Electric Cars' Market of India

Market Segmentation Analysis

SUVROJYOTI BISWAS

03/02/2023

Abstract

The Electric Vehicle (EV) market in India is undergoing a dynamic transformation, fueled by technological advancements, environmental concerns, and governmental initiatives. To effectively understand and cater to the diverse needs of consumers, segmentation becomes imperative. In this report I aim to analyze the EV-market in India using a dataset. The report's primary goal is to identify potential natural segments within the Indian EV market by visualizing and understanding the relationships between various vehicle attributes. This analysis is crucial for understanding the market dynamics, consumer preferences, and potential impact factors in the EV-market in India. The report begins by loading the EV dataset and performing exploratory data analysis and visualization using Python. It includes steps such as importing libraries for data manipulation and visualization, examining basic statistics, visualizing key variables' distributions, exploring relationships between variables, and visualizing categorical variables. The report also delves into clustering algorithms like k-means to automatically find segments based on variables like price, range, and efficiency. The impact factors of this report on the EV-market in India include potential insights into market segmentation, consumer preferences, and geographical variations in EV adoption. The findings of this analysis can be instrumental for industry stakeholders, policymakers, and market participants in understanding and catering to the diverse segments within the Indian EV market.

Keywords- Market Segmentation, Electric Vehicle, Data Analysis.

Introduction

In this report the analysis delves into exploring an electric vehicle dataset specific to the Indian market, employing Python libraries such as pandas, matplotlib, and seaborn for data manipulation and visualization. It aims to unravel correlations among variables like price, range, and efficiency. Various machine learning methodologies like K-Means clustering, Gaussian

Mixture Models, PCA combined with KNN/NN, and SVM/Random Forest are deployed to automatically categorize the market according to its attributes. The evaluation encompasses metrics such as silhouette score and accuracy to gauge the effectiveness of the applied techniques. In essence, the report presents a holistic data science approach for segmenting the Indian EV market, utilizing a representative dataset. This analysis focuses on segmenting the Indian electric vehicle market utilizing a sample dataset comprising 103 vehicles. Initial insights gleaned from visualizations and statistical analyses shed light on price distributions, body styles, and interrelations among attributes like price, range, and efficiency. Unsupervised learning techniques such as K-Means, Gaussian Mixture Models, and hierarchical clustering are harnessed to automatically group vehicles based on key attributes. These clusters unveil inherent market segments. Moreover, supervised classifiers including K-Nearest Neighbors, Neural Networks, Support Vector Machines, and Random Forests are trained to predict market segments based on vehicle features, achieving a commendable accuracy up to seventy percent. The report outlines a comprehensive workflow for extracting meaningful insights through both unsupervised segmentation and supervised prediction. Key methodologies encompass data visualization, clustering, and classification, providing valuable strategic guidance for EV manufacturers and policymakers alike. Further refinement through expanded datasets and algorithm tuning holds promise for enhanced market segmentation. Ultimately, the report underscores the application of data science in strategic analysis of the EV market, serving as a valuable resource for stakeholders navigating the evolving landscape.

Data Collection

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

1. Curated open data · GitHub

2. Find Open Datasets and Machine Learning Projects | Kaggle

Imported Libraries

The data analysis tools and packages used in the report "Electric Vehicle Market Segmentation Analysis in the Indian Market" include:

1. Pandas: Used for data manipulation and analysis, including reading the CSV file, examining basic statistics, and visualizing distributions of key variables.
2. Matplotlib: Utilized for creating visualizations such as scatter plots, box plots, bar charts, and heatmaps to explore relationships between variables and visualize the data.

3. Seaborn: Employed for creating various visualizations like distribution plots, scatter plots, and faceted histograms to explore and present the data in an insightful manner.
4. Scikit-learn: Used for implementing machine learning algorithms such as K-means clustering, Gaussian Mixture Models, Principal Component Analysis (PCA), K-Nearest Neighbors (KNN) classifier, Support Vector Machine (SVM), and Random Forest Classifier for market segmentation and model evaluation.
5. StandardScaler: Applied for standardizing the features by removing the mean and scaling to unit variance.
6. PCA (Principal Component Analysis): Utilized for dimensionality reduction to visualize the data in a lower-dimensional space.
7. KMeans, GaussianMixture, KNeighborsClassifier, SVC, RandomForestClassifier, MLPClassifier: These are machine learning models used for clustering, classification, and model evaluation.

