

# ANALYSIS OF ELECTRIC VEHICLE MARKET SEGMENTATION

By. Aditya Sangole

G.H. Raison College Of Engineering, Nagpur

## 1. *Abstract :*

This project analyzes India's electric vehicle market, focusing on segmentation from sales data, customer reviews, and technical specs. The study underscores the rapid growth of the two-wheeler market as a key revenue driver. Using customer review data and the k-means algorithm, the market was divided into four segments.

Segment 1 stands out, making up 39% of the consumer base, and is identified as the prime target for our strategy. Specific electric two-wheeler specs are recommended to meet the preferences of Segment 1, balancing affordability with competitive pricing. This strategic focus on Segment 1 positions our venture strongly within India's electric vehicle market.

## 2. **Introduction :**

India's transportation is transforming with the widespread adoption of electric vehicles (EVs), driven by urbanization, population growth, and rising incomes. Electric two-wheelers, in particular, are leading the way due to their affordability and popularity, offering a sustainable solution to pollution and emissions.

Government policies supporting local manufacturing and a robust supply network have been pivotal. By 2023, India's electric two-wheeler market has flourished, showcasing the success of these initiatives and the acceptance of clean mobility.

This study explores this transformation, focusing on electric two-wheelers. By analyzing consumer behavior, psychographics, and vehicle specifications, we provide EV price recommendations to guide consumers, policymakers, and industry stakeholders towards a sustainable and consumer-friendly electric transportation system in India.

## 3. **Problem Statement :**

The task at hand involves using data-driven insights from sales data, customer evaluations (including behavioral and psychographic data), and technical specifications of electric vehicles to strategically position our electric vehicle startup in the Indian market. Our goal is to use these insights to efficiently divide the market into

## 4. **Data Sources and Collection :**

For this project, data was collected from three distinct sources. The primary dataset, from the Society of Manufacturers of Electric Vehicles, covers sales figures for electric two-wheelers, three-wheelers, fourwheelers, and buses from 2017 to 2023, providing a comprehensive view of market trends and customer preferences.

The second dataset, from bikewale.com, includes customer reviews of electric two-wheelers, offering critical behavioral and psychographic insights. These qualitative inputs are essential for understanding customer behavior.

The third dataset, also from bikewale.com, details the technical specifications and pricing information of electric two-wheelers, allowing us to evaluate technical feasibility and pricing crucial for our market segmentation strategy.

By integrating these datasets, we developed a robust understanding of the electric vehicle market. Real sales data, customer sentiments, and technical specifics formed the foundation of our analysis, ensuring a data-driven and market-relevant segmentation approach.

## 5. Data Pre-processing :

The data pre-processing phase used Python libraries such as numpy, pandas, matplotlib, seaborn, and nltk. Sales data from 10 Excel sheets was merged using pandas, ensuring accurate electric vehicle maker names. Aggregation operations on electric two-wheeler sales data highlighted market trends.

Customer reviews and technical specifications were merged, with null values logically replaced to ensure data integrity. Sentiment analysis using nltk provided qualitative insights into customer sentiments. Key behavioral variables—Visual Appeal, Reliability, Performance, Service Experience, Extra Features, Comfort, Maintenance Cost, and Value for Money—were isolated and prepared for market segmentation, offering a nuanced understanding of customer preferences and attitudes toward electric vehicles.

In this notebook we will do market segmentation of 2W EV market in India as we have analysed the vehical type which is the most compatible for indian market in previous notebook please refer the "EV\_market\_Analysis.ipynb"

### import necessary libraries

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: data_bw=pd.read_csv("ev2_bikewale.csv")
model=pd.read_csv("ev_model_spec.csv")
```

```
In [19]: data = data_bw.merge(model, how = 'left', on = 'Model Name')
```

```
In [20]: data.head(4)
```

Out[20]:

	review	Used it for	Owned for	Ridden for	rating	Visual Appeal	Reliability	Performance	Service Experience	Extra Features	Comfort	Maintenance cost	Value for Money	Model Name	Price	Ridir Rang (kr)
0	We all checked the bike's capacity to be 150 k...	Daily Commute	Never owned	NaN	1	3.0	4.0	NaN	NaN	NaN	4.0	NaN	1.0	TVS iQube	137890	10

### Convert the text data into readable by analyzing sentiments using NLTK

