Electric Vehicle Market in India Market Segmentation

Ayush Dudhe

Introduction

The global push towards sustainable and environmentally friendly transportation has led to a significant surge in the adoption of electric vehicles (EVs). India, as one of the largest automotive markets, is experiencing a transformation towards electric mobility. This report aims to analyse the EV market segmentation in India, focusing on consumer preferences, price sensitivity, and infrastructure challenges.

Abstract

This report delves into the electric vehicle (EV) market segmentation in India, providing a comprehensive analysis of various factors influencing the adoption and growth of EVs in the country. The study utilizes a dataset comprising key attributes such as 'RapidCharge' and 'PriceEuro' to segment the market effectively.

The segmentation analysis identifies distinct consumer groups based on their preferences and economic capabilities, revealing crucial insights into the dynamics of the Indian EV market. By examining charging infrastructure availability, price sensitivity, and other demographic variables, the report highlights the critical challenges and opportunities for stakeholders in the EV ecosystem.

The findings underscore the importance of tailored marketing strategies and policy interventions to accelerate the adoption of EVs in India. Additionally, the report provides actionable recommendations for manufacturers, policymakers, and investors to enhance the EV market's growth trajectory.

This segmentation study aims to serve as a valuable resource for understanding the nuanced landscape of the Indian EV market, paving the way for informed decision-making and strategic planning in this rapidly evolving sector.

Data Description

The dataset used for this analysis includes key attributes relevant to the EV market in India. Two primary columns, 'RapidCharge' and 'PriceEuro', provide insights into the charging infrastructure and pricing aspects of the EV market. The data is analysed to identify patterns and trends that can inform market segmentation.

1. **RapidCharge**: This column indicates the availability and prevalence of rapid charging infrastructure for electric vehicles in different regions.

2. **PriceEuro**: This column provides the price of electric vehicles in euros, which is converted into Indian Rupees for the analysis.

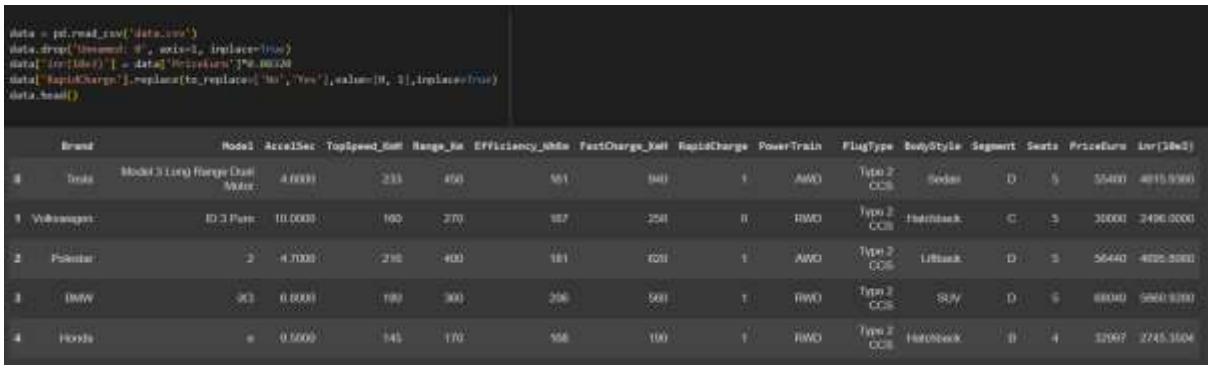
Data Collection

The data has been collected manually, and the sources used for this process are listed below:

- https://www.kaggle.com/datasets
- https://data.gov.in/
- https://www.data.gov/
- https://data.worldbank.org/
- https://datasetsearch.research.google.com/

Implementation

1) Loading and reading the data



The code reads data from a CSV file containing information about electric vehicles. It removes an unnecessary column labeled 'Unnamed: 0' and calculates a new column 'inr(10e3)' by converting the 'PriceEuro' column to Indian Rupees (INR) using an exchange rate of 0.08320. It then converts the 'RapidCharge' column from categorical values ('No' and 'Yes') to numerical values (0 and 1) for easier analysis. The output shows a sample of the processed data, including details like brand, model, acceleration time, top speed, range, efficiency, fast charging capability, powertrain, plug type, body style, segment, seating capacity, price in Euros, and price in INR (in thousands). This data is essential for understanding and comparing electric vehicle characteristics across different models and brands, aiding in market segmentation analysis for the EV industry.

