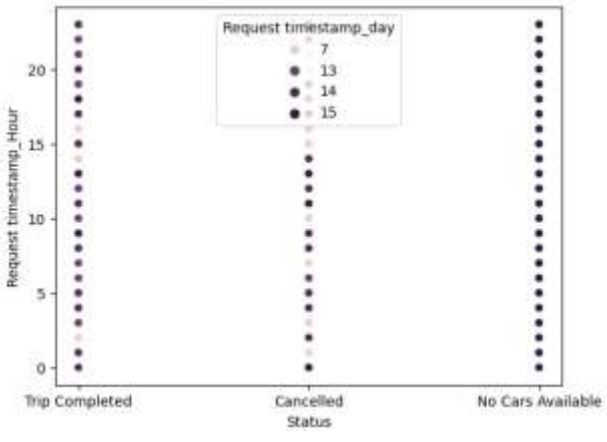


© All rights reserved by Fireblaze Technologies Pvt. Ltd.

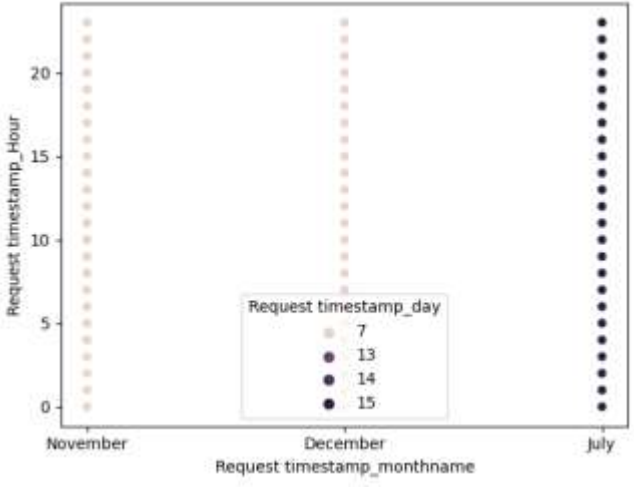
Descriptive Analysis



Pickup Point wrto request date on basis of Year



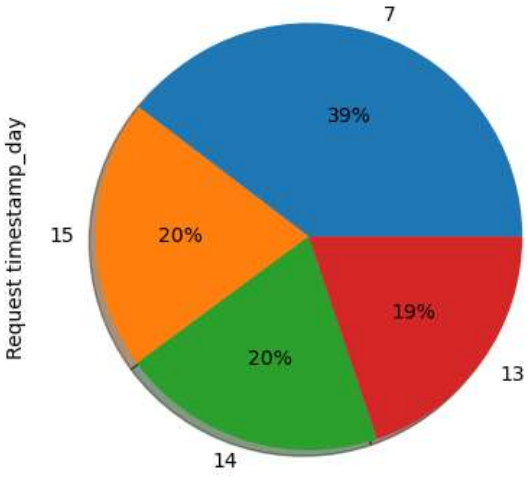
Status wrto request Hour on basis of dates

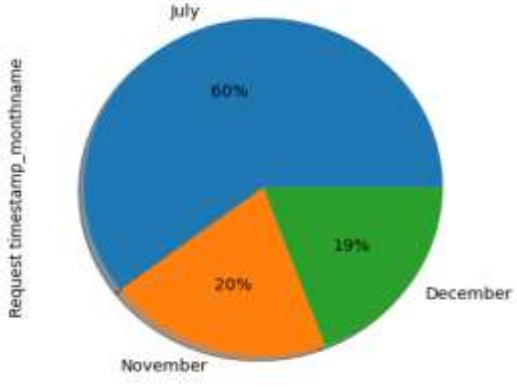


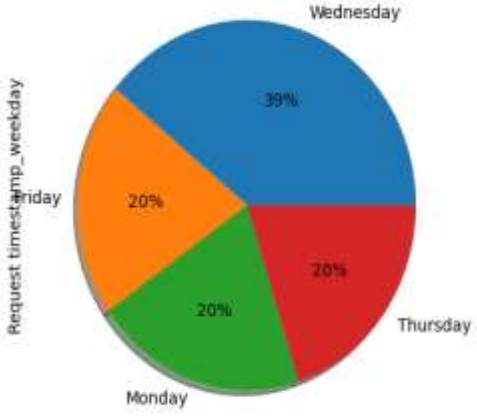
Hour wrto request month on basis of date

© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Descriptive Analysis





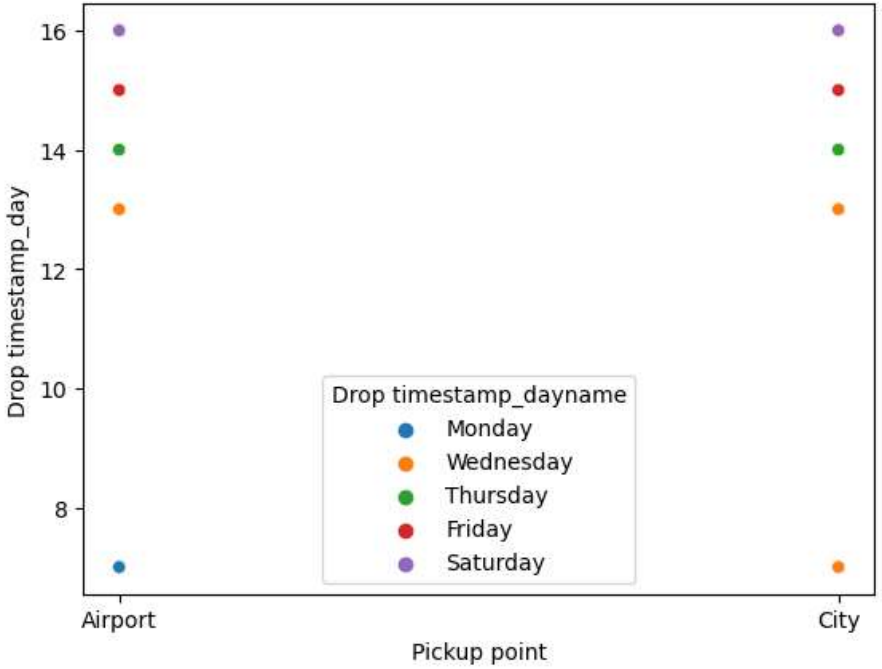


Pie Chart of Distribution of Request Booking on basis of Date ,Month , Weekdays

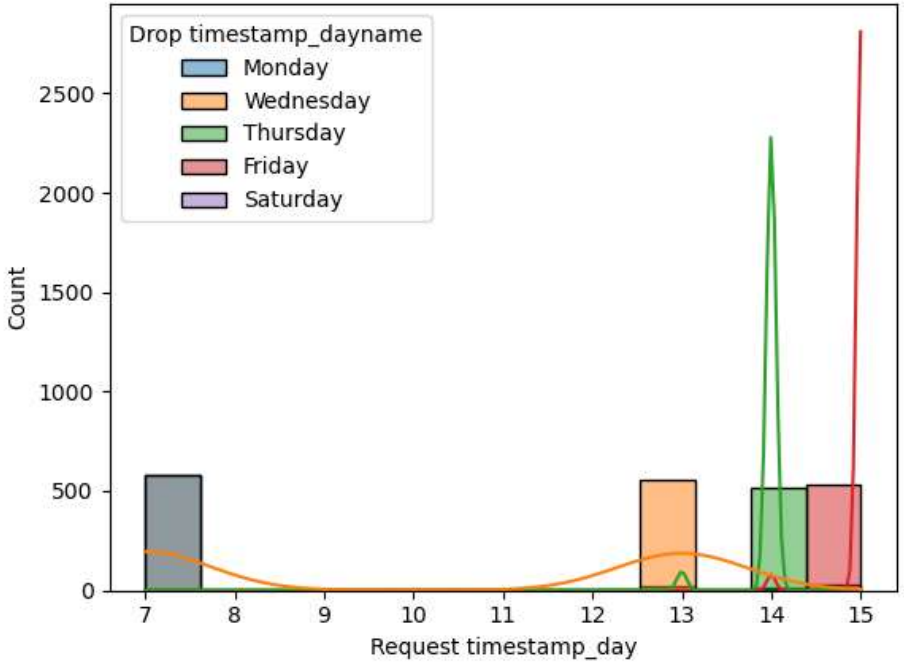
© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Descriptive Analysis

Lets us discuss about the Drop timestamp....



