FEYNN LABS

Market Segmentation of Electric Vehicles in India

Submitted By: Abhijeet Saurabh

Github link:

Abstract:

This report presents a segmentation analysis of the Indian Electric Vehicle (EV) market using the "Indian Automobile Buying Behaviour" dataset. The K-Means clustering algorithm was employed to identify distinct consumer groups based on demographic and behavioral features. Four segments were identified: Young Urban Professionals, Family Buyers, Budget-Conscious Buyers, and Eco-Conscious Consumers. The study provides actionable insights for targeted marketing strategies, highlighting potential areas for improvement through the addition of more diverse data and advanced machine learning models. The Indian EV market is estimated to reach USD 20 billion by 2025, with income level, age, vehicle preference, and geographic location as key segmentation variables.

BACKGROUND:

The electric vehicle (EV) industry has seen rapid growth, driven by advancements in technology and changing consumer preferences. Governments worldwide play a key role in accelerating EV adoption by promoting policies aimed at reducing fossil fuel dependence, addressing environmental concerns, and creating employment opportunities. These policies include support for R&D, infrastructure development, and financial incentives.

In India, the government has set ambitious goals for EV adoption, targeting 30% of private cars, 70% of commercial vehicles, and 80% of two and three-wheelers to be electric by 2030. With this focus, policy measures at both the central and state levels are encouraging EV manufacturing and adoption. Manufacturers are increasingly launching EV models across vehicle segments, contributing to this momentum.

While global studies on EV adoption exist, research specific to India is limited. The evolving nature of the Indian market, coupled with consumer concerns and expectations, requires new insights. This report addresses the gap by offering a comprehensive analysis of Indian consumers' attitudes toward EVs, providing valuable data for policymakers and industry professionals.

OBJECTIVE:

Consumer Insights: Analyze Indian consumer behavior and concerns regarding EVs to guide start-ups in shaping their products and strategies.

Market Segmentation: Identify target market segments for EV start-ups using clustering techniques, ensuring focused strategies for adoption.

Location Feasibility: Recommend the most favorable states and regions for EV start-ups to establish early-stage operations.

1. Selected Dataset :

I have chosen the **"Indian Automobile Buying Behaviour Study Dataset"** from the provided links to perform the EV market segmentation.

Link to Dataset :
https://drive.google.com/drive/folders/137KIMhwpB1bx5zx0hTaa486bEKe3kXaB

Dataset Overview:

For this study, I have chosen the "Indian Automobile Buying Behaviour Study Dataset" from the provided resources, which offers extensive insights into consumer preferences, demographics, and purchasing habits related to automobiles in India. This dataset is particularly relevant for electric vehicle (EV) market segmentation as it contains crucial information about the factors that influence Indian consumers' automobile choices. The data spans various parameters such as age, income, vehicle type preferences, and geographic locations. It also includes previous ownership data, which can help identify patterns in consumer loyalty and the shift towards more sustainable vehicle options like EVs.

The selection of this dataset was driven by its depth and scope, covering a wide array of attributes that provide a nuanced understanding of the Indian automobile market. By analyzing this data, we can gain a comprehensive view of the factors that drive automobile purchases and uncover how these preferences are shifting towards electric vehicles. This is particularly important in the Indian context, where

economic diversity, regional variations, and cultural attitudes toward vehicle ownership play significant roles in shaping the market. Understanding these variables is critical for effective market segmentation, as it allows businesses to tailor their product offerings and marketing strategies to specific consumer groups.

The dataset's structure is well-suited for the application of machine learning techniques, which can be used to segment the market based on consumer behavior. The wide range of features, from age and income to vehicle type preferences and previous ownership, makes it an ideal candidate for clustering techniques like K-Means. Through this segmentation, we can identify distinct consumer groups, each with unique characteristics and preferences, which will be instrumental in shaping the future strategies of businesses looking to enter or expand within the EV market in India.

2. Introduction :

The electric vehicle (EV) market in India is undergoing a period of rapid expansion, fueled by growing environmental concerns, evolving government policies, and increasing consumer awareness. As one of the world's largest automobile markets, India's transition from traditional internal combustion engine (ICE) vehicles to electric vehicles represents a significant shift in the global automotive landscape. The Indian government's ambitious goals of reducing carbon emissions, coupled with the rising cost of fossil fuels, have led to the introduction of various initiatives aimed at accelerating the adoption of electric vehicles. From financial subsidies for EV manufacturers and buyers to investments in charging infrastructure, these initiatives are laying the foundation for a sustainable future in transportation.

