Deepthi.V

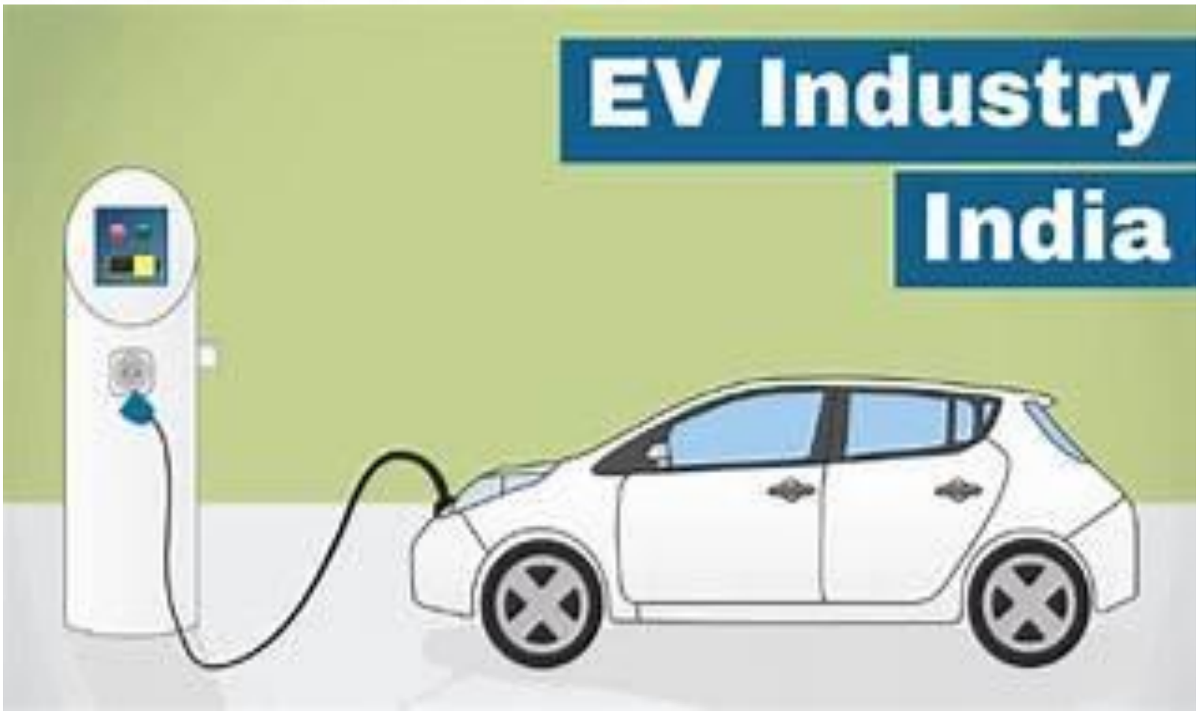
Feynn Labs-Project 2

# Market Segment Analysis of EV Market

*GitHub Link:*
*https://github.com/valasapalli/EV_project/blob/main/EV%20.ipynb*

## Electric Vehicle Market in India

## (Market Segmentation)



# Overview:

EVs are vehicles powered by one or more electric motors instead of traditional internal combustion engines.They can be powered by electricity from external sources or autonomously by batteries, sometimes supplemented by solar panels or fuel cells.EVs encompass various types of vehicles, including cars, trucks, trains, vessels, aircraft, and spacecraft.EVs have been around since the mid-19th century when electricity was a preferred propulsion method due to its comfort and ease of operation compared to gasoline cars.However, internal combustion engines dominated the automotive industry for about a century, while electric power remained common in other types of vehicles.In the 21st century, EVs have experienced a resurgence due to technological advancements and increased focus on renewable energy and mitigating climate change.The EV industry in India is at a nascent stage, representing less than 1% of total vehicle sales but with the potential to grow to over 5% in the coming years.Presently, there are over 500,000 electric two-wheelers and a few thousand electric cars on Indian roads.Most electric scooters in India operate at low speeds (<25km/hr) and do not require registration or licenses. They typically use lead batteries to keep prices low.However, battery failures and short battery life are major concerns, alongside challenges with charging infrastructure and government incentives.With initiatives like FAME-2 (Faster Adoption and Manufacture of Electric Vehicles), the industry is expected to witness significant growth in volumes and technology.Market segmentation is crucial for understanding and targeting potential buyers of EVs, considering factors such as psychographics, behavior, and socio-economic characteristics.Analytical techniques such as cluster analysis, discriminant analysis, and chi-square tests are employed to identify distinct consumer segments.The study identifies three consumer groups—conservatives, indifferent, and enthusiasts—expected to drive EV adoption among young consumers.The findings offer insights for scholars and policymakers to encourage EV adoption in emerging sustainable transport markets.Recommendations may include targeted marketing strategies, incentives, and infrastructure development to address consumer needs and preferences.Overall, the passage highlights the evolution of EVs, their current status in India, and the importance of market segmentation and consumer behavior analysis in promoting their adoption for a sustainable future.The main aim of this study is to explore and identify distinct sets of potential buyer segments for EVs based on psychographic, Behavioral, and distinct sets of potential buyer segments for EV's based on psychographic Behavioral, and socio-economic characterization by employing an integrated research framework of 'perceived benefits-attitude-intention'. The study applied robust analytical procedures including cluster analysis, multiple discriminant analysis and chi-square test to operationalize and validate segments from the data collected of 563 respondents using a cross-sectional online survey. The finding post the three distinct sets of young consumer groups have been identified and labelled as conservatives', 'indifferent' , and 'enthusiasts' which are deemed to be buddying EV buyers The implications are recommended, which may offer some pertinent guidance for scholars and policy marker to encourage EVs adoption in the backdrop of emerging sustainable transport market.

# MARKET SEMENTATION

## Target Market:

The target market of Electric Vehicle Market Segmentation can be categorized into Geographic, Sociodemographic, Behavioral, and Psychographic Segmentation.

**Behavioral Segmentation:** Behavioral Segmentation: searches directly for similarities in Behavioral or reported Behavioral. Example: prior experience with the product, amount spent on the purchase etc

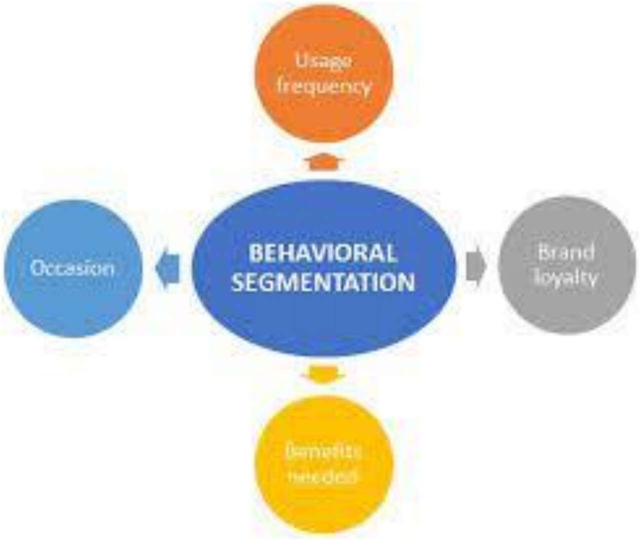


**Fig 1. Behavioral Segmentation**

Advantage: uses the very Behavioral of interest is used as the basis of segment extraction.
Disadvantage: not always readily available.

**Psychographic Segmentation**: Grouped based on beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product. Suitable for lifestyle segmentation. Involves many segmentation variables.

**Advantage**: generally, more reflective of the underlying reasons for differences in consumer Behavioral.

**Disadvantage**: increased complexity of determining segment memberships for consumers.

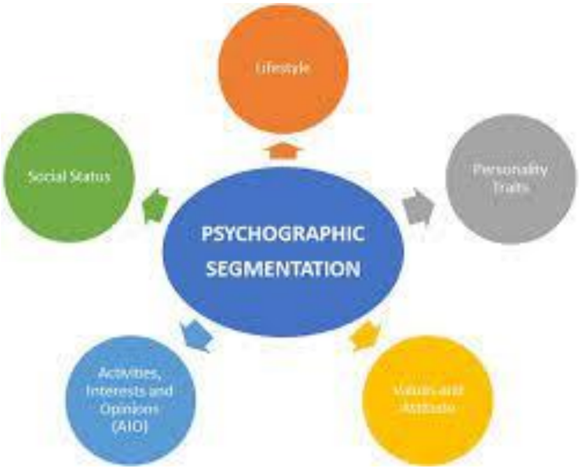

Lifestyle
Social Status
PSYCHOGRAPHIC SEGMENTATION
Personality Traits
Activities, Interests and Opinions (AIO)
Values and Attitude

**Fig 2 Psychographic Segmentation**

**Socio-Demographic Segmentation:** Includes age, gender, income and education. Useful in industries

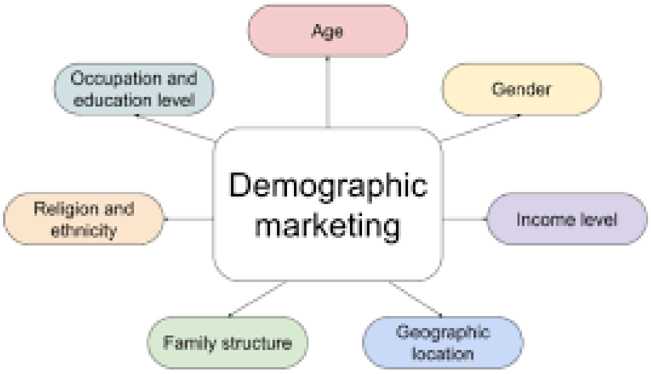


**Fig.3 Demographic Marketing**

**Advantage:** segment membership can easily be determined for every customer.

**Disadvantage**: if this criterion is not the cause for customers product preferences then it does not provide sufficient market insight for optimal segmentation decisions.

## Data Collection: -

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

- https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset
- https://www.kaggle.com/datasets
- https://data.worldbank.org/
- https://drive.google.com/drive/folders/137KIMhwpB1bx5zx0hTaa486bEKe3kXaB?usp=share_link

## Segmenting for Electric Vehicle Market:

The market segmentation approach aims at defining actionable, manageable, homogenous subgroups of individual customers to whom the marketers can target with a similar set of marketing strategies. In practice, there are two ways of segmenting the market-a-priori and post-hoc. An a-priori approach utilizes predefined characteristics such as age, gender, income, education, etc. to predefine the segments followed by profiling based on a host of measured variables (Behavioral, psychographic or benefit). In the post-hoc approach to segmentation on other hand, the segments are identified based on the relationship among the multiple measured variables. The commonality between both approaches lies in the fact that the measured variables determine the 'segmentation theme'. The present study utilizes an a-priori approach to segmentation so as to divide the potential EV customers into sub-groups.