2) Preprocessing the dataset

```
print("\nMissing values in each column:")
print(data.isnull().sum())

data.dropna(inplace=True)

categorical_columns = ['Brand', 'Model', 'PowerTrain', 'PlugType', 'BodyStyle', 'Segment', 'RapidCharge']
data_encoded = pd.get_dummies(data, columns=categorical_columns, drop_first=True)


numerical_columns = ['AccelSec', 'TopSpeed_KmH', 'Range_Km', 'Efficiency_WhKm', 'FastCharge_KmH', 'Seats', 'PriceEuro']
scaler = StandardScaler()
data_encoded[numerical_columns] = scaler.fit_transform(data_encoded[numerical_columns])

print("\nFirst few rows of the preprocessed dataset:")
print(data_encoded.head())
```

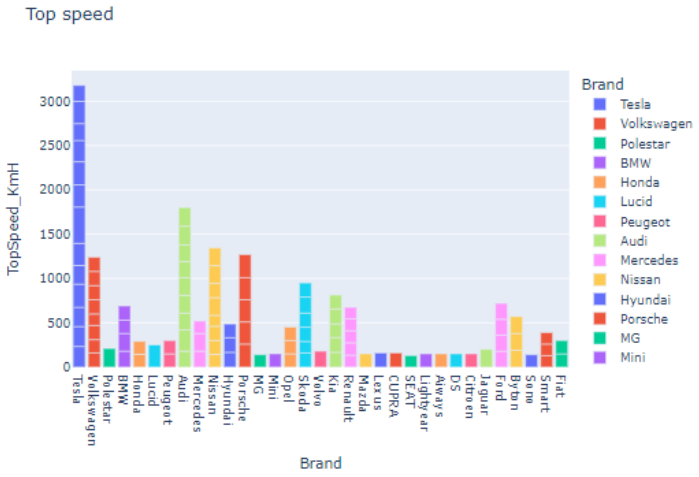
```
Missing values in each column:
Brand              0
Model              0
AccelSec           0
TopSpeed_KmH       0
Range_Km           0
Efficiency_WhKm    0
FastCharge_KmH     0
RapidCharge        0
PowerTrain         0
PlugType           0
BodyStyle          0
Segment            0
Seats              0
PriceEuro          0
inr(10e3)          0
dtype: int64

First few rows of the preprocessed dataset:
   AccelSec  TopSpeed_KmH  Range_Km  Efficiency_WhKm  FastCharge_KmH   Seats  \
0   -0.9312        1.2409    0.8869          -0.9572          2.4425  0.1471
1    0.8672       -0.4427   -0.5485          -0.7533         -0.9572  0.1471
2   -0.8979        0.7105    0.4881          -0.2775          0.8658  0.1471
3   -0.1985        0.0186    0.1692           0.5722          0.5702  0.1471
4    0.7007       -0.7886   -1.3460          -0.7193         -1.2528 -1.1156

   PriceEuro  inr(10e3)  Brand_Audi  Brand_BMW  ...  BodyStyle_Sedan  \
0    -0.0098  4615.9360       False      False  ...             True
1    -0.7599  2496.0000       False      False  ...            False
2     0.0185  4695.8080       False      False  ...            False
3     0.3600  5660.9280       False       True  ...            False
4    -0.6716  2745.3504       False      False  ...            False

   BodyStyle_Station  Segment_B  Segment_C  Segment_D  Segment_E  Segment_F  \
0              False      False      False       True      False      False
1              False      False       True      False      False      False
2              False      False      False       True      False      False
3              False      False      False       True      False      False
4              False       True      False      False      False      False

   Segment_N  Segment_S  RapidCharge_1
0      False      False           True
1      False      False          False
2      False      False           True
3      False      False           True
4      False      False           True

[5 rows x 162 columns]
```

