

A STUDY ON EXPLORATORY DATA ANALYSIS ON GETCAB USING PYTHON

Dr. Ch. Sahyaja*

Assistant Professor, School of Commerce, Presidency University, , Bangalore, Karnataka – 560064, Ph: 8019010226, Email: sahyaja.c@presidencyuniversity.in

Mr B.S.Rajath**

Assistant Professor, School of Commerce & Management studies, Dayanand Sagar University, Bangalore, Ph:9738530662, Email: rajathbelur-scms@dsu.edu.in

Lalith Priyan V***

Student, School of Commerce, Presidency University, PH:8300864308, Email:lalithpriyanvadel@gmail.com

Abstract:

This study is based on the Exploratory Data Analysis on GETCAB using Python and the sample size is in this study the data collected is Secondary data that is provided by the company where it is generated from their business functions of GETCABS. As it is startup established recently the study is targeted to rides covered in a month by the company. The main objective of the study is “Exploratory data analysis on get cabs using python”

Keywords: Exploratory Data Analysis; GETCAB; Python

1. Introduction

Call taxi system in India has grown significantly in India and infrastructure growth, growth of middle class, increasing disposable incomes and growing GDP are some of the factors responsible. The rise of the BPO industry is one of the reasons to growth of this sector because of odd working hours. This growth can be seen more in metropolitan cities of India (Rahman, 2014) and there is intense competition among various operators like Ola, Uber, Radio cabs, Yellow cabs and Meru etc. So, to sustain in this competitive market it is necessary to understand the users of the rental cab service. Various studies and researches have been done to understand the factors important while choosing a rental car. Call taxi app (CTA) helped in increasing perceived usefulness, ease of use, playfulness and subjective norms. Today transportation services play a main role in urban areas, changes in the taxi industry with innovation and downloading app has made it easy to travel safely to any designation point. The research carried with cross sectional study compared with model of SERVQUAL and IDCTP Model. A well-structured questionnaire was framed with five-point Likert scale based on ten dimensions. The study found variables as direct and indirect relation. The direct method was identified by unstandardized regression coefficient (β). The indirect method identified by multiple regressions. The study results customer satisfaction affects 99% with confidence level that's found positive.

2. Review of Literature

According to Hanif and Sagar (2016) had stated that there was demand for Call-a-Cab service offered by Meru Cab. The cab services are proving security through global positioning system (GPS) and women taxi drivers for women passengers especially during night times. According to Harding et al (2016) the auto-rickshaws (three wheelers) are more popular in urban transport before the advent of cars and cabs. Horsu and Yeboah (2015) had argued that driver behaviour has negative impact on customer satisfaction in Ghana. The variables like continuous service, comfort, reliability and affordability have an impact on customer satisfaction with regard to minicab taxi.

According to Lu et al (2015) the self-service mobile technologies help the commuters to access lot of data about cab services and such technologies had changed the role of both customers and companies. The adoption of call taxi app (CTA) is impacted perceived usefulness, perceived ease of use, subjective norms and

perceived playfulness.

According to Vaithianathan & Bolar (2013) The Meru cabs had become more popular and the demand for its cabs had exceeded that its supply which means technology had created huge demand for organized cab industry. The factors like accessibility, reliability and transparency are primary factors which have attracted customers towards branded cab services like Meru cabs the customer feedback in cab services industry is very important for attaining success in the competitive car rental industry.

3. Research Methodology

3.1 Sample size of the study

In this study the data collected is Secondary data that is provided by the company where it is generated from their business functions of GETCABS. As it is startup established recently the study is targeted to rides covered in a month by the company.

3.2 Objective of the study

The main objective of the study is “Exploratory data analysis on get cabs using python”

3.3 Contribution of the study

The main contribution of the study is to analyze how customers are responding to the services provided by GETCAB and to find different marketing strategies to bring awareness about the company and to attract new customers and provide insights for the better growth of the company. The study has been carried out to help them find new areas to focus like low transportation areas, travelling stations etc to improve rides and to be more available at any time and to decrease waiting time for customers and to increase the satisfaction towards the service. The study is based on the secondary data that is provided by the company. The study gives insights to the company that could help them to sustain in this competitive market

4. Data Analysis & Interpretation

Code1: Importing packages

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import pandas_profiling as pf
```

Code 2: For loading the dataset

