Market Segmentation for Online Vehicle Booking

-Rakesh Singh

Fermi Estimation (Breakdown of Problem Statement)

In the expansive landscape of India's vehicle booking service sector, a burgeoning startup envisions a pathway to substantial revenue generation through astute targeting of specific customer segments. Recognizing the pivotal role of segmentation analysis, the startup aims to unravel the most promising market niches, hitherto unexplored by competitors, and strategically tailor its services to meet the distinctive needs and preferences of these discerning segments. Setting its sights on differentiation, the startup contemplates forging strategic partnerships with local transportation providers, contemplating the infusion of value-added services such as in-car entertainment, Wi-Fi connectivity, and enticing loyalty programs. In a bold move to transcend conventional boundaries, the startup envisions venturing into untapped territories, specifically targeting rural areas with heightened cab frequency and identifying unconventional peak demand periods during the day. This strategic foray into uncharted segments forms the bedrock of the startup's market entry, a calculated move to disrupt the status quo of the well-established vehicle booking industry.

In a meticulous bid to meticulously sculpt its market presence, the startup not only seeks to redefine conventional notions of segmentation but also envisions delving into data-driven insights to uncover latent customer behaviors and preferences. The key imperative lies in identifying segments traditionally overlooked, yet possessing a substantial contribution potential to the overall sales dynamics of the car booking industry. A nuanced approach involves capitalizing on unique selling points, such as high cab frequency in rural pockets or capitalizing on periods of unusually high car booking requests during specific times of the day. This nuanced market entry strategy is underscored by an acute awareness that each member of the dataset brings a unique set of parameters and datasets to the table. As such, the segmentation process necessitates a judicious selection of segments, ensuring a balanced distribution of data and the discernibility of features within each cluster.

Furthermore, the startup recognizes the paramount importance of ensuring distinctiveness within each segment, juxtaposed with a pragmatic focus on features like higher or lower travel distances and minimized travel times. This deliberate emphasis on distinctive features aims not only to fortify the segmentation strategy but also to strategically position the startup's offerings in stark contrast to the existing cab booking services. In this intricate dance of data analytics and strategic foresight, the startup envisions a market presence that not only captivates the attention of the target audience but also redefines the contours of success in the dynamic and competitive realm of India's vehicle booking industry.

Data Sources

The datasets collected using database websites such as kaggle.com and each have different parameters relating to car booking service.

Data Pre-processing (Steps and Libraries used)

The data pre-processing was mainly done to remove duplicates, change values of blank or constant fields, and encode alphanumerical data into numerical data for clustering and the dropping of undesired columns and data. The main module used for data management was pandas.

Missing values in a dataset can affect k-means clustering by reducing the accuracy of the clustering results and potentially leading to biased or distorted clusters. K-means clustering algorithms typically require complete data, so missing values can disrupt the computation of distances between points and lead to incorrect clustering results. Hence, to remove the missing values, inbuilt functions of 'pandas' data frame management module.

Duplicate data in a dataset serves no extra purpose and will waste the computational powerof the program if there is too much duplicate data. It may also result in overfitting of data, which is undesirable. Hence, the duplicates are deleted using the drop_duplicates command, which is inbuilt in pandas.

Kmeans clustering requires data to be within a particular numerical range, which means that any categorical data would be of little to no use in a kmeans clustering algorithm. Hence, encoding or mapping of values to equivalent numerical data must be done before clustering can be applied. This is done using the df.map statement of pandas or using the replace command or using encoder statements such as label encoder, which is imported using the sklearn.preprocessing library.

Segment Extraction (ML techniques used)

K-Means Clustering:
- The main focus of the code is on applying the K-means clustering algorithm.
- Features like 'hour,' 'Status,' and 'Pickup point' are selected for clustering.
- Label encoding and handling of missing values are performed.
- K-means is initialized with two clusters (K=2) and fitted to the selected features.
- The resulting clusters are visualized using a scatter plot.

The code uses K-means clustering to group Uber request data into two clusters based on the specified features. The algorithm iteratively refines cluster assignments to minimize the sum of distances between data points and their corresponding centroids. The visualizations help in understanding the characteristics of different clusters and their distribution in the dataset.

Profiling and describing potential segments

In the code, profiling and describing potential segments involve analyzing various attributes to gain insights into different aspects of Uber request data. Here are the main attributes used for segment profiling:

1. Time-Related Attributes:

Request Hour: The hour of the day when the Uber request was made.

Drop Hour: The hour of the day when the trip was completed.

2. Status of Trips:

Cancelled Trips: Trips that were requested but later cancelled by either the driver or the passenger.

Trips with No Cars Available: Instances where users requested rides, but no cars were available.

Completed Trips: Successful trips that were completed.

3. Pickup Point:

City to Airport Trips: Trips originating in the city and headed to the airport.

Airport to City Trips: Trips originating at the airport and headed to the city.

4. Cluster Information:

K-Means Clusters: The data is segmented into clusters using the K-means algorithm based on features like request hour, status, and pickup point.

5. Demand and Supply Gap:

Supply: Number of completed trips.

Demand: Overall number of trips requested.