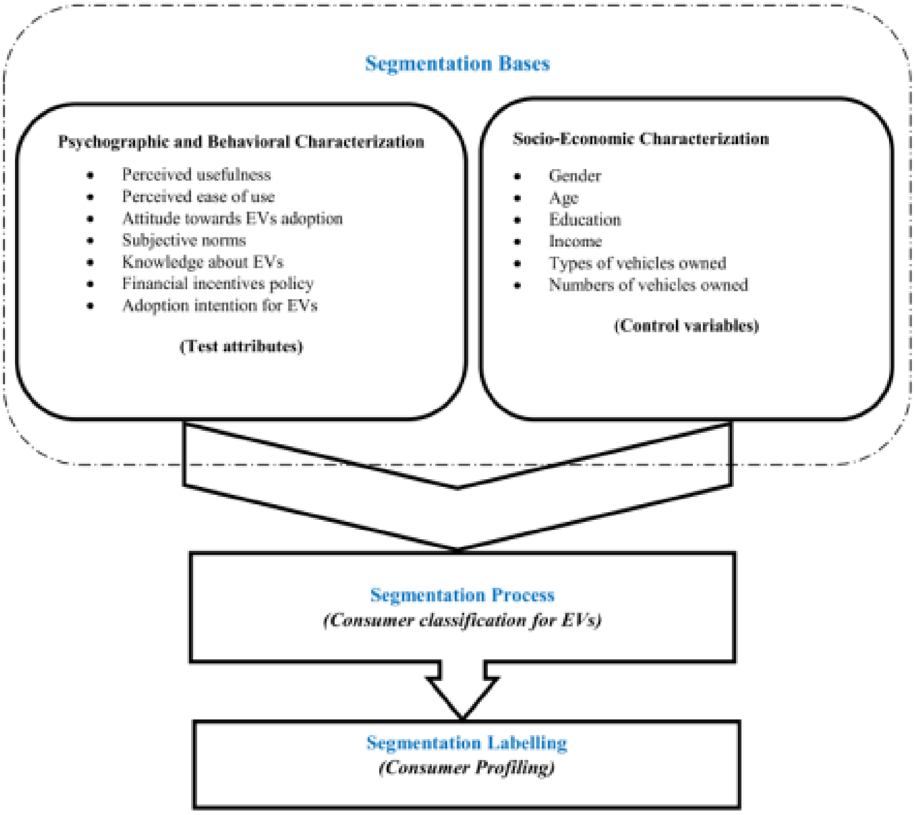


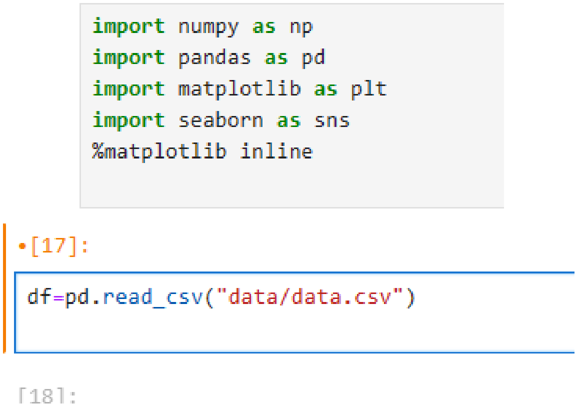
**Fig.4 Market Segmentation Electric Vehicles**

It is argued that the blended approach of psychographic and socioeconomic attributes for market segmentation enables the formulation of sub-market strategies which in turn satisfy the specific tastes and preferences of the consumer groups. Straughan and Roberts presented a comparison between the usefulness of psychographic, demographic, and economic characteristics based on consumer evaluation for eco-friendly products. Feynn Labs They pinpointed the perceived superiority of the psychographic characteristics over the socio-demographic and economic ones in explaining the environmentally-conscious consumer Behavioral and thus, the study recommended the use of psychographic characteristics in profiling the consumer segments in the market for eco-friendly products. The present study adds perceived-benefit characteristics guided by blended psychographic and socio-economic aspects for segmenting the consumer market.

# IMPLEMENTATION

## Packages/Tools used:

1. **NumPy** : To Calculate Various calculation related to arrays.
2. **Pandas** : To read or Load and manipulation data datasets.
3. **Matplotlib or Seaborn**: To Plot and all visualization of datasets.
4. **Scikit-learn**: We have sued LabelEncoder() to encode our values and used K-mean() unsupervised method.
5. **Python** : Python Language used to write all code in that Language.



```
import numpy as np
import pandas as pd
import matplotlib as plt
import seaborn as sns
%matplotlib inline
```

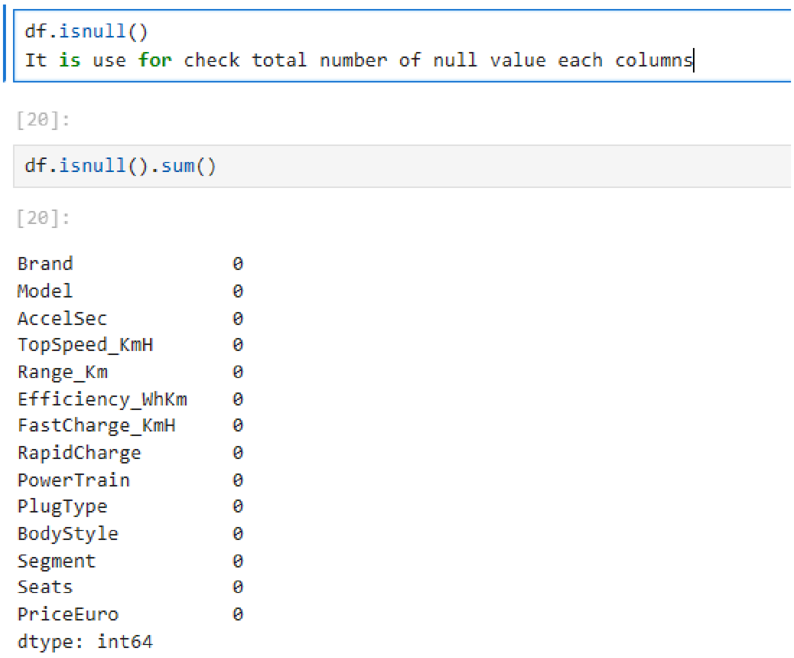
```
•[17]:
df=pd.read_csv("data/data.csv")

[18]:
```

**Fig.5. Library import and Dataset load in Jupyter**

## Data-Preprocessing:

The Data collected is compact and is partly used for visualization purposed and partly for clustering. The datasets does not have any null value or empty value cell in dataset so we must first pre-processed dataset check dataset does not have empty cell or null value.



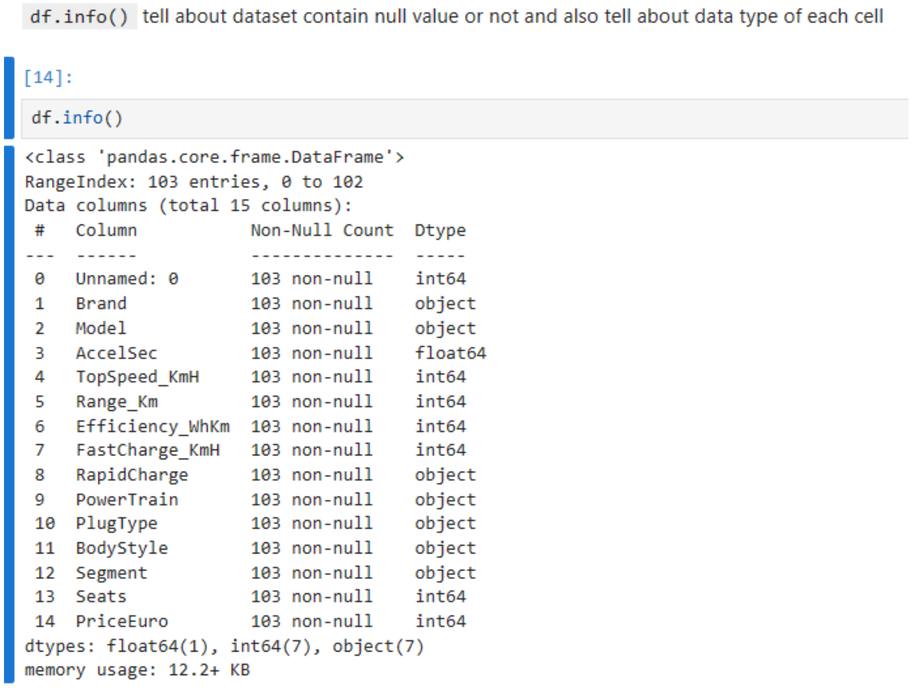
```
df.isnull()
It is use for check total number of null value each columns
```

```
[20]:
df.isnull().sum()

[20]:

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
dtype: int64
```

**Fig. 6. Check total null value**



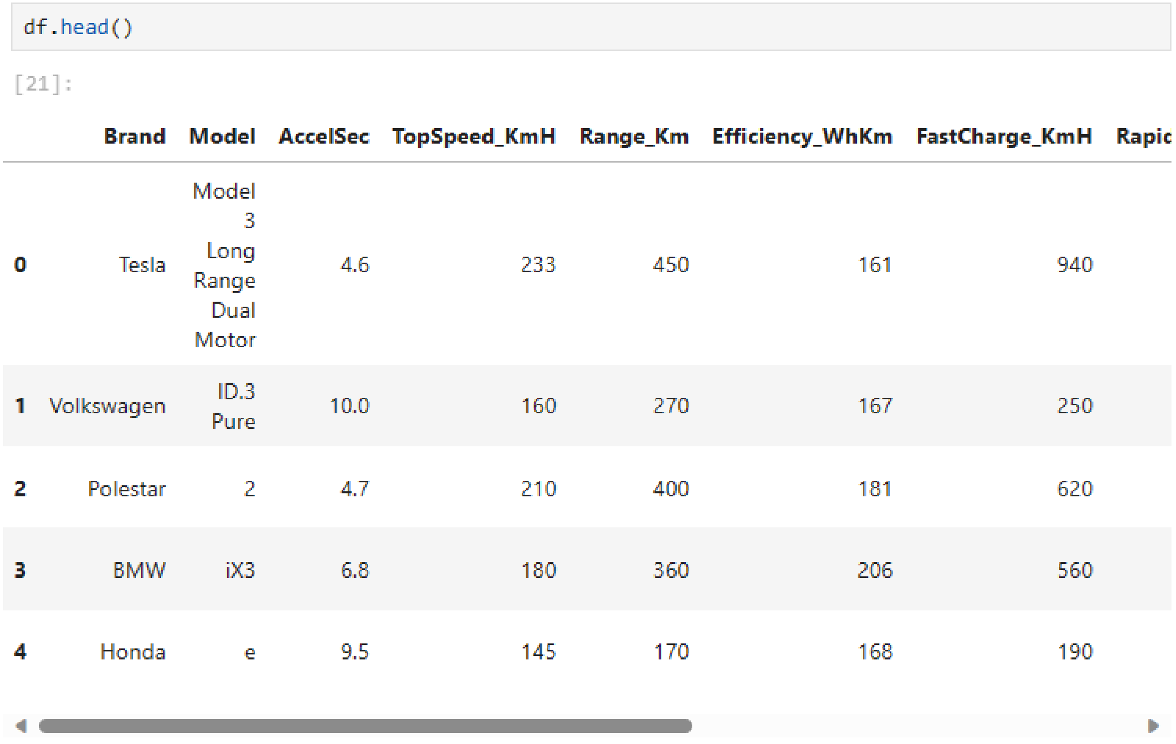
`df.info()` tell about dataset contain null value or not and also tell about data type of each cell

