EV market segmentation

Subhankar Pal

July 3, 2023

1 Introduction

Market segmentation plays a vital role in the advancement of transportation technology, particularly in emerging markets, such as electric vehicles (EVs), by facilitating widespread acceptance and implementation. The adoption of EVs is anticipated to experience exponential growth in the near future due to their environmentally friendly nature and cost-effective operation. This growth has sparked significant interest among researchers, prompting them to explore and identify distinct groups of potential EV buyers based on their psychographic, behavioral, and socio-economic characteristics. To achieve this, a comprehensive study was employed, and rigorous analytical techniques including cluster analysis were used to validate and operationalize the identified segments. The findings revealed three distinct segments of young consumers: 'Conservatives,' 'Indifferents,' and 'Enthusiasts,' who exhibit characteristics indicating their potential as future EV buyers. The implications of these findings are recommended to provide valuable insights for scholars and policymakers, facilitating the promotion of EV adoption within the emerging sustainable transport market.

- **KeyWords :** Electric vehicles, Market segmentation, Cluster analysis, KMeans Clustering, PCA, Dendogram, Attitude towards electric vehicles.

2 Problem Statement

The popularity of electric vehicles (EVs) is surging worldwide as an eco-friendly alternative to traditional gasoline cars. In India, the demand for EVs has been steadily rising due to concerns about air pollution, escalating fuel prices, and government incentives. As an Electric Vehicle Startup, it is crucial for us to thoroughly examine the Indian EV market and develop a viable strategy that focuses on the segments most inclined to embrace EV usage.

Our objective is to conduct an extensive analysis of the Indian EV market utilizing segmentation analysis. We will take into account various factors such as geographic, demographic, psychographic, and behavioral data to pinpoint the most suitable location to establish an early market in accordance with the Innovation Adoption Life Cycle. Additionally, we will scrutinize available

datasets to identify potential segments that are likely to adopt EVs, and tailor the marketing mix to effectively target these segments.

3 Data Collection and Preprocessing

The data has been collected from `https://www.kaggle.com/`. The data collected is compact and is used for visualization purposes and for clustering.

Features: The dataset used in this project contain 1000 rows and 15 features. The features are as follows: Age, City, Profession, Marital Status, Education, No. of Family members, Annual Income, Preference for replacing all vehicles to Electronic vehicles?, If Yes/Maybe what type of EV would you prefer?, Do you think Electronic Vehicles are economical?, Which brand of vehicle do you currently own?, How much money could you spend on an Electronic vehicle?, Preference for wheels in EV, Do you think Electronic vehicles will replace fuel cars in India?

First we load the dataset after importing necessary libraries.

```
data=pd.read_csv("Electric_Vehicle_Data.csv")
data_copy=data
data.head()
```

Next we look at some information and statistics of the dataset.

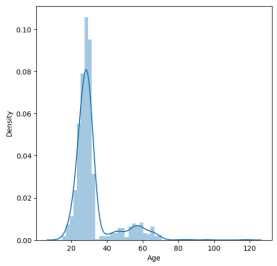
```
data.describe()
```

	Unnamed: 0	Age	No. of Family members	Annual Income	Preference for wheels in EV
count	1000.000000	1000.000000	1000.000000	1.000000e+03	1000.000000
mean	499.500000	31.800000	4.118000	2.258342e+06	3.349000
std	288.819436	11.294847	1.469774	9.993558e+05	0.887686
min	0.000000	15.000000	0.000000	-3.761509e+05	2.000000
25%	249.750000	26.000000	4.000000	1.782116e+06	2.000000
50%	499.500000	29.000000	4.000000	2.329246e+06	4.000000
75%	749.250000	31.000000	5.000000	2.753170e+06	4.000000
max	999.000000	118.000000	8.000000	1.282128e+07	4.000000

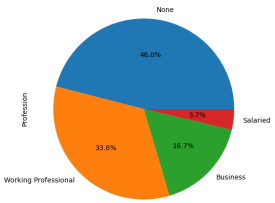
Figure 1: Data Statistics

Before conducting the analysis, the datasets were preprocessed to ensure data quality and accuracy. This involved data cleaning, handling missing values, resolving inconsistencies, and transforming categorical variables into numerical representations if required. The first unnammed column containing the serial number of the rows was dropped by writing the code

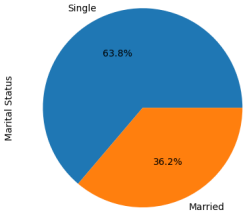
```
data=data.drop("Unnamed: 0",axis=1)
```