```
In [25]: from nltk.sentiment import SentimentIntensityAnalyzer
sentiment= SentimentIntensityAnalyzer()
```

```
In [26]: sentiments = SentimentIntensityAnalyzer()
sentiments_list = []

for review_text in data['review']:
    if pd.isna(review_text):
        sentiments_list.append('neutral')
        continue
    sentiment_scores = sentiments.polarity_scores(review_text)

    positive_score = sentiment_scores['pos']
    negative_score = sentiment_scores['neg']

    if positive_score > negative_score:
        sentiment_label = 'positive'
    elif positive_score < negative_score:
        sentiment_label = 'negative'
    else:
        sentiment_label = 'neutral'

    sentiments_list.append(sentiment_label)

data['sentiment'] = sentiments_list
```

```
In [27]: data['sentiment'].value_counts()
```

## Perform PCA operation for dimension reduction and better result

```
In [35]: from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data_segment)

In [39]: from sklearn.decomposition import PCA
pca = PCA(random_state = 36)
pca.fit(data_scaled)

Out[39]: PCA(random_state=36)
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [40]: data_pca=pca.transform(data_scaled)

In [41]: df_pca = pd.DataFrame(data_pca, columns = [f'PC{x +1}' for x in range(len(data_segment.columns))])

In [43]: df_pca.head()

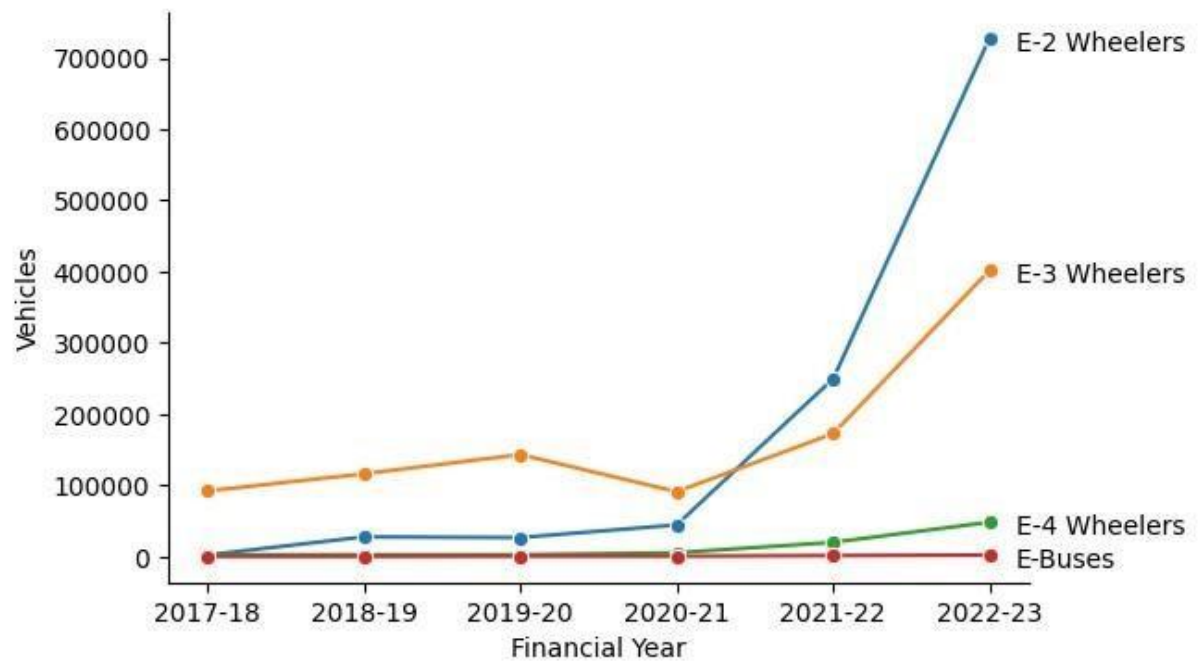
Out[43]:
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
0	0.291227	-1.038055	0.354864	-0.623469	-1.102720	-0.169287	-0.438012	0.957827
1	0.710801	-1.394405	-0.360466	-0.621671	0.320899	-0.086053	-0.426279	-0.149917
2	-0.849149	-1.189765	0.167683	-0.410898	-0.409054	-0.191904	-0.329993	0.830738
3	1.967022	-0.878935	-0.100197	0.330003	-0.075822	0.069599	0.013068	0.011328