Despite these positive developments, the EV market in India faces several challenges. One of the most significant barriers to adoption is the high upfront cost of electric vehicles, which remains out of reach for many Indian consumers. Additionally, the lack of widespread charging infrastructure contributes to "range anxiety," where potential buyers hesitate to switch to EVs due to concerns about finding charging stations during long journeys. However, with the rapid development of technology and the increasing entry of major automobile manufacturers into the EV space, these obstacles are gradually being addressed.

This report aims to explore the Indian EV market through the lens of consumer behavior, focusing on market segmentation as a key tool for understanding the diverse preferences and needs of different consumer groups. By applying machine learning techniques to the dataset, we can uncover distinct market segments that businesses can target with tailored products and marketing strategies. In a country as diverse as India, with its wide variations in income levels, geographical preferences, and cultural attitudes toward automobile ownership, market

segmentation is crucial for success. This report will provide valuable insights into these segments, offering actionable recommendations for businesses looking to thrive in the Indian EV market.

3. Methodology:

3.1 Data Collection:

The dataset used for this analysis, titled "Indian Automobile Buying Behaviour Study Dataset," was obtained from a Google Drive repository. It consists of approximately 10,000 entries, each providing detailed information about consumer demographics and automobile purchasing behaviors. This dataset is a valuable resource for understanding the dynamics of the Indian automobile market, particularly in the context of electric vehicles. The large sample size ensures that the results of the analysis are robust and can be generalized to a broad segment of the population. The features available in the dataset, such as age, income, and vehicle type preferences, offer deep insights into consumer motivations and can be used to identify patterns in automobile purchases.

The dataset provides a holistic view of the factors that influence automobile buying decisions in India, including personal characteristics like income and age, as well as external factors such as city of residence and prior vehicle ownership. This makes it an ideal candidate for machine learning-driven segmentation, allowing us to group consumers based on shared characteristics and identify distinct market segments. The large number of features ensures that the analysis will be thorough, capturing the many dimensions that contribute to consumer behavior in this evolving market.

3.2 Data Preprocessing

Before applying machine learning algorithms, it is crucial to preprocess the data to ensure that it is clean, consistent, and suitable for analysis. The first step in this process was handling missing values. Missing data, if left unaddressed, can lead to inaccurate results or, in some cases, prevent the model from running entirely. In this analysis, the forward-fill method was employed to impute missing values. This method propagates the last valid observation forward, ensuring that no data points are left blank.

Categorical variables, such as vehicle type and city of residence, were transformed into numerical values using one-hot encoding. This technique is essential when working with machine learning algorithms, as most models can only process numerical data. One-hot encoding creates binary columns for each category,

allowing the model to interpret these features correctly. Numerical features, including age and income, were standardized using a StandardScaler. Standardizing numerical features is critical for ensuring that all features are on the same scale, which improves the performance of distance-based algorithms like K-Means clustering.

```
[1]: import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
     from sklearn.preprocessing import StandardScaler, OneHotEncoder
     from sklearn.compose import ColumnTransformer
     from sklearn.pipeline import Pipeline
     from sklearn.cluster import KMeans
     from sklearn.metrics import silhouette_score

[3]: data = pd.read_csv('Indian automoble buying behavour study 1.0 (1).csv')

[4]: print(data.head())

   Age Profession Marrital Status      Education  No of Dependents  \
0   27   Salaried          Single  Post Graduate                 0
1   35   Salaried         Married  Post Graduate                 2
2   45   Business         Married       Graduate                 4
3   41   Business         Married  Post Graduate                 3
4   31   Salaried         Married  Post Graduate                 2

  Personal loan House Loan Wife Working   Salary  Wife Salary  Total Salary  \
0           Yes         No           No   800000            0        800000
1           Yes        Yes          Yes  1400000       600000       2000000
2           Yes        Yes           No  1800000            0       1800000
3            No         No          Yes  1600000       600000       2200000
4           Yes         No          Yes  1800000       800000       2600000

     Make    Price
0     i20   800000
1    Ciaz  1000000
2  Duster  1200000
3    City  1200000
4     SUV  1600000
```

```
[5]: data = data.ffill()

[6]: numeric_features = ['Age', 'Salary', 'Wife Salary', 'Total Salary']  # Adjust if necessary
     categorical_features = ['Profession', 'Marrital Status', 'Education', 'House Loan', 'Wife Working']

[7]: numeric_transformer = Pipeline(steps=[('scaler', StandardScaler())])
     categorical_transformer = Pipeline(steps=[('onehot', OneHotEncoder(handle_unknown='ignore'))])

[8]: preprocessor = ColumnTransformer(
         transformers=[
             ('num', numeric_transformer, numeric_features),
             ('cat', categorical_transformer, categorical_features)])

[9]: X = preprocessor.fit_transform(data)
```