(a) *AgeDensity*



(b) *Profession*



(c) *MartialStatus*

Figure 2: Three simple plots

Then we check for missing values. But the dataset contained no missing values.

```
data.isnull().sum()
```

3.1 Packages/Tools used:

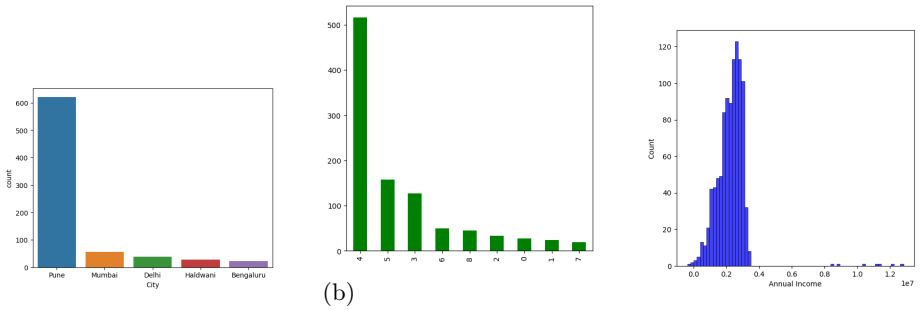
- **Numpy**: To calculate various calculations related to arrays.
- **Pandas**: To read or load the datasets.
- **Matplotlib**: To create static, animated, and interactive visualizations.
- **SKLearn**: To model the KMeans algorithm and PCA. We also have used LabelEncoder() to encode our values.
- **Seaborn**: It provides a high-level interface for creating informative and attractive statistical graphics.
- **Scipy**: To solve the complex scientific and mathematical problems.

4 Data Visualization

By using various data visualization methods like pie-plot, distplot, countplot, histplot we visualize three important characteristics, namely age, profession and martial status, of the dataset.[Figure 2]

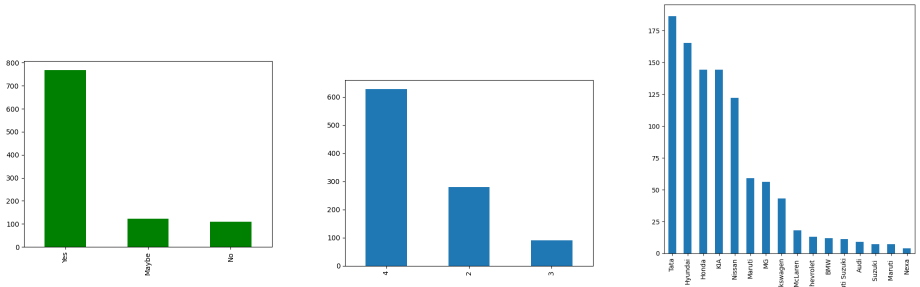
```
sns.distplot(data["Age"])
data["Profession"].value_counts().plot(kind="pie",autopct="%0.1f%%")
data["Marital Status"].value_counts().plot(kind="pie",autopct="%0.1f%%")
```

Then we use the countplot and histplot to visualize the counts based on city, number of family members and family income.[Figure 3]



(a) *City* (b) *NumberOfFamilyMembers* (c) *AnnualIncome*

Figure 3: Counts



(a) *EVsAreEconomicalORnot* (b) *PreferenceForWheelsInEV* (c) *CurrentBrand*

Figure 4: Preferences

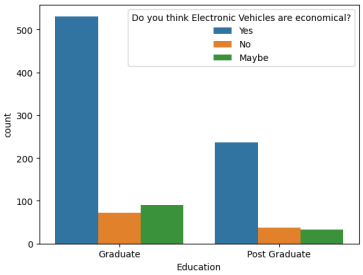
```
sns.countplot(x="City",data=data,order=data.City.value_counts().iloc[:5].index)
data["No. of Family members"].value_counts().plot(kind="bar",color="green")
sns.histplot(data["Annual Income"],color="blue")
```

Now, we find the counts of various features based on customer preferences.[Figure 4]

