Market Segmentation

Analysing the respective market in India using Segmentation analysis for an

Electrical vehicle Startup by

Ankit Chaudhary

Overview

The Electric Vehicle (EV) market in India is experiencing rapid growth as the country transitions toward sustainable transportation, driven by government initiatives such as the Faster Adoption and Manufacturing of Hybrid and Electric Vehicles (FAME) scheme, which provides subsidies and encourages infrastructure development, alongside state-specific policies like tax exemptions. The market is segmented by vehicle type, including passenger vehicles, commercial vehicles, two-wheelers, and three-wheelers, as well as charging infrastructure (public and private chargers) and battery types such as lithium-ion.

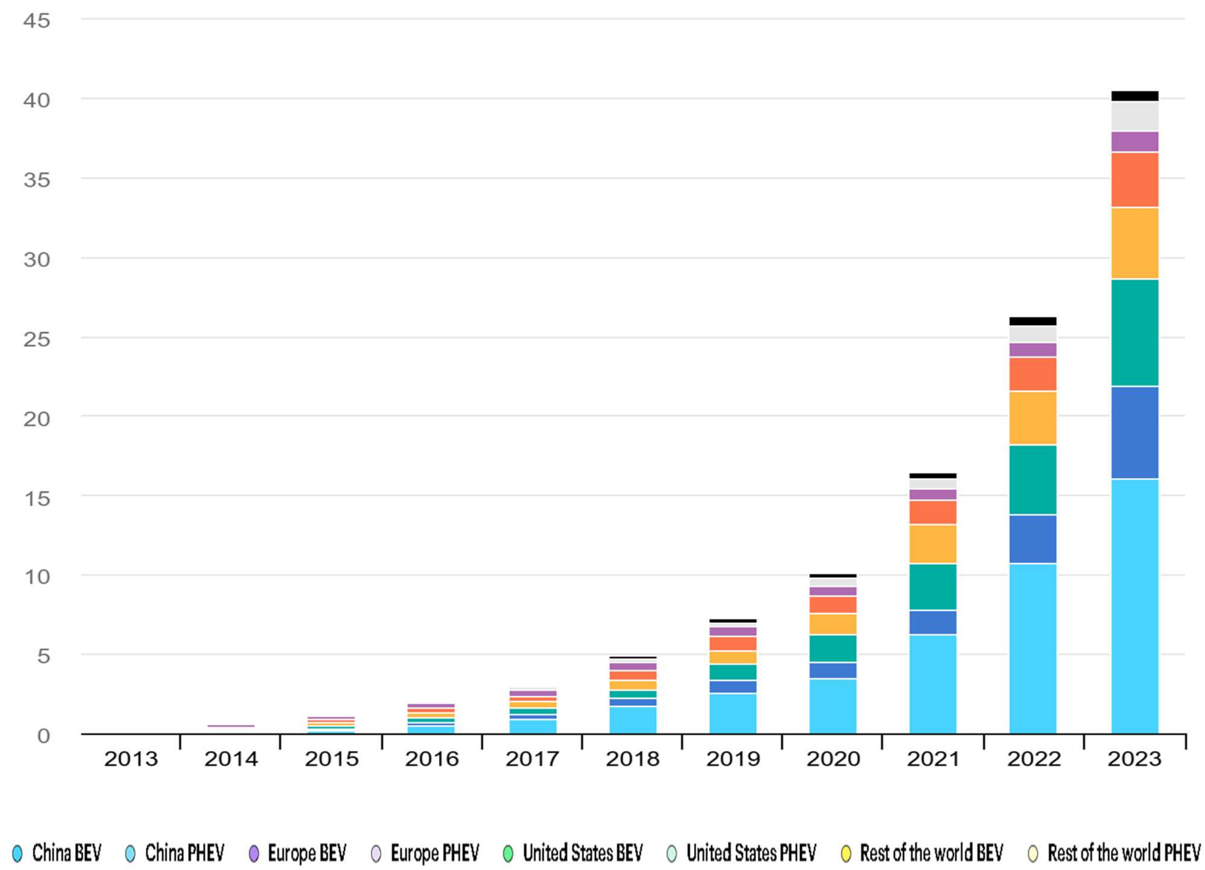
Technological advancements like improved battery ranges and smart connectivity, combined with rising fuel prices and environmental concerns, are pushing consumers toward electric mobility. Private investments by automakers and startup further bolster the ecosystem, focusing on vehicle manufacturing, battery production, and charging network development.

However, challenges persist, including limited charging infrastructure, a lack of standardized charging protocols, dependence on imported raw materials like lithium, and the need for affordable battery solutions. Despite these challenges, significant opportunities exist in urban centres for two- and three-wheelers, last-mile connectivity, fleet electrification, and integration with renewable energy for sustainable charging solutions. Moreover, expansion into Tier-2 and Tier-3 cities supported by government policies offers additional growth avenues. With these factors, India is well-positioned to emerge as one of the largest EV markets globally by 2030, aligning with its sustainability goals and growing demand for eco-friendly transportation options.