```
df=pd.read_excel("C:/Users/ajay kumar/Desktop/SALMA project/GETCAB RIDES DATA.xlsx")
```

Code3: To know the shape of the dataset

```
df.shape
output[] (149, 24)
```

Code 4: To show all the columns available in the dataset

```
df.columns
output[]
Index(['User_Id', 'user_name', 'Gender', 'email', 'country_code', 'phone_number', 'Driver ID', 'Cab Type', 'Price per Km',
```

'Request time stamp', 'Drop time stamp', 'Pickup point', 'Drop point','distance(km)', 'Distance CHG/KM', 'Ride Time', 'Ride Time CHG/MIN', 'price before tax and comission', 'Tax', 'Comission', 'Discount', 'Total Fare', 'Status'],

dtype='object')

4.1: To show Exploratory data analysis using pandas profiling

report=pf.ProfileReport(df)report

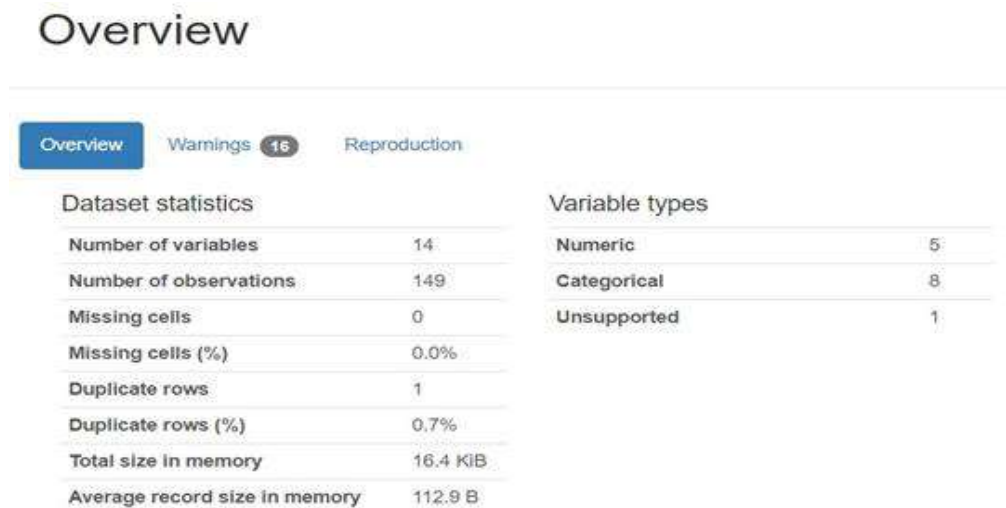


Figure 4.1: The Figure shows the Descriptive statistics of the dataset

Interpretation

Here from the above figure, we can clearly identify that the dataset contains 14 variables where five are numerical variables and eight are categorical variables and it contains 149 observations with no missing vales.

Statistical Analysis for the data

4.2: To find the correlation for the data

df.corr()

output[]

	User_Id	distance(km)	Ride Time	Total Fare	Time
User_Id	1.000000	-0.087931	0.060160	-0.102238	0.059418
distance(km)	-0.087931	1.000000	0.667198	0.937667	0.033096
Ride Time	0.060160	0.667198	1.000000	0.661905	0.070270
Total Fare	-0.102238	0.937667	0.661905	1.000000	0.043518
Time	0.059418	0.033096	0.070270	0.043518	1.000000

Interpretation

From the above output it can be said that there is positive correlation of value 0.937667 percentbetween Total Fare and distance (km).

4.3: To plot the relationship between distance and the fare.

```
fig = px.scatter(df, x='distance(km)', y='Total Fare', opacity=0.65,trendline='ols',  
trendline_color_override='darkblue')
```