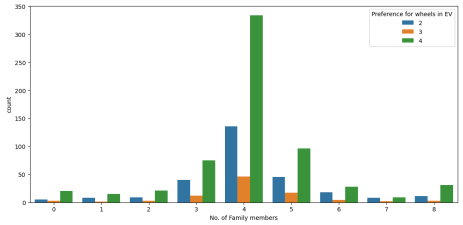
```
data["Do you think Electronic Vehicles are economical?"].value_counts().plot(kind="bar",
data["Preference for wheels in EV"].value_counts().plot(kind="bar")
data["Which brand of vehicle do you currently own?"].value_counts().plot(kind="bar")
```

Next, we use the countplot to visualize some features by combining them.[Figure 5]

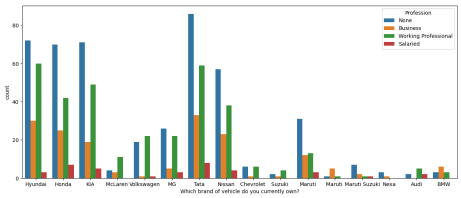
```
sns.countplot(x="Education",data=data,hue="Do you think Electronic Vehicles are economic
sns.countplot(x="No. of Family members",data=data,hue="Preference for wheels in EV")
sns.countplot(x="Which brand of vehicle do you currently own?",data=data,hue="Profession
```



(a) $EducationVSEconomical$



(b) $FamilyMembersVSWheels$



(c) $BrandVSProfession$

Figure 5: Combined Features

5 Segmentation using KMeans Clustering Algorithm

5.1 Correlation of the Features

A correlation matrix is simply a table that displays the correlation. It is best used in variables that demonstrate a linear relationship between each other. The matrix depicts the correlation between all the possible pairs of values through the heatmap in the below figure. The correlation of the above features can be found by writing the code segment-

```
corr_matrix=data.corr()
sns.heatmap(corr_matrix,annot = True, cmap= "Spectral")
```

The correlation matrix between the features is attached in Figure 6.

5.2 KMeans Clustering

K Means algorithm is an iterative algorithm that tries to partition the dataset into pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It tries to make the intra-cluster data points as similar as possible while also keeping the clusters as different (far) as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the cluster's centroid (arithmetic mean of all the data points that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data points are within the same cluster. The way k means algorithm works is as follows:

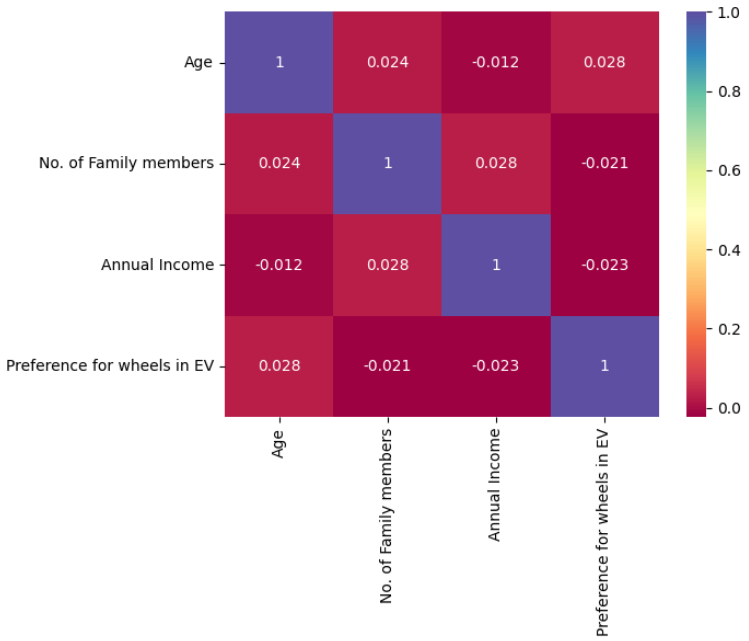


Figure 6: Correlation Matrix

- Specify number of clusters K.
- Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
- Compute the sum of the squared distance between data points and all centroids. Assign each data point to the closest cluster (centroid).
- Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster.
- Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.

The approach k-means follows to solve the problem is expectation maximization The E-step is assigning the data points to the closest cluster. The M-step is computing the centroid of each cluster.

1. **Determining the Number of Clusters:** The Elbow method is a popular method for determining the optimal number of clusters. The method is based on calculating the Within-Cluster-Sum of Squared Errors (WSS) for a different number of clusters (k) and selecting the k for which change in WSS first starts to diminish. The idea behind the elbow method is that the explained variation changes rapidly for a small number of clusters and

then it slows down leading to an elbow formation in the curve. The elbow point is the number of clusters we can use for our clustering algorithm.