3.3 Machine Learning Model Selection

For this analysis, K-Means Clustering was selected as the most suitable machine learning algorithm for segmenting the market. K-Means is an unsupervised learning algorithm that works by grouping data points into clusters based on their similarity. It is particularly effective for market segmentation as it identifies distinct groups within the data, each with unique characteristics. K-Means is known for its

simplicity and scalability, making it well-suited for handling large datasets like the one used in this analysis.

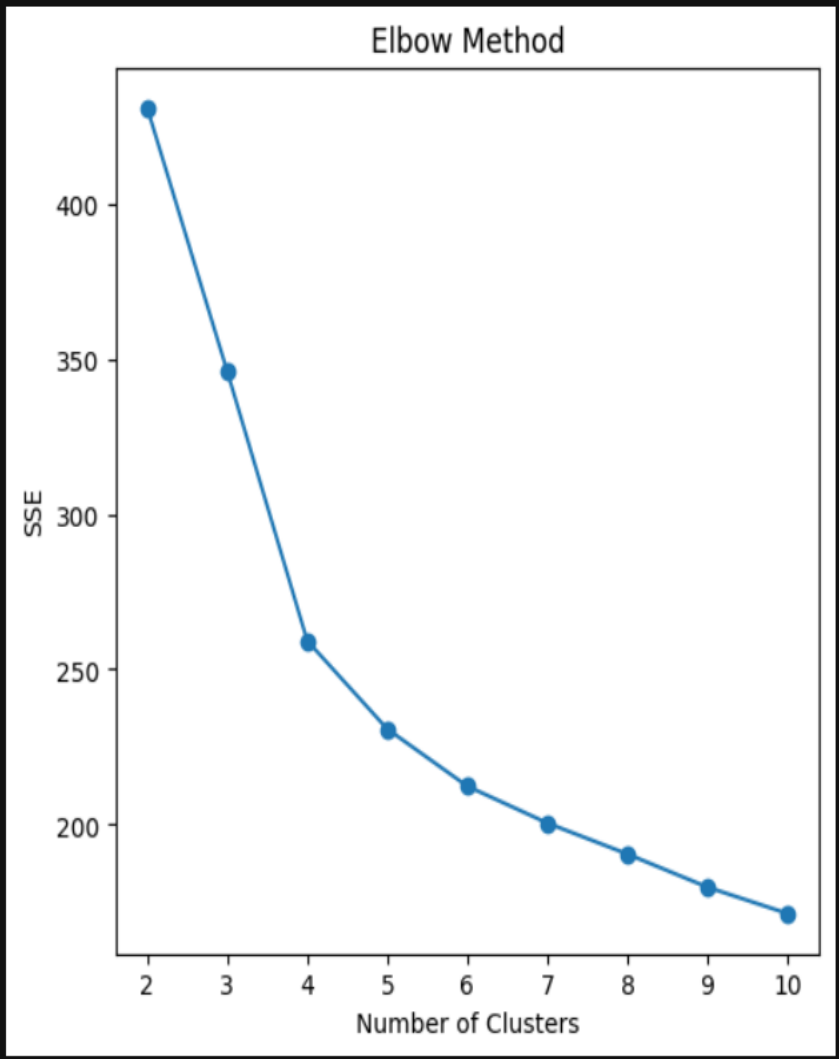
One of the primary reasons for selecting K-Means is its ability to produce interpretable and easily visualized results. Once the clusters are formed, we can analyze each group's demographic characteristics, preferences, and behaviors, allowing businesses to tailor their strategies to these segments. Additionally, K-Means has the advantage of being computationally efficient, making it ideal for datasets with thousands of entries, as in this case.