```
fig.show()
```

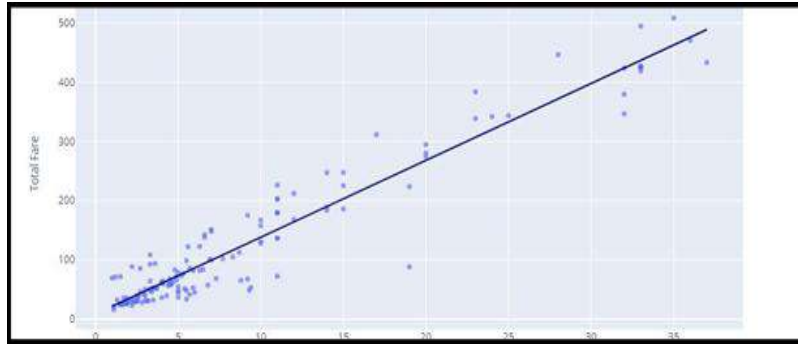


Figure 4.3: The Figure shows the relationship between distance and total fare

Interpretation

Here the above figure shows that there is positive correlation between total fare and the distance travelled by the users of GETCAB i.e., as distance increases then fare increases. It states the positive correlation between these two variables.

4.4: To plot the rides done in GetCab by each gender fig=px.bar(df,x='Cab Type',color='Gender',title='Rides done by the Gender') fig.update_xaxes(type='category')

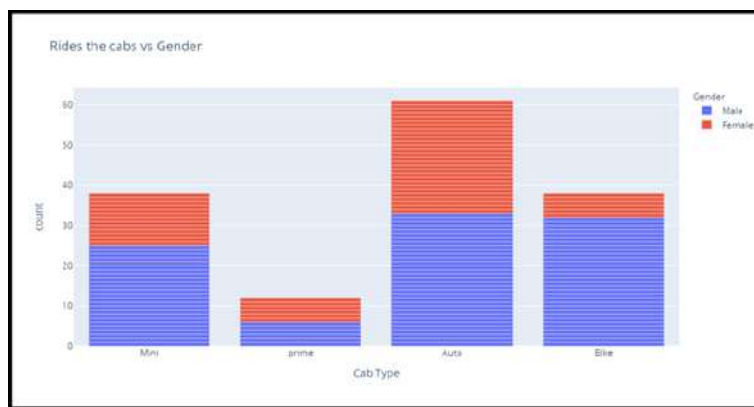


Figure 4.4: The Figure shows the Plot of Rides of the cab's vs Gender

Interpretation

From the analysis of 149 rides, it is clear that from the total 38 rides taken by the cab type Mini it is Visualized that 25 rides are requested by Male customers and 13 rides are requested by Female customers. From the analysis of 149 rides, it is clear that from the total 12 rides taken by the cab type prime it is identified that 6 rides are requested by Male customers and 6 rides are requested by Female customers. From the analysis of 149 rides, it is clear that from the total 61 rides taken by the cab type auto it is identified that 33 rides are requested by Male customers and 28 rides are requested by Female customers.

From the analysis of 149 rides, it is clear that from the total 38 rides taken by the cab type auto it is identified that 32 rides are requested by Male customers and 6 rides are requested by Female customers. Thus, it is clear that the customers most preferred type of ride is through Auto and Male are more to request the rides and it is parallely observed that very few women request through bike.

4.5: To plot the frequency of the users using GETCAB `fig=px.histogram(df,x="User_Id",title='Frequency of the Users') fig.update_xaxes(type="category")`

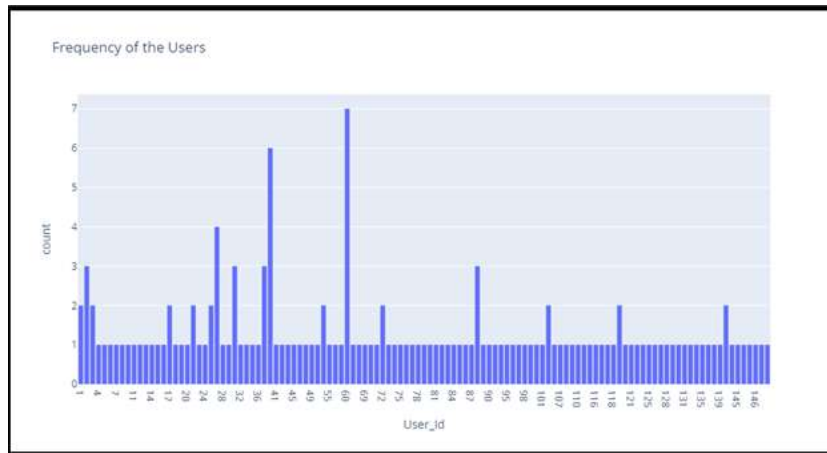


Figure 4.5 : The Figure shows the Plot of Frequency of the Users

Interpretation

From the analysis of 149 rides taken by the GETCAB the above figure clearly identifies the frequency of the customer bookings. It is identified that customer with user id's 39 and 60 have more bookings than compare to other customers. Hence it is proved that customers with user id's 39 and 60 have done frequent booking.

4.6 To plot the total ride travelled by each type of cab.

`sns.catplot(x='Cab Type',y='Ride Time',data=df,kind='bar') sns.set(style="whitegrid")`

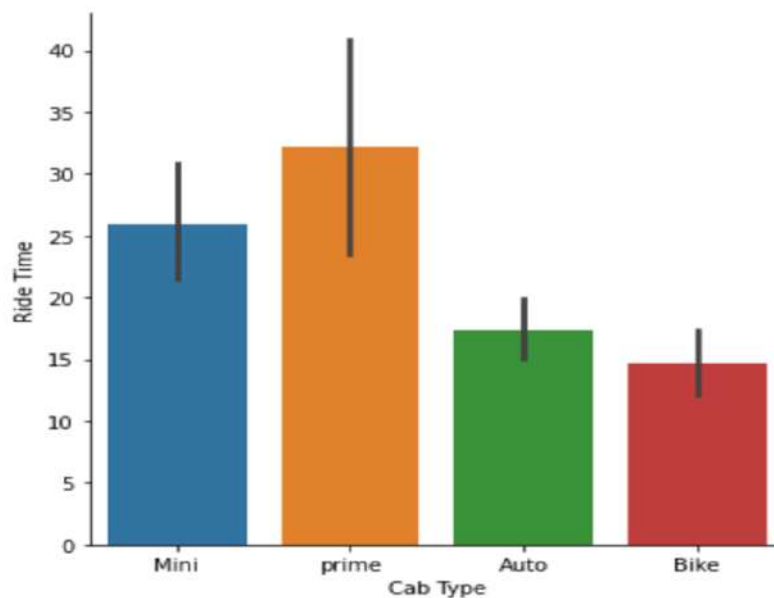


Figure 4.6: The Figure shows the plot of average ride time travelled by users in respective cab type.

Interpretation

From the analysis of 149 rides, it is clear that average time travelled in the cab type of mini is 30 and the average time travelled in the cab type of prime if 40 and average time travelled in the cab type auto is 20 and the average time travelled by the cab type bike is 15. Thus, it is clear that average time travelled in the cab type of prime is more than compare to other cab types. Hence it is proved that users are requesting prime if they have to travel for more time or long distance.

4.7: To plot the relation of Total Fare vs Distance

