Market Segmentation – Online Vehicle Booking

Team Leader – Abhik Gupta

Team Members

- Yash Dawande
- Saikat Basu Majumdar
- Khatija Tul Kubra Latheef

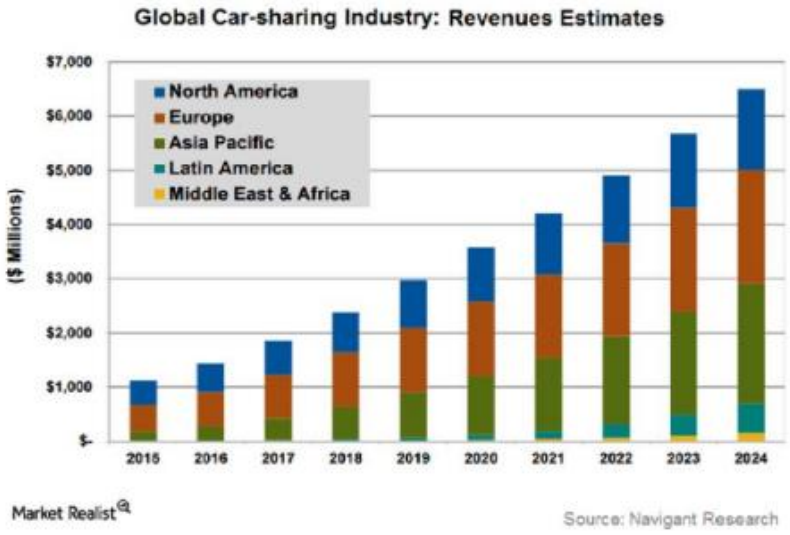
Problem Statement

The task is to apply segmentation analysis to analyze the Online Vehicle Booking Market in India and come up with a plausible strategy to enter the market, targeting the segments most likely to use their product in terms of Geographic, Demographic, Psychographic, and Behavioral.

Market Overview – Global Market

The online cab booking apps are one of the best technology innovations that brilliant ideas and vision have provided, and they have completely changed the fleet business. Taxi booking applications have you covered for any destination.

The market for online car booking service is anticipated to expand between 2020 and 2025 at a CAGR of 10.08%. The main elements that have dominated online taxi booking categories are low cab fares and the simplicity of booking through websites and applications. Urban residents, in particular, think that it would be simpler for them to reserve a cab online than to wait for cars at the public parking lot.



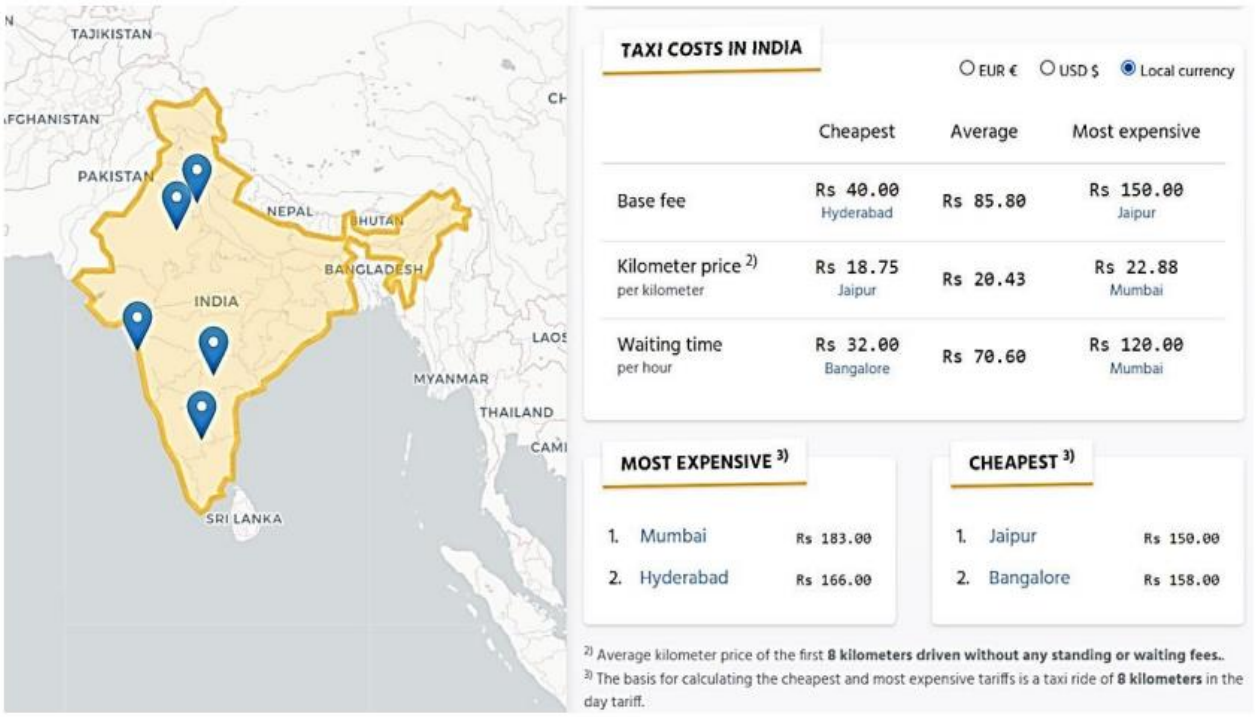
The aforementioned image illustrates how Asian, American, and European nations are turning to online cab booking services due to a steady increase in revenue. Additionally, individuals have started choosing ride-sharing services over traditional taxi services in developing nations like Vietnam, Malaysia, China, India, etc. because to the rise in smartphone usage.

Market Overview – Indian Market

The organized taxi market in India experienced tremendous development with the switch from the radio signals-based conventional tracking system to the GPS or GPRS-enabled system. 11.5% of the total revenue made by the Indian taxi business as a whole in FY 2019 came from the organized taxi market. In FY 2019, the Indian market for

online cab services was valued at INR 29.75 billion. By 2024, the market is projected to grow at a CAGR of around 16.60% to reach a sales value of INR 61.59 Bn.

