

Team-Santosh

Electric Vehicle Segmentation analysis

Team Members

1. Santosh kumar
2. Dhruv Gala
3. Salvador Ron Nathaniel
4. Abhishek Pal
5. Keerthana S K
6. Aanchal Gupta

Background

The electric vehicle is a fast growing industry in the vehicle market. It is an emerging market which started in the late nineteenth century.EVs first came into existence in the late 19th century, when electricity was among the preferred methods for motor vehicle propulsion, providing a level of comfort and ease of operation that could not be achieved by the gasoline cars of the time. Internal combustion engines were the dominant propulsion method for cars and trucks for about 100 years, but electric power remained commonplace in other vehicle types, such as trains and smaller vehicles of all types.An electric vehicle can be classified on the basis of their attributes such as charging time, driving range, and the maximum load it can carry. Of these attributes, the two most important characteristics of an electric vehicle of concern to the consumer are: Driving range(i.e. Maximum distance an EV can run when fully charged) Charging time of batteries(i.e. the time required to fully charge the battery) and Charging time depends on the input power characteristics (i.e. input voltage and current), battery type and battery capacity. Therefore, such a vehicle is seen as a possible replacement for current-generation automobiles, in order to address the issue of rising pollution, global warming, depleting natural resources, etc

Problem Statement

You are a team working under an Electric Vehicle Startup. The Startup is still deciding in which vehicle/customer space it will develop its EVs. You have to analyse the Electric Vehicle market in India using Segmentation analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use Electric vehicles.

Fermi Estimation

The main task is to find out the vehicle space in which the EV startup will develop its products. For this, we use a dataset which has the preferred car characteristics of the customers in the market. As we are using the preferences and interests of the customers and also considering the frequency of purchase (as it is selling well in the market) and amount spent on purchasing, we can say that the segmentation criterion is psychographic and behavioural segmentation. And we use machine learning algorithms for segmentation of the dataset, after we have done with profiling it, we can then find the most optimal market segments to open in the markets.

Data Sources

We have found the data from various popular cars in the market and selected this dataset because it includes all the specifications of the property of the different cars available in the market. It includes the customer satisfaction and the price paid by them. It includes the fourteen attributes which are of two types categorical and numerical of best selling cars in the market.

Categorical Attributes

- Brand
- Model
- Rapid Charge
- Power train
- Plug Type
- Body Style
- Segment

Numerical Attributes

- Acceleration per second
- Maximum Speed
- Range
- Efficiency
- Charging Speed
- Seats
- Price

Data Preparation and Preprocessing

Data collection

We collected data from the Kaggle site.

[Electric Vehicles - India | Kaggle](#)

[Electric & Alternative Fuel Vehicles US \[2022\] | Kaggle](#)

[Cheapest Electric Cars | Kaggle](#)

[Electric Vehicles User Reviews India | Kaggle](#)

[Electric Vehicle Analysis | Kaggle](#)

Checking the details

In this step we are loading the data file and checking the details of tabular data like number of columns, number of rows, type of features (categorical data or numerical data), and getting descriptive information like Mean, Standard Deviation, Median, Max Value, Min Value.

Glimpse of data used for clustering

Model	AccelSec	TopSpeed_KmH	Range_Km	Efficiency_WhKm	FastCharge_KmH	RapidCharge	PowerTrain	PlugType	BodyStyle	Segment	Seats	PriceEuro
Model 3 g Range al Motor	4.6	233	450	161	940	Yes	AWD	Type 2 CCS	Sedan	D	5	55480
I3 Pure	10.0	160	270	167	250	Yes	RWD	Type 2 CCS	Hatchback	C	5	30000
I2	4.7	210	400	181	620	Yes	AWD	Type 2 CCS	Liftback	D	5	56440
iX3	6.8	180	360	206	560	Yes	RWD	Type 2 CCS	SUV	D	5	68040
e	9.5	145	170	168	190	Yes	RWD	Type 2 CCS	Hatchback	B	4	32997
...
Ia 63kWh	7.5	160	330	191	440	Yes	FWD	Type 2 CCS	Hatchback	C	5	45000
e-tron S portback 5 quattro	4.5	210	335	258	540	Yes	AWD	Type 2 CCS	SUV	E	5	96050
Ariya e-4ORCE 63kWh	5.9	200	325	194	440	Yes	AWD	Type 2 CCS	Hatchback	C	5	50000
Ariya e-4ORCE	5.4	200	375	200	450	No	AWD	Type 2 CCS	Hatchback	C	5	55000

Libraries used in python

Numpy

This library consists of multidimensional array objects. Using NumPy, mathematical and logical operations on arrays can be performed.