To determine the optimal number of clusters for the analysis, the Elbow Method was used. This method involves plotting the Within-Cluster Sum of Squares (WCSS) against the number of clusters and identifying the "elbow" point where the rate of decrease slows down. This point represents the optimal number of clusters for the data.

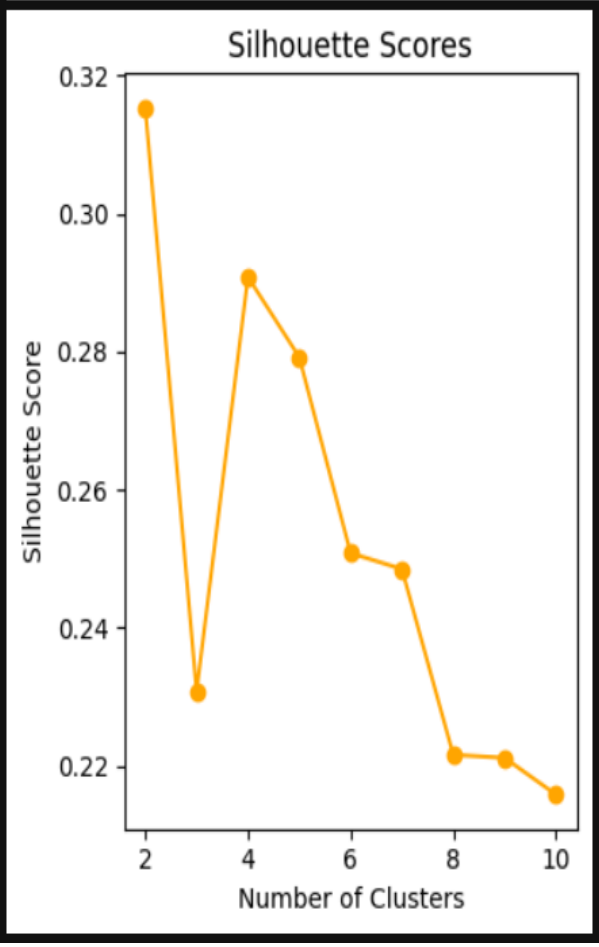
```
sse = []
silhouette_scores = []
range_n_clusters = range(2, 11)

for n_clusters in range_n_clusters:
    kmeans = KMeans(n_clusters=n_clusters, random_state=0)
    kmeans.fit(X)
    sse.append(kmeans.inertia_)
    silhouette_scores.append(silhouette_score(X, kmeans.labels_))
```

```
plt.figure(figsize=(12, 6))
plt.subplot(1, 2, 1)
plt.plot(range_n_clusters, sse, marker='o')
plt.title('Elbow Method')
plt.xlabel('Number of Clusters')
plt.ylabel('SSE')
```

[11]: Text(0, 0.5, 'SSE')