The primary factors driving this increase include an increase in smartphone adoption in India, better high-speed internet access options, rising foreign institutional investor (FII) involvement in the online taxi business, and rising disposable income.



Two major players, ANI Technologies Private Limited (Ola Cabs) and Uber India Systems Private Limited (Uber India), dominate the Indian market for online cab services.

In FY 2019, Ola Cabs generated almost 72.44% of the entire revenue from the Indian market for online taxi services, while Uber India held a share of about 21.01%.

Meru Mobility Tech Private Limited, Mega Cabs Private Limited, and Cars on Rent (India) Private Limited are other market participants. Together, they controlled 6.55% of the overall market in FY 2019.

ANI Technologies Private Limited has transitioned from a radio signal-based tracking system to GPS and GPRS-based tracking system, giving them a large market share due to

their extensive geographical reach within India (152 cities) and wide range of services available for customers of all income levels.

Further investments by private equity firms and customer-friendly features in applications are expected to further drive their market growth.

Fermi Estimation

Parameters to predict right cabs for right customers

- Types of cabs customer preferred,
- Lifestyle of customers (indicating behavior of customer),
- Cancellations by customers,
- Distance travelled by customers,
- Period of customer availing the cab services

In this study, we analyzed current market scenarios, strategy applied by market giants in India using segments such as location, price, availability of Cabs, Demands and Supply market for online cabs, and cancellation frequency per cab booking, and so on.

Data Sources

The Datasets that we came across and found to be beneficial are:

- https://www.kaggle.com/datasets/arashnic/taxi-pricing-with-mobility-analytics?select=sigma_cabs.csv
- https://www.kaggle.com/datasets?search=vehicle+booking

Data Preprocessing

Steps taken to pre-process the data:

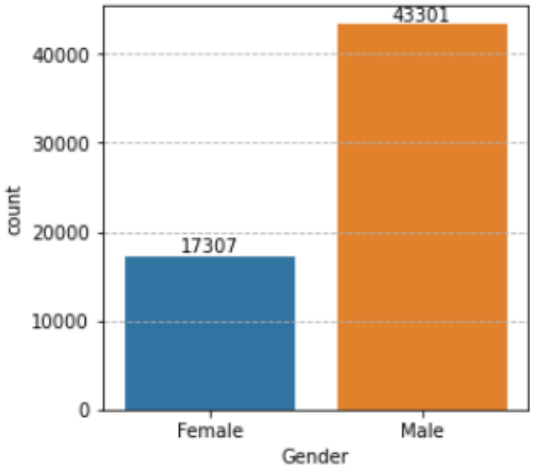
- Treating the Null values
- Outlier Handling
- Categorization data based on data types, continuous or categorical.
- Performing EDA and Data Visualizations.

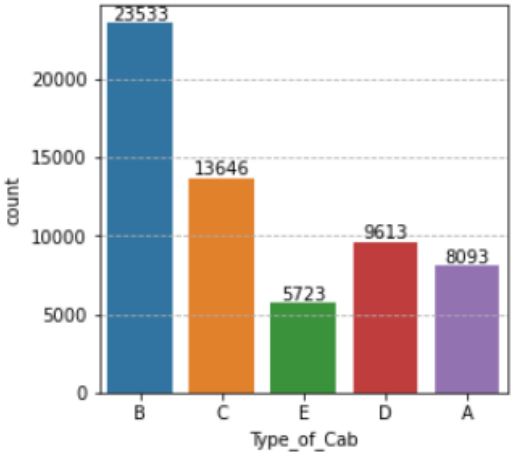
- Feature scaling by StandardScaler for numerical features
- Label Encoding for Categorical features.

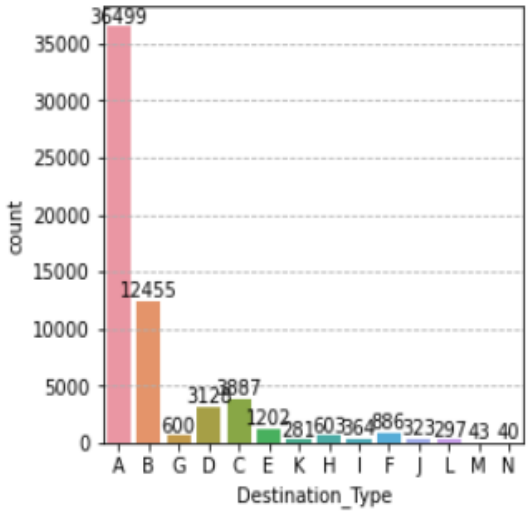
Exploratory Data Analysis (EDA)

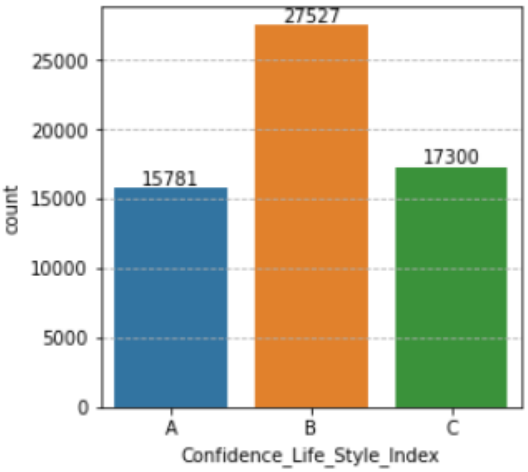
We performed an Exploratory Data Analysis to look closely at our data set and find out if there were any patterns, trends, or relationships. We analyzed our dataset using univariate (analyze data over a single variable/column), bi-variate (analyze data by taking two variables/columns into consideration) and multivariate (analyze data by taking more than two variables/columns into consideration) analysis.