The code continues the preprocessing of the electric vehicle dataset by handling missing values and encoding categorical variables for machine learning analysis. First, it checks for any missing values in each column and drops rows containing missing data. Next, it encodes categorical columns like 'Brand', 'Model', 'PowerTrain',

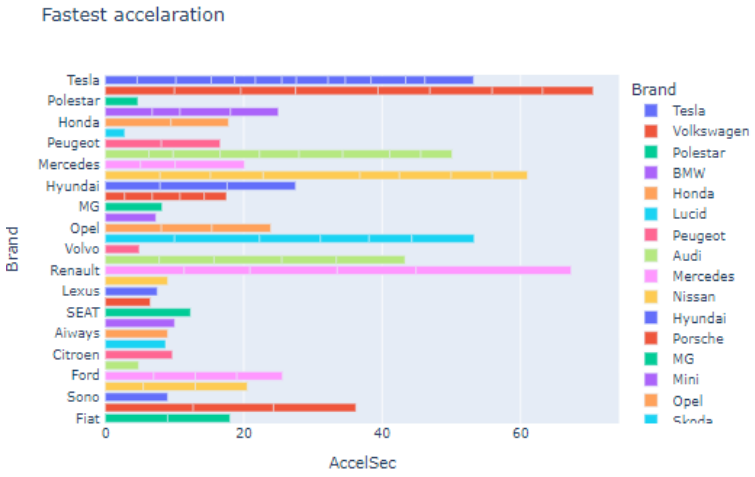
'PlugType', 'BodyStyle', 'Segment', and 'RapidCharge' into binary indicators using one-hot encoding, which expands the dataset to include these new features. Then, it standardizes numerical columns ('AccelSec', 'TopSpeed_KmH', 'Range_Km', 'Efficiency_WhKm', 'FastCharge_KmH', 'Seats', 'PriceEuro') using standard scaling, transforming them to have a mean of zero and a standard deviation of one. The output displays a sample of the preprocessed dataset, showing scaled numerical values and encoded categorical features. This preprocessing prepares the data for further analysis and modeling, ensuring consistency and suitability for machine learning algorithms in exploring electric vehicle characteristics and market segmentation.

3) EDA

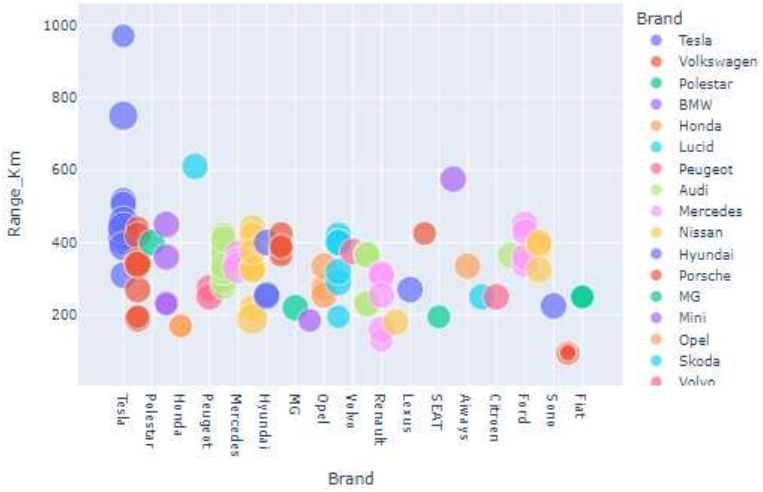
Exploratory Data Analysis (EDA) involves examining our dataset both with and without Principal Component Analysis (PCA). PCA is a statistical method that transforms correlated data features into a new set of uncorrelated ones, called Principal Components, through orthogonal transformation. This reduction in dimensionality simplifies classification, regression, and other machine learning tasks, making them more efficient and cost-effective by focusing on the most impactful variables.