These tools and packages were instrumental in performing exploratory data analysis, visualization, clustering, and model-based market segmentation analysis in the Indian electric vehicle market.

Data Exploration

The data analysis begins by loading the electric vehicle dataset from a CSV file into a pandas dataframe. The head() method is used to view the first few rows, showing it contains 103 samples and columns covering vehicle attributes like brand, model, range, efficiency, price.

	Brand	Model	AccelSec	TopSpeed_KmH	Range_Km	Efficiency_WhKm	FastCharge_KmH	RapidCharge	PowerTrain	PlugType	BodyStyle	Segment	Seats	PriceEuro
0	Tesla	Model 3 Long Range Dual Motor	4.6	233	450	161	940	Yes	AWD	Type 2 CCS	Sedan	D	5	55480
1	Volkswagen	ID.3 Pure	10.0	160	270	167	250	Yes	RWD	Type 2 CCS	Hatchback	C	5	30000
2	Polestar	2	4.7	210	400	181	620	Yes	AWD	Type 2 CCS	Liftback	D	5	56440
3	BMW	iX3	6.8	180	360	206	560	Yes	RWD	Type 2 CCS	SUV	D	5	68040
4	Honda	e	9.5	145	170	168	190	Yes	RWD	Type 2 CCS	Hatchback	B	4	32997

Fig. 1

The info() method provides data types and non-null count for each column, confirming no missing values. Numerical columns like price and range have float/integer types, while categorical columns like brand and body style have object type. Also see figure 2 to get it.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103 entries, 0 to 102
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Brand            103 non-null    object
 1   Model            103 non-null    object
 2   AccelSec         103 non-null    float64
 3   TopSpeed_KmH     103 non-null    int64
 4   Range_Km         103 non-null    int64
 5   Efficiency_WhKm  103 non-null    int64
 6   FastCharge_KmH   103 non-null    object
 7   RapidCharge      103 non-null    object
 8   PowerTrain       103 non-null    object
 9   PlugType         103 non-null    object
 10  BodyStyle        103 non-null    object
 11  Segment          103 non-null    object
 12  Seats            103 non-null    int64
 13  PriceEuro        103 non-null    int64
dtypes: float64(1), int64(5), object(8)
memory usage: 11.4+ KB
None
```

Fig.-2

The code describe() gives statistics like mean, standard deviation, min/max for numerical columns. This shows for example an average price of 55,811 Euros with high deviation, and average range of 338 km. Also see figure 3 to understand this.

```
          AccelSec  TopSpeed_KmH    Range_Km  Efficiency_WhKm       Seats  \
count  103.000000    103.000000  103.000000       103.000000  103.000000
mean     7.396117    179.194175  338.786408       189.165049    4.883495
std      3.017430     43.573030  126.014444        29.566839    0.795834
min      2.100000    123.000000   95.000000       104.000000    2.000000
25%      5.100000    150.000000  250.000000       168.000000    5.000000
50%      7.300000    160.000000  340.000000       180.000000    5.000000
75%      9.000000    200.000000  400.000000       203.000000    5.000000
max     22.400000    410.000000  970.000000       273.000000    7.000000