Univariate Analysis

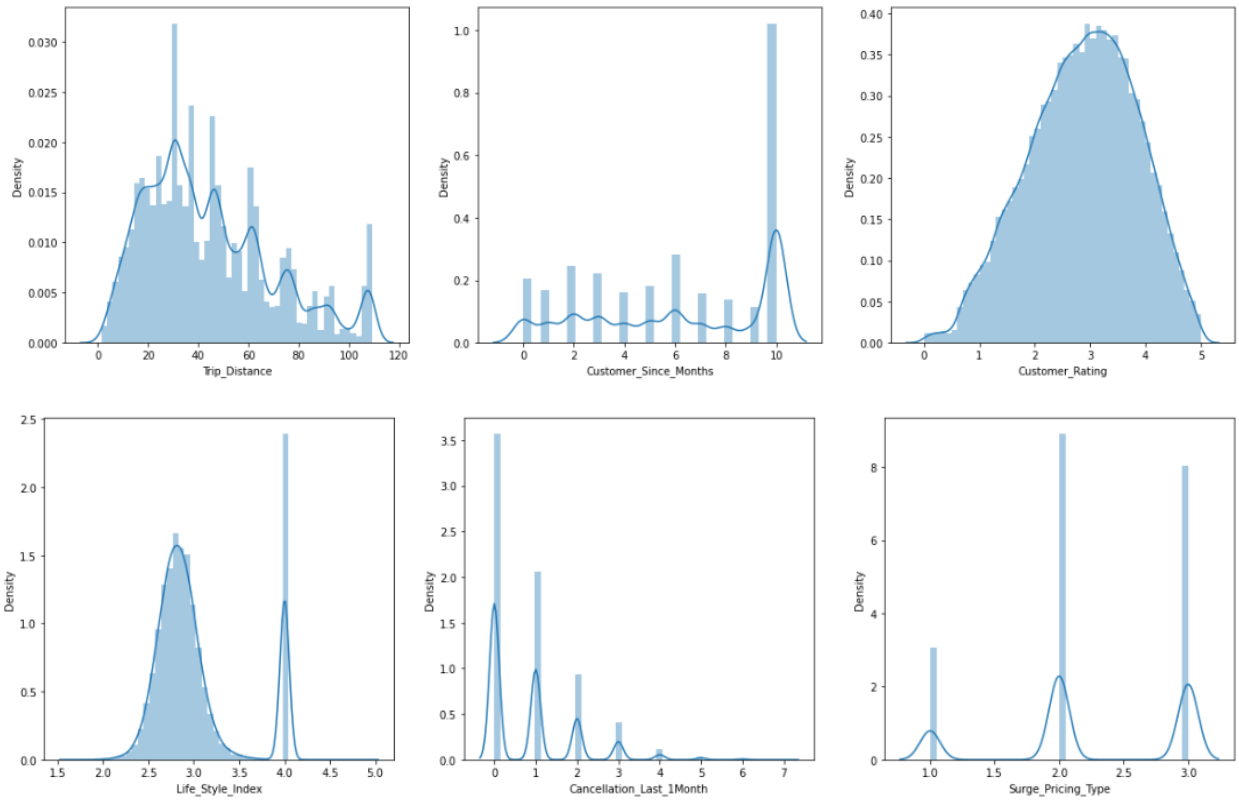




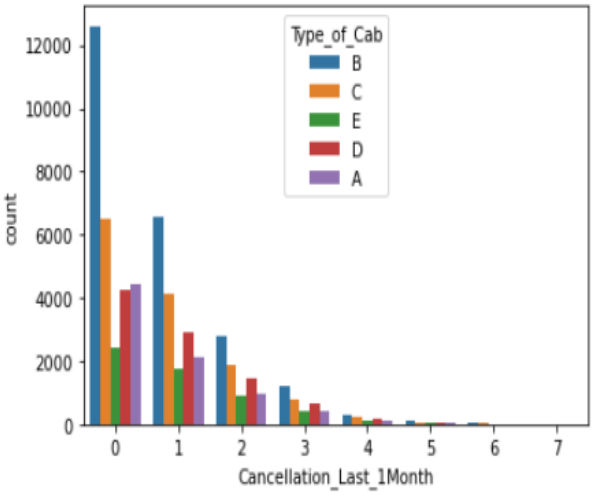


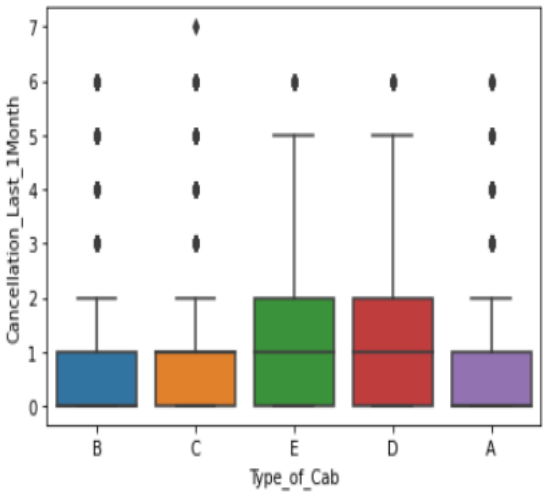


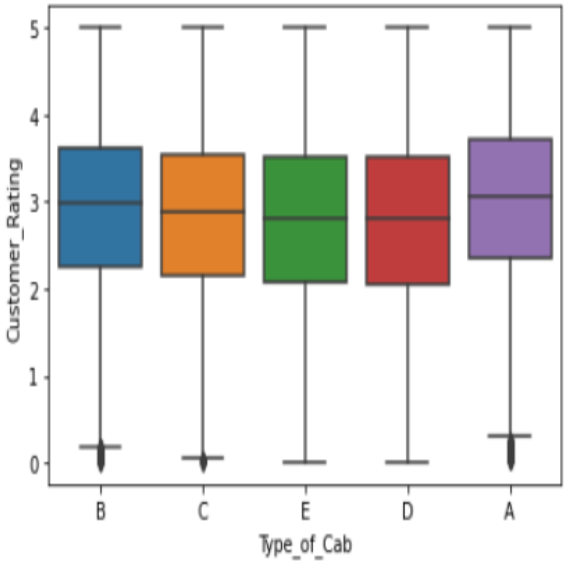
- Around 72% of the customers are Male
- Type of cab preffered is usually B
- Most customers have preffered ‘A’ Destination type
- Category B accounts for most counts in Confidence life style index