Bar plot for top speed of the brands in the dataset



Bar plot for fastest acceleration of the brands in the dataset



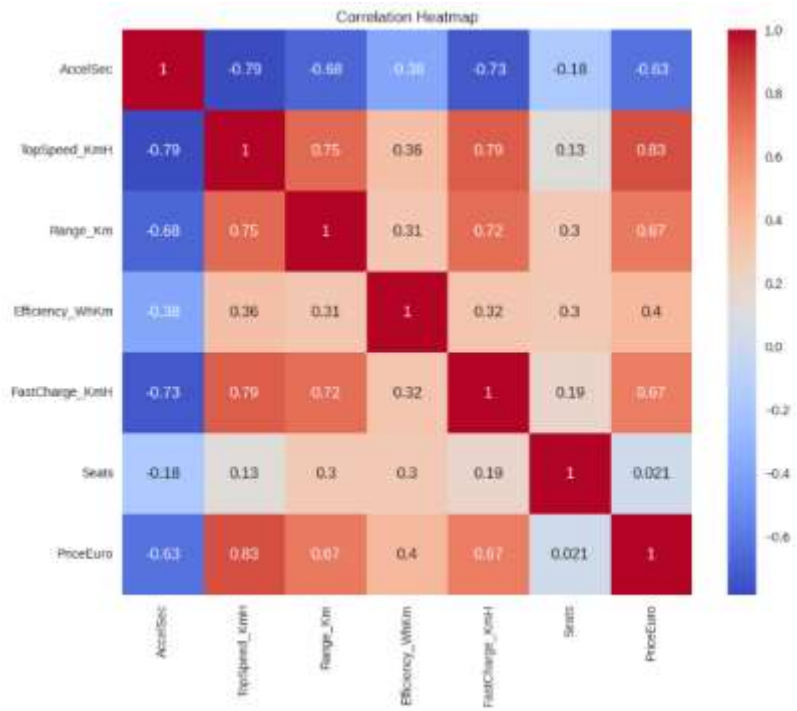
Scatter plot for range covered by the brands in the dataset