The Elbow method was used to identify the optimal number of clusters in the dataset. This technique involved running K-means clustering with a varying number of clusters and selecting the number that resulted in an elbow point in the plot of the sum of squared distances. The appropriate number of clusters was found to be $n = 4$ according to the Elbow method.[Figure 7]

```
from sklearn.cluster import KMeans
# Determine the number of clusters using the Elbow method
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42)
    kmeans.fit(data)
    wcss.append(kmeans.inertia_)

# Plot the Elbow curve
plt.plot(range(1, 11), wcss)
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.title('Elbow Method')
plt.grid()
plt.show()
```

2. **K-means Clustering:** K-means clustering was applied to segment the dataset into distinct clusters based on the selected features. The identified clusters represented different consumer segments within the EV market in India.

```
# Perform K-means clustering with the optimal number of clusters
kmeans = KMeans(n_clusters=4, init='k-means++', random_state=42)
kmeans.fit(data)
```

3. **PCA implementation and visualization:** Principal Component Analysis (PCA) was employed to visualize the clusters in a lower-dimensional space. This technique reduced the dimensionality of the data while retaining the most significant information. The clusters were then plotted using PCA for visualization purposes.

```
from sklearn.decomposition import PCA

# Perform PCA
pca = PCA(n_components=4)
pca_data = pca.fit_transform(data)
```

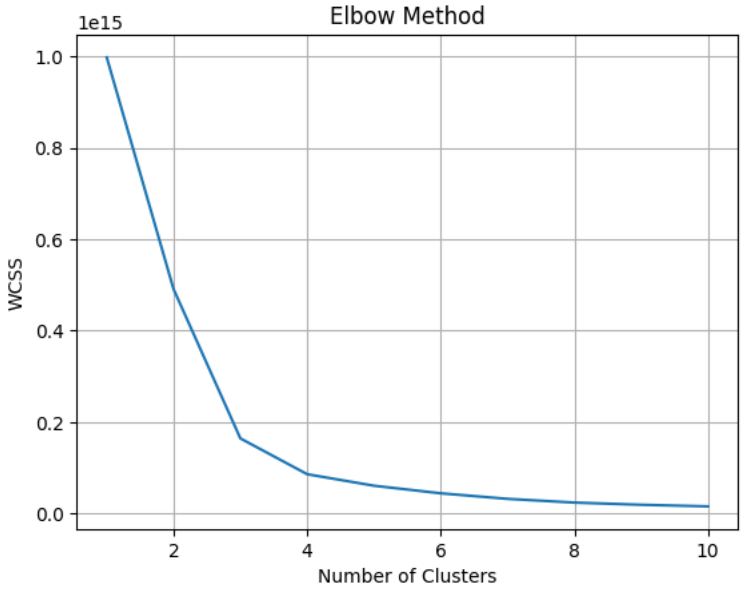


Figure 7: Find the No. of Clusters using the Elbow method

```
data2 = pd.DataFrame(pca_data, columns=['PC1', 'PC2','PC3','PC4'])
data2.head()
```

Next, to visualize [Figure 8] the clusters in a two-dimensional space using PCA we have the following code segment-

```
# Add the cluster labels to the dataset
data['Cluster'] = kmeans.labels_
# Visualize the clusters
plt.scatter(pca_data[:, 0], pca_data[:, 1], c=data['Cluster'], cmap='viridis')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.title('EV Market Segmentation - Cluster Analysis')
plt.show()
```

4. **Dendrogram** This technique is specific to the agglomerative hierarchical method of clustering. The agglomerative hierarchical method of clustering starts by considering each point as a separate cluster and starts joining points to clusters in a hierarchical fashion based on their distances. To get the optimal number of clusters for hierarchical clustering, we make use of a dendrogram which is a tree-like chart that shows the sequences of merges or splits of clusters. If two clusters are merged, the dendrogram will join