Market Overview

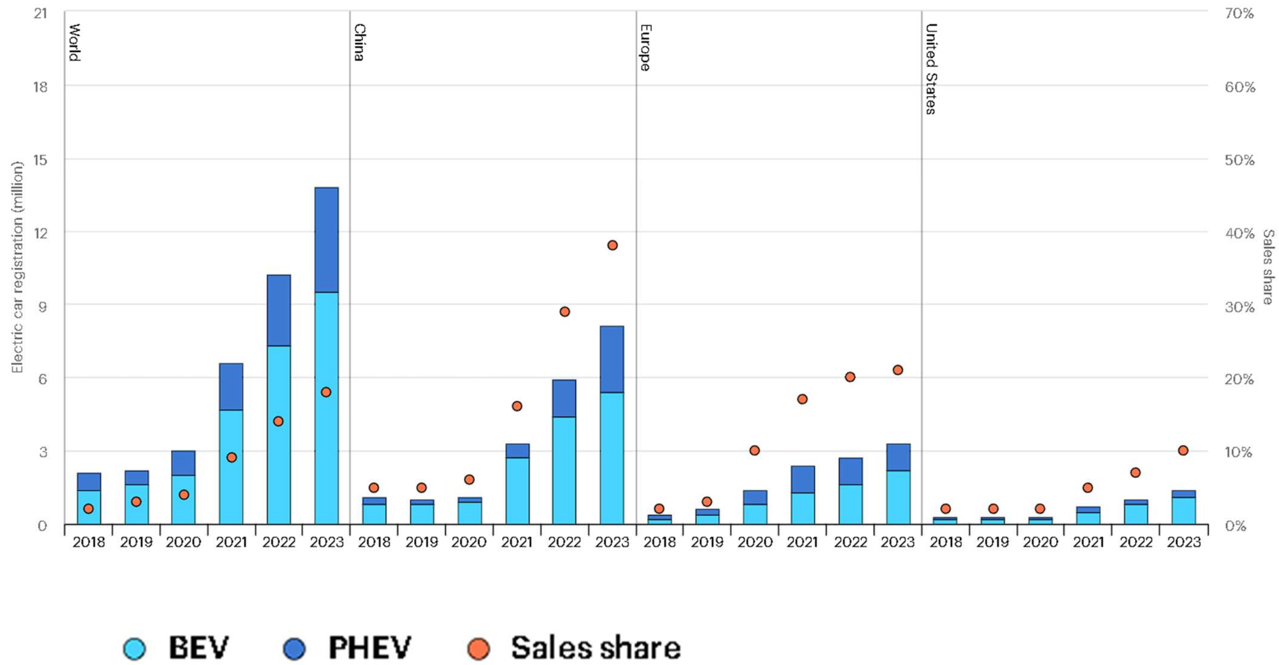
Electric car sales neared 14 million in 2023, 95% of which were in China, Europe and the United States

Almost 14 million new electric cars were registered globally in 2023, bringing their total number on the roads to 40 million, closely tracking the sales forecast from the 2023 edition of the Global EV Outlook (GEVO-2023). Electric car sales in 2023 were 3.5 million higher than in 2022, a 35% year-on-year increase. This is more than six times higher than in 2018, just 5 years earlier. In 2023, there were over 250 000 new registrations per week, which is more than the annual total in 2013, ten years earlier. Electric cars accounted for around 18% of all cars sold in 2023, up from 14% in 2022 and only 2% 5 years earlier, in 2018. These trends indicate that growth remains robust as electric car markets mature. Battery electric cars accounted for 70% of the electric car stock in 2023.



While sales of electric cars are increasing globally, they remain significantly concentrated in just a few major markets. In 2023, just under 60% of new electric car registrations were in the People's Republic of China (hereafter 'China'), just under 25% in Europe,[2] and 10% in the United States – corresponding to nearly 95% of global electric car sales combined.

In these countries, electric cars account for a large share of local car markets: more than one in three new car registrations in China was electric in 2023, over one in five in Europe, and one in ten in the United States. However, sales remain limited elsewhere, even in countries with developed car markets such as Japan and India. As a result of sales concentration, the global electric car stock is also increasingly concentrated. Nevertheless, China, Europe and the United States also represent around two-thirds of total car sales and stocks, meaning that the EV transition in these markets has major repercussions in terms of global trends

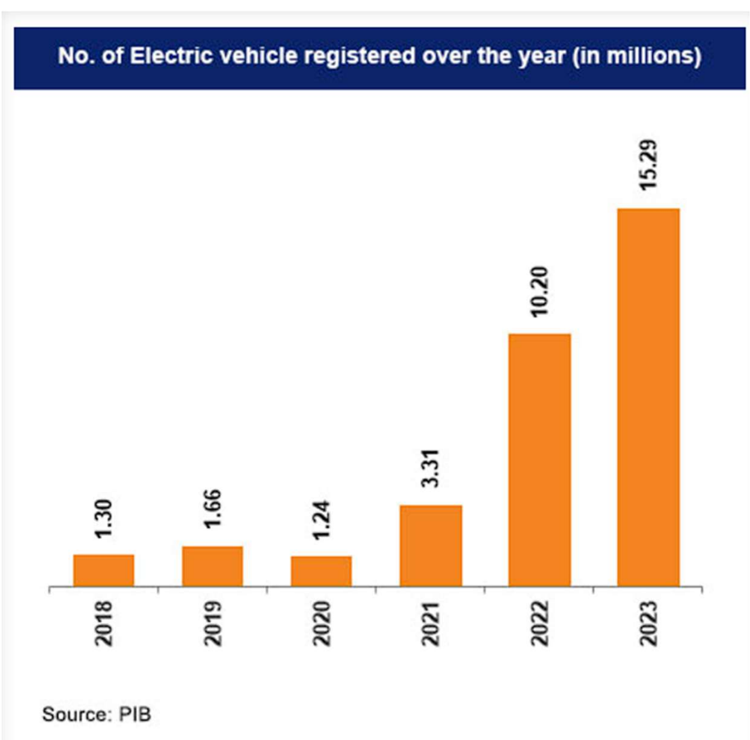


INDIAN Market Overview

According to Fortune Business Insights, the Indian EV market is forecasted to expand from US$ 3.21 billion in 2022 to US$ 113.99 billion by 2029, with a 66.52% CAGR.