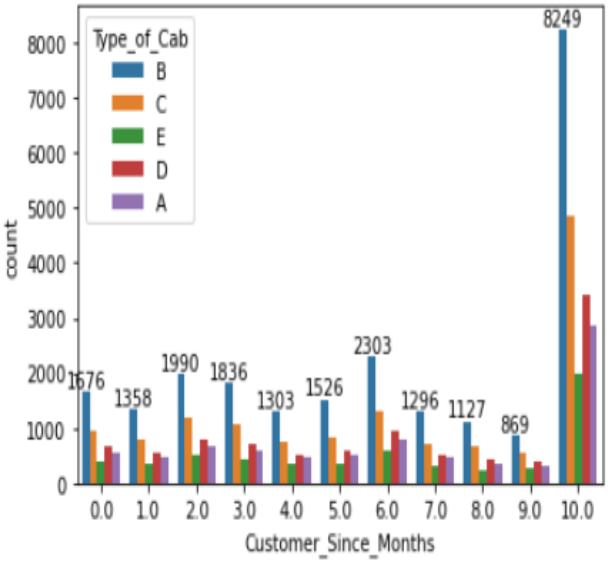


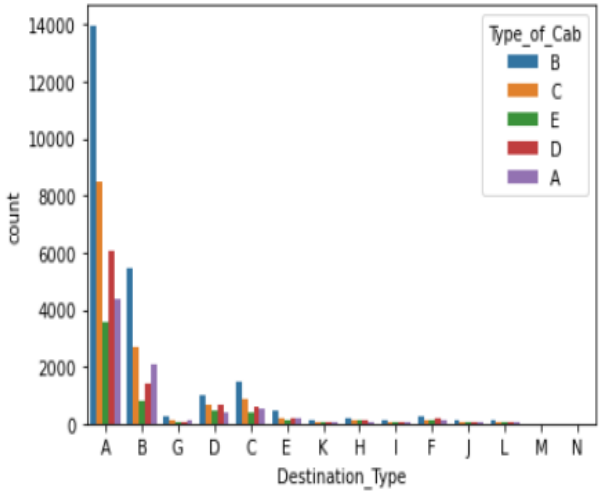
Bivariate Analysis

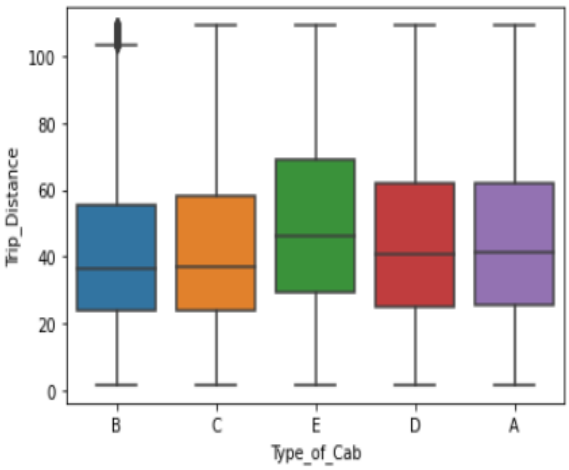


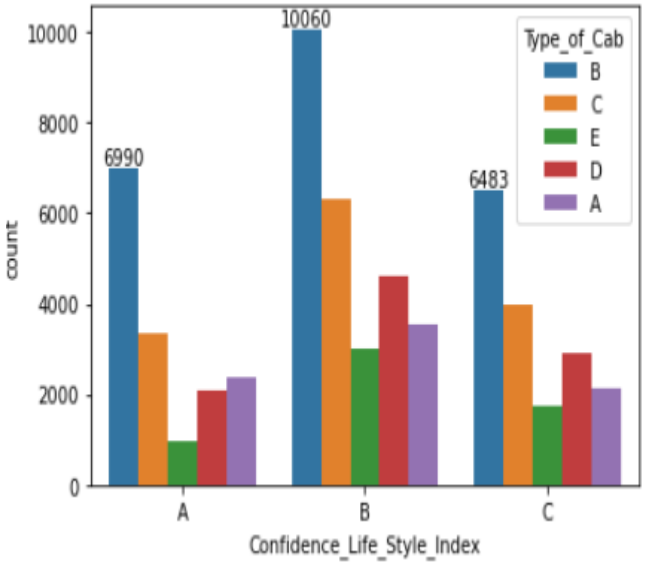


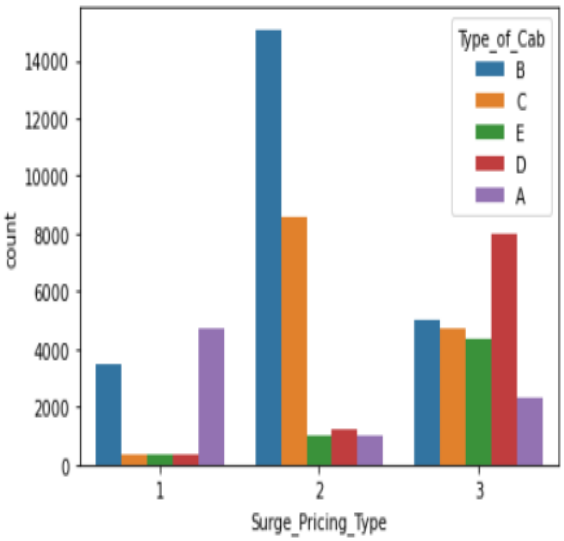

Customer_Rating
Type_of_Cab
B
C
E
D
A

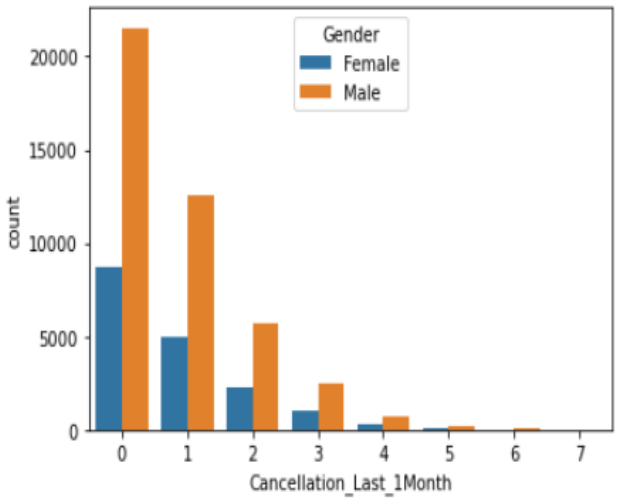

Type_of_Cab
B
C
E
D
A
count
Customer_Since_Months
1676
1358
1990
1836
1303
1526
2303
1296
1127
869
8249