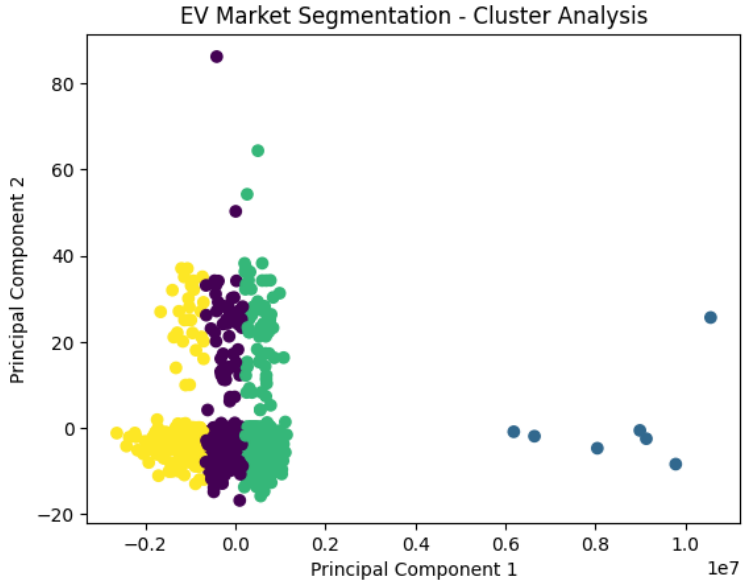


Figure 8: Visualize the clusters using PCA

them in a graph and the height of the join will be the distance between those clusters. As shown in Figure, we can chose the optimal number of clusters based on hierarchical structure of the dendrogram.[Figure 9]

```
from scipy.cluster.hierarchy import dendrogram, linkage

linked = linkage(data2, 'complete')
plt.figure(figsize=(13, 9))
dendrogram(linked, orientation='top')
plt.show()
```

6 Market Segmentation Analysis

The market segmentation approach aims to identify distinct and manageable subgroups of customers who can be targeted with specific marketing strategies. In practice, there are two main methods of segmenting the market: a-priori and post-hoc. The a-priori approach involves predefining segments based on characteristics like age, gender, income, and education, followed by profiling using various measured variables such as behavior, psychographics, or benefits. On the other hand, the post-hoc approach identifies segments based on relationships among multiple measured variables. In both approaches, the measured variables determine the segmentation theme. This study adopts the a-priori approach to segment potential EV customers into sub-groups.

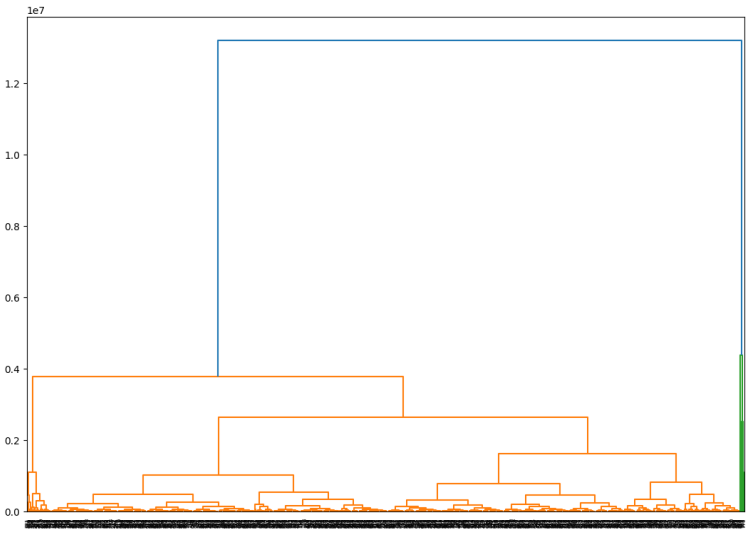


Figure 9: Dendogram

It is argued that utilizing a combination of psychographic and socioeconomic attributes for market segmentation allows for the development of sub-market strategies that cater to specific consumer preferences. The present study expands on this by incorporating perceived-benefit characteristics guided by a blend of psychographic and socio-economic aspects for segmenting the consumer market The target market of Electric Vehicle Market Segmentation can be categorized into Demographic, Geographic, Psychographic, and Behavioral Segmentation.

6.1 Demographic Analysis

Demographic analysis was performed to identify patterns based on variables such as age group, education and income group, profession, education, and annual income. This analysis helped uncover trends specific to different demographic segments.[figure 10]

```
Cluster_0=data_copy[data_copy.Clusters==0]
Cluster_1=data_copy[data_copy.Clusters==1]
Cluster_2=data_copy[data_copy.Clusters==2]
Cluster_3=data_copy[data_copy.Clusters==3]

    [Cluster_0["Age"].value_counts().head(),
    Cluster_1["Age"].value_counts().head(),
    Cluster_2["Age"].value_counts().head(),
    Cluster_3["Age"].value_counts().head()]

    [Cluster_0["Annual Income"].mean(),
```

```
[29    45
 28    42
 30    37
 31    36
 27    35
 Name: Age, dtype: int64,
 29    48
 28    43
 30    43
 25    39
 26    39
 Name: Age, dtype: int64,
 30    2
 26    1
 29    1
 56    1
 28    1
 Name: Age, dtype: int64,
 31    27
 30    23
 29    22
 27    18
 28    16
 Name: Age, dtype: int64]
```