The Indian EV battery market is projected to surge from US$ 16.77 billion in 2023 to a remarkable US$ 27.70 billion by 2028.

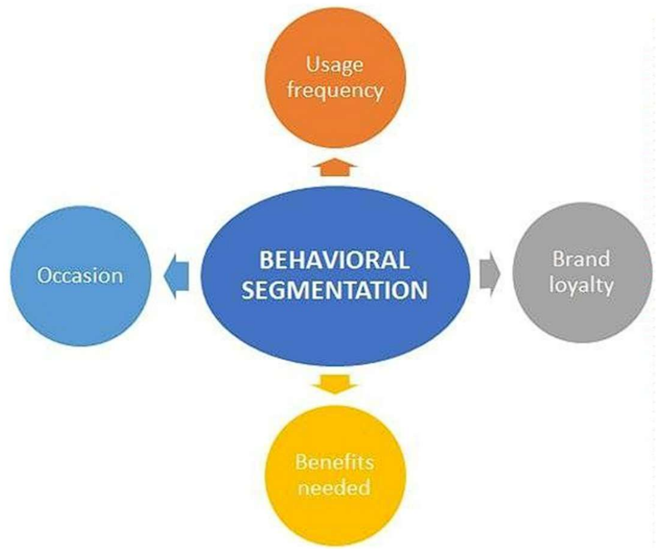
According to NITI Aayog and Rocky Mountain Institute (RMI), India's EV finance industry is likely to reach Rs. 3.7 lakh crore (US$ 50 billion) in 2030



Target Market

The target market of Electric Vehicle Market Segmentation can be categorized into Geographic, Socio Demographic, Behavioural, and Psychographic Segmentation.

Behavioural Segmentation : searches directly for similarities in behaviour or reported behaviour.



Example: prior experience with the product, amount spent on the purchase, etc.

Advantage: uses the very behavior 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 behaviour.

Disadvantage : increased complexity of determining segment memberships for consumers.

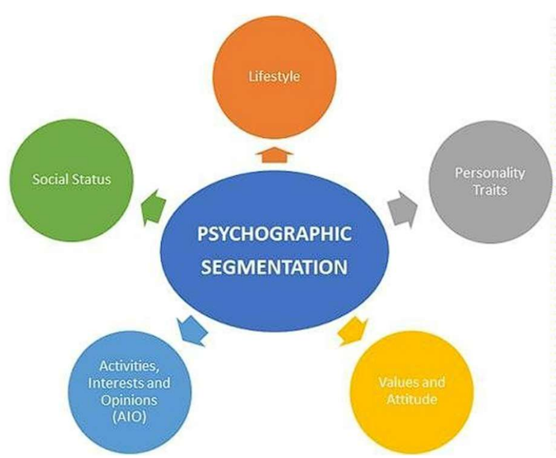
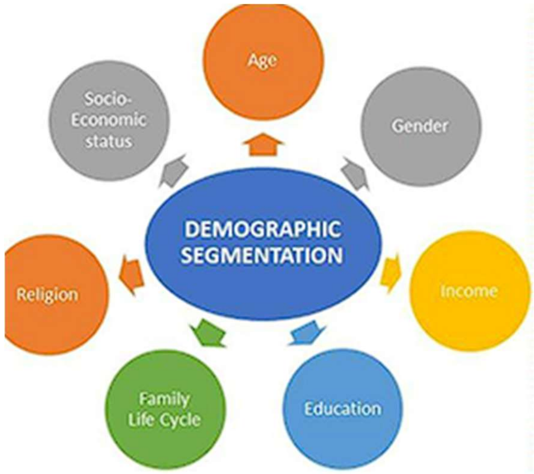


Figure 2: *Psychographic Segmentation*

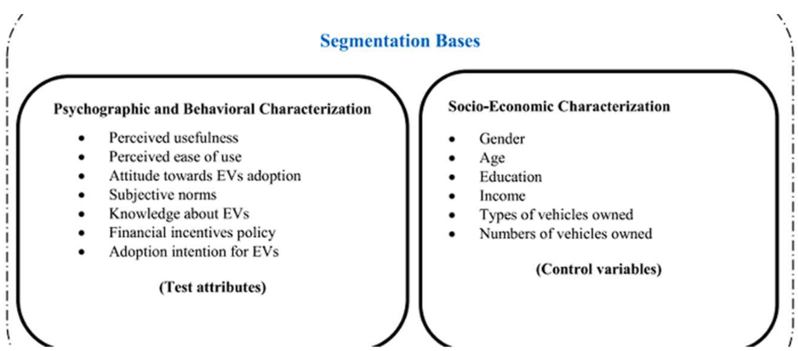
Socio-Demographic Segmentation: includes age, gender, income and education. Useful in industries.



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

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

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 (behavioural, 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.



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. Straight and Roberts presented a comparison between the usefulness of psychographic, demographic, and economic characteristics based on consumer evaluation for eco-friendly products.

They pinpointed the perceived superiority of the psychographic characteristics over the socio-demographic and economic ones in explaining the environmentally-conscious con sumer behavior and thus, the study recommended the use of psychographic character istics 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 calculations related to arrays.