[14]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103 entries, 0 to 102
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   Unnamed: 0       103 non-null    int64  
 1   Brand            103 non-null    object 
 2   Model            103 non-null    object 
 3   AccelSec         103 non-null    float64
 4   TopSpeed_KmH     103 non-null    int64  
 5   Range_Km         103 non-null    int64  
 6   Efficiency_WhKm  103 non-null    int64  
 7   FastCharge_KmH   103 non-null    int64  
 8   RapidCharge      103 non-null    object 
 9   PowerTrain       103 non-null    object 
 10  PlugType         103 non-null    object 
 11  BodyStyle        103 non-null    object 
 12  Segment          103 non-null    object 
 13  Seats            103 non-null    int64  
 14  PriceEuro        103 non-null    int64  
dtypes: float64(1), int64(7), object(7)
memory usage: 12.2+ KB
```

**Fig .7. Dataset contain null value or datatype of column**



```
df.head()
```

[21]:

| | Brand | Model | AccelSec | TopSpeed_KmH | Range_Km | Efficiency_WhKm | FastCharge_KmH | Rapic |
|---|---|---|---|---|---|---|---|---|
| **0** | Tesla | Model 3 Long Range Dual Motor | 4.6 | 233 | 450 | 161 | 940 | |
| **1** | Volkswagen | ID.3 Pure | 10.0 | 160 | 270 | 167 | 250 | |
| **2** | Polestar | 2 | 4.7 | 210 | 400 | 181 | 620 | |
| **3** | BMW | iX3 | 6.8 | 180 | 360 | 206 | 560 | |
| **4** | Honda | e | 9.5 | 145 | 170 | 168 | 190 | |