(a) *AgeClusters*

```
[Graduate         260
 Post Graduate    106
 Name: Education, dtype: int64,
 Graduate         300
 Post Graduate    134
 Name: Education, dtype: int64,
 Graduate         5
 Post Graduate    2
 Name: Education, dtype: int64,
 Graduate         128
 Post Graduate     65
 Name: Education, dtype: int64]
```

(b) *EducationClusters*

```
[None                   158
 Working Professional   133
 Business                61
 Salaried                14
 Name: Profession, dtype: int64,
 None                   210
 Working Professional   136
 Business                73
 Salaried                15
 Name: Profession, dtype: int64,
 None                   5
 Working Professional   1
 Business               1
 Name: Profession, dtype: int64,
 None                   87
 Working Professional   66
 Business               32
 Salaried                8
 Name: Profession, dtype: int64]
```

(c) *ProfessionClusters*

Figure 10: Demographic Analysis

```
Cluster_1["Annual Income"].mean(),
Cluster_2["Annual Income"].mean(),
Cluster_3["Annual Income"].mean()]
```

From the Demographic analysis, it is revealed that

- Age group of 28-31 needs to be targeted.
- The persons having profession as "Working Professionals" are more likely to buy EVs.
- Graduate people needs to be targeted most.
- Cluster 1 has the highest average income, followed by cluster 0,3,2.

6.2 Geographic Analysis

Geographic analysis was conducted to explore regional variations in EV adoption and preferences. The "City" variable was used to analyze the concentration of EV users and potential market opportunities across different cities in India.[Figure 11]

```
[Cluster_0["City"].value_counts().head(),
Cluster_1["City"].value_counts().head(),
Cluster_2["City"].value_counts().head(),
Cluster_3["City"].value_counts().head()]
```

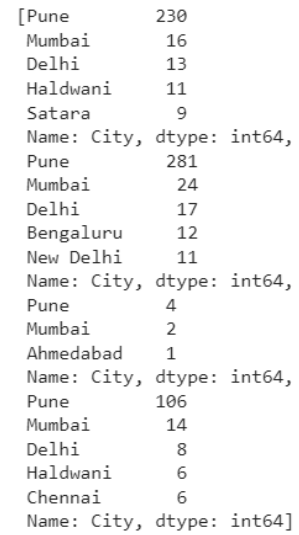


Figure 11: Geographic Clusters

From the Geographic analysis we have the following decisions:

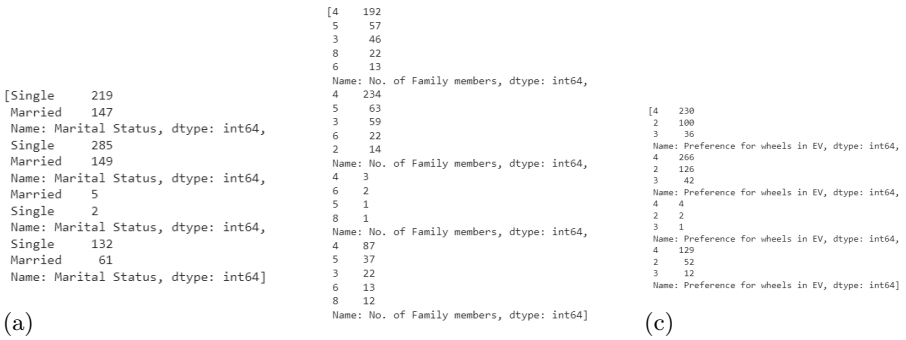
- Pune and Mumbai are present in all the cluster and they are more suitable to target the EV Market.
- Delhi, Haldwani and Bengaluru should be the next city to target followed by Satara and Chennai.