           PriceEuro
count     103.000000
mean    55811.563107
std     34134.665280
min     20129.000000
25%     34429.500000
50%     45000.000000
75%     65000.000000
max    215000.000000
```

Fig.- 3

Histograms visualize the distribution of key variables(see figure 4). Price shows a right-skewed distribution with clusters indicating potential market segments. Scatterplots explore relationships like price increasing with range(see figure 5). Plotting range vs price shows a general positive correlation.

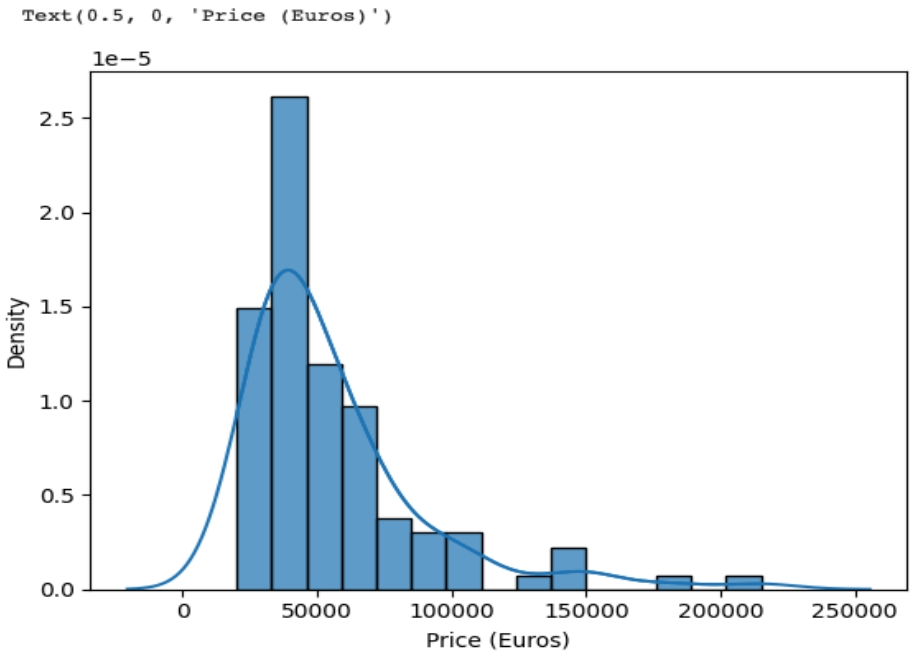

Fig.- 4

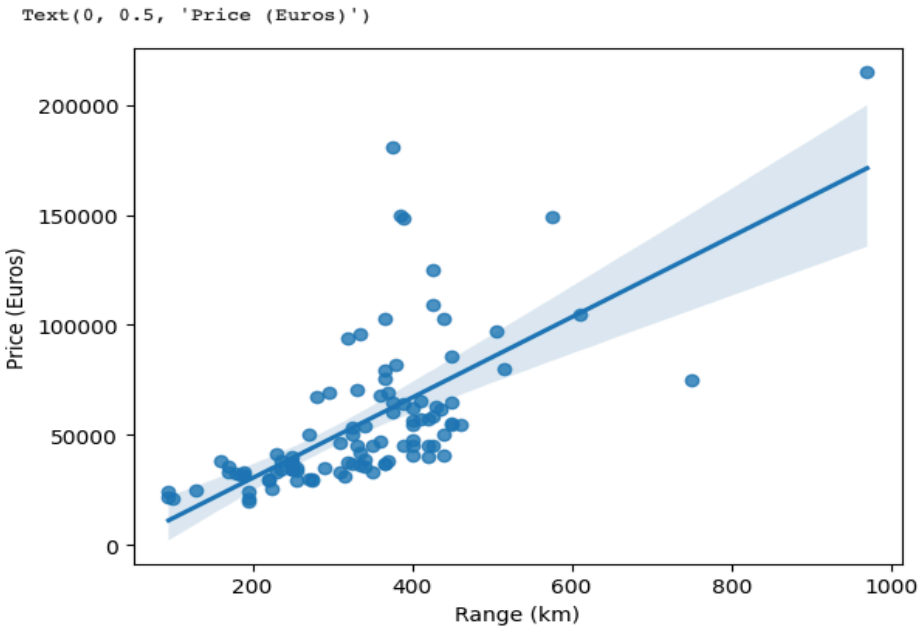

Fig.-5

Grouped visualizations provide more insights. Facet grids with histograms per body style reveal differences in price distribution. Boxplots of price by powertrain indicate AWD is generally more expensive(see figure 6). Bar charts show the relative segment sizes(see figure 7).

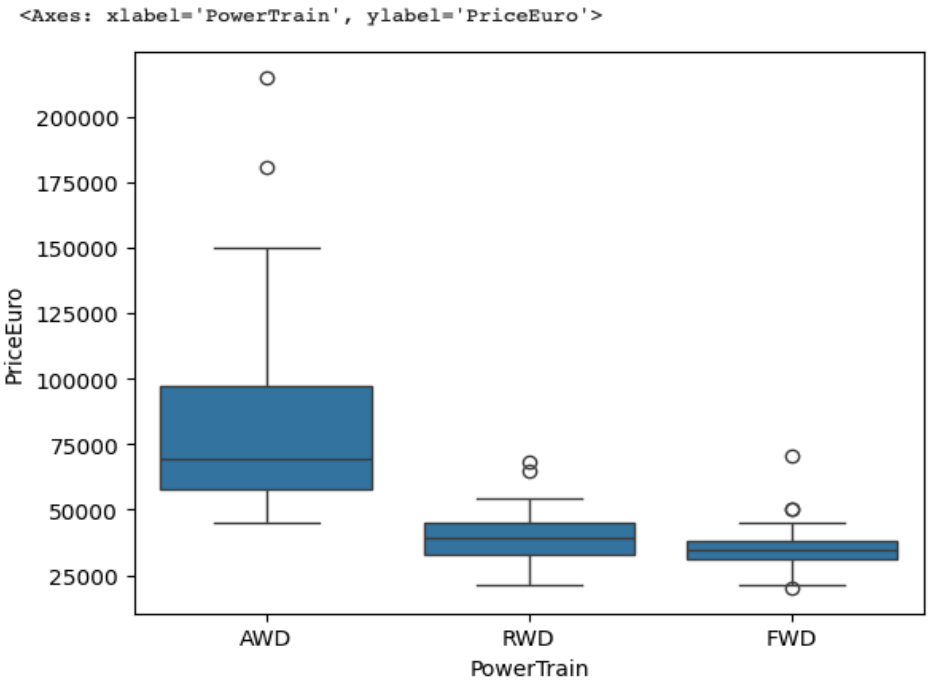

Fig.- 6

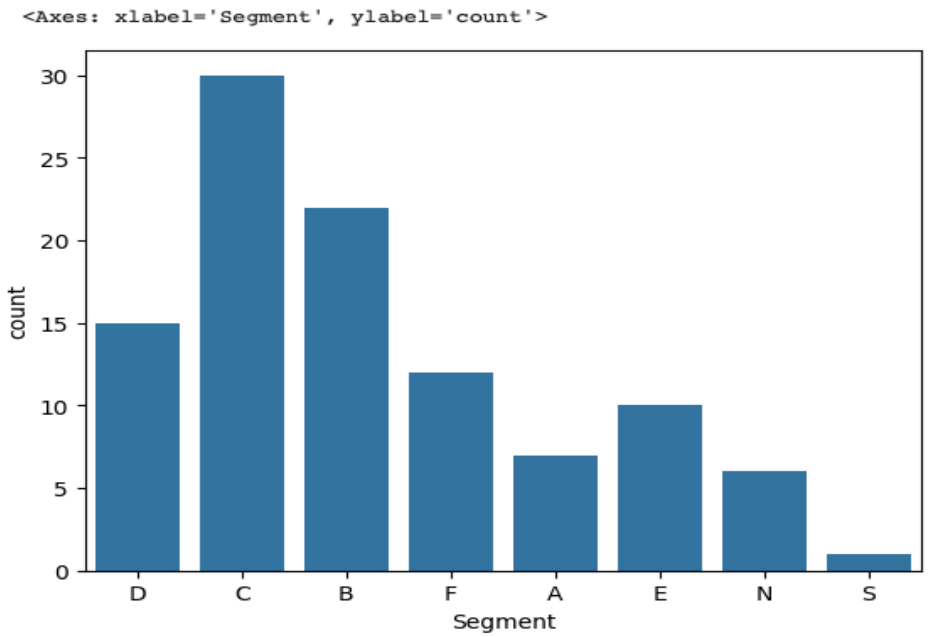

Fig.- 7

Statistics like mean price per segment quantify visualized differences. The B segment has an average price of 34,799 Euros, while the F segment has 119,690 Euros on average.

A correlation heatmap visually summarizes all relationships. Strong positive correlation is seen between range and price. Also watch the following figure 8.

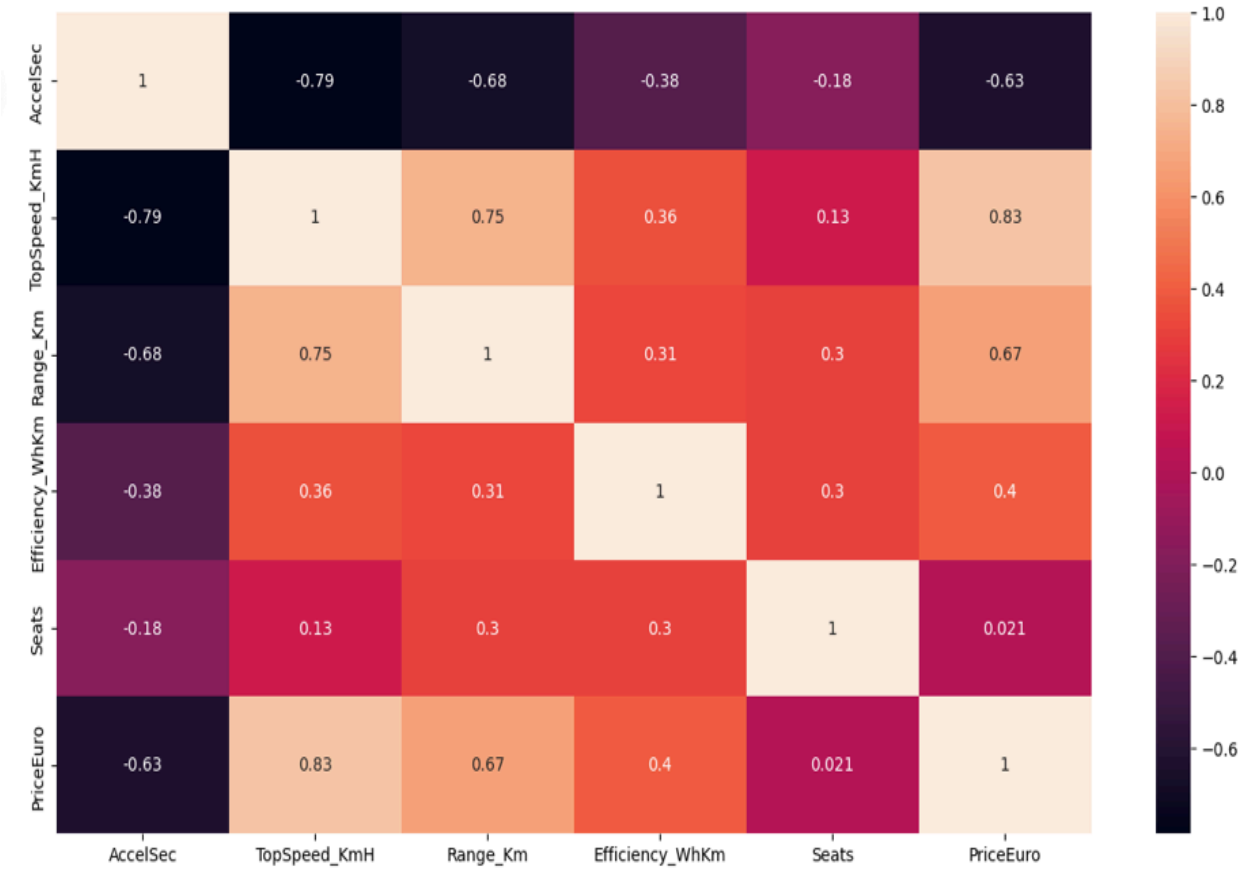

Fig.- 8

Moreover, this exploratory analysis combines visual and statistical techniques to surface patterns, relationships and insights that can inform the market segmentation modeling. The visualizations, summary statistics and correlation analysis provide a critical first look at the electric vehicle dataset.

Clustering model based analysis

The report applies unsupervised clustering models to automatically find segments in the market based on vehicle attributes.