**Fig 8. Dataset Print using Head() function**

# EDA(EXPLORING DATA)

First we explore the key characteristic of the data set by loading the set and inspecting basic feature such as the variables names, the sample size, and the first three rows of the data. We start the Exploratory Data Analysis with some data Analysis drawn from the data without Principal Component analysis and with some Principle Component Analysis in the dataset obtained from the combination of all the data we have. PCA is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. These new transformed features are called the Principal Components. The process helps in reducing dimensions of the data to make the process of classification/regression or any form of machine learning, cost-effective.

## Comparison of cars in our data

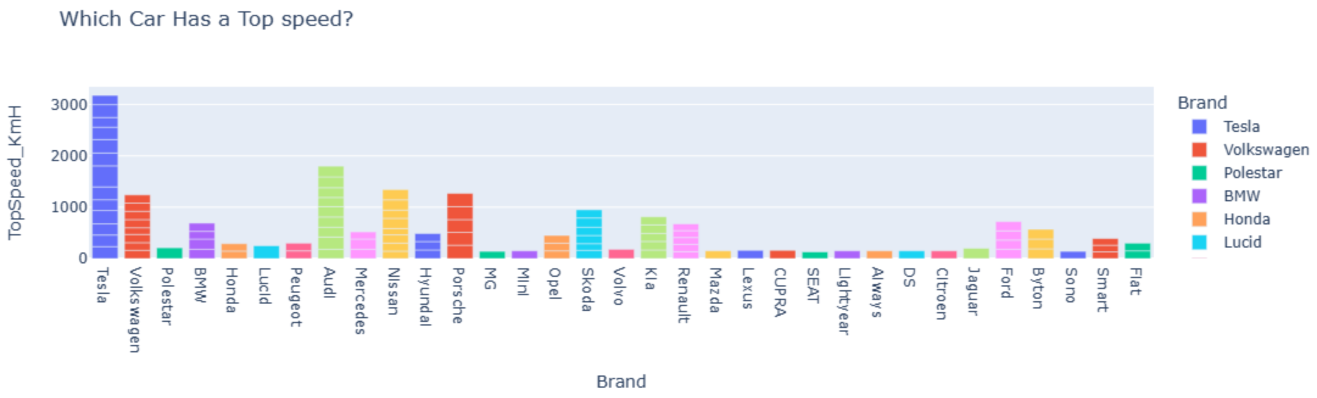


**Fig.9. which car has top speed?**

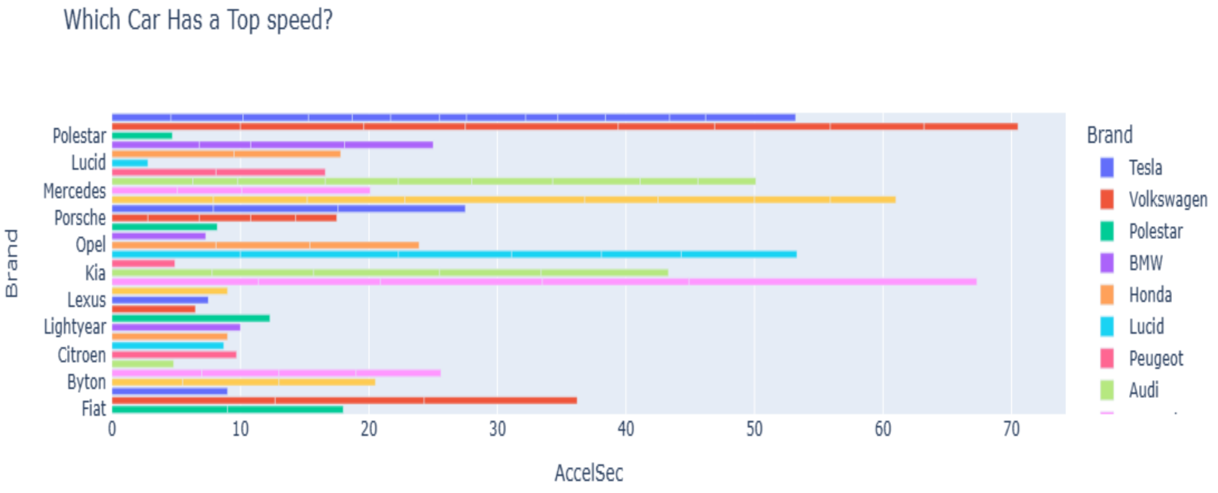
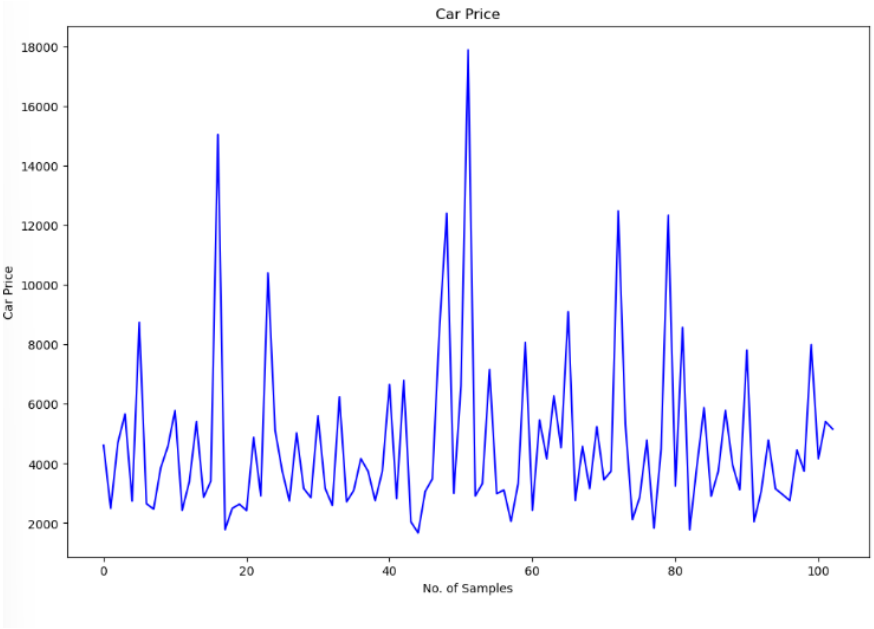
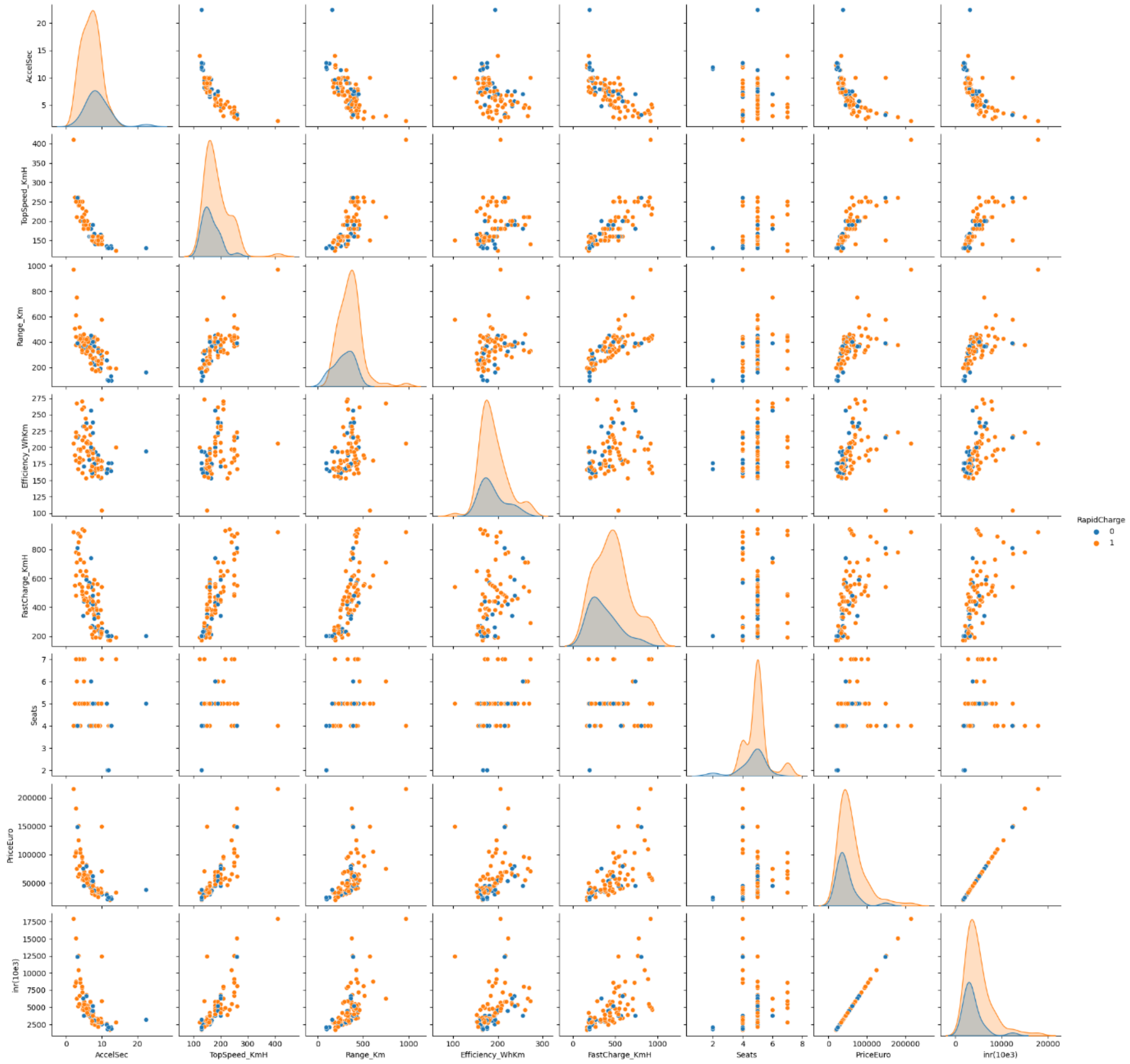


**Fig.10. Which car has a Top acceleration**



## For Electric Vehicle Market one of the most important key is Charging:



## Correlation Matrix:

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. Coefficients for different variables. The matrix depicts the correlation between all the possible pairs of values through the heatmap in the below figure. The relationship between two variables is usually considered strong when their correlation coefficient value is larger than 0.7

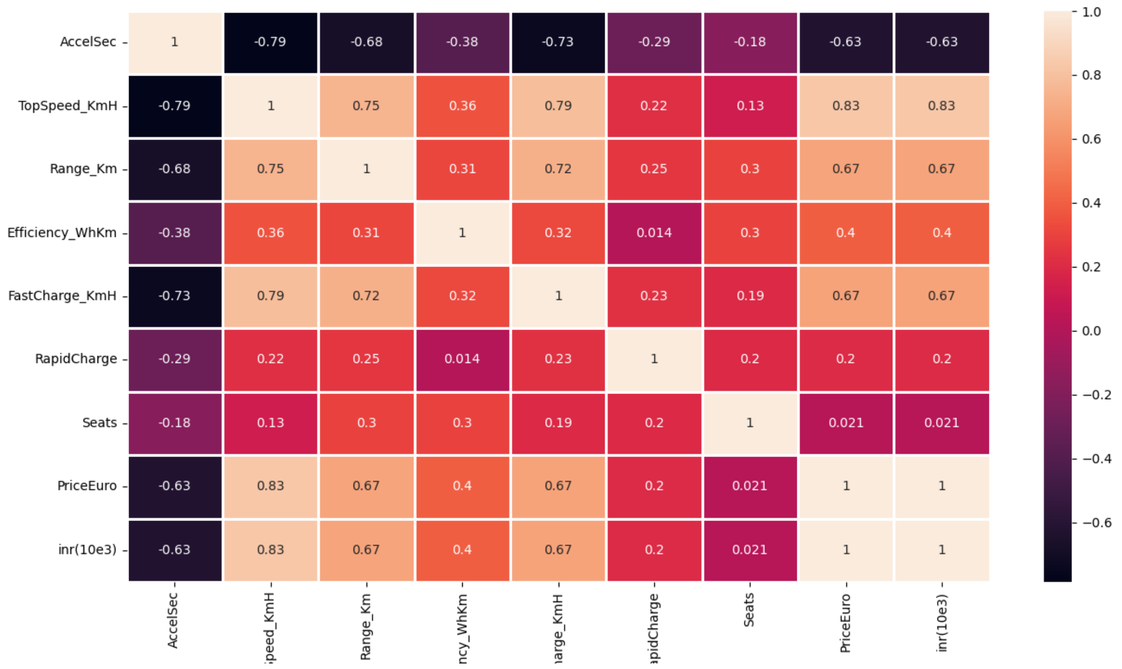
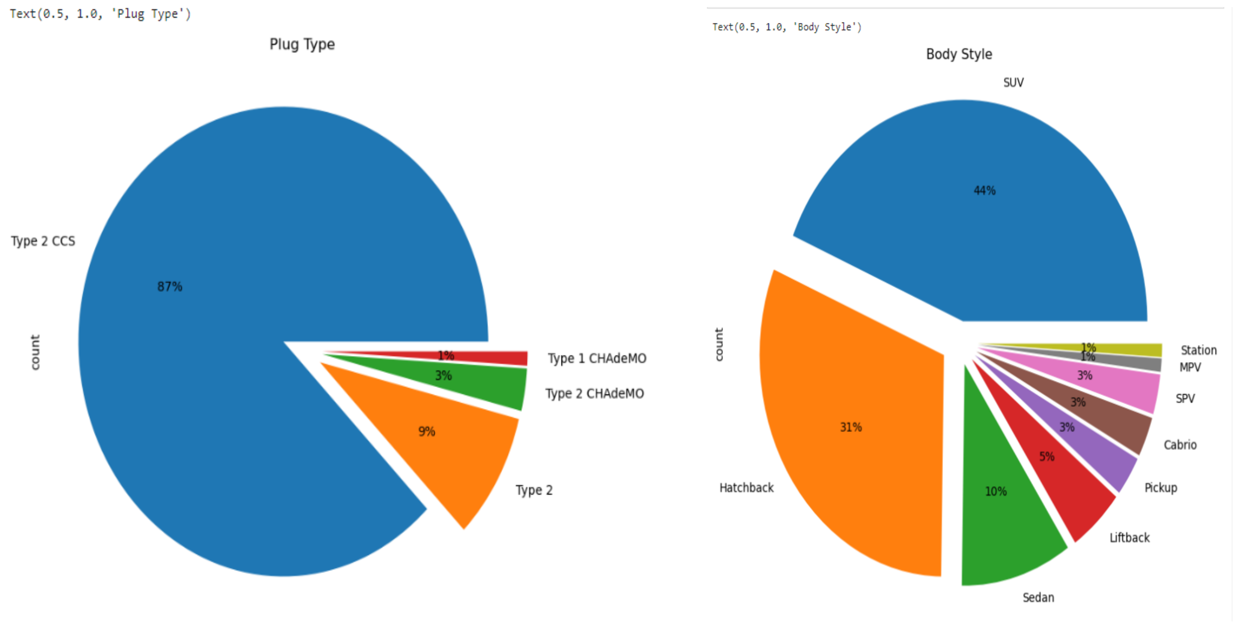


**Fig.11. Correlation Matrix**