## 6. Segment Extraction :

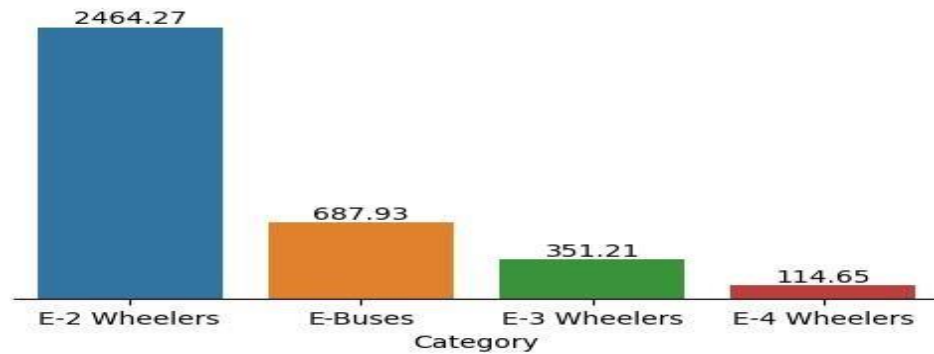
### 6.1. Using Sales Data :

In this segment, a detailed analysis was conducted based on three significant figures representing India's electric vehicle market.



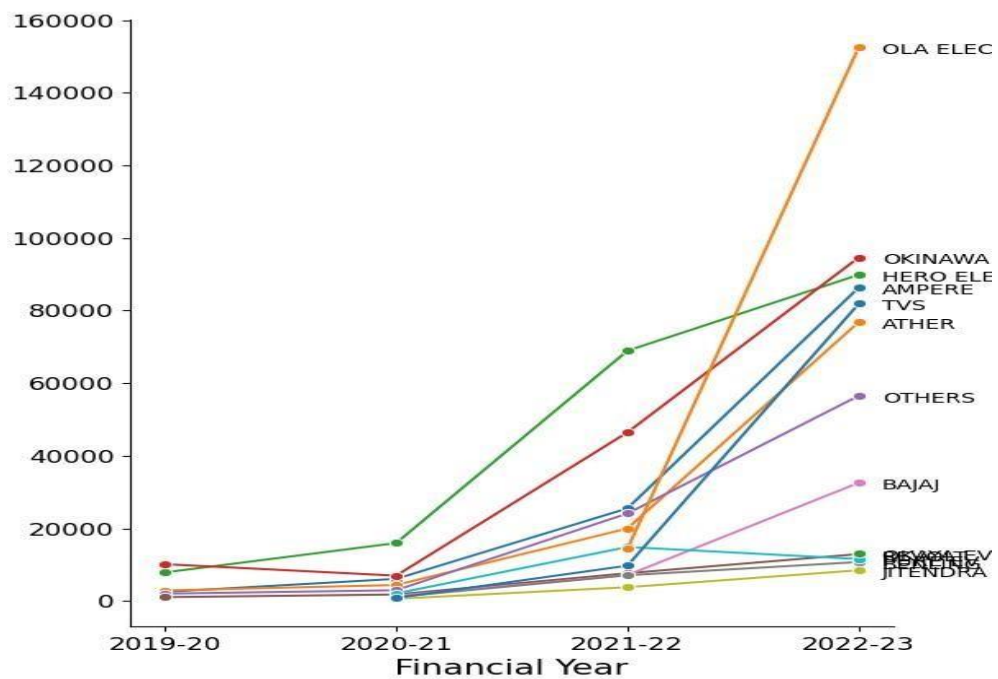
**Figure 5.1 India's electric vehicle market**

Figure 5.1 showcased the remarkable growth trajectory of India's two-wheeler market in 2023, underscoring its leading position within the industry.