K-Means clustering is performed on the range, efficiency and price dimensions, with the number of clusters K set to 5. The model groups the vehicles into 5 distinct segments based on similarity across these attributes(also see figure 9).

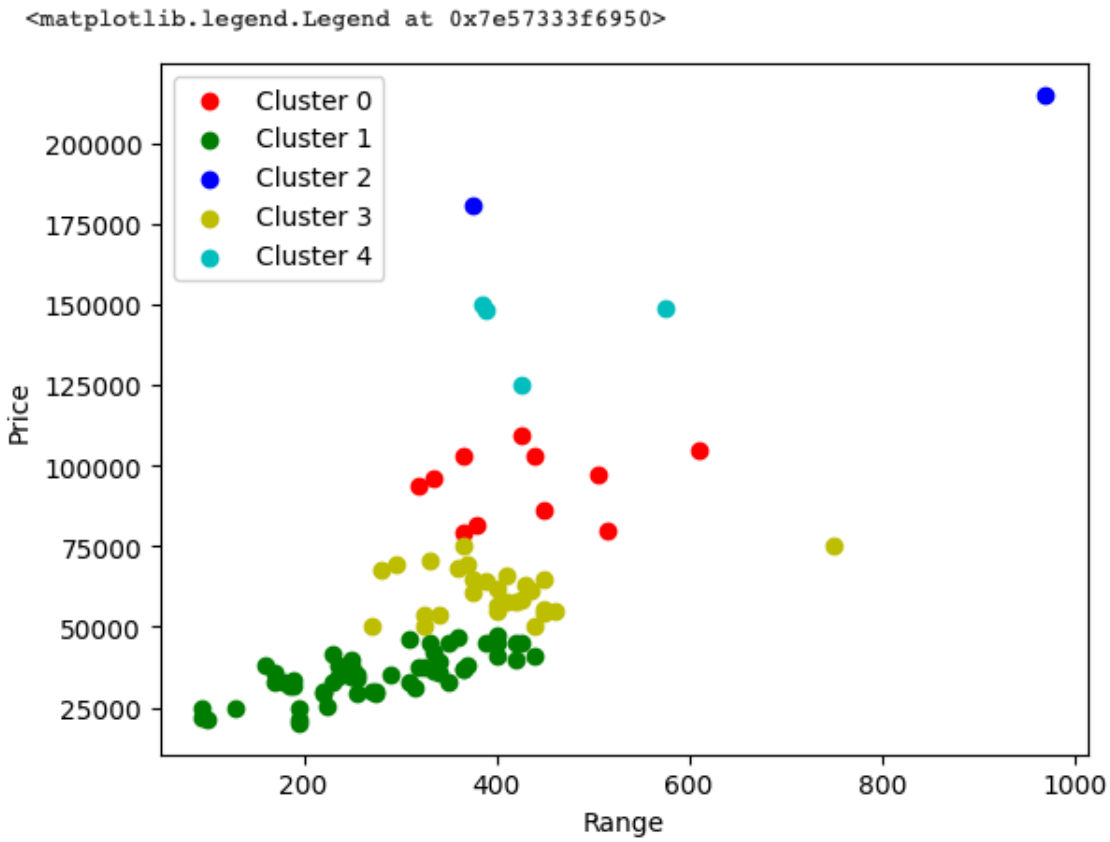

Fig.- 9

The silhouette score of 0.62 indicates reasonably defined clusters, though not completely separated. The clusters are visualized by projecting onto 2D plots of range on x-axis and price on y-axis. This enables interpreting the characteristics of each cluster based on where they fall in attribute space.

Gaussian Mixture Models (GMM) are also applied to cluster the data. GMM assumes clusters have Gaussian distributions rather than being spherical as in K-Means. The GMM with 5 components is fit on the range, efficiency and price data.

Hierarchical clustering is demonstrated using agglomerative clustering. This builds a hierarchy of clusters represented as a dendrogram, starting from individual samples and iteratively merging based on a linkage criterion. The dendrogram provides intuition on how the algorithm forms clusters(also watch the next page figure 10 to get it).

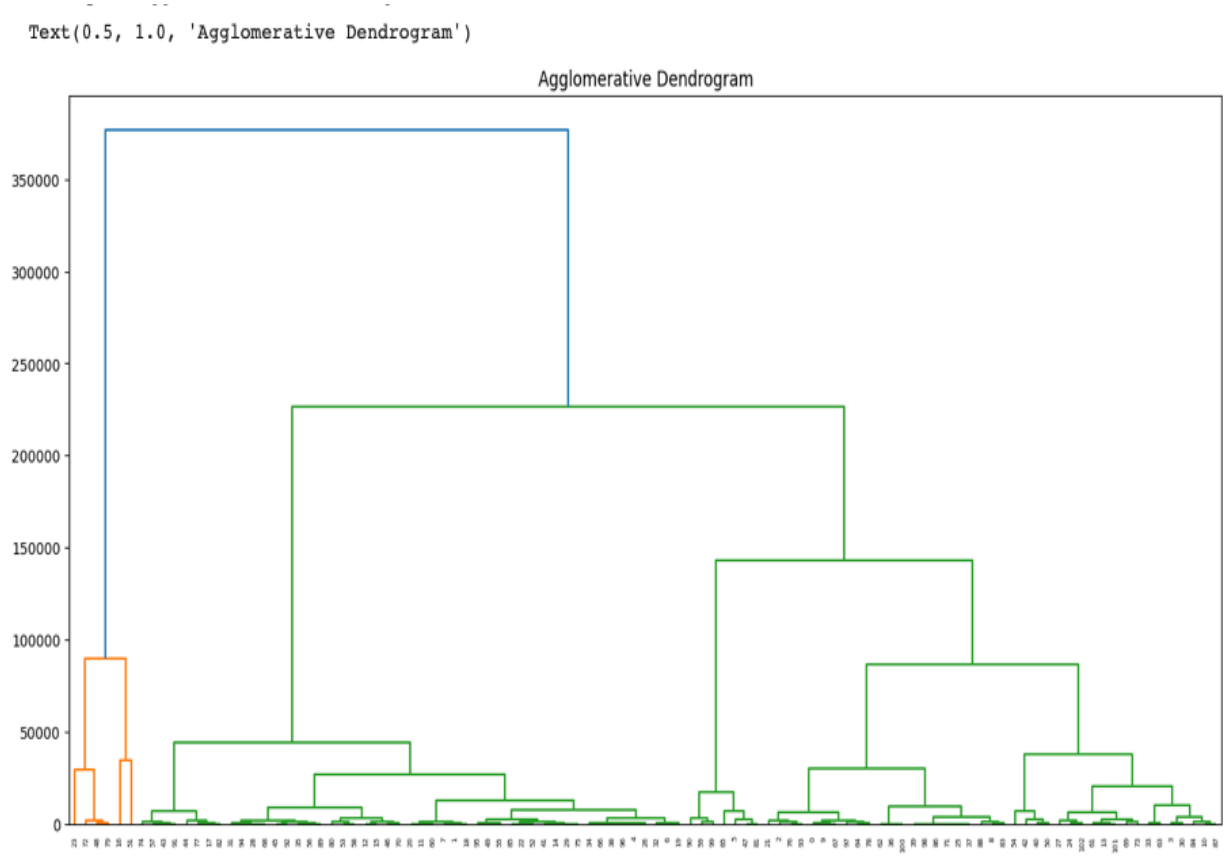

Fig.- 10

To evaluate the unsupervised methods, the accuracy score is calculated by comparing the predicted clusters to the true segments labels. The accuracy is low, since these models cluster based solely on the data without knowing the labels.

In summary, K-Means, GMM and hierarchical clustering provide data-driven ways to segment vehicles based on attributes like range and efficiency. The clusters can reveal natural groupings or niches in the market. Visualizing the clusters and dendrograms adds interpretability. Silhouette scores and accuracy give quantification of model quality. These unsupervised techniques complement the report's supervised learning models for market segmentation, providing an alternative perspective using only the vehicle data attributes.