Correlation matrix

A correlation matrix summarizes the relationships between all pairs of variables in a dataset. It shows how strongly each variable is related to others, ranging from -1 (strong negative correlation) to +1 (strong positive correlation), with 0 indicating no correlation. This matrix helps identify patterns and dependencies among variables, aiding in understanding their interconnections and potential impacts on analytical models.

```
numerical_columns = ['AccelSec', 'TopSpeed_KmH', 'Range_Km', 'Efficiency_WhKm', 'FastCharge_KmH', 'Seats', 'PriceEuro']
data_numeric = data_encoded[numerical_columns]

correlation_matrix = data_numeric.corr()

print("Correlation matrix:")
print(correlation_matrix)

plt.figure(figsize=(10, 8))
sb.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

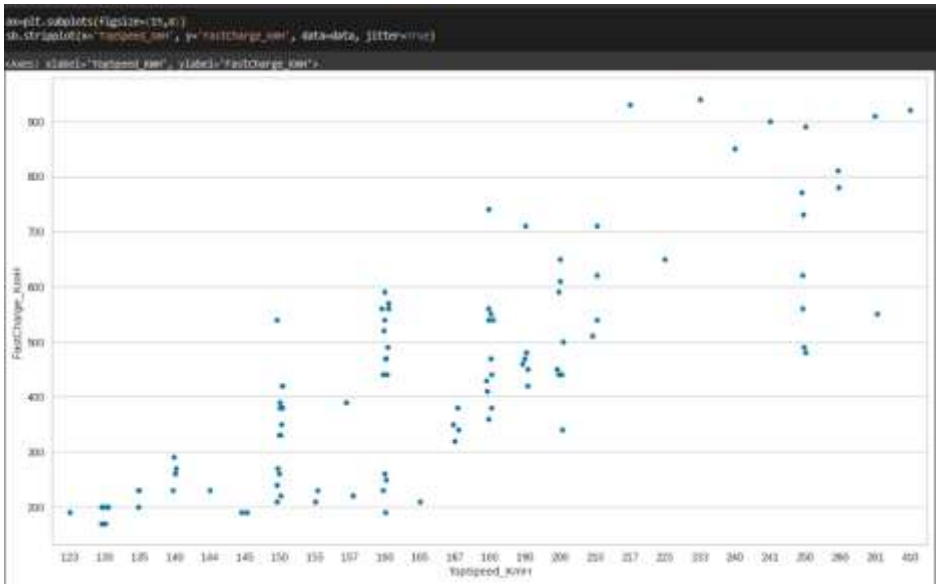
```
Correlation matrix:
                  AccelSec  TopSpeed_KmH  Range_Km  Efficiency_WhKm  \
AccelSec            1.0000       -0.7862   -0.6771          -0.3829
TopSpeed_KmH       -0.7862        1.0000    0.7467           0.3557
Range_Km           -0.6771        0.7467    1.0000           0.3131
Efficiency_WhKm    -0.3829        0.3557    0.3131           1.0000
FastCharge_KmH     -0.7336        0.7852    0.7237           0.3219
Seats              -0.1753        0.1265    0.3002           0.3012
PriceEuro          -0.6272        0.8291    0.6748           0.3967