```
Text(0.5, 1.0, 'Plug Type')
```

```
Text(0.5, 1.0, 'Body Style')
```

**Fig.12. Pie Char According Plug Type and Body Style.**

```
Text(0.5, 1.0, 'Segment')
```

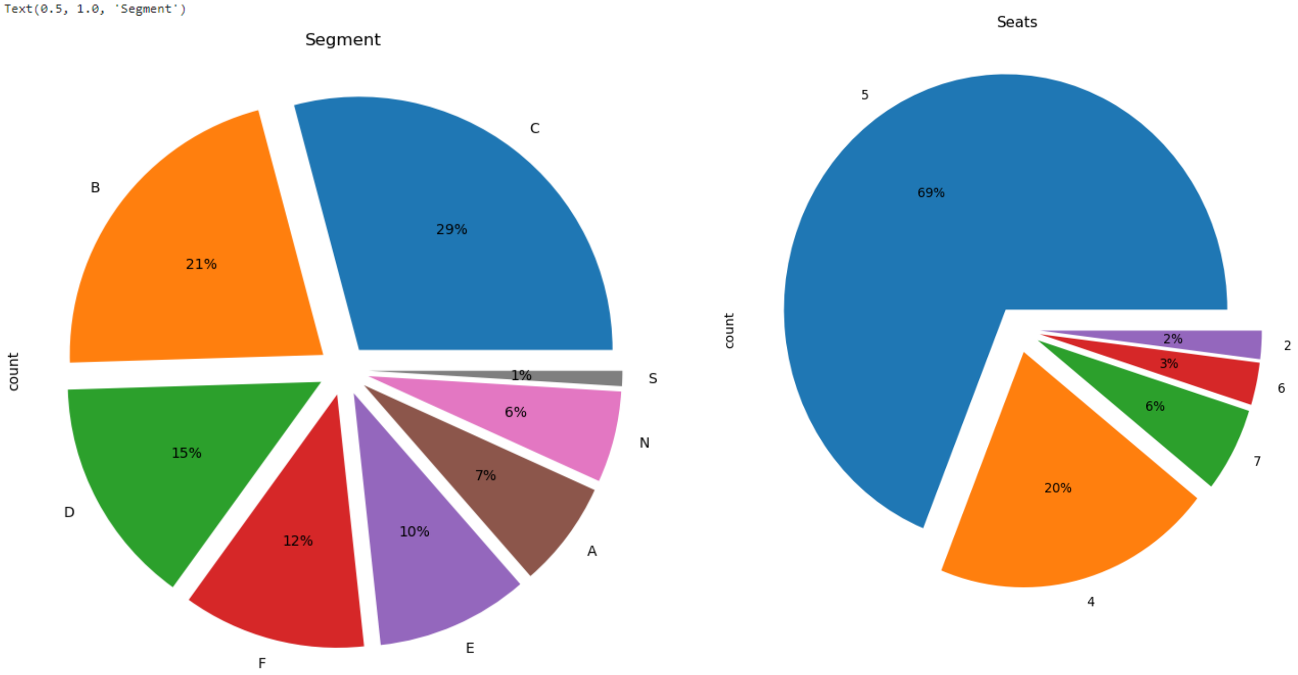


**Fig.13. Pie Chart Seat and Segment**

Now we can see that the requirements of what type of cars are most need for customer’s and from the past 10 year there is a rapid growth of Electric Vehicles usages in India

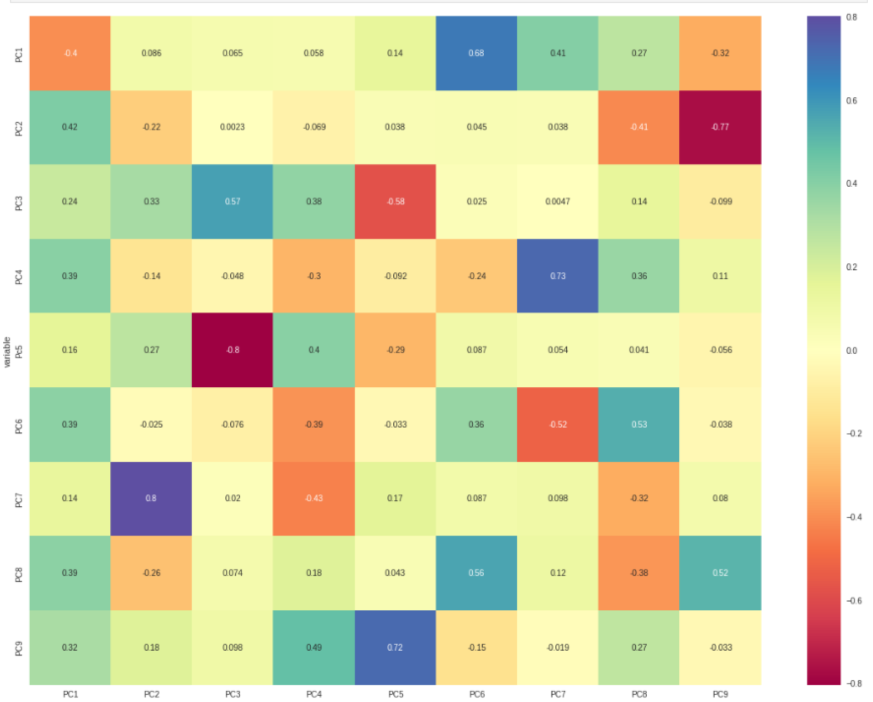


**Fig.14. Correlation matrix plot for loading**

**Scree Plot:** is a common method for determining the number of PCs to be retained via graphical representation. It is a simple line segment plot that shows the eigenvalues for each individual PC. It shows the eigenvalues on the y-axis and the number of factors on the x-axis. It always displays a downward curve. Most scree plots look broadly similar in shape, starting high on the left, falling rather quickly, and then flattening out at some point. This is because the first component usually explains much of the variability, the next few components explain a moderate amount, and the latter components only explain a small fraction of the overall variability. The scree plot criterion looks for the “elbow” in the curve and selects all components just before the line flattens out. The proportion of variance plot: The selected PCs should be able to describe at least 80% of the variance.

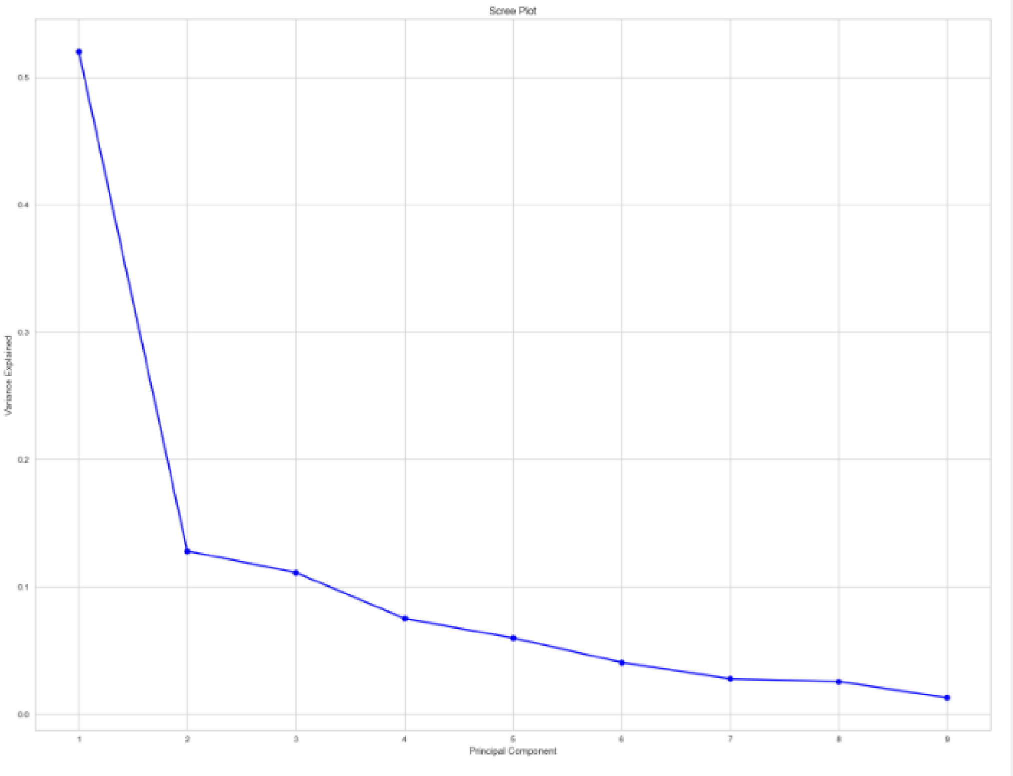


**Fig.15. : Screen Plot for our dataset**

## Extracting Segments

**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 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. As highlighted by other cluster validation metrics, four to five clusters can be considered for the agglomerative hierarchical as well.

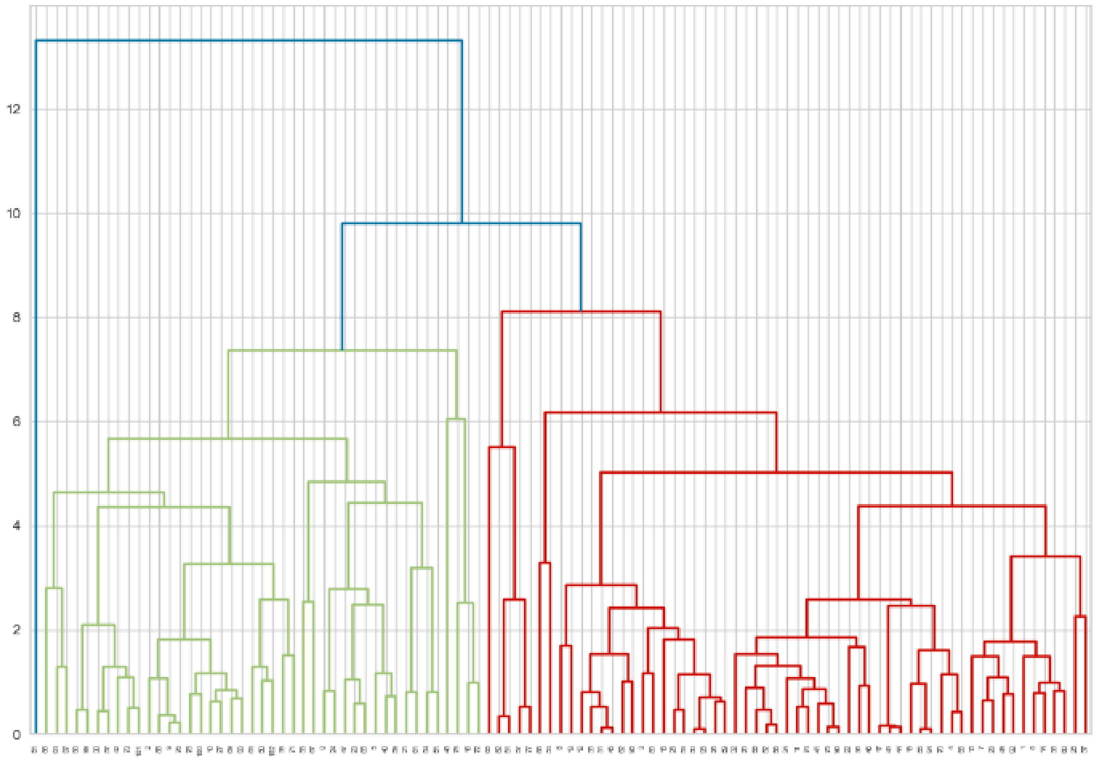
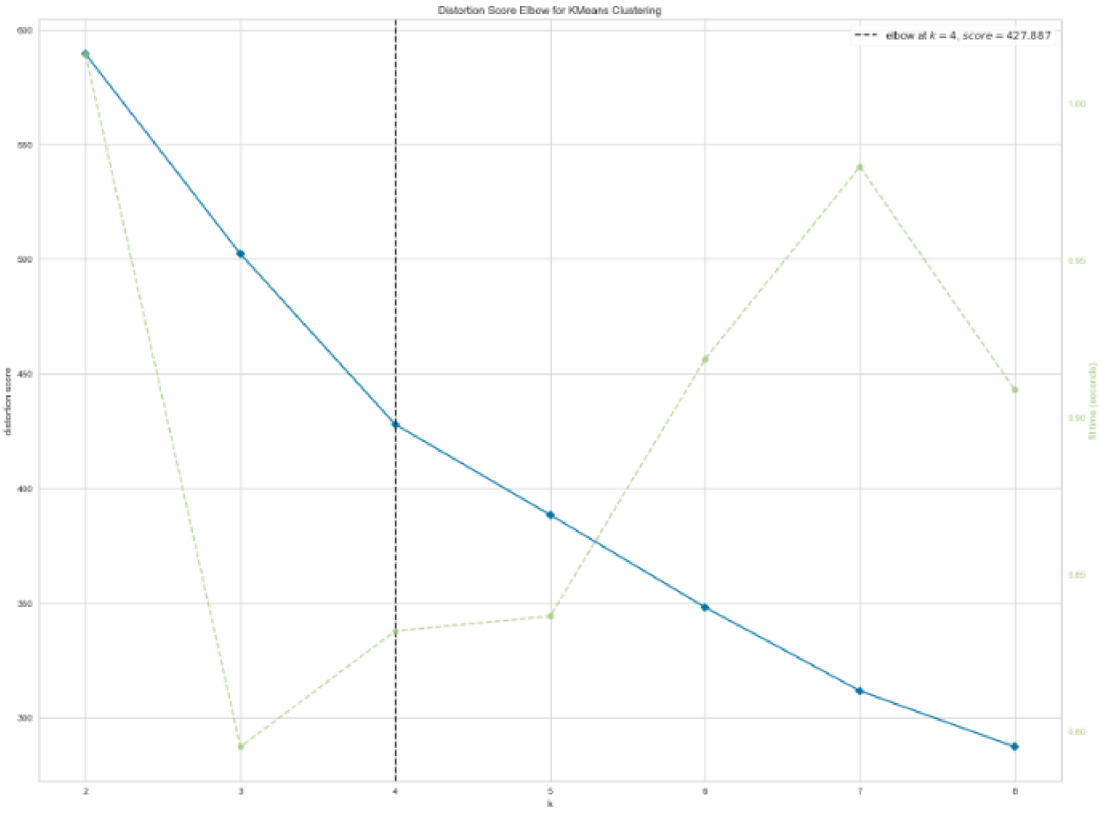


Fig.16. Dendrogram Plot for our Dataset