```
fig = px.scatter(df, x="distance(km)", y="Total Fare",title='Diatance vs Total Fare') fig.show()
```

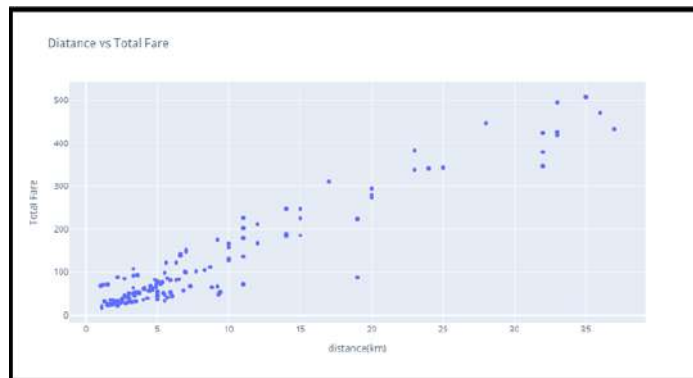


Figure 4.7: The Figure shows the plot to tell the relationship between Total Fare and distance

Interpretation

From the analysis of 149 rides, the above figure show that the points are very close to each other and as the distance increases the total fare increases. Thus, it is clear that there is positive relationship between distance and total fare. Hence it is proved that as the distance increases total fare also increases.

4.8 To map the boarding point and dropping point of rides

