Market Segment Analysis of Online Vehicle Booking

by

Neha Mahagaonkar

Project Link:

[Market-Segmentation-of-online-vehicle-booking]

1.Fermi Estimation: Estimating Potential Market Size and Revenue for a New Cab Service

Fermi estimation involves making rough approximations to understand the scale of a problem using logical assumptions. Based on the data and analysis, we can perform a Fermi estimation to predict the potential early market size, number of customers, and expected revenue for a new cab service in India targeting specific segments.

Estimate Summary

Parameter	Estimated Value
Target Cities Population	~57 million
Addressable Market (15%)	~8.55 million
Early Adopters (16%)	~1.37 million users
Avg. Monthly Rides per User	8
Total Monthly Rides	~10.94 million
Avg. Fare per Ride	₹180
Monthly Revenue	₹197 crore
Annual Revenue	₹2,363 crore
Estimated Annual Profit (15%)	**₹354.45 crore**

This Fermi estimate gives a high-level projection of the early market size and potential profitability for launching a cab service in major Indian metros,

providing a data-backed foundation for strategic planning and investment decisions.

2. Data Preprocessing

Data preprocessing is a vital step in preparing raw data for analysis and machine learning. It ensures the dataset is clean, consistent, and suitable for downstream tasks such as segmentation.

The first step involves importing the raw data into a structured format like a data frame. Then, null values are identified and removed to avoid inconsistencies or errors during analysis.

One important transformation used in preprocessing is **one-hot encoding**, which converts categorical variables into a numeric format. This allows machine learning models to interpret and process the data effectively.

3. Visualization

Data visualization simplifies complex datasets, making patterns, relationships, and trends easier to understand. It helps in communicating insights effectively and can make data more engaging for deeper exploration.

Visual representations also support decision-making by providing a clearer understanding of key metrics that may not be evident through raw data.

Observations:

- **Source and destination points** are almost equally distributed, indicating a wide customer spread across locations. This is beneficial for optimizing fleet allocation and coverage.
- **Cloudy weather** correlates with the highest number of bookings, while **foggy days** see the least. Fog may discourage travel due to safety concerns.
- **Most cabs are booked after 10 P.M.**, showing strong late-night demand.
- **Month-ends are busier**, likely due to salary disbursement or increased errand runs.
- Customers **prefer budget-friendly cabs**, suggesting high price sensitivity among users.

4. Geometric Analysis

Geometric analysis involves studying the shape, size, and position of data points, offering insights into spatial patterns and regional behaviors.
Observations:
Most cab bookings originate from the Boston area, which may be due to:

- Higher population density
- More business activities
- Better reputation for reliable and affordable cab services

This makes Boston a strategic area for expanding or concentrating services.

5. Psychographic Analysis

Psychographic analysis explores customer values, interests, and lifestyles to understand their motivations. It helps businesses create targeted, customer-centric strategies.
Observations:

- Temperature distribution is roughly normal, with most values between 35°C to 45°C.
- Precipitation intensity peaks around 0.00, with high booking activity when precipitation is above 0.01 and humidity exceeds 0.8.
- The busiest booking days occur under cloudy weather, while foggy days result in fewer rides.

These patterns indicate that weather conditions significantly influence customer behavior.

6. Demographic Analysis

Demographic analysis focuses on population characteristics like age, gender, and income to tailor services effectively.
Observations:

- Peak booking hours are between 10 A.M. to 6 P.M., and after 10 P.M.
- End-of-month periods consistently see higher demand.
- Fewer bookings occur between the 4th and 13th of each month.
- Source and destination points are evenly spread, suggesting broad user adoption across different regions.

---

7. Behaviour Analysis

Behavior analysis examines customer actions and decisions, offering insights into their preferences and decision-making factors.
Observations:

- Most bookings are for budget cabs, priced between ₹5 to ₹25.