```
<Axes: title={'center': 'Distortion Score Elbow for KMeans Clustering'}, xlabel='k', ylabel='di
stortion score'>
```

Fig.17. Evaluating The Cluster using Distortion

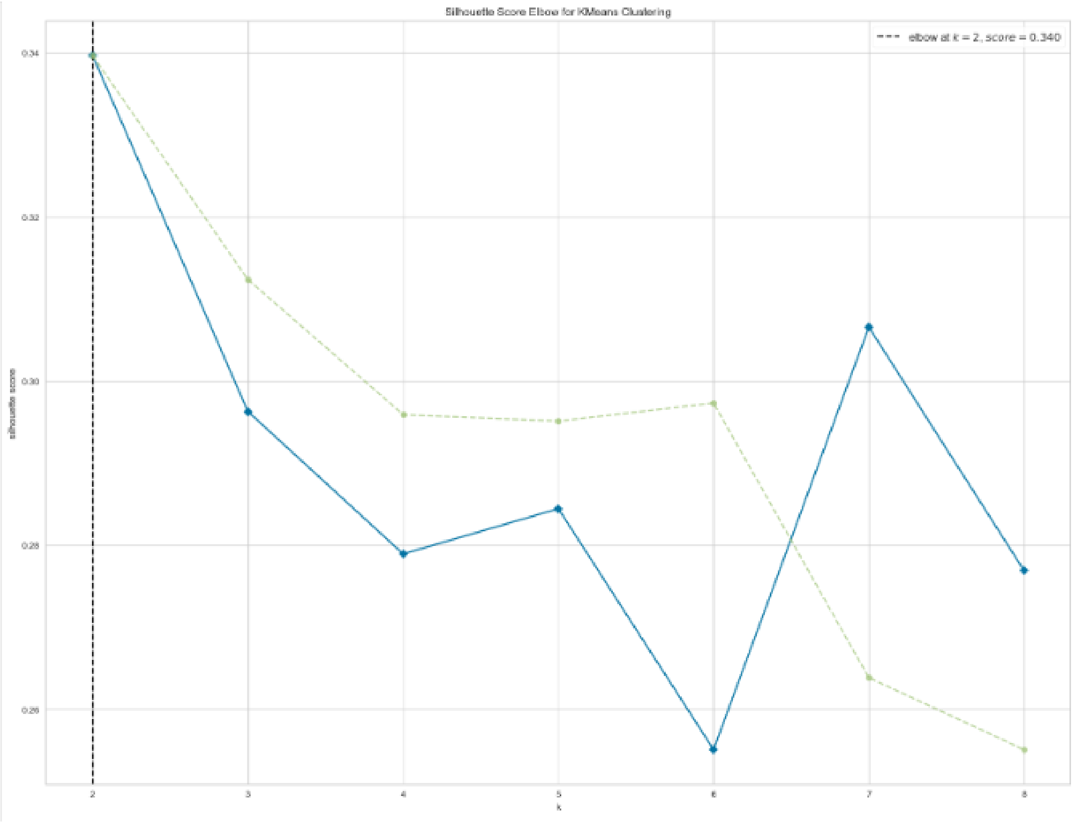


**Fig.18. Evaluating the clusters using silhouette**

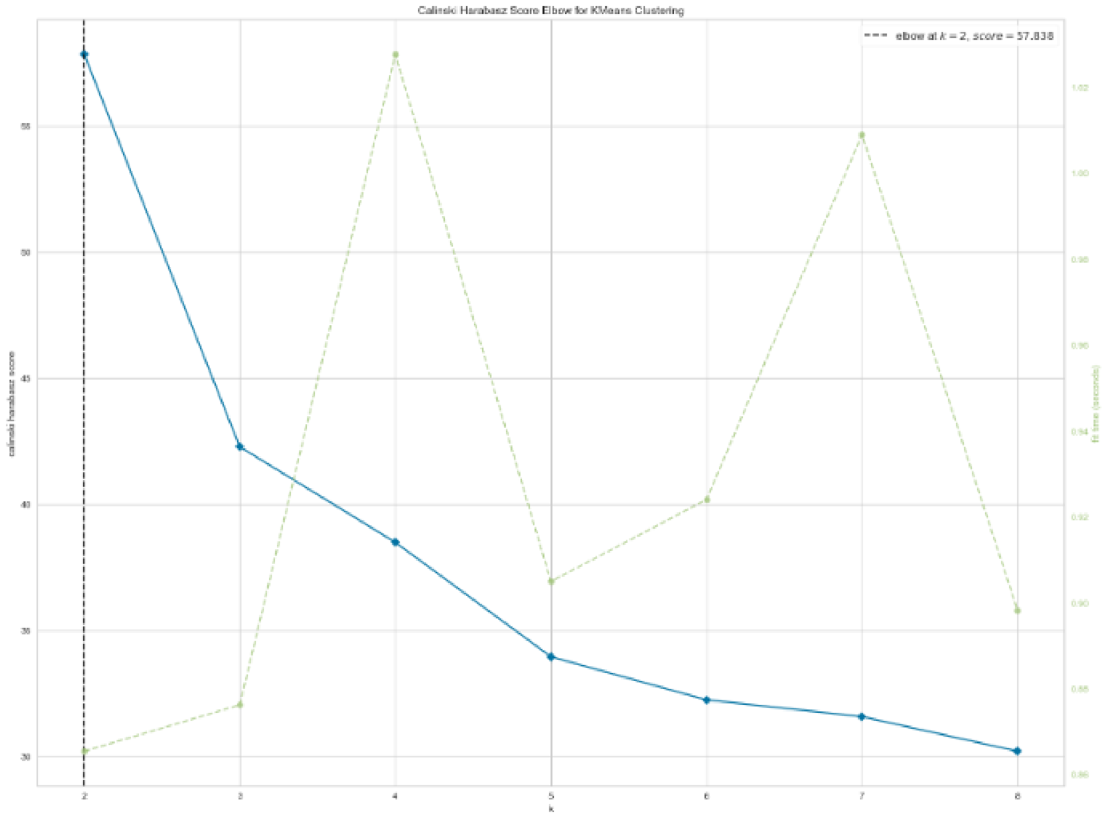


**Fig.19. Evaluating the clusters using Calin ski harabaz**

# ANALYSIS AND APPROACHES USED FOR SEGMENTATION

## Clustering

**Clustering** is one of the most common exploratory data analysis techniques used to get an intuition about the structure of the data. It can be defined as the task of identifying subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different. In other words, we try to find homogeneous subgroups within the data such that data points in each cluster are as similar as possible according to a similarity measure such as Euclidean based distance or correlation-based distance. The decision of which similarity measure to use is application-specific. Clustering analysis can be done on the basis of features where we try to find subgroups of samples based on features or on the basis of samples where we try to find subgroups of features based on samples.

### K-Means Algorithm

**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:

- Specify number of clusters K.
- Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
- 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. Below is a breakdown of how we can solve it mathematically

The Objective function is

$$\boxed{J=\sum^{m}_{i=0}\sum^{k}_{i=1} W_{ik} \| x^{i} - \mu_{k} \|}$$

And M-Step is

$$\frac{\partial J}{\partial \mu_k} = 2\sum_{i=1}^{m} w_{ik}(x^{i} - \mu_k) = 0$$

$$= > \mu_k = \sum^{m}_{i=1} W_{ik}\, x^i / \sum^{m}_{i=1} W_{ik}$$

**Applications**

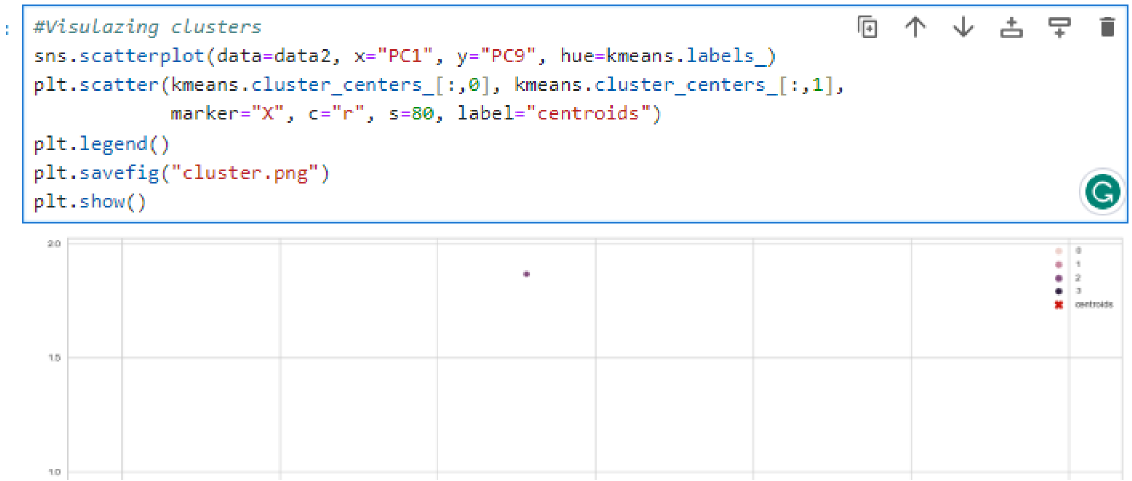
K means algorithm is very popular and used in a variety of applications such as market segmentation, document clustering, image segmentation and image compression, etc. The goal usually when we undergo a cluster analysis is either:

1. Get a meaningful intuition of the structure of the data we're dealing with.

2. Cluster-then-predict where different models will be built for different subgroups if we believe there is a wide variation in the Behavioral of different subgroups.

The k-means clustering algorithm performs the following tasks:

- 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. According to the Elbow method, here we take K=4 clusters to train K-Means model. The derived clusters are shown in the following figure