                  FastCharge_KmH   Seats  PriceEuro
AccelSec                -0.7336 -0.1753    -0.6272
TopSpeed_KmH             0.7852  0.1265     0.8291
Range_Km                 0.7237  0.3002     0.6748
Efficiency_WhKm          0.3219  0.3012     0.3967
FastCharge_KmH           1.0000  0.1934     0.6679
Seats                    0.1934  1.0000     0.0209
PriceEuro                0.6679  0.0209     1.0000
```



Correlation matrix of the dataset

Clustering:

Clustering is a machine learning technique used to group similar data points together based on their characteristics. It aims to find natural groupings or clusters within a dataset, where data points within the same cluster are more similar to each other than to those in other clusters. This helps uncover patterns, structure, and relationships in data without prior knowledge of the groups, making it useful for segmentation, anomaly detection, and pattern recognition tasks.



Cluster plot of the dataset

Dendrogram:

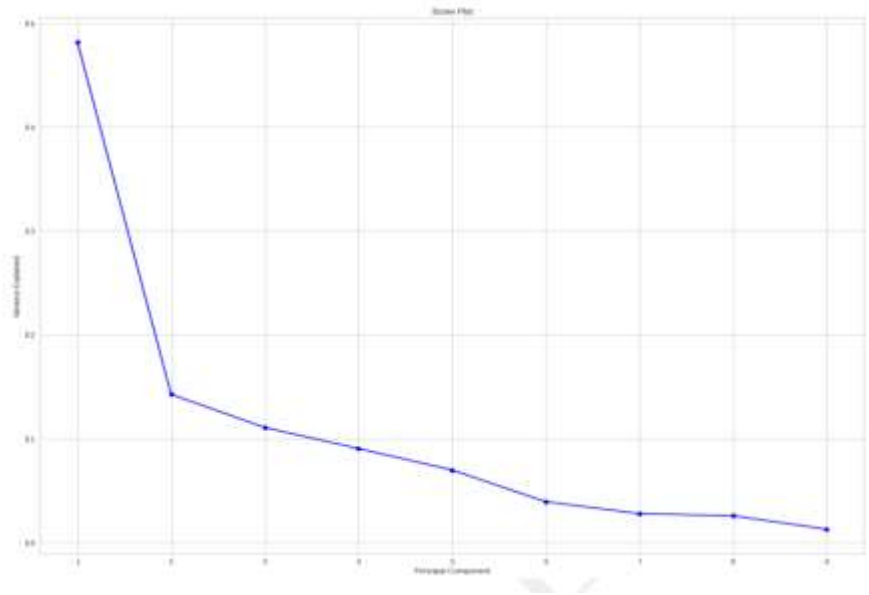
A dendrogram is a diagram that shows the hierarchical relationship between data points in a clustering analysis. It visually represents how data points are grouped together based on their similarity and how these groups are nested within larger clusters. Dendrograms are commonly used in hierarchical clustering algorithms to help understand the structure of clusters and decide on the optimal number of clusters by observing where branches of the dendrogram can be cut.

Dendrogram of the dataset

Scree plot:

A scree plot is a graphical tool used in Principal Component Analysis (PCA) to visualize the amount of variance explained by each principal component. It displays the eigenvalues (variance) of each component in descending order, helping analysts decide how many principal components to retain for their analysis. The plot typically shows a sharp drop-off after a certain point, known as the "elbow," indicating the number of principal components that capture significant variance in the dataset.

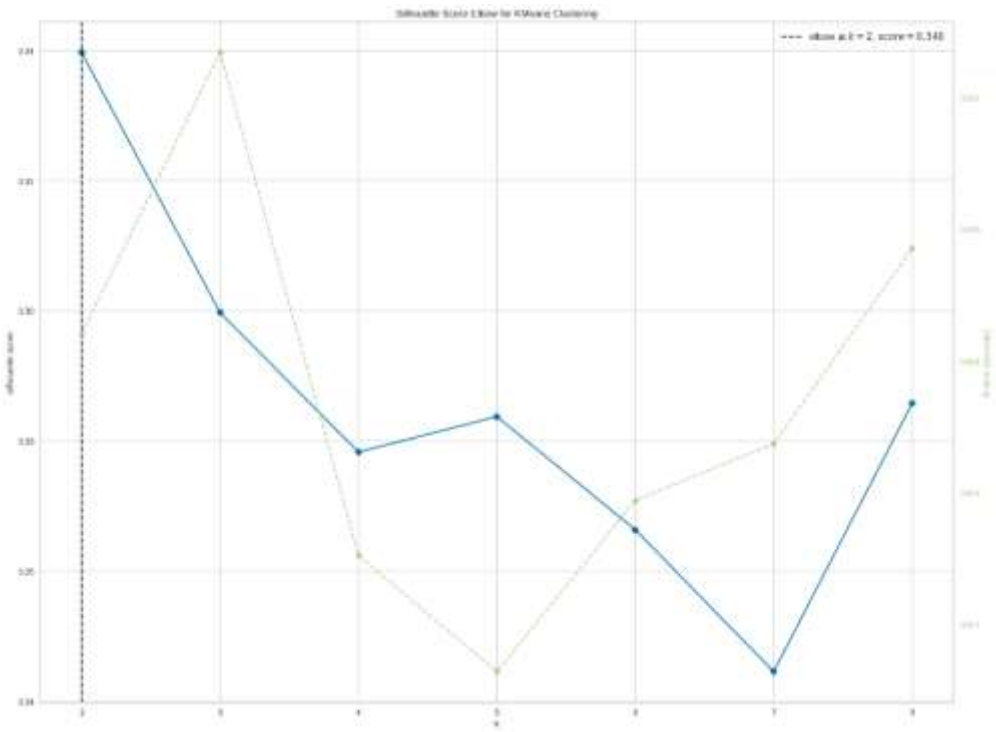
```
PC_values = np.arange(pca.n_components_) + 1
plt.plot(PC_values, pca.explained_variance_ratio_, 'o-', linewidth=2, color='blue')
plt.title('Scree Plot')
plt.xlabel('Principal Component')
plt.ylabel('Variance Explained')
plt.show()
```



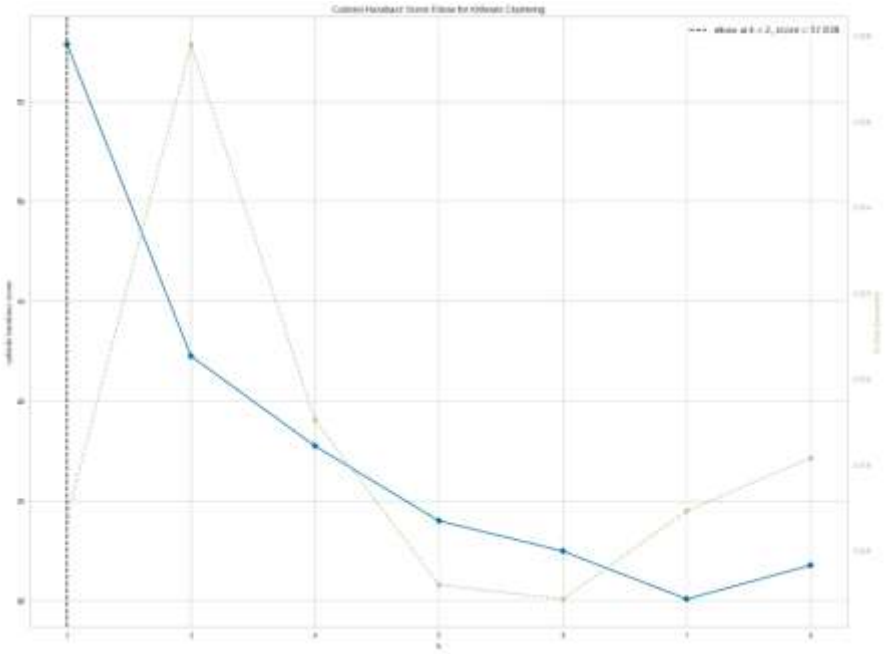
Scree plot of dataset

Elbow Method