6.3 Psychographic Analysis

This segmentation grouped based on beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product. Suitable for lifestyle segmentation. The analysis was performed to identify patterns based on variables such as marital status, number of family members, type of preferred EV, current brand, EV brand preference and expected spending on EVs. This analysis helped uncover trends specific to different psychographic segments.[Figure 12,13]

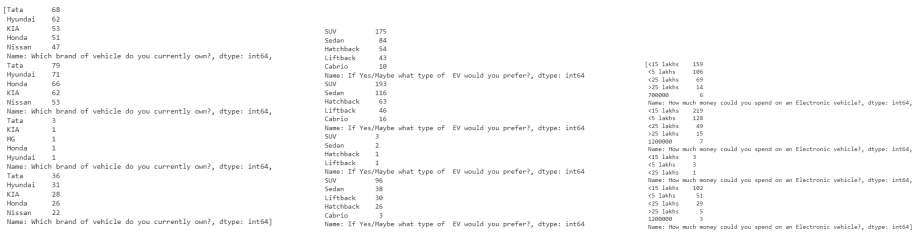
```
[Cluster_0["Marital Status"].value_counts().head(),
Cluster_1["Marital Status"].value_counts().head(),
Cluster_2["Marital Status"].value_counts().head(),
Cluster_3["Marital Status"].value_counts().head()]
```

```
[Cluster_0["Preference for wheels in EV"].value_counts().head(),
Cluster_1["Preference for wheels in EV"].value_counts().head(),
Cluster_2["Preference for wheels in EV"].value_counts().head(),
Cluster_3["Preference for wheels in EV"].value_counts().head()]
```



(a) $MartialStatusClusters$ (b) $FamilySizeClusters$ (c) $PreferredEVClusters$

Figure 12: Psychographic Analysis-1



(a) $CurrentBrand$ (b) $PreferredEVBrand$ (c) $ExpectedSpending$

Figure 13: Psychographic Analysis-2

```
[Cluster_0['Which brand of vehicle do you currently own?'].value_counts().head(),
Cluster_1['Which brand of vehicle do you currently own?'].value_counts().head(),
Cluster_2['Which brand of vehicle do you currently own?'].value_counts().head(),
Cluster_3['Which brand of vehicle do you currently own?'].value_counts().head()]
```

From the Psychographic analysis, we have the following decisions:

- Single People should be the targeted.
- Family having 3 to 5 should be targeted as they are more interested in buying EV.
- 4 wheeler are more preferred EV vehicle in the market followed by 2 wheeler and 3 wheeler.
- Tata is the most preferred brand in all the cluster.
- Hyundai and KIA is the 2nd best preferred brand in the market followed by Honda and Nissan.

- Suv and Sedan are the most preferred EV in all the cluster followed by Hatchback and Liftback being 2nd preferred EV. Cabrio being the least preferred.
- Out of 1000 people, 483 people wants to spend less than 15 lakh for EV - 48.3%, 288 people wants to spend less than 5 lakh for EV - 28.8%, 148 people wants to spend less than 25 lakh for EV - 14.8%, and only 34 people only wants to spend greater than 25 lakh for EV - 3.4%.

6.3.1 Behavioral Analysis

A behavioral analysis was conducted to gain insights into consumers' preferences and attitudes towards EVs. Behavioral Segmentation searches directly for similarities in behavior or reported behavior. It has advantage as it uses the very behavior of interest is used as the basis of segment extraction. Variables such as "Preference for replacing all vehicles to Electronic vehicles?", "If Yes/Maybe what type of EV would you prefer?", "Do you think Electronic Vehicles are economical?", and "Do you think Electronic vehicles will replace fuel cars in India?" were examined to understand the consumer perspective. From the Behavioral Analysis on customer outputs, we have the following decisions:

- To answer the question "Would you prefer replacing all your vehicles to Electronic vehicles?", 66.3% seems optimistic, 20.1% were neutral, and 13.6% said "No".
- In response to the the question "If Yes/Maybe what type of EV would you prefer?", 46.7% choose SUV, 24% choose Sedan, 14.4% choose Hatchback followed by Liftback (12%) and Cabrio (2.9%).
- In response to the the question "Do you think Electronic Vehicles are economical?", 76.8% said "Yes", 12.3% were neutral, and 10.9% said "No".
- In response to the the question "Do you think Electronic vehicles will replace fuel cars in India?", 42% said "in less than 10years", 38.6% said "in less than 20 years", 13.2% said "yes, possibly after 20 years", whereas 6.2% said "I don't think so".

7 Customizing the Marketing Mix

The marketing mix refers to the set of actions, or tactics, that a company uses to promote its brand or product in the market. The $4P$s make up a typical marketing mix are - Price, Product, Promotion and Place.