```
fig = px.treemap(df, path=['Drop point', 'Pickup point'], values='Ride Time') fig.show()
```

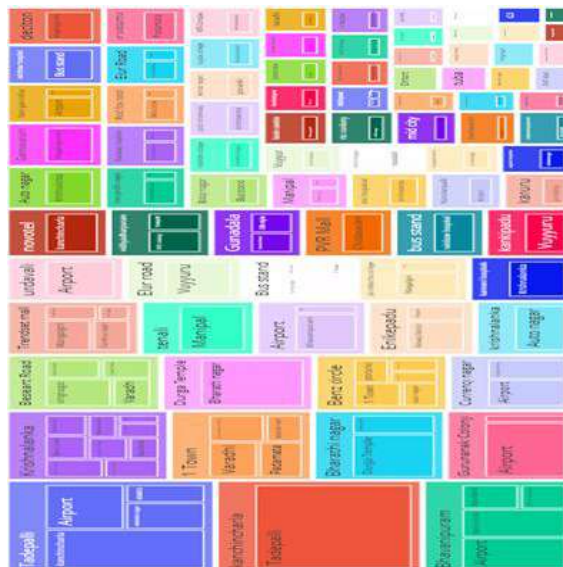


Figure 4.8: The Figure shows the map of boarding and dropping points of the ride

Interpretation

From the analysis of 149 rides the above tree map gives the full map of from where to where the users are travelling. It is visualized that the users boarding from Krishnalanka are deport to more locations than compare to other boarding locations.

5. Conclusion

The project is done with GetCab to study their market share and the demand they have in Vijayawadaby using Pandas profiling and some visualization techniques like seaborn and plotly in Python. In this study it clearly states the demand of cab aggregation is high in Vijayawada but require to create more awareness about the

company. I have put my efforts to gain knowledge about the Cab Aggregation sector and to know about the activities they implement in their business operations for growth of their company my moving with them. For this study the data is secondary data that Collected from the company generate by the rides travelled by the users where it is taken for making proper data analysis and to find hidden insights from it.

After the proper analysis and Interpretation of the study I finally conclude the following facts that here from this study, it can be clearly stated that most of the rides are been requested on Fridays at evenings and on weekends and most rides are preferred at the mornings in each day between 9am to 11am and at evenings between 17pm to 20pm. It can also be concluded that most of the rides are been requested from Krishnalanka, Airport regions where it can be said that that regions with low transportation facilities and at the travelling stations there be high request of pickups and droppings. As it is a start-up, more awareness about the company is to be created among the people for better growth of the company and for better sales. In the above the study, it is stated that the trend of revenue generated for each day in the period of 1st January 2021 to 31st January 2021. It is found that it is irregular trend with full of ups and downs.

References

- [1] Abdul Wahid Khan et al., (2016), Factors Affecting Customer Satisfaction in the Taxi Service Market in India, Journal of Entrepreneurship & Management, Vol 5 – Issue 3, PublishingIndia.com.
- [2] Dr. Mukund H. Khupse (2017), A Study of passengers' motives for using mobile app-based cabs, KAAV International Journal of Economics Commerce & Business Management, KIJECEBM/JUL-SEP (2017)/VOL-4/ISS-3/A68 Page No.474-480, ISSN:2348-4969 Impact Factor (2017)-7.8902
- [3] Emmanuel Nondzor Horsu et al., (2015), Influence of Service Quality on Consumer Satisfaction: A study of Minicab Taxi Services in Cape Coast, Ghana, International Journal of Economics, Commerce and Management, United Kingdom, Vol.III, issue 5, May2015.
- [4] G. Venkatesh et al., (2015), Measuring the Performance of Taxi Aggregator Service Supply Chain, SAMVAD: SIBM Pune Research Journal, Vol X,26-36, December 2015.
- [5] Zi Yu Liu (2014), An Analysis of Technology Acceptance Model, Exploring user acceptance and intension of taxi-hailing app in Shanghai, Bachelor of Science Thesis in the Programme Software Engineering and Management.
- [6] Mr.M. Sreenivas et al., (2016), the role of Transportation in Logistics chain, PP 1-10.
- [7] Mukti Advani et al., (2006), Review of Capacity Improvement Strategies for Bus transit service, October – December 2006, Transportation Research & Injury Prevention Programme (TRIPP), Indian Institute of Technology, Delhi.
- [8] Prof.Caroline Saunders et al., (2014), Economic Development and Transport Summary report, Ministry of Transport, New Zealand. Prof.Ken Button (2008), The impacts of Globalisation on International Air Transport Activity, Global forum on Transport and Environment in a Globalising World 10-12,November 2008, Guadalajara, Mexico