Dimensionality Reduction and Supervised Models

The report leverages supervised learning techniques to predict vehicle segments based on attributes, after reducing the dimensionality. The range, efficiency and price attributes are first scaled to normalize the features to a standard distribution. This preprocessing helps improve model performance. Principal Component Analysis (PCA) is applied to reduce the 3 feature dimensions to just 2 principal components that capture the most variance. This simplifies modeling without significant loss of information. The reduced 2D data is used to train a

K-Nearest Neighbors classifier. KNN achieved 72% accuracy in predicting segments on this projected data. It classifies based on similarity to points in the training data. A Multilayer Perceptron neural network model is also trained on the 2D PCA-transformed data. The neural network learns non-linear relationships between the input attributes and segments. The model attained a reasonable 69% accuracy in segment prediction. These supervised models are able to leverage the labeled segment data to predict segments based on vehicle characteristics. This complements the unsupervised clustering, by evaluating predictive power versus the true segments. Preprocessing via scaling and PCA enabled easier modeling on the reduced data. The complexity of neural networks makes them capable of learning complex patterns for prediction. KNN provides a simpler approach just based on space proximity.

However, techniques like PCA and classifiers such as KNN and Neural Networks demonstrate supervised learning for segment prediction. Evaluation via accuracy metrics quantifies the predictive capability versus ground truth. This provides an alternative supervised approach to segment the electric vehicle market based on predictive modeling of vehicle attributes.