2. Pandas: To read or load the datasets.

3. SKLearn: We have used LabelEncoder() to encode our values.

Data-Preprocessing

Data Cleaning

The data collected is compact and is partly used for visualization purposes and partly for clustering. Python libraries such as NumPy, Pandas, Scikit-Learn, and SciPy are used for the workflow, and the results obtained are ensured to be reproducible.

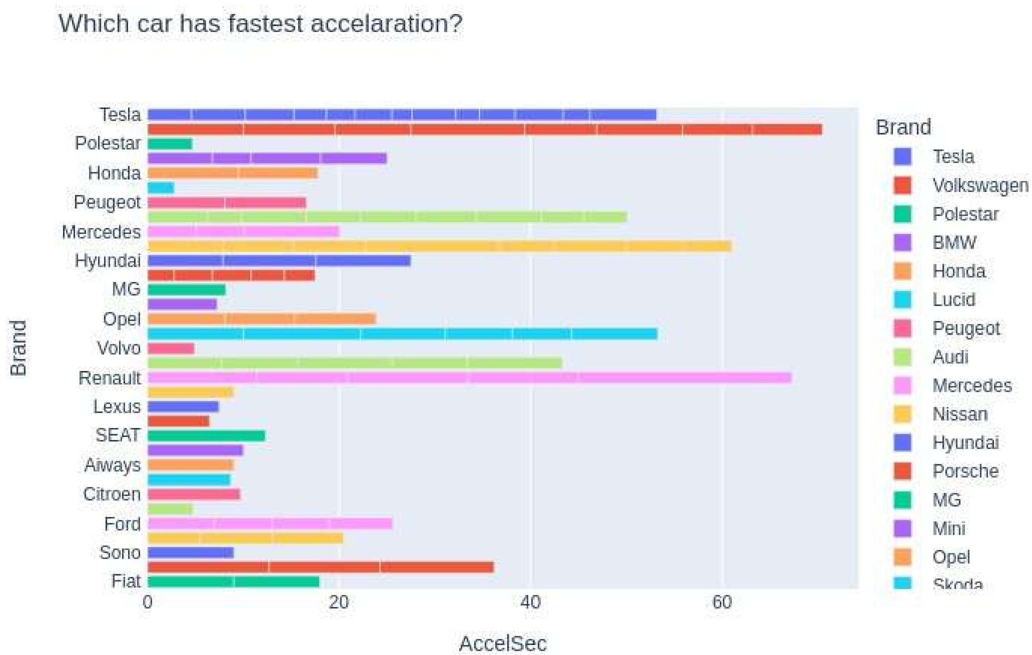
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
import statsmodels.api as sm
import plotly.express as px
```

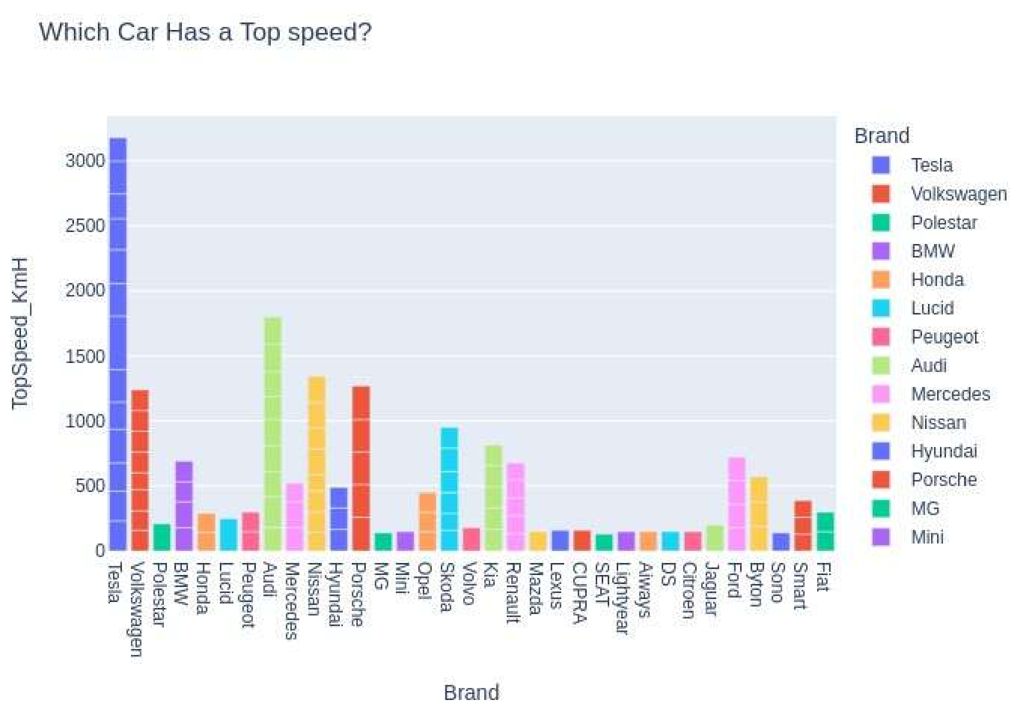
```
df = pd.read_csv('ElectricCarData_Clean.csv')
df.head()
```

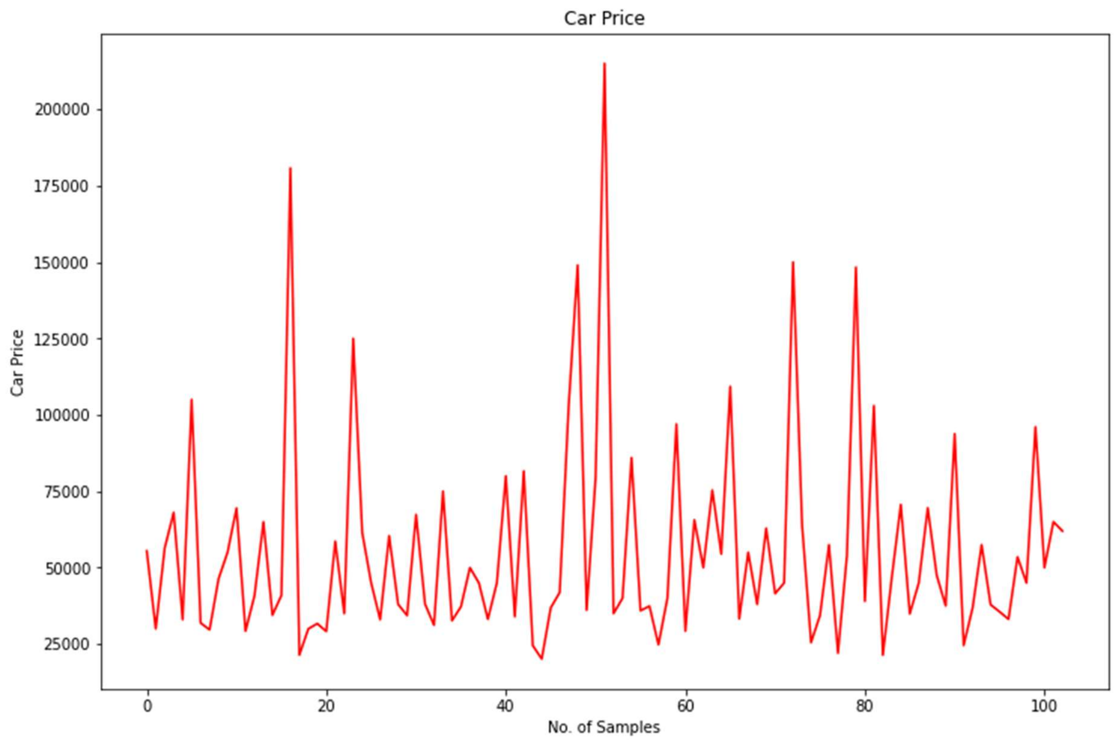
	Brand	Model	AccelSec	TopSpeed_KmH	Range_Km	Efficiency_WhKm	FastCharge_KmH	RapidCharge	PowerTrain	PlugType	BodyStyle	Segment	Seats	PriceEuro
0	Tesla	Model 3 Long Range Dual Motor	4.6	233	450	161	940	Yes	AWD	Type 2 CCS	Sedan	D	5	55480
1	Volkswagen	ID.3 Pure	10.0	160	270	167	250	Yes	RWD	Type 2 CCS	Hatchback	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

EDA (Exploratory Data Analysis)