- **Price** refers to the value that is put for a product. It depends on segment targeted, ability of the companies to pay, ability of customers to pay supply - demand and a host of other direct and indirect factors. In our project, Out of 1000 people, about 48.3% people wants to spend less than 15 lakh for EV whereas 3.4% people wants to spend greater than 25 lakh.

- **Product** refers to the product actually being sold – In this case, the service. The product must deliver a minimum level of performance; otherwise even the best work on the other elements of the marketing mix won't do any good. In this study 46.7% choose SUV to be the brand of their EV vehicle, suggests the popularity of SUV in current market scenario. Among the remaining, 24% choose Sedan, 14.4% choose Hatchback followed by Liftback (12%) and Cabrio (2.9%). This suggests the Brand like Cabrio, Liftback and Hatchback need to take necessary actions in order to compete in the market.

- **Place** refers to the point of sale. In every industry, catching the eye of the consumer and making it easy for her to buy it is the main aim of a good distribution or 'place' strategy. Retailers pay a premium for the right location. In our study, Pune and Mumbai are more suitable to target the EV market.

- **Promotion** refers to all the activities undertaken to make the product or service known to the user and trade. This can include advertising, word of mouth, press reports, incentives, commissions and awards to the trade. It can also include consumer schemes, direct marketing, contests and prizes.

8 Conclusion

The market segmentation analysis of the EV market in India provided valuable insights into consumer preferences and behaviors. The analysis revealed distinct clusters within the dataset, representing different consumer segments based on demographic, geographic, psycographic, and behavioral factors. This information can guide marketing strategies, product development, and target market selection to enhance EV adoption in India.

- The electric vehicle (EV) market segments in India are experiencing significant growth and potential. Several key factors contribute to the development and adoption of EVs in the country.

- The findings suggest the need for targeted campaigns to address specific segments' concerns and preferences. The analysis also highlighted the importance of factors such as affordability, perception of economic viability, and regional variations in driving EV adoption.

- **Increasing Awareness:** There is a growing awareness among Indian consumers regarding the environmental benefits of EVs, including reduced emissions and lower pollution levels. This awareness is driving the demand for EVs, especially among environmentally conscious individuals and organizations.

- **Urban Commuting and Ride-Sharing:** India's urban areas, particularly major cities, are witnessing a surge in demand for EVs for commuting and ride-sharing purposes. The lower operating costs and the ability to

navigate through congested traffic make EVs an attractive option for urban dwellers.

- **Four and Two-Wheeler Dominance:** The two-wheeler segment holds significant potential for EV adoption in India. About 63% of population reliant on four-wheelers and about 30% suggests two-wheelers for transportation. The introduction of affordable electric scooters and motorcycles is gaining traction, particularly in urban and semiurban areas.

8.1 Limitations and Areas for Further Study:

The interdependency of all the components in the marketing mix contributes to the overall business plan of a company and, when managed effectively, can lead to remarkable success. The marketing mix necessitates a comprehensive understanding of the market, extensive research, and consultation with various stakeholders, including customers, trade partners, and manufacturers.

- The Indian government has implemented several policies and initiatives to encourage the adoption of electric vehicles (EVs), such as offering financial incentives, subsidies, and tax benefits to EV manufacturers. Additionally, efforts have been made to develop a robust charging infrastructure throughout the country. To conduct a study in this area, relevant data collection would be necessary.

- In recent years, the cost of EVs, including batteries, has gradually decreased, making them more affordable for the Indian market. This reduction in costs, along with the availability of government incentives, has expanded the potential consumer base for EVs. A future study could be conducted to further explore this trend.

- Government agencies and public transport operators are progressively transitioning to electric buses, taxis, and rickshaws. This aspect could also be the subject of a research project.

- While the charging infrastructure for EVs in India is still developing, significant improvements have been made. Ongoing efforts are focused on establishing a greater number of charging stations, both public and private, in major cities and along highways, addressing the concerns of potential EV buyers regarding range anxiety. This could be a future scope of study as well.

While the EV market in India is promising, advancements in battery technology, expansion of infrastructure, and raising public awareness, the EV market in India shows promise. With continued government support, technological advancements, and increasing consumer interest, the future of the EV market in India looks bright.

References

[1] A. Razmjoo, A. Ghazanfari, M. Jahangiri, E. Franklin, M. Denai, M. Marzband, D. Astiaso Garcia, and A. Maheri, "A comprehensive study on the expansion of electric vehicles in europe," *Applied Sciences*, vol. 12, p. 11656, 11 2022.