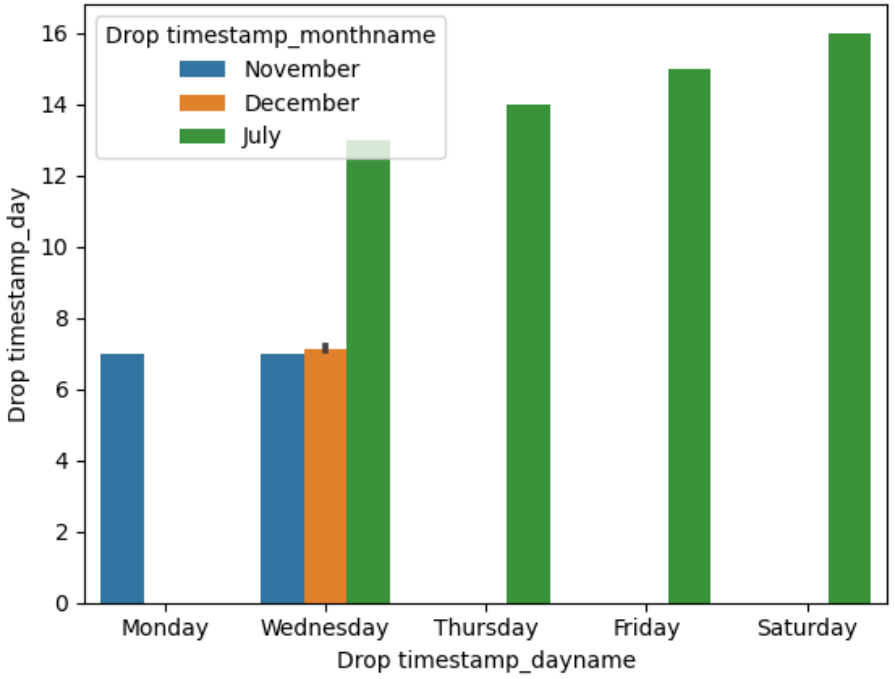
Pickup point wrto Date on basis of Days



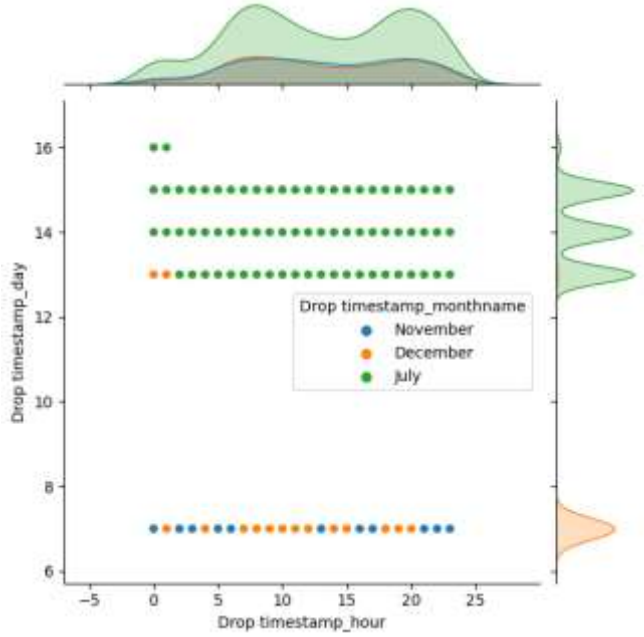
Request Date On basis of Drop Days

© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Descriptive Analysis



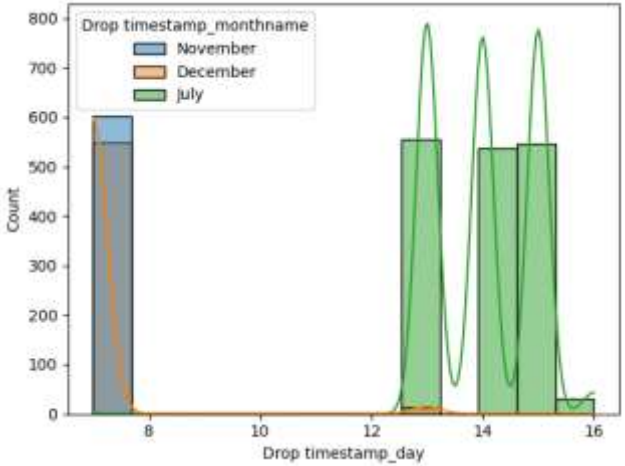
Drop Weekdays wrto Drop Data on basis of Drop Month