**Figure 5.2 India's electric vehicle industry in crores**

Figure 5.2 delved into the market's financial perspective, representing the industry's total value in crores. Notably, two-wheelers emerged as the primary revenue generators, highlighting their economic significance.



**Figure 5.3 Top electric two-wheeler companies**

Focusing on certain electric two-wheeler firms, Figure 5.3 shows market leadership and competitiveness in the sector, with Ola Electric emerging as the market leader in 2023.

After a thorough examination of these numbers, it was clear that the electric two-wheeler market offered the most potential for our in-depth investigation. Its strong growth, market leadership, and revenue domination all pointed to its importance and potential, which made it the perfect subject for our in-depth investigation.

### Perform clustering

```
In [51]: from sklearn.cluster import KMeans  
data_km28 = [KMeans(n_clusters = k, random_state = 42).fit(data_scaled) for k in range(2, 9)]
```

```
In [52]: kmeans = data_km28[2]
```

```
In [53]: data_clust = pd.DataFrame(data_scaled, columns = data_segment.columns)  
data_clust['cluster'] = kmeans.labels_
```

```
In [54]: data_pca = df_pca.copy()  
data_pca['cluster'] = kmeans.labels_
```

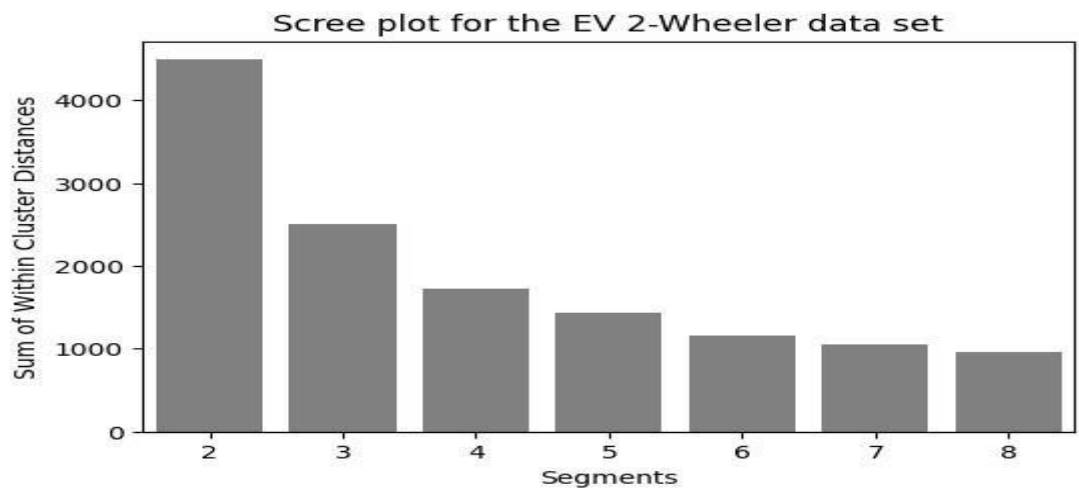
### Segment Profiling

```
In [55]: data_profile = data_segment.copy()  
data_profile['cluster'] = kmeans.labels_
```

```
In [56]: columns_segment = data_clust.columns  
data_pivot = data_profile[columns_segment].groupby('cluster').mean().T
```

```
In [57]: data_pivot_mean = data_pivot.mean(axis = 1).reset_index()  
data_pivot_mean.columns = ['Variable', 'Value']
```