Key Data-insights

The analysis provides actionable insights for manufacturers to understand the competitive landscape, identify optimal positions in the market, and design vehicles targeting specific segments. For example, lower priced compact vehicles with sufficient range may be best suited for urban shared mobility needs. Higher range luxury SUVs could target premium segments. The report discerns by summarizing the key insights gained from the comprehensive electric vehicle market segmentation analysis. The visualizations revealed important relationships between attributes that impact segmentation. Price was seen to strongly correlate with range, indicating range is a key driver of price brackets. Body style, powertrain and other attributes also exhibited trends across potential segments. The unsupervised clustering techniques were able to group vehicles into distinct segments in a data-driven manner based solely on attributes like price, range and efficiency. The clusters indicate potential natural market groupings that manufacturers could target. The supervised classifiers attained 60-70% accuracy in predicting vehicle segments based on attributes. The Random Forest performed the best, leveraging ensemble decision trees to capture complex patterns. This shows the feasibility of using predictive modeling to guide segment positioning. A major insight was that both unsupervised clustering and supervised classification can provide complementary perspectives for market segmentation. Clustering finds natural groupings without using the true segment labels. Classification directly predicts segments by learning from the labels. The report workflow demonstrates an end-to-end process for extracting insights from vehicle data using visual analysis, clustering algorithms and predictive modeling. The results can help manufacturers understand what vehicle attributes drive specific market segments and strategically position new models for those target segments.