```
#Visulazing clusters
sns.scatterplot(data=data2, x="PC1", y="PC9", hue=kmeans.labels_)
plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1],
            marker="X", c="r", s=80, label="centroids")
plt.legend()
plt.savefig("cluster.png")
plt.show()
```

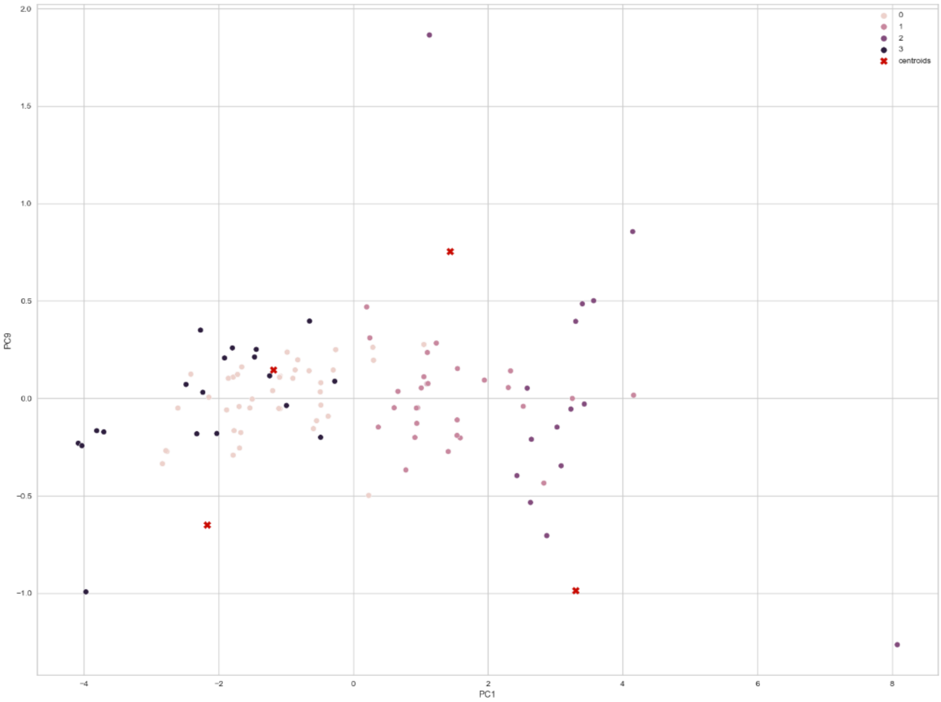


**Fig.20.Visualize Cluster using K-mean**

## Prediction of Prices most used cars

Linear regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models targets prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Here we use a linear regression model to predict the prices of different Electric cars in different companies. X contains the independent variables and y is the dependent Prices that is to be predicted. We train our model with a splitting of data into a 4:6 ratio, i.e. 40% of the data is used to train the model.

**LinearRegression().fit(Xtrain,ytrain)** command is used to fit the data set into model. The values of intercept, coefficient, and cumulative distribution function (CDF) are described in the figure.

## Regression for data2

02]:

```
X=data2[['PC1', 'PC2','PC3','PC4','Pc5','PC6', 'PC7','PC8','PC9']]
y=df['inr(10e3)']
```

03]:

```
X_train, X_test, y_train, y_test = train_test_split(X, y,test_size=0.4, random_state=101)
lm=LinearRegression().fit(X_train,y_train)
```

04]:

```
print(lm.intercept_)
```

```
4643.522050485438
```

05]:

```
lm.coef_
```

05]:

```
array([ 1101.5872075 ,  -741.20904198,   208.53617452,   508.32245827,
         122.35330123,  1579.00685826,   333.61147115, -1079.99511501,
        1461.7226913 ])
```

06]:

```
X_train.columns
```

06]:

```
Index(['PC1', 'PC2', 'PC3', 'PC4', 'Pc5', 'PC6', 'PC7', 'PC8', 'PC9'], dtype='object')
```

07]:

```
cdf=pd.DataFrame(lm.coef_, X.columns, columns=['Coeff'])
cdf
```

07]:

|  | Coeff |
|---|---|
| **PC1** | 1101.587208 |
| **PC2** | -741.209042 |
| **PC3** | 208.536175 |
| **PC4** | 508.322458 |
| **Pc5** | 122.353301 |
| **PC6** | 1579.006858 |
| **PC7** | 333.611471 |
| **PC8** | -1079.995115 |
| **PC9** | 1461.722691 |

After completion of training the model process, we test the remaining 60% of data on the model. The obtained results are checked using a scatter plot between predicted values and the original test data set for the dependent variable and acquired similar to a straight line as shown in the figure and the density function is also normally distributed.

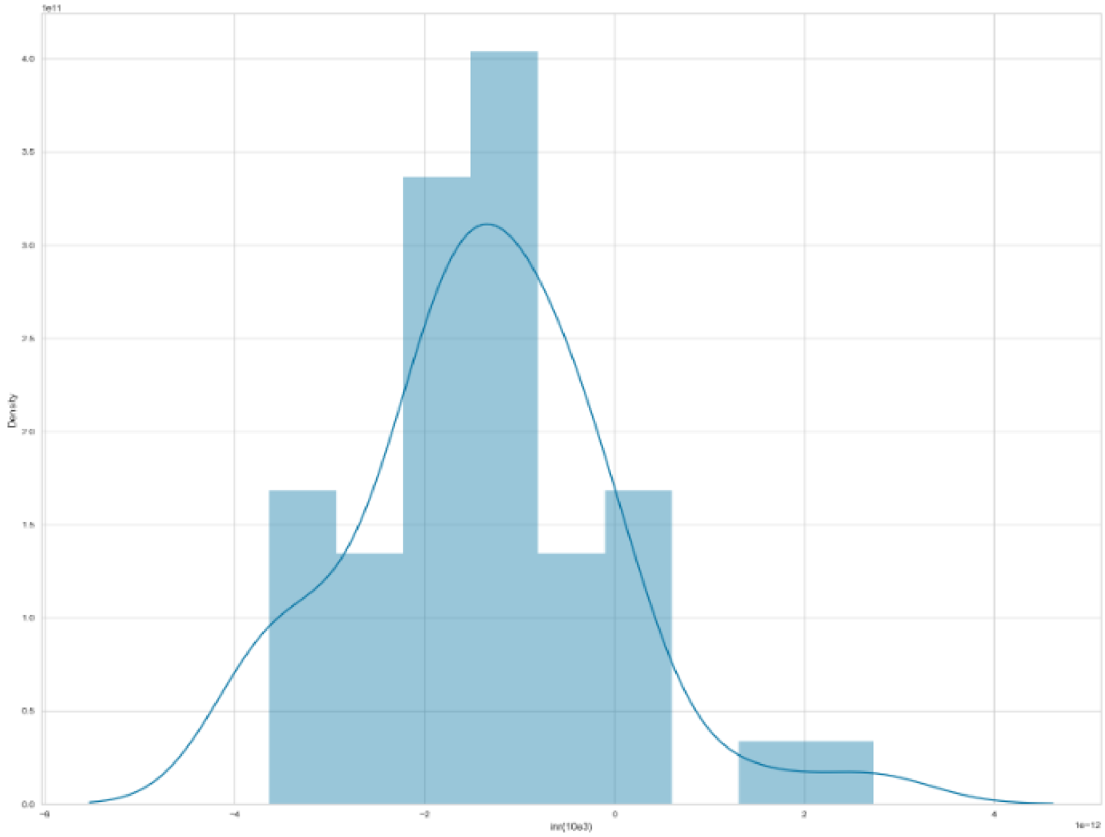
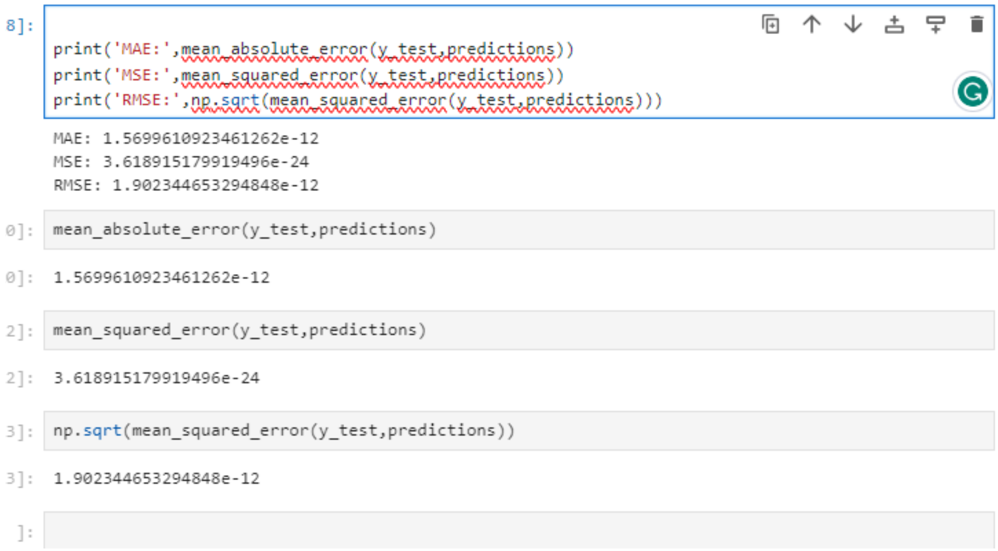


**Fig.21. Density Function Is Also Normally Distributed**

The metrics of the algorithm, Mean absolute error, Mean Squared error and mean square root error are described in the below figure:



```
8]:  print('MAE:',mean_absolute_error(y_test,predictions))
     print('MSE:',mean_squared_error(y_test,predictions))
     print('RMSE:',np.sqrt(mean_squared_error(y_test,predictions)))

     MAE:  1.5699610923461262e-12
     MSE:  3.618915179919496e-24
     RMSE:  1.902344653294848e-12

0]:  mean_absolute_error(y_test,predictions)

0]:  1.5699610923461262e-12

2]:  mean_squared_error(y_test,predictions)

2]:  3.618915179919496e-24

3]:  np.sqrt(mean_squared_error(y_test,predictions))