## 6.2. Using k-Means :



**Figure 5.4 Scree plot for the electric vehicle data set**

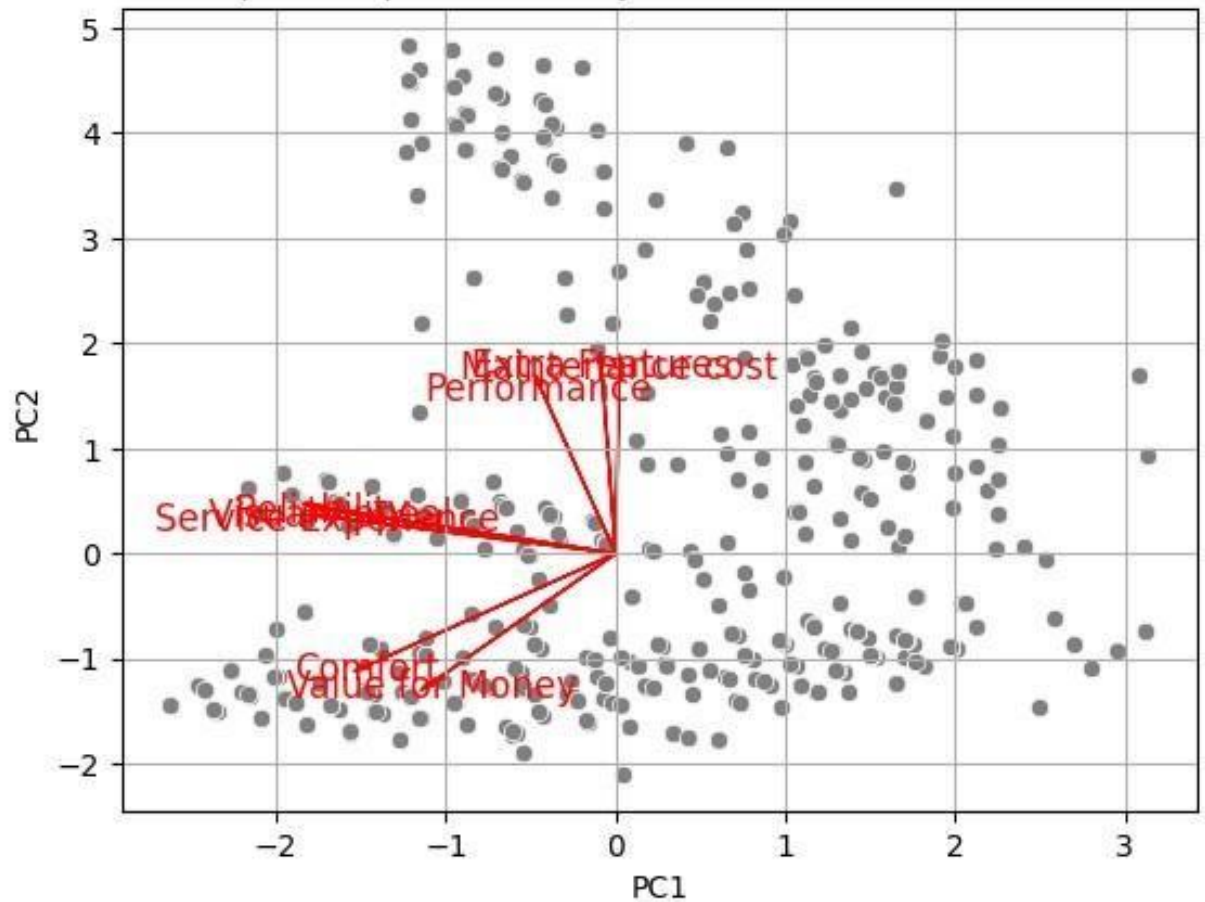
The conventional k-means technique was used in the following research to investigate potential market segmentation within the customer reviews data for electric two-wheelers. Two to eight market categories were the subjects of methodical testing of the solutions. The scree plot Figure 5.4, which shows a clear elbow at four segments, greatly aided in the decision-making process. This highlighted point showed a significant decrease in distances and represented the ideal number of segments for our investigation. We maintained a laser-like focus on the electric two-wheeler market category by integrating the results from these investigations, which guaranteed accuracy and applicability in our market segmentation strategy.