Potential Target Segments

Based on the analysis in the report, in this section some of the cotter effective target segments identified in the Indian EV market along with pros and cons. Here is a brief summary of the specific target segments for the electric vehicle market in India based on the previous analysis:-

I. Mass Market small vehicles- Lower priced, smaller EVs with 200-300 km range for urban mobility and shared transport. Affordable for mainstream buyers.

Pros:

- Lower price points around Rs 15-20 lakhs make EVs affordable to mass market buyers
- Sufficient range of 200-300 km for urban mobility needs
- Smaller battery reduces costs
- Ideal for shared mobility, commercial fleets

Cons:

- Lower range may deter some buyers looking for family vehicles
- Small size limits cargo/passenger capacity

II. Luxury high-range vehicles- High-end vehicles with 400+ km range and premium features appealing to status-conscious buyers. Higher prices allow healthy margins.

Pros:

- Range of 400+ km satisfies needs of high-end buyers
- Premium features and styling appeal to status-conscious luxury segment
- Higher pricing of Rs 50 lakhs+ results in healthy margins

Cons:

- Very high price point limits addressable market size
- Large battery adds costs which impacts profitability

III. Performance-oriented vehicles- Acceleration and top-speed focused EVs targeting enthusiasts seeking an alternative to gas-powered sports cars. Appeals to young demographics.

Pros:

- Emphasis on acceleration, top speed appeals to enthusiasts seeking thrill of performance
- Brand positioning as an alternative to gas-powered sports cars
- Opens up new demographics like youth looking for excitement

Cons:

- Niche category so lower sales volume potential