- Customers prefer shorter distances, typically from 0.5 to 3.5 units.
- As distance and fare increase, the likelihood of booking decreases. This shows a clear preference for affordability and convenience. Businesses should focus on pricing strategies and route optimization for short-distance travelers to retain customers.

8.Segment Extraction

Segment extraction is the process of dividing a broad market into smaller, more manageable subgroups based on shared characteristics. These characteristics may include demographics (age, gender, income), psychographics (lifestyle, values), behavioral patterns (purchase history, usage), and geographic location (city, region).
This approach helps businesses better understand their target audience, enabling them to design tailored marketing strategies, improve product offerings, and enhance customer experiences. By catering to the unique needs of each segment, businesses can improve customer satisfaction, foster loyalty, and ultimately drive higher profitability.

8.1. Clustering

Clustering is a type of **unsupervised machine learning** technique used to group similar data points together based on features they share. It is particularly useful in **market segmentation**, as it helps reveal patterns and structures within a dataset without predefined labels.
In this case, **K-Means Clustering** was used. K-Means works by:

- Selecting *K* initial centroids (number of clusters).
- Assigning each data point to the nearest centroid based on distance (usually Euclidean).
- Recomputing centroids based on the current cluster members.
- Repeating the assignment and update steps until convergence.

To determine the optimal number of clusters (K), we used the **Elbow Method**. This technique involves plotting the **within-cluster sum of squares (WCSS)** for different values of K. The “elbow point” on the graph—where the rate of WCSS decrease slows—suggests the most suitable number of clusters. Based on this method, **4 clusters** were identified.
Why K-Means?

- It is simple and fast, making it efficient for large datasets.
- It works well when clusters are spherical and equally sized.
- It gives a clear structure to the data which can be easily interpreted.

In [5]:
```
numeric_cols = df.select_dtypes(include=['float64', 'int64']).columns.tolist()
df_numeric = df[numeric_cols]

scaler = StandardScaler()
scaled_df = scaler.fit_transform(df_numeric)

pca = PCA(n_components=40)
pca_df = pca.fit_transform(scaled_df)
pca_df = pd.DataFrame(pca_df)
```