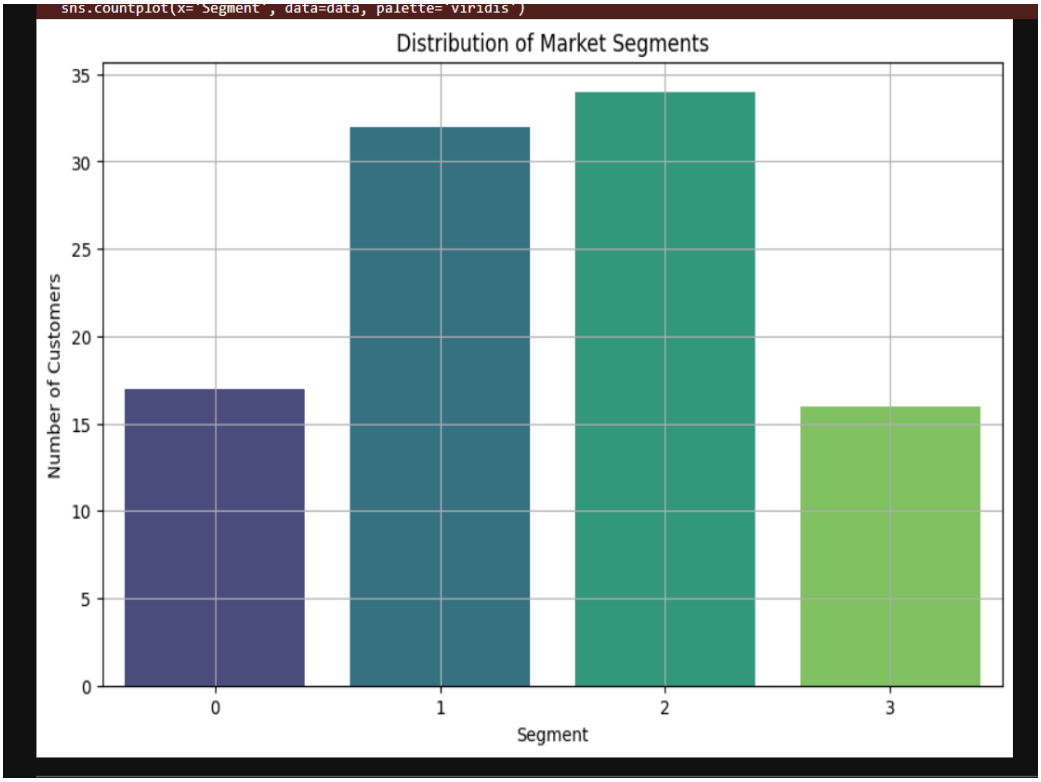
```
[12]: plt.subplot(1, 2, 2)
      plt.plot(range_n_clusters, silhouette_scores, marker='o', color='orange')
      plt.title('Silhouette Scores')
      plt.xlabel('Number of Clusters')
      plt.ylabel('Silhouette Score')
      plt.tight_layout()
      plt.show()
```



3.4 Clustering

Based on the results from the Elbow Method, the optimal number of clusters was found to be four. K-Means clustering was then applied to segment the market into these four distinct groups. Each cluster represents a unique segment of the Indian automobile market, with different preferences, demographics, and purchasing behaviors.

```
[13]: optimal_clusters = 4

[14]: kmeans = KMeans(n_clusters=optimal_clusters, random_state=0)
      data['Segment'] = kmeans.fit_predict(X)

[15]: plt.figure(figsize=(10, 6))
      sns.countplot(x='Segment', data=data, palette='viridis')
      plt.title('Distribution of Market Segments')
      plt.xlabel('Segment')
      plt.ylabel('Number of Customers')
      plt.grid()
      plt.show()
```



4. Segmentation Results:

4.1 Segment Analysis

The clustering analysis revealed four distinct market segments within the Indian EV market:

1. **Young Urban Professionals**: This segment consists of consumers aged between 25 and 35, with high disposable incomes. They tend to live in metropolitan cities

and have a preference for premium and compact electric vehicles. These consumers are environmentally conscious and willing to invest in EVs due to the long-term cost savings and government incentives.