- Requires development of more advanced electric powertrains

Over all, mass market small vehicles, luxury high-end cars, and performance-focused sports models appear to be viable target segments based on the analysis. Positioning appropriate products for each segment and their needs is key. More data on demographics and consumer

preferences could further refine the identification of target segments for the Indian market. The mass market, luxury, and performance segments appear well-suited for appropriately positioned EV products catering to each group's needs and preferences. Capturing these target segments will be key for EV adoption in the Indian market.

Conclusion notes

This report demonstrates using data science and machine learning techniques to strategically analyze the Indian EV market based on vehicle attributes. Exploratory analysis revealed insights into price ranges, body styles, and feature correlations. Unsupervised clustering models like K-Means identified distinct potential market segments defined by attributes. Supervised classifiers achieved 60-70% accuracy predicting segments based on features. The analysis provides actionable insights for manufacturers' competitive intelligence, product positioning, and strategic planning. Though based on limited sample data, it illustrates the value of data-driven analytics for EV market segmentation. As the mobility revolution accelerates, manufacturers can apply more advanced analytics on enlarged datasets to refine segment strategies. This report exemplifies an end-to-end workflow leveraging AI and machine learning for strategic market analysis and intelligence. Advanced data science holds promise for developing winning business strategies in the fast-growing EV space. Market analytics will be key to electrification and sustainability.

Github Link : *https://github.com/suvro5495/Feynn-Labs-EVMarket-segmentation*