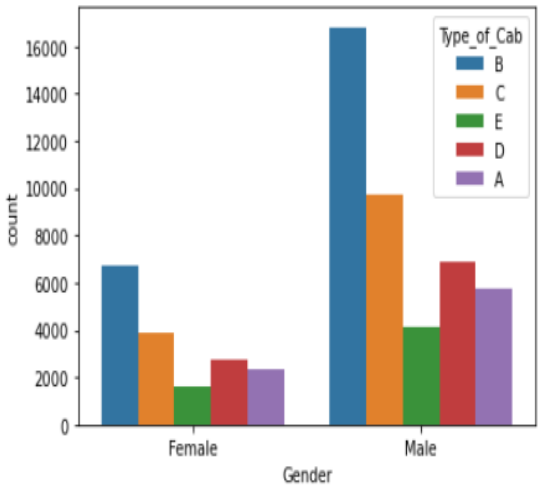

Type_of_Cab
B
C
E
D
A
count
Destination_Type
A B G D C E K H I F J L M N

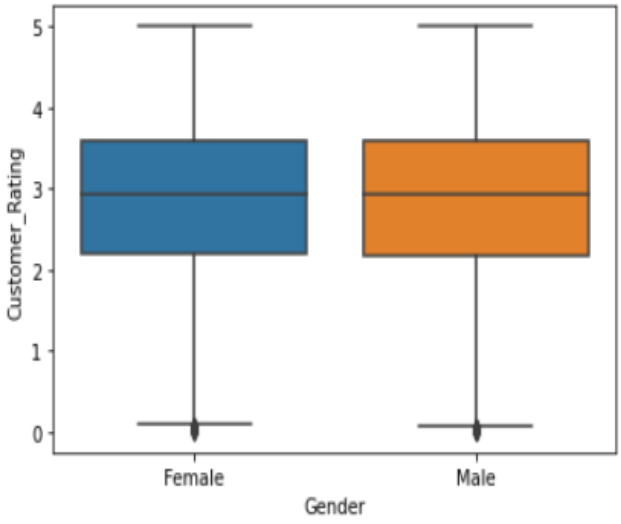

Trip_Distance
Type_of_Cab
B
C
E
D
A

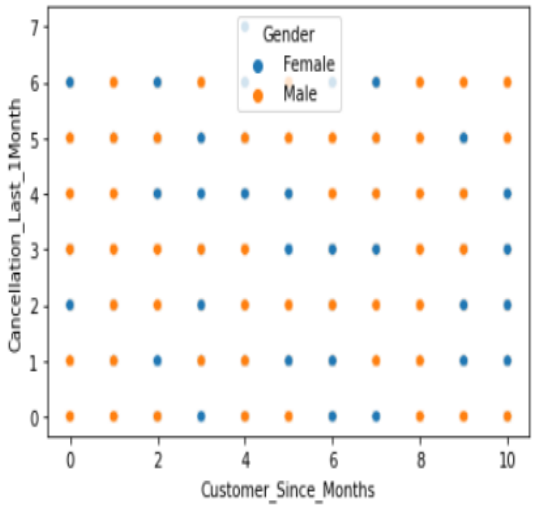

Type_of_Cab
B
C
E
D
A
count
Confidence_Life_Style_Index
6990
10060
6483
A
B
C

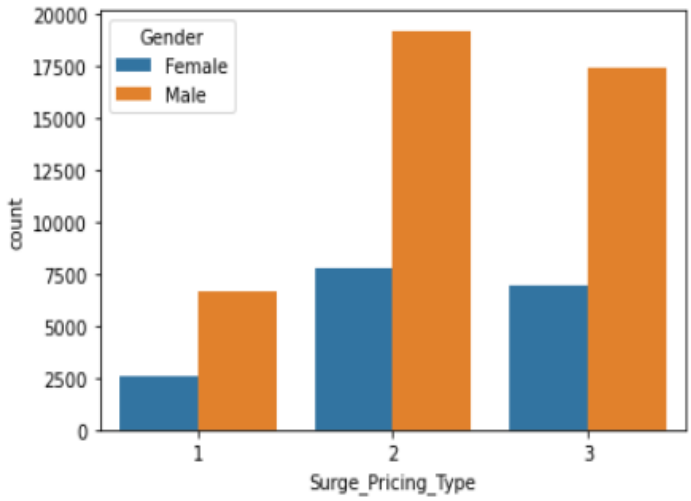

Type_of_Cab
B
C
E
D
A
count
Surge_Pricing_Type
1
2
3


Gender
Female
Male
count
Cancellation_Last_1Month


Type_of_Cab
B
C
E
D
A
count
Female
Male
Gender

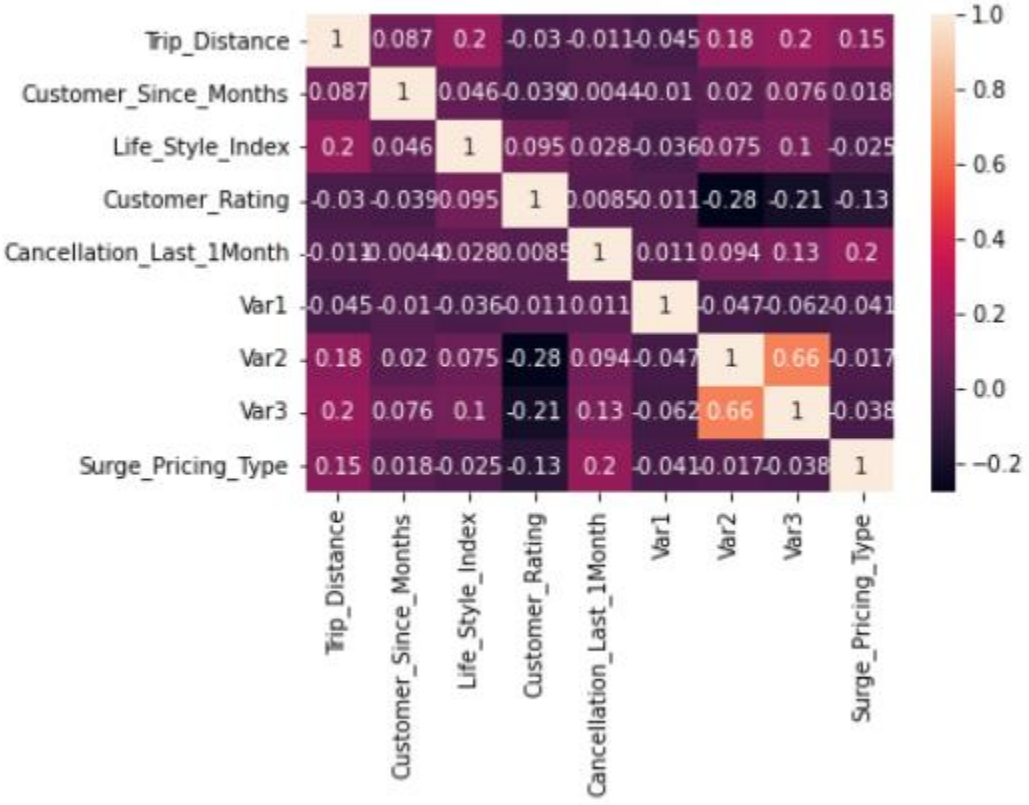

Customer_Rating
Female
Male
Gender


Gender
Female
Male
Cancellation_Last_1Month
Customer_Since_Months


Gender