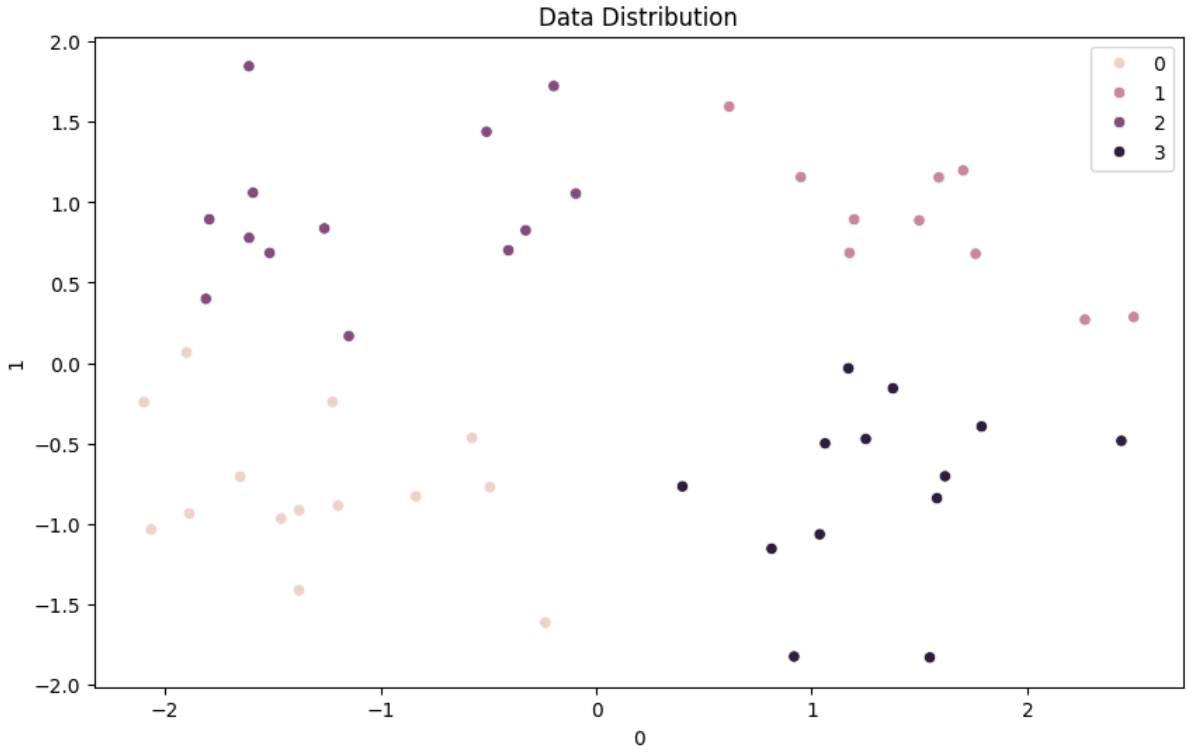
In [6]:
```
pca_df.head()
```

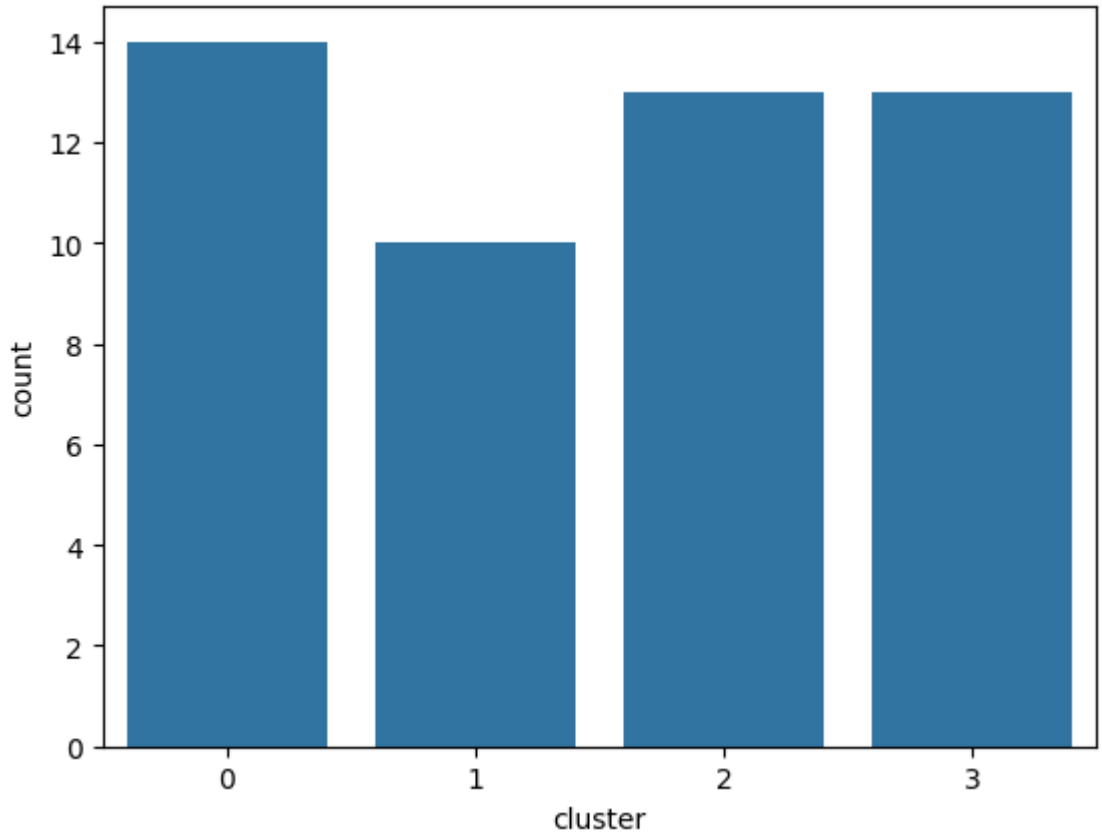
Out[6]:

	0	1	2	3	4	5	6	7	8	9	...	30	31	32	33	
0	4.250572	2.337611	1.241928	0.691632	0.582285	2.322285	-2.067313	-2.618848	-0.325984	0.789120	...	-0.087545	-0.021262	-0.020030	0.067880	0.04
1	-6.035866	1.161779	4.821680	-1.759994	0.901835	1.463514	-1.500048	-2.021841	2.005597	0.786049	...	0.017192	0.019638	0.008228	-0.011297	-0.00
2	-3.711734	-2.060587	-0.322583	0.180432	2.000880	0.675256	-1.697733	-2.845596	0.436642	0.773736	...	0.032695	0.006346	-0.000417	-0.002700	-0.00
3	-2.049622	-2.858534	-1.255705	-0.771054	2.393263	1.021622	-0.306492	-2.402638	0.414025	0.734479	...	0.068318	-0.036385	-0.003710	-0.007223	0.02
4	-2.565964	-3.698104	0.127494	0.273182	2.084639	0.933597	-1.581177	-2.241277	-0.275702	0.752221	...	0.014979	-0.026694	-0.018628	0.010833	0.00

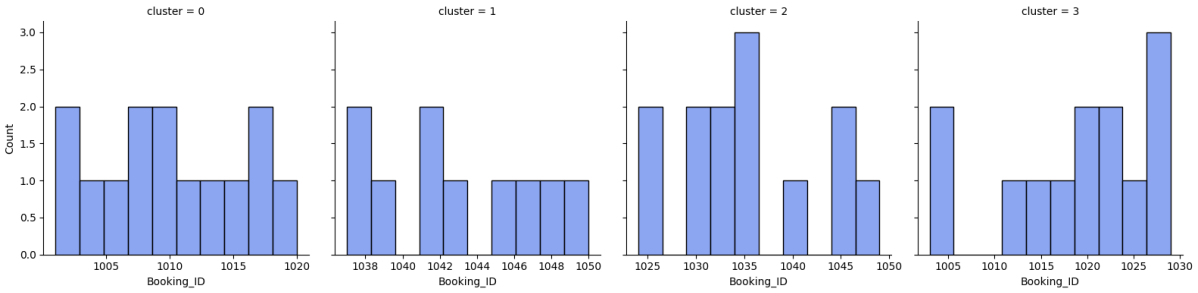
5 rows × 40 columns

In [8]:
```
kmeans = KMeans(n_clusters=4)
kmeans.fit(pca_df)
KMeans (n_clusters=4)

np.random.seed(42)
preds = kmeans.predict(pca_df)

plt.figure(figsize=(10,6))
sns.scatterplot(x=pca_df[0],y=pca_df[1],hue=preds)
plt.title('Data Distribution')
plt.show()
```





In [ ]:
```
sns.set_palette('coolwarm')

for i in df.drop('cluster', axis=1):
    grid = sns.FacetGrid(df, height=4, col='cluster', sharex=False)
    grid = grid.map(sns.histplot, i, bins=10)
    plt.show()
```



[Market-Segmentation-of-online-vehicle-booking]

8.2. Profiling Segments

Profiling segments involves describing and analyzing the characteristics of each cluster created by the K-Means algorithm. This analysis includes:

- **Demographic attributes** like age, gender, and income.
- **Psychographic traits** like preferences, interests, and lifestyles.
- **Behavioral data** such as frequency of usage, booking times, fare preferences, and weather conditions.