We start the Exploratory Data Analysis with some data Analysis drawn from the data without Principal Component Analysis and with some Principal Component Analysis in the dataset obtained from the combination of all the data we have. PCA is a statis tical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. These new trans formed 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. Comparision of cars in our data .







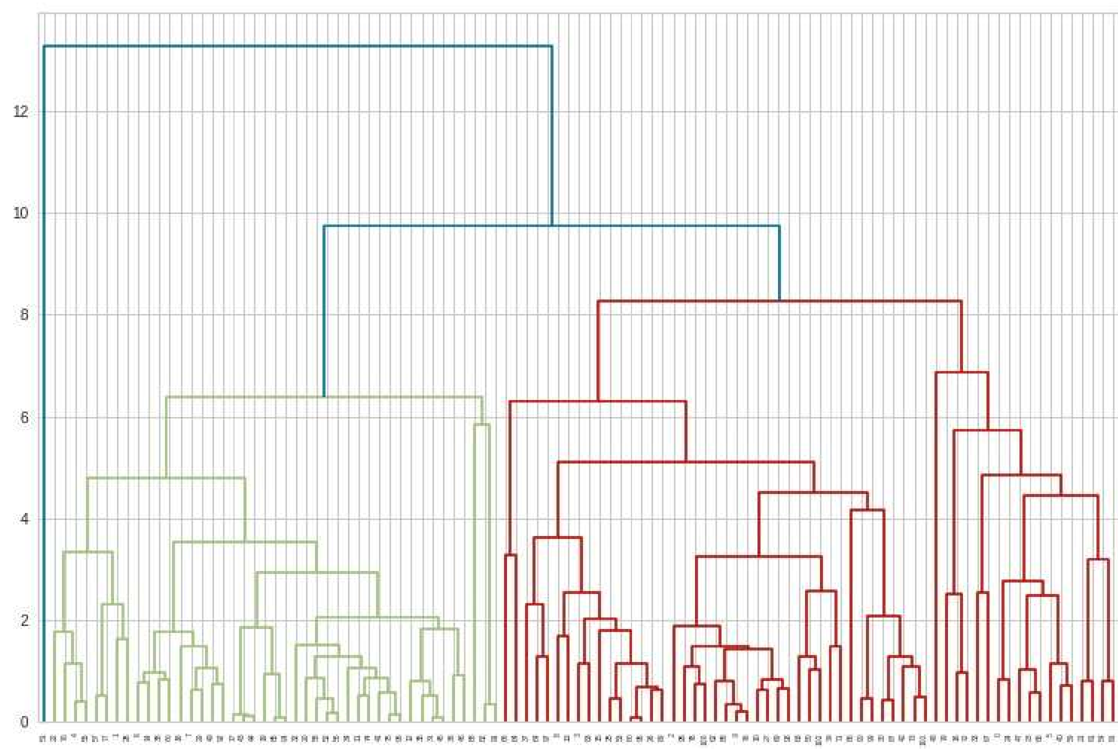
Correlation Matrix

A correlation matrix is simply a table that displays the cor relation. It is best used in variables that demonstrate a linear relationship between each other. Coefficients for different variables. The matrix depicts the correlation be tween 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.

Extracting Segments

Dendrogram

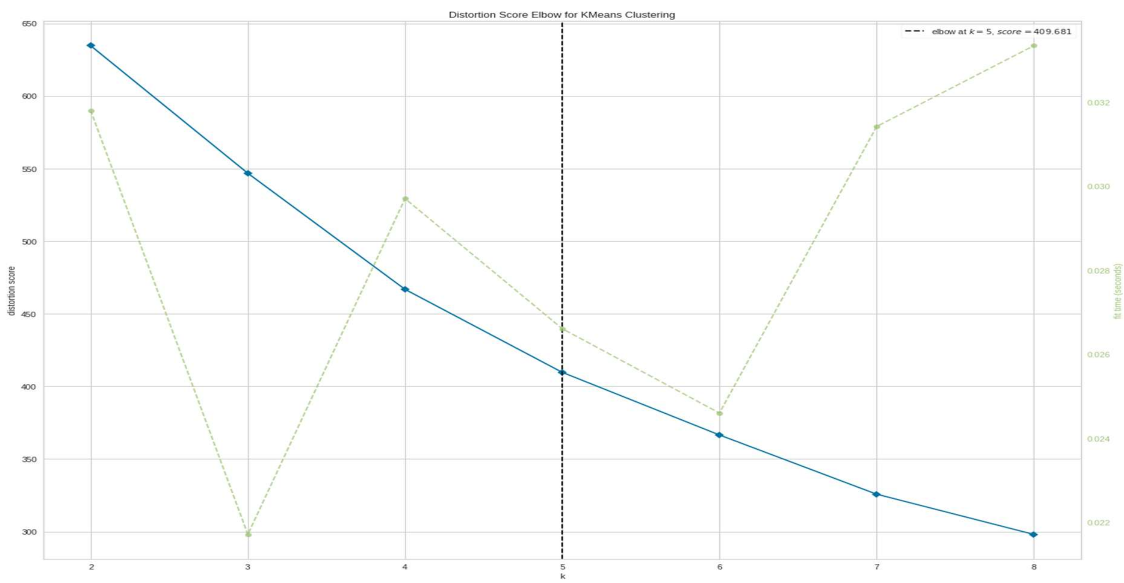
This technique is specific to the agglomerative hierarchical method of clustering. The agglomerative hierarchical method of clustering starts by considering each point as a separate cluster and starts joining points to clusters in a hierarchical fashion based on their distances. To get the optimal number 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.



Elbow Method