## 7. Profiling and Describing Segmentation :

As seen in Figure 6.1, this part provides a thorough study of our consumer categories. The graph illustrates the various perspectives among the various parts. 15% of customers, or Segment 0, place a high value on the aesthetic appeal, performance, dependability, comfort, and ease of servicing of electric two-wheelers. On the other hand, 39% of customers in Segment 1 express unhappiness with all factors, making them the largest but least satisfied segment. Customers in Segment 2 (33%), in particular, prefer comfort, visual attractiveness, dependability, and a good value for their money. Last but not least, the smallest category, represented by 13% of customers, emphasizes visual appeal,

performance, dependability, service experience, extra features, and maintenance costs. This segment exhibits unique perspectives, especially with regard to features and pricing.

## A Principal components analysis of the EV 2-Wheeler data set



### Segment Profiling

```
In [55]: data_profile = data_segment.copy()
data_profile['cluster'] = kmeans.labels_

In [56]: columns_segment = data_clust.columns
data_pivot = data_profile[columns_segment].groupby('cluster').mean().T

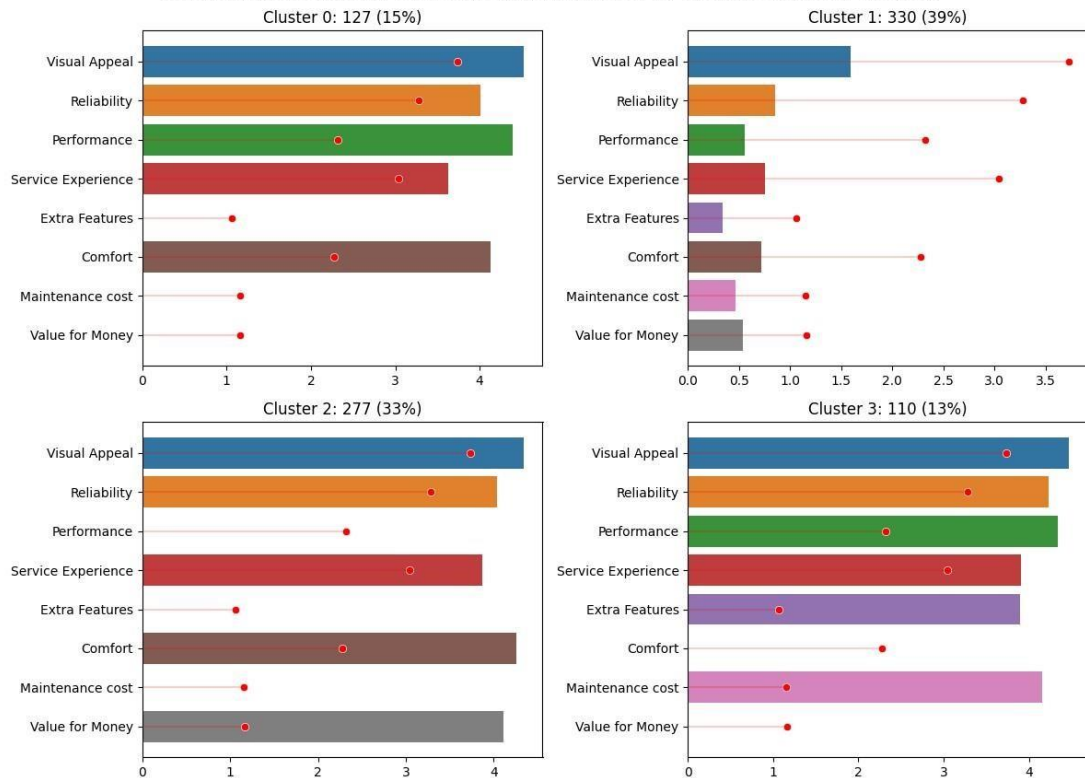
In [57]: data_pivot_mean = data_pivot.mean(axis = 1).reset_index()
data_pivot_mean.columns = ['Variable', 'Value']

In [58]: plt.figure(figsize = (12, 9))
for i in range(4):
    plt.subplot(2, 2, i+1)
    sns.barplot(data_pivot, x = i, y = data_pivot.index)
    sns.scatterplot(data_pivot_mean, x = 'Value', y = 'Variable', color = 'red')
    for index, row in data_pivot_mean.iterrows():
        plt.hlines(y=row['Variable'], xmin=0, xmax=row['Value'], colors='red', alpha = 0.2)
    plt.ylabel("")
    plt.xlabel("")
    plt.title(f"Cluster {i}: {data_profile['cluster'].value_counts()[i]} ({data_profile['cluster'].value_counts()[i]*100/len(data_profile['cluster'])}% of data)")
plt.suptitle("Segment profile plot for the four-segment solution for the EV 2-Wheeler data set", fontsize = 15)
plt.tight_layout()
plt.savefig("segment_profile.png")
plt.show()
```