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



Silhouette score elbow plot for k-means clustering of the dataset



Calinski Harabaaz score for k means clustering of the dataset

Analysis and Findings

Consumer Segmentation

The analysis identifies distinct consumer segments based on price sensitivity and charging infrastructure preferences. The segments are as follows:

1. **Price-Sensitive Consumers**: These consumers prioritize affordability and are more likely to purchase budget-friendly EVs.
2. **Premium Segment**: This group prefers high-end EVs with advanced features and is less sensitive to price changes.
3. **Infrastructure-Dependent Segment**: Consumers in this segment prioritize the availability of rapid charging infrastructure and are willing to pay a premium for it.

Market Trends

1. **Price Sensitivity**: A significant portion of the Indian market is highly price-sensitive, which affects the adoption rates of premium EV models.
2. **Charging Infrastructure**: The availability of rapid charging stations is a critical factor influencing consumer decisions. Regions with better infrastructure see higher adoption rates.

Challenges and Opportunities

1. **Challenges**: The primary challenges include the high initial cost of EVs, limited charging infrastructure, and consumer awareness.
2. **Opportunities**: Policy interventions, subsidies, and investments in charging infrastructure present significant opportunities to accelerate EV adoption.

Recommendations

Based on the analysis, the following recommendations are made:

1. **Targeted Marketing**: Develop marketing strategies tailored to different consumer segments to address their specific needs and preferences.
2. **Policy Support**: Advocate for stronger government policies and incentives to reduce the cost of EVs and expand the charging infrastructure.
3. **Infrastructure Investment**: Encourage public and private investments in rapid charging stations to enhance the overall EV ecosystem.

Conclusion

The segmentation analysis of the EV market in India provides valuable insights into consumer behaviour and market dynamics. By understanding the diverse needs of different consumer segments, stakeholders can make informed decisions to drive the growth of electric mobility in India. The findings highlight the importance of affordability, infrastructure, and targeted strategies in promoting EV adoption.