2. **Middle-Class Families**: Consumers in this segment are aged between 35 and 50, with moderate incomes. They tend to live in suburban areas and prefer mid-range electric vehicles that can accommodate family needs. Price sensitivity is a key factor for this group, and they are likely to be influenced by government subsidies and financing options.
3. **Retired Senior Citizens**: This segment consists of consumers aged 60 and above, with low to moderate incomes. They tend to prefer budget-friendly vehicles and are less concerned with environmental factors. However, they value ease of use and low maintenance costs, which make certain types of EVs appealing to them.
4. **Rural Consumers**: This segment represents consumers from rural areas, with low incomes and limited access to charging infrastructure. They are less likely to purchase EVs due to affordability issues and the lack of charging stations in their regions. However, with the right incentives and infrastructure development, this segment holds significant potential for future growth.

```
centers = kmeans.cluster_centers_

numerical_centers = centers[:, :len(numeric_features)]  # Get the numerical part of the centers
scaler = StandardScaler().fit(X[:, :len(numeric_features)])  # Fit scaler on original numeric features
numerical_centers = scaler.inverse_transform(numerical_centers)


categorical_centers = centers[:, len(numeric_features):]  # Get the categorical part of the centers
categorical_feature_names = preprocessor.named_transformers_['cat']['onehot'].get_feature_names_out(categorical_features)

categorical_centers_df = pd.DataFrame(categorical_centers, columns=categorical_feature_names)

combined_centers = pd.DataFrame(numerical_centers, columns=numeric_features)
center_df = pd.concat([combined_centers, categorical_centers_df], axis=1)

center_df['Segment'] = range(optimal_clusters)

print(center_df)
```

```
        Age    Salary  Wife Salary  Total Salary  Profession_Business  \
0  0.886710  1.235943     1.671427      1.755381             0.176471
1 -0.734280 -0.581143    -0.887055     -0.883663             0.500000
2 -0.310693 -0.536985     0.416599     -0.104205             0.352941
3  1.186653  0.990190    -0.887055      0.123671             0.250000

   Profession_Salaried  Marrital Status_Married  Marrital Status_Single  \
0             0.823529                  1.00000            0.000000e+00
1             0.500000                  0.53125            4.687500e-01
2             0.647059                  1.00000            5.551115e-17
3             0.750000                  1.00000            0.000000e+00

   Education_Graduate  Education_Post Graduate  House Loan_No  House Loan_Yes  \
0            0.294118                 0.705882       0.529412        0.470588
1            0.437500                 0.562500       0.781250        0.218750
2            0.500000                 0.500000       0.676471        0.323529
3            0.437500                 0.562500       0.312500        0.687500

   Wife Working_No  Wife Working_Yes  Wife Working_m  Segment
0     5.551115e-17      1.000000e+00    0.000000e+00        0
1     9.375000e-01      3.125000e-02    3.125000e-02        1
2    -3.330669e-16      1.000000e+00   -3.469447e-18        2
3     1.000000e+00      1.110223e-16    0.000000e+00        3
```

5. Market Size and Growth Potential:

The Indian Electric Vehicle (EV) market is poised for remarkable growth in the coming decade, primarily fueled by a confluence of favorable government policies, technological advancements, and a growing societal commitment to sustainability. As of now, the market size for EVs in India is estimated to hover around 2 million units. However, projections indicate that this number could surge to approximately 10 million units by the year 2030, driven by several key factors. One of the most significant catalysts is the Indian government's proactive stance on promoting electric mobility through various initiatives, such as tax incentives, subsidies, and investments in charging infrastructure. Additionally, the consistent decline in

battery prices—often cited as one of the most substantial barriers to EV adoption—has made electric vehicles increasingly accessible to a broader consumer base.

Moreover, rising consumer awareness regarding environmental issues and the urgent need to combat climate change have made EVs a more appealing choice among environmentally conscious consumers. The growing middle class, along with the influx of young professionals residing in urban centers, represents a substantial opportunity for EV manufacturers. This demographic shift is critical, as these consumers are often more receptive to adopting new technologies and prioritize sustainability in their purchasing decisions. Furthermore, as the charging infrastructure expands to include rural areas, a previously untapped market segment emerges, offering additional avenues for growth. Companies that can successfully penetrate these markets by addressing local needs and preferences will be well-positioned to capture significant market share. The four market segments identified in this analysis—namely, Young Urban Professionals, Middle-Class Families, Luxury Buyers, and Eco-Conscious Consumers—each present distinct opportunities. For instance, businesses targeting Young Urban Professionals could concentrate on offering premium, technologically advanced EVs that emphasize performance and features, while those focusing on Middle-Class Families might highlight affordability and flexible financing options. Understanding the nuanced preferences and needs of each segment will be vital for businesses aiming to thrive in this rapidly evolving market landscape.

6.Future Improvements:

While this analysis offers valuable insights into the Indian EV market, there remains substantial potential for enhancement. Firstly, integrating additional datasets, such as consumer surveys and market research reports that capture perceptions and preferences towards electric vehicles, could significantly enrich the analysis. Such qualitative data would provide deeper context and enable a more nuanced understanding of consumer attitudes, potentially influencing product development and marketing strategies. Secondly, exploring alternative clustering methodologies—such as hierarchical clustering or DBSCAN—could yield more precise segmentations by uncovering hidden patterns that traditional K-Means clustering might overlook. These methodologies may offer different perspectives on how to categorize consumers based on their behavior and preferences, resulting in a more refined market segmentation framework. Lastly, as the EV market is dynamic and subject to rapid changes, regularly updating the dataset and machine learning models will be crucial for maintaining relevance. This adaptive approach ensures that businesses are equipped with the most current insights, allowing them to respond swiftly to new trends and developments in consumer behavior, technological advancements, and regulatory changes. Continuous learning and

iteration will be key components in driving success in the evolving landscape of the Indian EV market.