Segment profile plot for the four-segment solution for the EV 2-Wheeler data set



Segment profile plot for the four-segment solution for the EV 2-Wheeler data set



## 8. Selection of Target Segment :

Out of all the target segments in the electric vehicle industry, Segments 1 and 2 are particularly noteworthy as possible centers of attention. With 39% of the market, segment 1 represents a sizable market base with a wide range of opinions and tastes. Our data reveals that this category has a range of feelings that reflect their unique needs and objectives. There is an opportunity to understand their distinct perspectives, such as discontent with several things. Directly addressing these issues can increase brand loyalty and consumer satisfaction within this sizeable market share.

With 33% of the customer base, Segment 2 offers yet another alluring prospect. Their expectations are shaped by their unique views, which include how much they value comfort, dependability, visual attractiveness, and service experience. The input from this part offers priceless information that help us customize our electric cars.

### Segment Description

```
In [60]: data_desc = data.copy()
data_desc['cluster'] = kmeans.labels_

In [61]: data_desc.columns
Out[61]: Index(['review', 'Used it for', 'Owned for', 'Ridden for', 'rating',
               'Visual Appeal', 'Reliability', 'Performance', 'Service Experience',
               'Extra Features', 'Comfort', 'Maintenance cost', 'Value for Money',
               'Model Name', 'Price', 'Riding Range (km)', 'Top Speed (kmph)',
               'Weight (kg)', 'Battery Charging Time (hrs)', 'Rated Power (W)',
               'sentiment', 'cluster'],
              dtype='object')

In [62]: def mosaicplot(df, color_list, figsize=(5,5), barwidth = 1, title = "Simple Mosaic Plot", xlabel = 'cat'):
    fig, ax = plt.subplots(figsize =figsize)

    # Convert to ratio of each row sum and do a cumsum along row
    p_cnt_ratio_df = df.div(df.sum(1), axis =0)
    ratio_cum_df = p_cnt_ratio_df.cumsum(axis=1)

    # Get column order.
    col_order = ratio_cum_df.columns.tolist()

    x = [str(n) for n in ratio_cum_df.index.tolist()]

    # On the actual plots
```

## 9. Potential Early Market Customer Base:

Two main categories are identified from the study of the prospective early market customer base: Segment 1, which includes 330 members (39% of customers), and Segment 2, which includes

277 participants, or 33% of customers. The logical target price for Segment 1 is between ₹51,094 and ₹1,67,844, and for Segment 2, it is between ₹51,094 and ₹1,37,890, according to an analysis of the pricing range data.

In this early market scenario, the number of possible clients in each sector is multiplied by our planned price range to determine the potential sales (profit). For example, our potential profit from section 1 alone would be ₹39.60 crores if our target price for this section is set at ₹1,20,000. In a similar vein, Segment 2's potential profit at a target price of ₹1,10,000 would be ₹30.47 crores.

## 10. Conclusion :

In summary, our in-depth analysis of India's electric vehicle market identified Segment 1 as the optimal target, representing 39% of the consumer base and a significant market opportunity. By tailoring our electric two-wheeler specifications to meet this segment's preferences, we ensure our products align with a large customer base's demands. This strategic decision is based on a thorough understanding of market segmentation, consumer behavior, and technical specifications.

These insights provide clear direction for market entry, emphasizing precision and relevance in product development and marketing strategies. This approach equips us with a solid foundation, ensuring our offerings resonate effectively within India's evolving electric vehicle landscape.

## 11. Github Link : <https://github.com/AdityaSangole/Feynn-Labs-Internship-Task/tree/main/Task%2003>