Female
Male
count
Surge_Pricing_Type

- Customers who are male and the customers with more than 10 months are less prone to cancellation of cabs.
- 'A' type of cab has the highest median customer rating
- E and D type cabs seems to have more median cancellations last one month
- Median Trip distance covered by E Type cab seems highest.

Multivariate Analysis



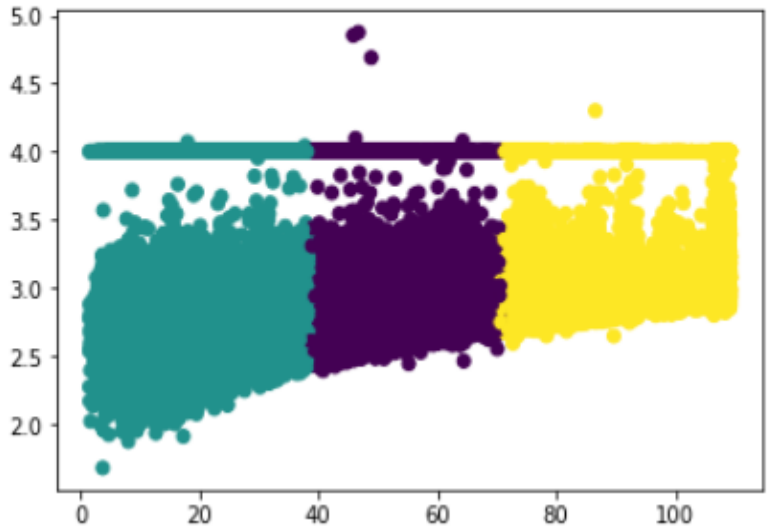
Segment Extraction

K means is one of the most popular Unsupervised Machine Learning Algorithms Used for Solving Classification Problems. K Means segregates the unlabelled data into various groups, called clusters, based on having similar features, common patterns