Supply-Demand Gap: Visualized through pie charts, indicating the proportion of completed trips to total requested trips.

6. Visualizations:

Pie Charts: Used to show the distribution of trip statuses (Cancelled, No Cars Available, Trip Completed).

Bar Charts: Illustrate the supply-demand gap in terms of completed trips.

Histograms: Display the distribution of cancelled and unavailable trips across different time slots.

By analyzing these attributes, the code aims to profile and describe potential segments within the Uber request data. The insights derived from these attributes help in understanding patterns, identifying issues (such as high cancellation rates or no cars available), and ultimately segmenting the data for further analysis and targeted strategies.

Selection of target segment

The selection of the target segment involves a meticulous analysis of both provided codes in the context of the cab services industry. The first code extensively profiles potential segments, considering various trip statuses and pickup points. It emphasizes selecting segments with high completion rates and low cancellation rates. The second code employs KMeans clustering on start hours and miles, revealing distinct user preferences. The optimal target segment should ideally exhibit favorable values in completion rates and specific user behaviors identified through clustering. By combining insights from both codes, a holistic approach is recommended, taking into account completion rates, low cancellation rates, and unique user preferences. This approach ensures the selection of a target segment that aligns with the company's goals and enhances the effectiveness of marketing strategies and service customization.

Customizing the Marketing Mix

From the First Code:

- The code analyzes a dataset related to cab services and primarily focuses on understanding the distribution of trips based on various factors such as status, hours, and pickup points.
- It provides insights into potential segments like canceled trips, unavailable cars, and completed trips but doesn't use segmentation variables for targeted marketing strategies.

From the Second Code:

- The code involves the analysis of Uber trip data, focusing on start hours, miles, and purposes of trips.
- It performs K-Means clustering on start hours and miles, creating clusters that could represent different user behaviors.
- Visualizations are generated to understand the distribution of trip distances and purposes.
- The clusters derived from the analysis could be utilized for customizing the marketing mix based on distinct user segments identified through clustering.

Customizing the Marketing Mix:

- Both codes could benefit from an explicit customization of the marketing mix based on the segments identified in the analysis.
- In the second code, the clusters formed using KMeans could be used to tailor marketing strategies for different user behaviors.

For example:

- Targeting users in clusters with specific start hours, emphasizing the availability of cabs during busy hours.
- Offering promotions or incentives for users in clusters with longer trip distances.

- Designing marketing messages based on the purposes of trips identified in the 'purpose' column.
- Both codes could explore innovative ways to differentiate services from competitors and focus on providing high-quality, personalized experiences for customers.

In conclusion, while the first code provides insights into potential segments, the second code, with its clustering analysis, offers a more explicit opportunity to customize the marketing mix based on identified clusters.

Potential customer base in the early market

The potential market size and revenue for a vehicle booking service in India:

1. Total population of India: 1.4 billion
2. Percentage of population with access to smartphones and internet connectivity: 30%
3. Percentage of smartphone users who use ride-hailing services: 30%
4. Total addressable market (TAM) for vehicle booking service in India: 126 millionusers (1.4 billion * 30% * 30%)
5. Average frequency of rides per user per month: 5
6. Average fare per ride: Rs. 250
7. Estimated monthly revenue for the vehicle booking service: Rs. 15.750 billion (126 million * 5 rides * Rs. 250 fare)

Price range could vary depending on the type of service chosen and distance covered. For example shared cab could cost much cheaper, let's say 50 Rs per person using the same cab Otherwise 20- 22 Rs per kilometer would be an ideal price range.

The most optimal market segments to open

Let's analyze both codes to identify potential optimal market segments for opening services:

From the First Code:

1. The code analyzes data related to cab services, focusing on canceled trips, unavailable cars, and completed trips.
2. It categorizes trips based on different statuses such as canceled, unavailable, and completed, and also considers pickup points (City, Airport).
3. Optimal market segments could be identified by considering the frequency and characteristics of completed trips. For example:
   - Trips completed from the City to the Airport or vice versa might represent a lucrative market segment.
   - Analyzing successful trip patterns can help tailor services to meet the needs of these segments.

From the Second Code:

1. The code analyzes Uber trip data using K-Means clustering on start hours and miles to identify clusters representing different user behaviors.
2. Optimal market segments could be determined based on the characteristics of each cluster:
   - Clusters with short distances and specific start hours may represent users with distinct preferences.
   - Clusters with long distances and different start hours may indicate a separate market segment.
3. By understanding these clusters, the company can customize its services and marketing strategies for each identified segment.

Combined Insights:

1. Combining insights from both codes, the most optimal market segments to open would be those where completed trips are frequent, ensuring a stable user base.
2. Optimal segments might exhibit specific patterns, such as high demand during particular hours or for certain distances, allowing for targeted strategies.
3. The first code's insights provide a broad overview of overall market dynamics, while the second code's clustering analysis offers a more nuanced understanding of user behaviors.

Conclusion:

In conclusion the most optimal market segments to open would be those where completed trips are frequent and user behaviors within these segments align with the characteristics identified through a combination of both codes.

Github Links:

https://github.com/711Rakesh/Market-Segmentation-for-Online-Vehicle-Booking.git