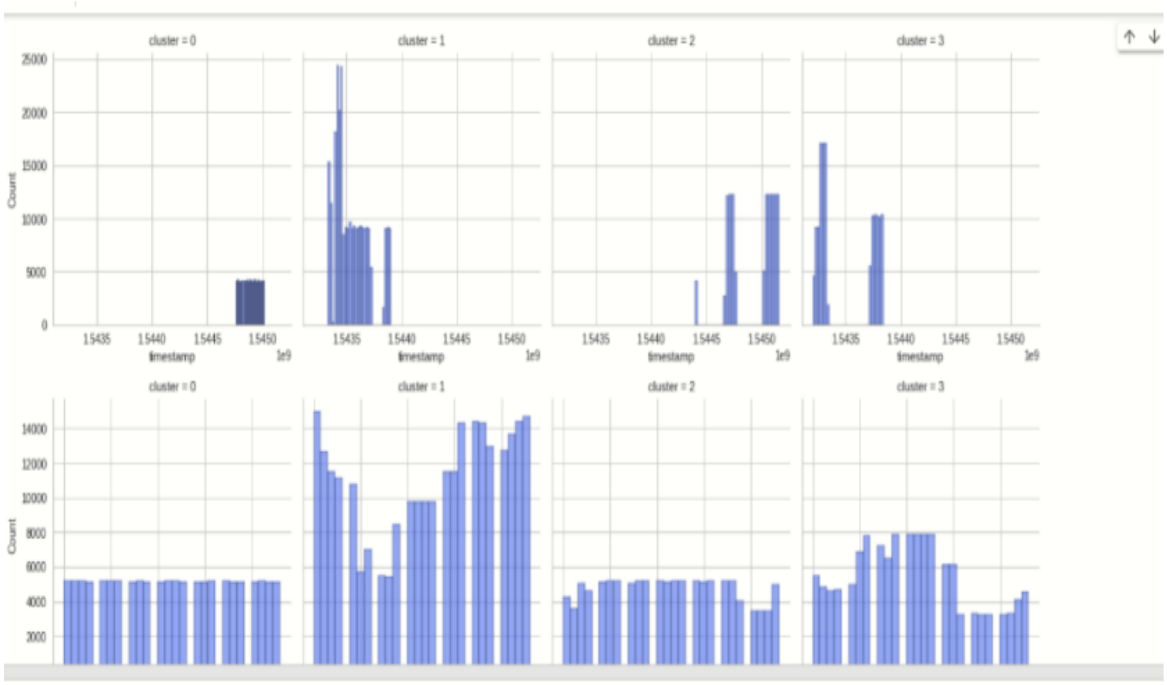
- **Geographic patterns**, like the most frequent pickup and drop-off locations.

By creating these profiles, businesses can gain insights into:

- What motivates each group.
- When and why they are most likely to make a purchase.
- What types of offers or services will appeal to them most.

For example, one segment may consist of **late-night budget riders in urban areas**, while another might include **high-income occasional users who prefer premium services**.

Profiling empowers companies to make **data-driven decisions** in product development, customer service, and marketing communication—ensuring that each segment receives targeted and relevant engagement.



9.Customizing the Marketing Mix

For a successful launch in the business market, the marketing mix must be carefully tailored to the needs and expectations of different customer segments. The product offering should include both premium and budget cab options. Business travelers may prefer executive vehicles with trained drivers, while students and budget-conscious riders will be attracted to affordable, reliable transportation.

Pricing strategies should be flexible. Premium pricing can be applied to business travelers and high-demand zones, while discounts and student fares can be used to attract price-sensitive segments. A penetration pricing strategy in the early stages can help acquire customers rapidly.
The place or distribution channel should rely heavily on a seamless, user-friendly mobile app, supported by a web portal. The app must support easy ride booking, real-time tracking, feedback systems, and payment options to ensure smooth service delivery.
Promotion should be focused on digital platforms and should include targeted campaigns, social media advertising, influencer collaborations, and strategic partnerships with universities, hotels, and corporate hubs. Referral programs, seasonal offers, and first-ride discounts can boost early adoption.