The Elbow method is a popular method for determining the optimal number of clus ters. 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. The K Elbow Visualizer function fits the K Means model for a range of clusters values between 2 to 8. As shown in Figure, the elbow point is achieved which is highlighted by the function itself. The function also informs us about how much time was needed to plot models for various numbers of clusters through the green line.



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 anal ysis 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 be longs 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 break down of how we can solve it mathematically.

The Objective function is :

$$J = \sum_{i=1}^{m} \sum_{k=1}^{K} w_{ik} ||x^i - \mu_k||$$

M step is

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

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:

- Get a meaningful intuition of the structure of the data we're dealing with.
- Cluster-then-predict where different models will be built for different subgroups if we believe there is a wide variation in the behaviours 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.

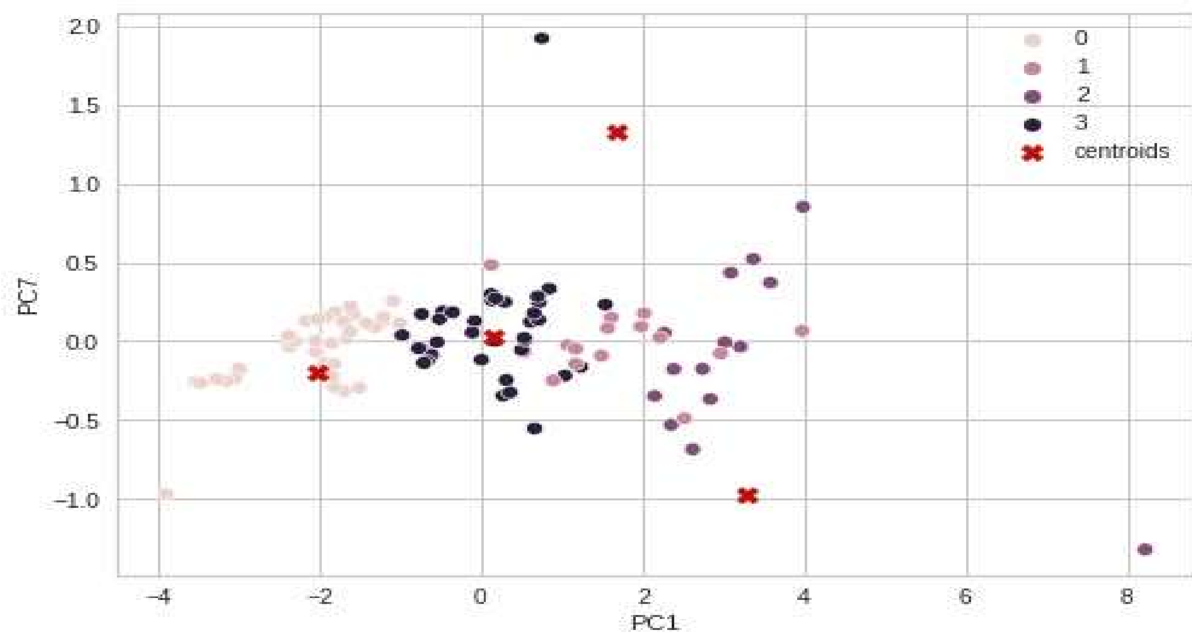
• Assign each data point to the closest cluster (centroid).

• 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 .