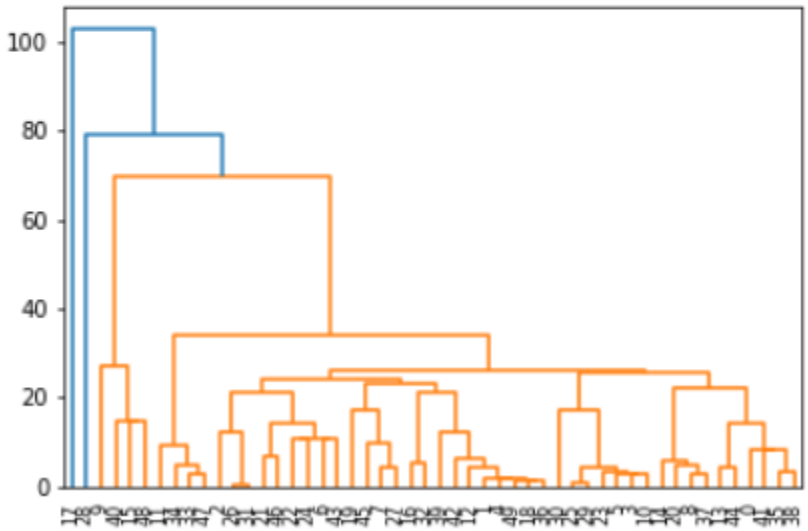
The technique to segregate Datasets into various groups, on the basis of having similar features and characteristics, is called Clustering. The groups being Formed are known as Clusters. Clustering is being used in Unsupervised Learning Algorithms in Machine Learning as it can segregate

multivariate data into various groups, without any supervisor, on the basis of a common pattern hidden inside the datasets

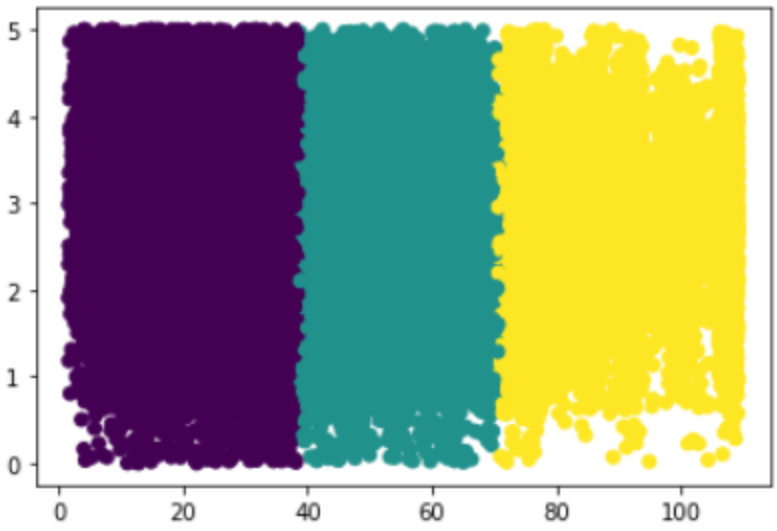
```
# making scatter plot b/w Trip_distance and Life_Style_Index
plt.scatter(c1.Trip_Distance, c1.Life_Style_Index, c=model.labels_)
plt.show()
```



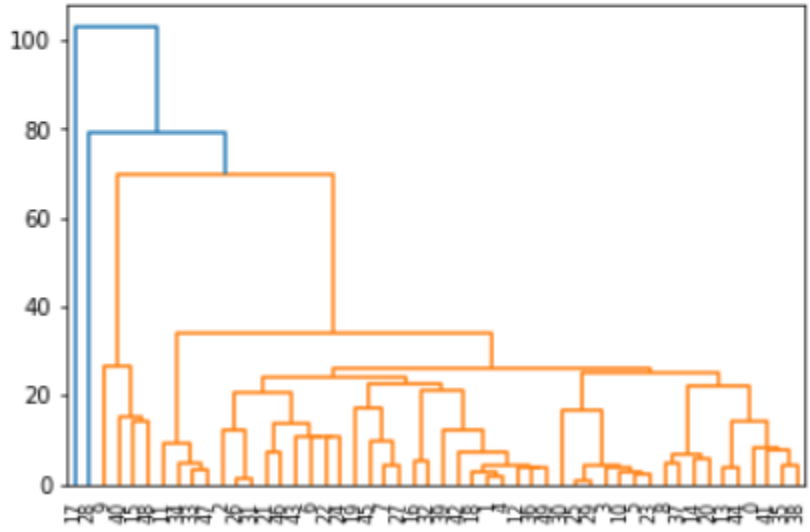

```
from scipy.spatial import distance_matrix
DM = pd.DataFrame(distance_matrix(c1.values, c1.values))
from scipy.cluster.hierarchy import dendrogram, linkage
dendrogram(linkage(DM))
plt.show()
```



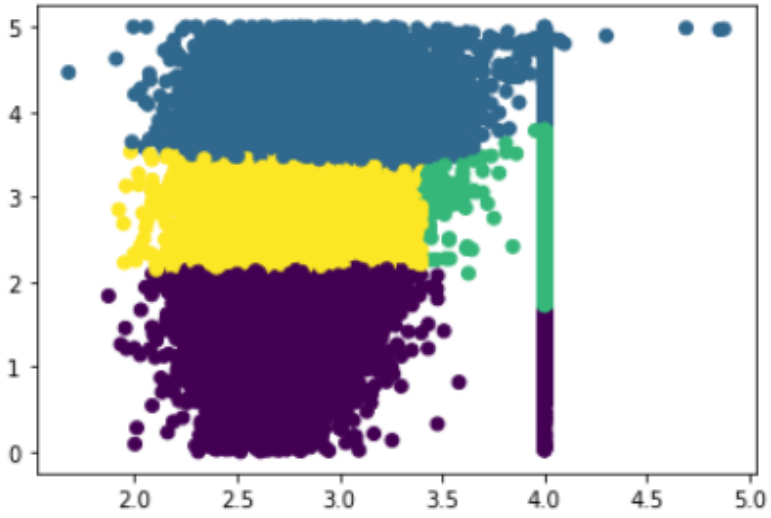
```
# scatter plot b/w Trip distance and Customer rating
plt.scatter(c2.Trip_Distance, c2.Customer_Rating, c=model.labels_)
plt.show()
```