10.Estimating the Potential Customer Base and Early Market Profit

Based on Fermi estimation and segmentation analysis, cities like Bangalore, Mumbai, Delhi, and Hyderabad present the most promising early markets. With a combined urban population of approximately 57 million, and assuming 15% of this population engages in regular cab usage, the addressable market stands at roughly 8.55 million people.
Focusing on early adopters and innovators, who typically constitute 16% of the addressable market, the early market comprises around 1.37 million potential users.
Assuming an average fare of ₹180 and each user completes 8 rides per month, the potential monthly sales are:

- 1.37 million users × 8 rides × ₹180 = ₹1,969 million (or ₹197 crore/month)

Assuming a 15% profit margin, the monthly profit in the early market would be approximately ₹29.5 crore, and annual profit close to ₹354 crore

11.Most Optimal Market Segments

Based on market research and data-driven segmentation, the most optimal customer segments to target at launch are:

1. Business Travelers – Require premium service, willing to pay more for reliability and professionalism.
2. Tourists – Prefer affordable, convenient travel options and are more likely to book ahead for airport transfers and sightseeing.
3. College Students – Highly price-sensitive, prefer budget rides and frequent short-distance trips.
4. Elderly Customers – Value safety, comfort, and assistance, ideal for tailored services with accessible vehicles and trained drivers.

These segments show the highest potential in both volume and profitability, and targeting them with customized service offerings will maximize early traction and customer satisfaction.

12.Conclusion

Based on a detailed analysis of cab booking data from late 2018, valuable insights have emerged to guide the market entry of an online vehicle booking startup in India. A key finding is that customers strongly prefer budget-friendly rides, particularly for short distances between 0.5 and 3.5 kilometers. Most fares fall within the ₹5 to ₹25 range.

Demand for cab services peaks between 10 A.M. and 6 P.M., with another spike after 10 P.M. Additionally, the end of each month consistently sees the highest volume of bookings, while the first half of the month remains relatively quiet. Bookings are uniformly distributed across source and destination points, suggesting that demand is widespread rather than concentrated.

Environmental conditions also influence customer behavior. Cab bookings tend to increase on days with precipitation above 0.01 and humidity greater than 0.8. Cloudy weather corresponds with the highest number of rides, while foggy conditions see the least. Temperature values are normally distributed, mostly ranging between 35°C and 45°C. These trends highlight the importance of incorporating weather data into demand forecasting.

To better understand customer groups, segmentation analysis was conducted using Principal Component Analysis (PCA) and KMeans clustering. This revealed distinct user segments: business travelers who need premium, punctual service; tourists seeking safe and affordable transportation; college students who prioritize low-cost and convenient commutes; and elderly individuals who value comfort and assistance.
For initial market entry, the startup should focus on cities that align with the early phases of the Innovation Adoption Lifecycle. Bangalore, Mumbai, Delhi, and Hyderabad are ideal due to their large populations of innovators and early adopters. These urban centers offer a fertile ground for introducing new technology-driven services.

The go-to-market strategy should leverage a balanced marketing mix. The product range must cater to both budget-conscious users and those seeking premium services. Special attention should be given to student and senior citizen offerings. Pricing should remain competitive yet adaptable to various customer needs and service costs.

The platform should be delivered through a reliable and easy-to-use mobile app. Promotion strategies should focus on digital channels, discounts, and partnerships with hotels, universities, and tourist services to build trust and expand reach.

In conclusion, the use of data analytics enables precise segmentation and smarter decision-making. By aligning offerings with customer behaviors—especially price sensitivity, time patterns, and weather conditions—the startup can build a strong market presence. Emphasizing customer feedback and adaptability will be key to long-term success in India’s fast-growing mobility market.