3]:  1.902344653294848e-12

 ]:
```

# PROFILING AND DESCRIBING THE SEGMENTS

Sorting the Top Speeds and Maximum Range in accordance to the Price with head () we can view the Pie Chart.

## Pie Chart:

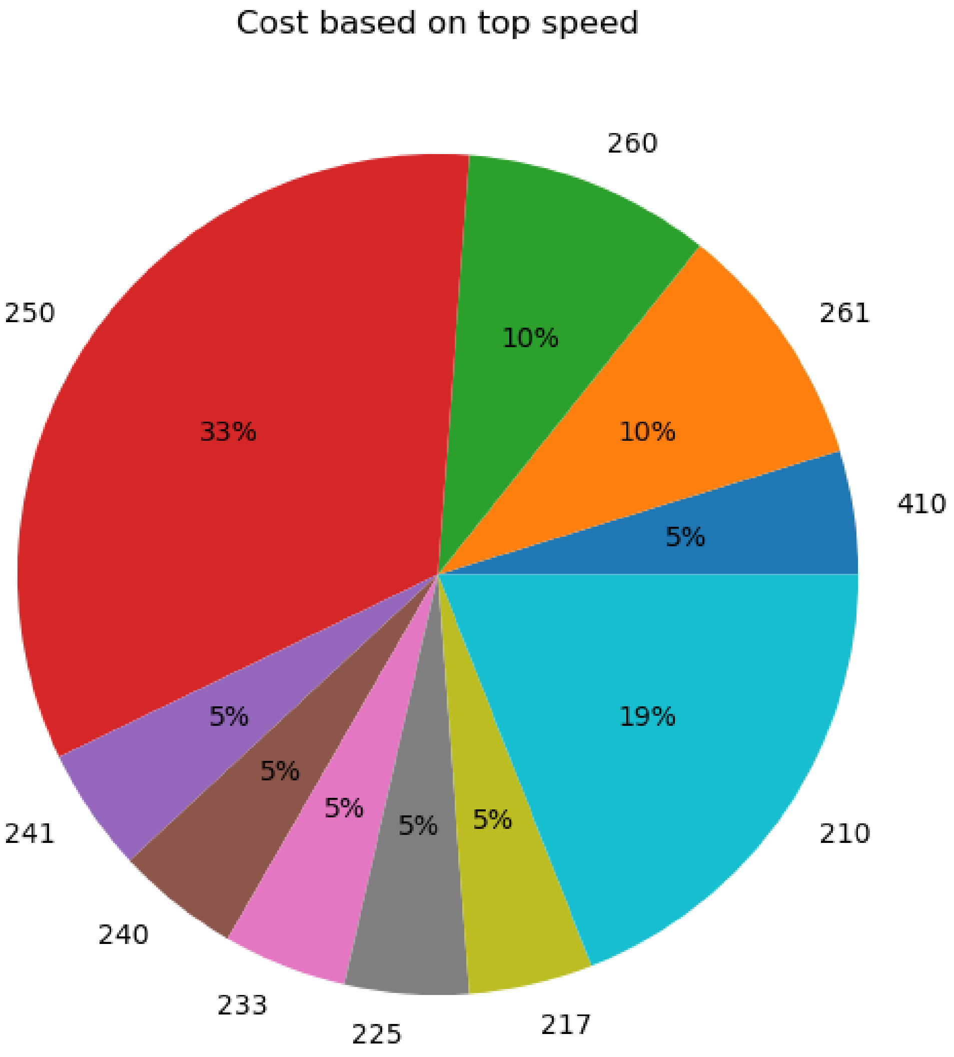


**Fig.22. Cost Based on Top Speed**

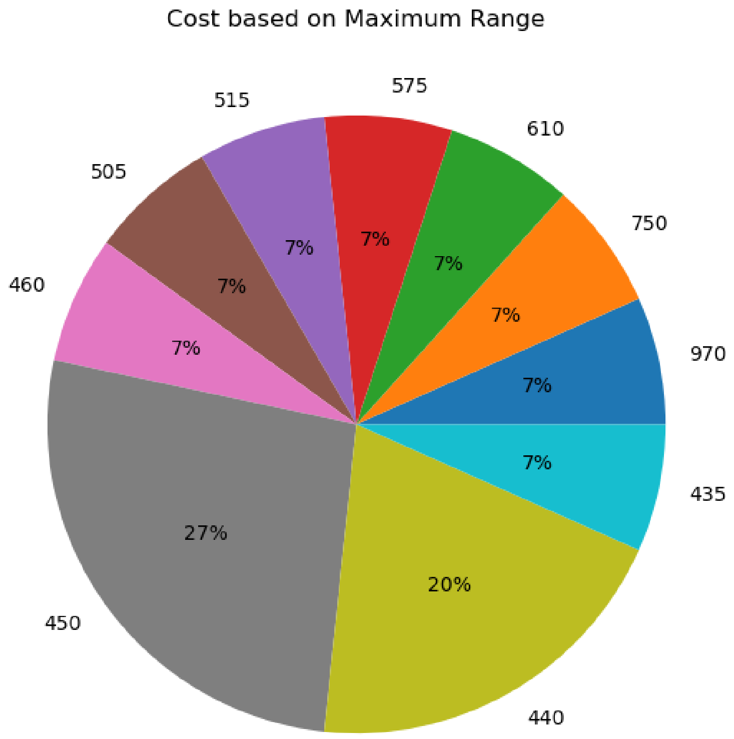


**Fig.23. Cost Based on Maximum Range**

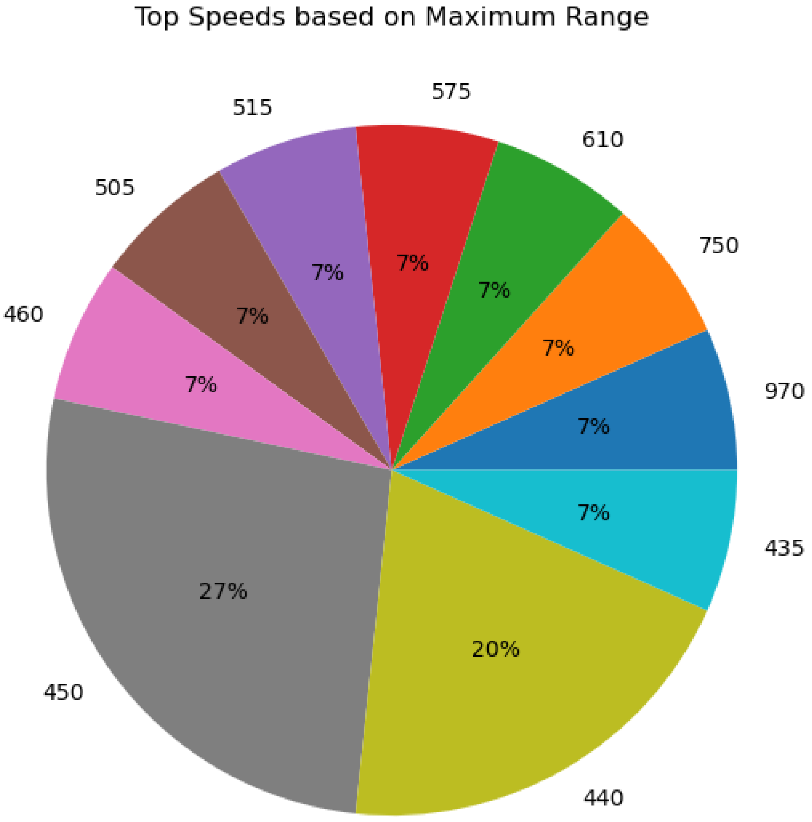


**Fig.24. Top Speed Based on Maximum Range**

# TARGET SEGMENTS:

So, from the analysis we can see that the optimum targeted segment should be belonging to the following categories:

## Behavioral:

Mostly from our analysis there are cars with 5 seats.

## Demographic:

• Top Speed & Range : With a large area of market the cost is dependent on Top speeds and Maximum range of cars.

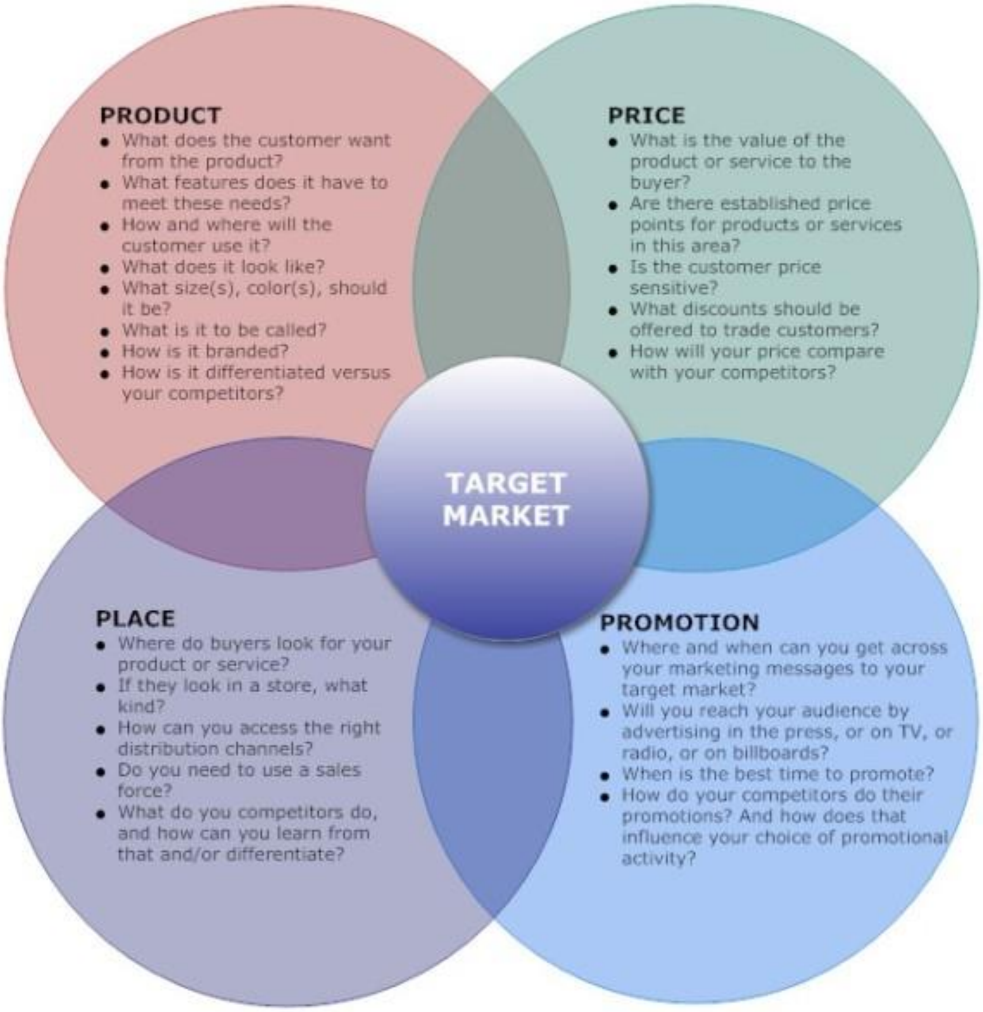
• Efficiency : Mostly the segments are with most efficiency.

## Psychographic:

• Price : From the above analysis, the price range is between 16,00,000 to 1,80,00,000.

Finally, our target segment should contain cars with most Efficiency, contains Top Speed and price **between 16 to 180 lakhs** with mostly with **5 seats**.

### Customizing the Marketing Mix



**PRODUCT**

- What does the customer want from the product?
- What features does it have to meet these needs?
- How and where will the customer use it?
- What does it look like?
- What size(s), color(s), should it be?
- What is it to be called?
- How is it branded?
- How is it differentiated versus your competitors?

**PRICE**

- What is the value of the product or service to the buyer?
- Are there established price points for products or services in this area?
- Is the customer price sensitive?
- What discounts should be offered to trade customers?
- How will your price compare with your competitors?

**TARGET MARKET**

**PLACE**

- Where do buyers look for your product or service?
- If they look in a store, what kind?
- How can you access the right distribution channels?
- Do you need to use a sales force?
- What do you competitors do, and how can you learn from that and/or differentiate?

**PROMOTION**

- Where and when can you get across your marketing messages to your target market?
- Will you reach your audience by advertising in the press, or on TV, or radio, or on billboards?
- When is the best time to promote?
- How do your competitors do their promotions? And how does that influence your choice of promotional activity?

# Reference

- https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset
- https://c570blog.wordpress.com/2014/05/31/the-four-ps-of-marketing-mix/
- https://en.wikipedia.org/wiki/K-means_clustering
- View: Time for India to move into top gear with an eye on 2030 EV public infra goal - The Economic Times (indiatimes.com)