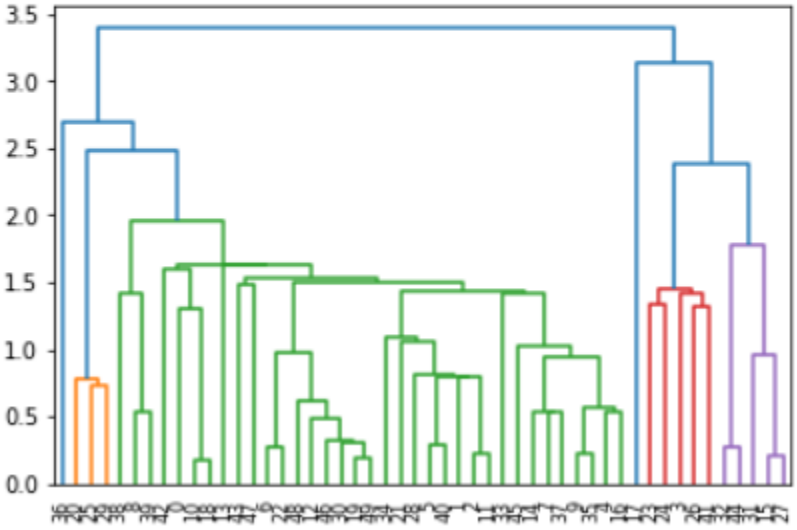

```
from scipy.spatial import distance_matrix
DM = pd.DataFrame(distance_matrix(c2.values, c2.values))
from scipy.cluster.hierarchy import dendrogram,linkage
dendrogram(linkage(DM))
plt.show()
```



```
# scatter plot b/w Life_Style_Index and Customer rating
plt.scatter(c3.Life_Style_Index, c3.Customer_Rating, c=model.labels_)
plt.show()
```




```
from scipy.spatial import distance_matrix
DM = pd.DataFrame(distance_matrix(c3.values,c3.values))
from scipy.cluster.hierarchy import dendrogram,linkage
dendrogram(linkage(DM))
plt.show()
```



Profiling of segments

Identifying Key Characteristics of Market Segments:

The aim of the profiling step is to get to know the market segments resulting from the extraction step. Profiling is only required when data-driven market segmentation is used. For common sense segmentation, the profiles of the segments are predefined. At the profiling stage, we inspect a number of alternative market segmentation solutions. This is particularly important if no natural segments exist in the data, and either a reproducible or a constructive market segmentation approach has to be taken.

Traditional Approaches to Profiling Market Segments:

Data-driven solutions are usually presented in two ways:

- high-level summaries simplifying segment characteristics to a point where they are misleadingly trivial
- large tables that provide, for each segment, exact percentages for each segmentation variable.

Segment Profiling with Visualizations:

Graphics, or visualizations, are important in statistical analysis because they can help to understand the relationships between variables and facilitate the interpretation of data segments. Visualizations can be useful in the process of dividing data into groups, or segments, to better understand and analyze them. The ability to see and compare different segmentation solutions through visualizations can help in choosing the best solution.

Describing Segments

The seventh step in market segment analysis is describing the segments that have been identified. This step involves gathering additional information about the segments, such as their demographics, past behavior, and attitudes towards products and brands. This information, known as descriptor variables, can be used to understand the characteristics of each segment in more detail and to develop a customized marketing strategy for each segment. There are two main ways to study the differences between segments using descriptor variables: through descriptive statistics and visualizations, or through inferential statistics. Visualizations can be especially useful in making the results of this analysis clearer and more accessible to users.

Selecting the Target Segment(s)

In order to select the best target market segment, a company will consider a number of factors such as the attractiveness of the segment and the company's competitiveness within that segment. To make this decision, the company will need to consider both which

segments they would most like to target and which segments are most likely to choose their products or services.

Market Segment Evaluation

The attractiveness of a segment is based on how appealing it is to the company, while competitiveness is based on how likely the company is to be chosen by the segment. To determine the value of these factors for each segment, the company will consider the profiles and descriptions developed in earlier steps of the market segmentation process and assign values to each segment based on these criteria. The resulting values are then plotted on a chart to help the company compare the different segments and make a decision about which one to target.

Customizing the Marketing Mix

Implications for Marketing Mix Decisions

The marketing mix is a set of tools used by businesses to create and promote their products or services. The most common version of the marketing mix includes four elements: product, price, promotion, and place. Product decisions relate to the design and features of a product, while price decisions involve setting the price and deciding on discounts. Place decisions involve deciding how and where the product will be sold, and promotion decisions involve creating and communicating an advertising message. Market segmentation, which involves dividing a larger market into smaller groups with similar characteristics, is often part of the overall process of designing a marketing strategy, along with positioning and competition.

Potential Target Market

Uber has two main target markets, demand-side and supply-side, and more specific use cases, demographics and service types. Analysts predict that the Uber stock price could reach $48.54 by Aug 3, 2023, with a potential upside of 74.49%. To maximize potential, businesses should create a model with extra segments and target a large customer base.

Suitable Early Market Strategy

Now we try to analyze which location in India is most suitable to create the early market in accordance with the Innovation Adoption Life Cycle.
The technology Adoption Life Cycle is a sociological model that describes the adoption or acceptance of a new product or innovation, according to the demographic and psychological characteristics of defined adopter groups. The process of adoption over time is typically illustrated as a classical normal distribution or "bell curve". The model indicates

that the first group of people to use a new product is called "innovators", followed by "early adopters". Next come the early majority and late majority, and the last group to eventually adopt a product are called "Laggards" or "phobic."
We defined demographic and psychological characteristics of our dataset they were:
a. Demographic: Trip Distance, Life Style Index, Destination Type etc.

b. Psychological: Confidence Life Style Index, Type of Cab, Customer Rating, Cancellation Last Month

Large Cities with remote areas where local transportation is readily available can be good areas to target. Metro cities offers a lot of business opportunities since a lot of potential customers comes to these places for studies, jobs and businesses and hence these areas can be considered as a suitable market to target.

GitHub Link:

https://github.com/abhikgupt/Online_Vehicle_Booking-MarketSegmentation