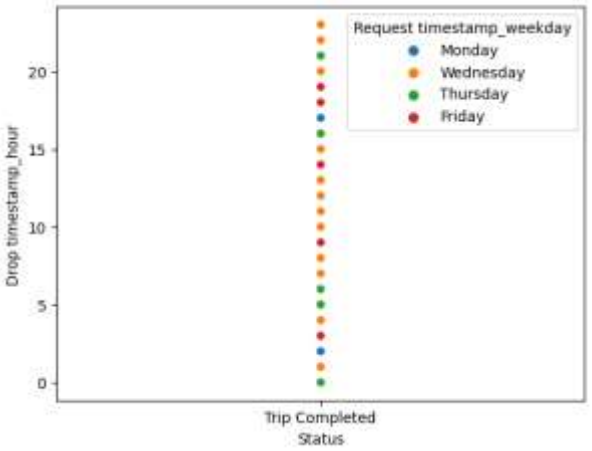
Joint plot of Drop Hour wrto Drop Day

© All rights reserved by Fireblaze Technologies Pvt. Ltd.

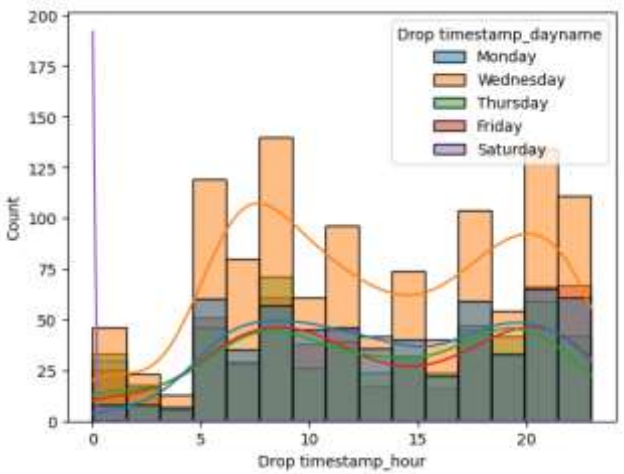
Descriptive Analysis



Drop time Monthname



Week Day wise Status of trip



Drop Day name in particular hour

© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Conclusion

- From this we can get the idea about the timing, the day , date ,months , year according to their status ,Pick up area wrto their Drop and Pickup time.....
- And we conclude that the Pickup area is from Airport is less than the City, Status of Cars is quite good , Trip Completed is more than the Cancelled ,No Car available category...
- For the Pick up Friday , July and 13,14,15 date are the busiest day of the week and in terms of time 1 to 4 o' clock is the busiest ever...
- The whole data are from Year 2016
- From this we can also conclude that the in term of Drop timestamp.. Pickup area is from Airport is less than the City, Status of Cars is quite good, their is no cancellation of the trip....
- For the Drop Monday , July and 7,15 date are the busiest day of the week and in terms of time 1 to 4 o' clock is the busiest ever....
- So From this we can get the best time ,day, month ,date to travel and to avoid unwanted disruption....while Travelling

© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Future Scope

Commuting is now considerably easier than it was previously. This is due to technological advancements and the introduction of taxi booking applications. People have started using taxi apps instead of taking public transit in the previous three years. Uber is one of the most well-known names in the ride-hailing and taxi industries, with millions of users worldwide.

With no time to wait, the on-demand services serve as the best choice and have become the perfect solution in this modern world to get exactly what they want. Hence, there are many other several factors that are responsible for the growth and development of the on-demand industry.

Here are some of the prominent reasons for the success and growth of the on-demand industry,

•Advancements in technologies
•Gaining competitive advantage
•Providing a unique experience
•Easy availability and use
•More choices for business space
•Flexibility and affordability

© All rights reserved by Fireblaze Technologies Pvt. Ltd.

Your Car Arrive.....

© All rights reserved by Fireblaze Technologies Pvt. Ltd.