```
#K-means clustering

kmeans = KMeans(n_clusters=4, init='k-means++', random_state=0).fit(t)
df['cluster_num'] = kmeans.labels_ #adding to df
print (kmeans.labels_) #Label assigned for each data point
print (kmeans.inertia_) #gives within-cluster sum of squares.
print(kmeans.n_iter_) #number of iterations that k-means algorithm runs to get a minimum within-cluster sum of squares
print(kmeans.cluster_centers_) #Location of the centroids on each cluster.
```



Prediction of Prices most used cars

Linear regression is a machine learning algorithm based on supervised learning. It per forms a regression task. Regression models targets prediction value based on indepen dent 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.

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

```
[86] 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)
```

```
[87] print(lm.intercept_)
```

```
4643.522050485437
```

```
[88] lm.coef_
```

```
array([1144.95884,  530.09473,   54.50586, -843.38276, -306.27756,
       1449.94438,  595.62449, 1005.47168, 1455.75874])
```

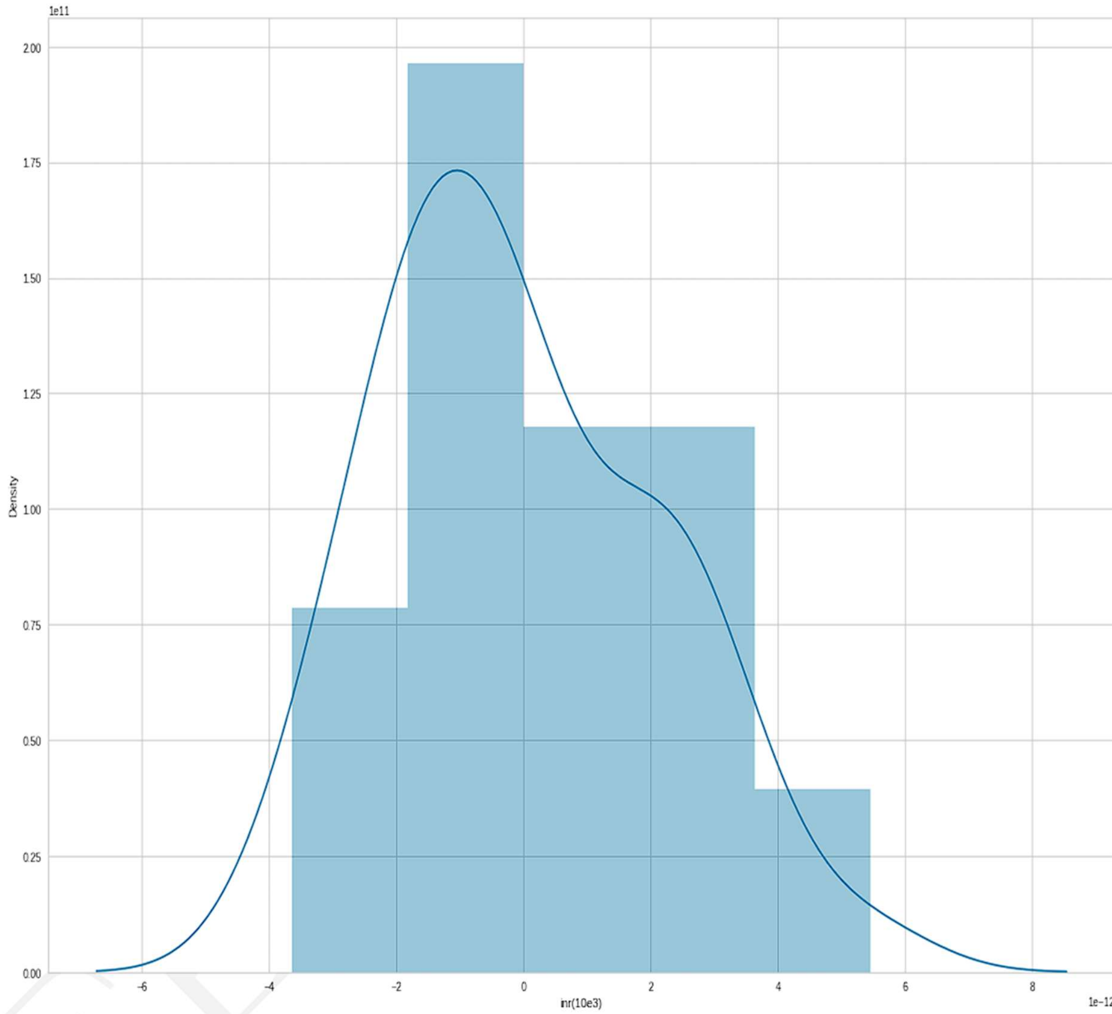
```
[89] X_train.columns
```

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

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

	Coeff
PC1	1144.9588
PC2	530.0947
PC3	54.5059
PC4	-843.3828
Pc5	-306.2776
PC6	1449.9444
PC7	595.6245
PC8	1005.4717
PC9	1455.7587

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 dis tributed

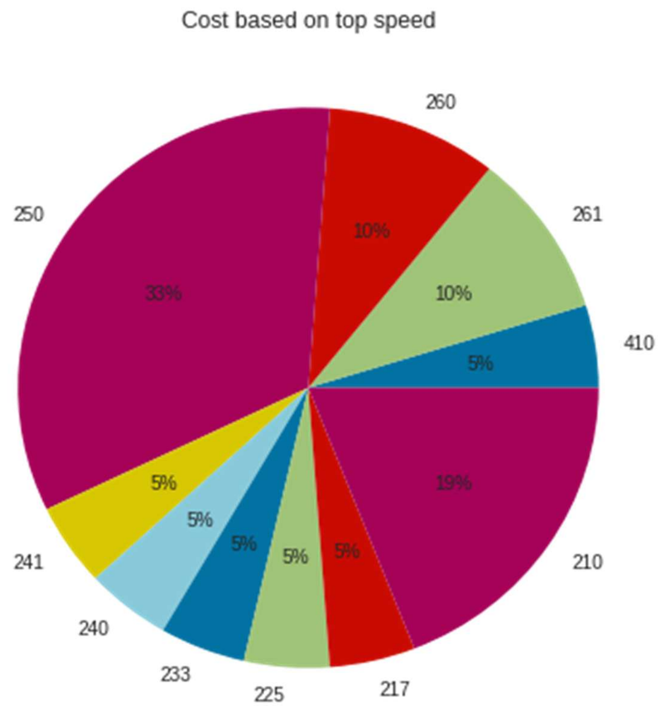


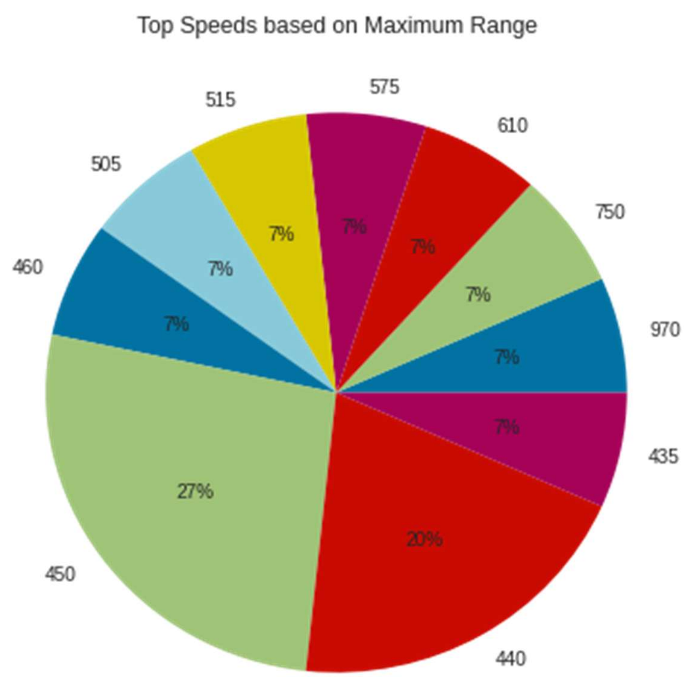
The metrics of the algorithm, Mean absolute error, Mean squared error and mean square root error are described in the below figure:

```
[99]  1  print('MAE:',metrics.mean_absolute_error(y_test,predictions))
      2  print('MSE:',metrics.mean_squared_error(y_test,predictions))
      3  print('RMSE:',np.sqrt(metrics.mean_squared_error(y_test,predictions)))

   MAE: 1.7540254962763616e-12
   MSE: 4.588882922020368e-24
   RMSE: 2.142167809024393e-12

[100] 1  metrics.mean_absolute_error(y_test,predictions)

   1.7540254962763616e-12

[101] 1  metrics.mean_squared_error(y_test,predictions)

   4.588882922020368e-24

[102] 1  np.sqrt(metrics.mean_squared_error(y_test,predictions))

   2.142167809024393e-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:


Cost based on top speed
260
10%
261
10%
250
33%
410
5%
5%
5%
5%
5%
5%
19%
241
210
240
233
225
217


Top Speeds based on Maximum Range
575
515
610
505
750
7%
7%
7%
460
7%
7%
7%
970
7%
7%
435
27%
20%
450
440

Target Segments: So from the analysis we can see that the optimum targeted segment should be belong ing to the following categories:

Behavioural: 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

The marketing mix refers to the set of actions, or tactics, that a company uses to pro mote its brand or product in the market. The 4Ps make up a typical marketing mix-Price, Product, Promotion and Place.

• **Price:** refers to the value that is put for a product. It depends on segment tar geted, ability of the companies to pay, ability of customers to pay supply- de mand and a host of other direct and indirect factors.

• **Product:** refers to the product actually being sold– In this case, the service. The product must deliver a minimum level of performance; otherwise even the best work on the other elements of the marketing mix won't do any good.

• **Place:** refers to the point of sale. In every industry, catching the eye of the con sumer and making it easy for her to buy it is the main aim of a good distribution or 'place' strategy. Retailers pay a premium for the right location. In fact, the mantra of a successful retail business is 'location, location, location'.

• **Promotion:** this refers to all the activities undertaken to make the product or service known to the user and trade. This can include advertising, word of mouth, press reports, incentives, commissions and awards to the trade. It can also include consumer schemes, direct marketing, contests and prizes.

All the elements of the marketing mix influence each other. They make up the business plan for a company and handle it right, and can give it great success. The marketing mix needs a lot of understanding, market research and consultation with several people from users to trade to manufacturing and several others.

Data Sources

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

Github : https://github.com/Ankit141731/Electric_Vehicle_MS.git