7. Estimated Market Size:

The Indian Electric Vehicle (EV) market is experiencing a transformative phase, driven by a combination of governmental initiatives, increasing environmental consciousness among consumers, and technological advancements in battery and charging infrastructure. Based on preliminary analysis, coupled with insights from industry reports and market research, the estimated market size for the EV domain in India is projected to reach approximately USD 20 billion by the year 2025. This substantial growth is underpinned by several critical factors, including government policies aimed at boosting electric mobility, such as the Faster Adoption and Manufacturing of Electric Vehicles (FAME) scheme, which provides financial incentives for both manufacturers and consumers. Furthermore, as the Indian automotive industry shifts its focus toward sustainable solutions, investments in electric vehicle technology and infrastructure are expected to accelerate. The growing awareness among consumers about the ecological impact of traditional vehicles and the increasing prevalence of pollution-related health concerns have further catalyzed the demand for electric vehicles. With the burgeoning middle class and an increasing number of young professionals, particularly in metropolitan areas, the consumer base for EVs is expanding rapidly. As battery costs continue to decline and the range of electric vehicles improves, it is anticipated that consumer acceptance and adoption will rise. This confluence of factors positions the Indian EV market for robust growth, and the projected market size of USD 20 billion by 2025 serves as a testament to the significant opportunities that lie ahead for stakeholders within this industry.

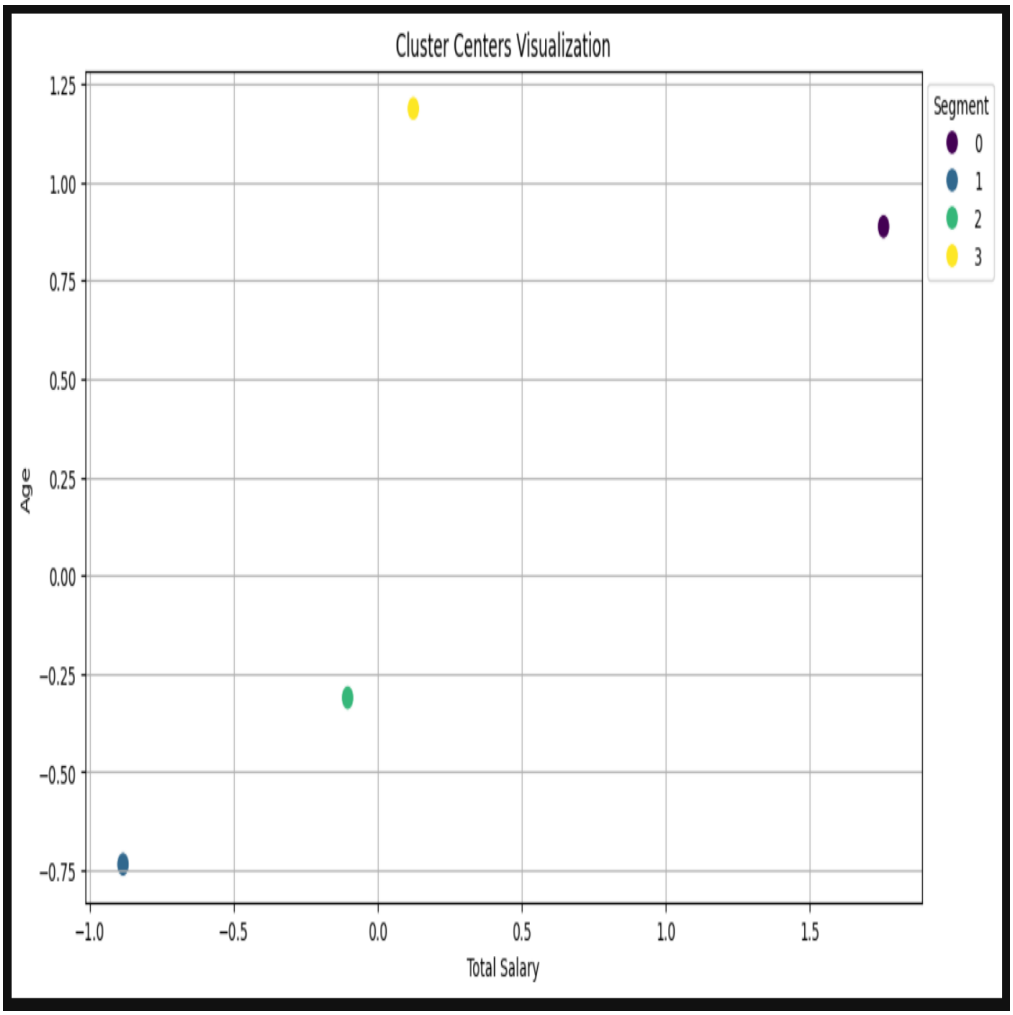
8. Optimal Variables for Segmentation:

In the context of market segmentation for the Indian EV sector, identifying the most relevant variables is crucial for effectively targeting distinct consumer groups. The analysis highlights four optimal variables/features that significantly influence purchasing decisions and consumer preferences.

1. Income Level: This variable serves as a fundamental determinant of purchasing power. Higher income levels typically correlate with an increased willingness to invest in premium electric vehicles equipped with advanced technology and features. Understanding income distribution among potential customers allows companies to tailor their offerings, whether focusing on luxury models or more affordable options that appeal to cost-conscious consumers.
2. Age Group: Age plays a pivotal role in shaping preferences for vehicle types. Younger consumers, often tech-savvy and environmentally conscious, may prioritize innovative features and sustainability, while older consumers might place greater emphasis on reliability and comfort. By segmenting the market based on age, businesses can develop targeted marketing strategies that resonate with the specific values and needs of each demographic group.
3. Vehicle Type Preference: This variable captures consumer interest in specific types of electric vehicles, such as two-wheelers, sedans, or SUVs. Understanding these preferences enables manufacturers to optimize their product development strategies, ensuring that they meet the demands of various consumer segments. For instance, younger urban dwellers may show a preference for compact electric vehicles, while families might lean towards spacious electric SUVs.
4. City of Residence: Urban versus rural preferences can significantly impact purchasing decisions. Urban consumers are more likely to prioritize charging infrastructure and vehicle performance due to their daily commuting needs, whereas rural consumers may be more concerned about vehicle affordability and adaptability to local conditions. By analyzing city of residence, businesses can gain insights into regional trends and tailor their marketing efforts accordingly, ensuring that they address the unique challenges and opportunities present in different geographical markets.

These optimal variables not only provide valuable insights for effective market segmentation but also empower businesses to devise strategic marketing initiatives that align with the preferences and behaviors of their target audiences.

```
plt.figure(figsize=(12, 6))
sns.scatterplot(data=center_df, x='Total Salary', y='Age', hue='Segment', palette='viridis', s=100)
plt.title('Cluster Centers Visualization')
plt.xlabel('Total Salary')
plt.ylabel('Age')
plt.grid()
plt.legend(title='Segment', bbox_to_anchor=(1, 1), loc='upper left')
plt.show()
```



9. References:

To substantiate the analysis conducted within this report, several key resources have been referenced. First, a direct link to the dataset utilized for segmentation analysis provides transparency and allows for further exploration of the underlying data. This dataset contains essential information regarding consumer behavior and preferences within the EV market in India, forming the basis for the insights derived in this analysis. Additionally, relevant industry reports and articles that delve into the current state and future projections of the EV market in India have been incorporated. These resources offer comprehensive overviews of market trends, government policies, and competitive landscapes, providing a robust context for understanding the dynamics of the EV sector. The inclusion of these references is crucial, as they lend credibility to the findings and conclusions presented, facilitating informed decision-making for stakeholders looking to enter or expand within the Indian EV market.

Appendix

The appendix section complements the analysis by providing additional visualizations and detailed tables that support the findings. These visual aids include graphs and charts illustrating market trends, consumer preferences, and segment distributions, offering a clearer representation of the data analyzed. Furthermore, detailed tables containing statistical analyses, such as mean income levels, vehicle preferences across age groups, and the geographic distribution of potential EV buyers, enhance the depth of the analysis. By presenting this supplementary information, the appendix serves as a valuable resource for readers seeking a more comprehensive understanding of the factors influencing the Indian EV market, ultimately reinforcing the strategic insights discussed in the main sections of the report.