Pandas

Pandas is used for manipulating the data in different ways. It can handle missing data, cleaning up the data and it supports multiple file formats.

Visualisation libraries

Matplotlib, seaborn are libraries which are used to plot different kinds of graphs and charts like, bar, scatter, histogram, line etc.

Libraries/Packages used in R

flexclust

This library is used to segment the market into specific segments using standard k -means clustering using nrep number of restarts.

purrr

This library basically enhances R's functional programming by providing consistent tools for working with functions and vectors.

factoextra

This package in R helps to easily extract and visualise the output of exploratory multivariate analysis.

Exploratory Data Analysis

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns,to spot anomalies,to test hypotheses and to check assumptions with the help of summary statistics and graphical representations.

So, we perform EDA on this dataset, to understand the dataset and draw rough estimations from the data which might be further useful in the pre-processing step.

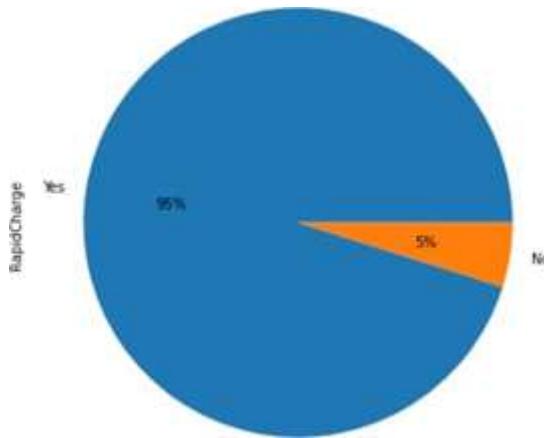
First we use a unique() function for each categorical attribute , to find unique classes for each attribute. So that we will be able to understand what type of data is present in the dataset.

And then we use seaborn library for pair plot, this helps us analyse the rough relation between 2 attributes. Some observations based on the pair plots (roughly):

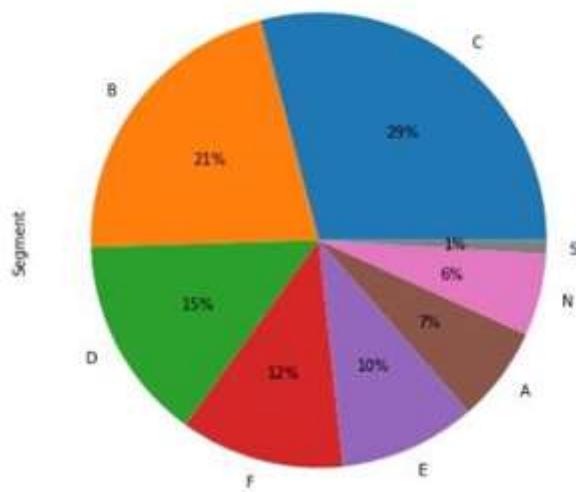
- TopSpeed and AccelSec has linear relation with negative correlation
- TopSpeed and PriceEuro has polynomial relation with probably degree 2
- FastCharge and Range has a linear relation with positive correlation
- Efficiency vs other attributes plots has no specific pattern which indicates no relation between efficiency and other attributes

Plotting Pie-Charts for Categorical Variables:

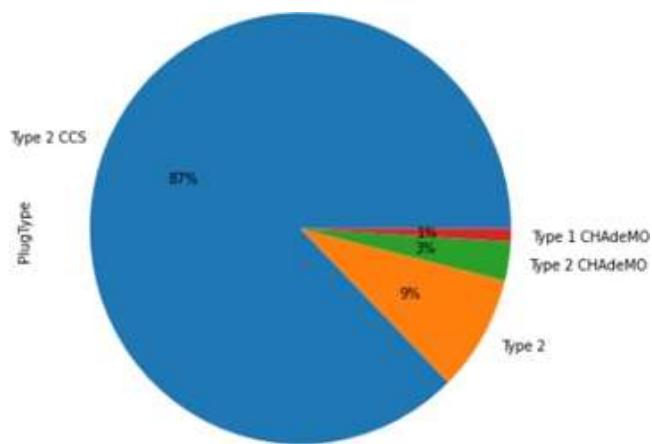
After plotting a pie chart for RapidCharge, we can see that a disproportionate number of the instance values for these attributes is of one class which is Yes, so this does not provide us with much information. Hence, we can remove it from the dataset.



And then we plot a pie chart for the attribute Segment, we can see that Segment C is the majority class. And all other classes are pretty much evenly distributed.



Then, we plot a pie chart for the attribute PlugType, and we can see more than 80 % of the instances is of class Type 2 CCS. Hence as a disproportionate number of the instances is of one class, it provides us with minimal information and we can safely remove it.

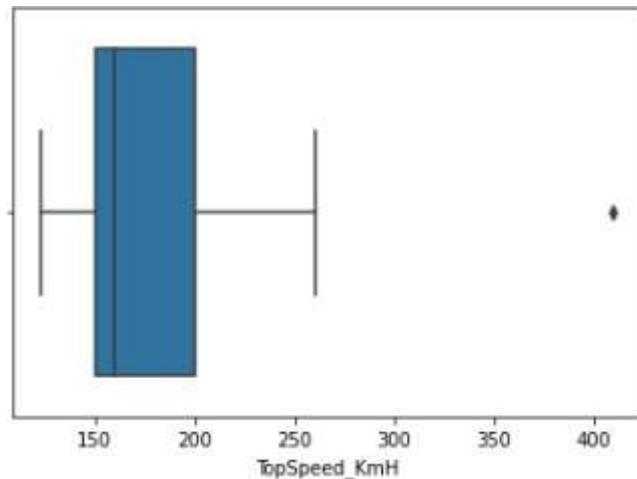


Plotting Bar Plot for numerical variable

To have a visualisation of numerical data, we basically convert a range of numerical values into categorical classes and then we plot a bar plot, where x axis is the categorical classes that we have synthesised and y axis is the frequency of the particular class in the attribute.

In this dataset, we convert each numerical attribute into three categorical values (Low, Medium and High). And for deciding which range of values should be chosen for each categorical value, we take the help of a box plot.

For instance, take the attribute TopSpeed_KmH, after plotting a box plot for it.



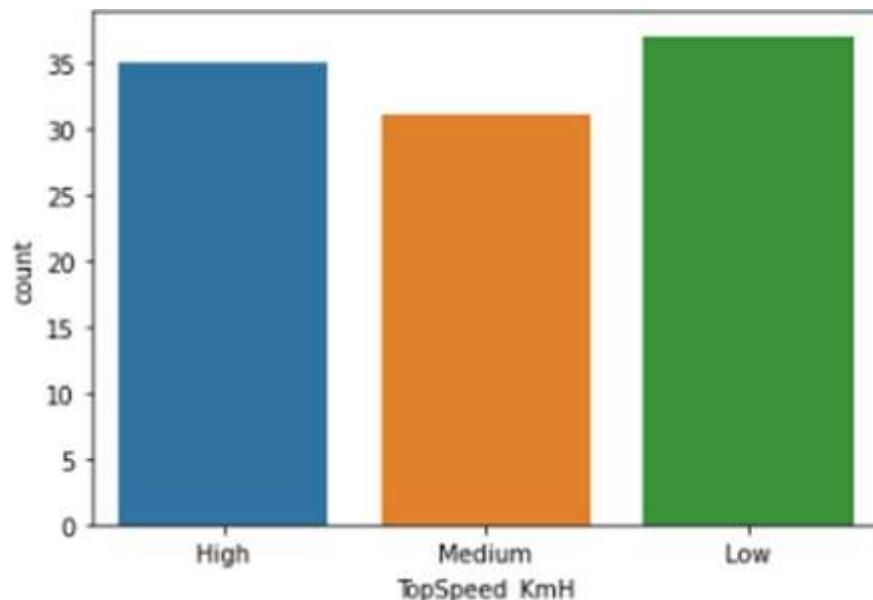
From the above box plot we can analyse assign the following: less than 160: Low

between 160 and 190:

Medium greater than 190:

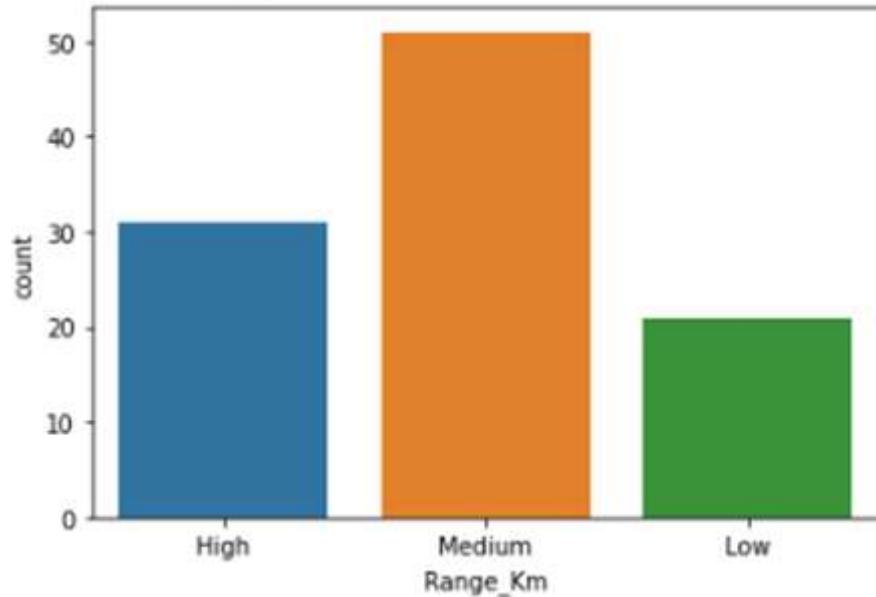
High

Based on this we can convert the instances of numerical attributes to a list of values with categorical values. From the Bar plot we can see that the values in each range are pretty evenly distributed and the number of values in the class Low (<160) are slightly higher. The bar plot for this processed attribute can be seen below:

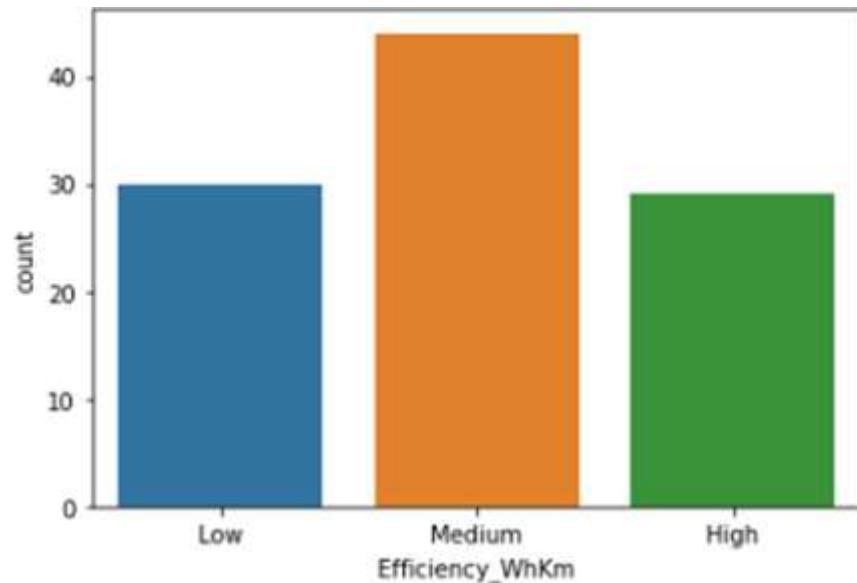


The similar thing can be done for numerical attributes such as Range_Km, Efficiency_WhKm and AccelSec, and rough conclusions can be drawn based on each bar plot.

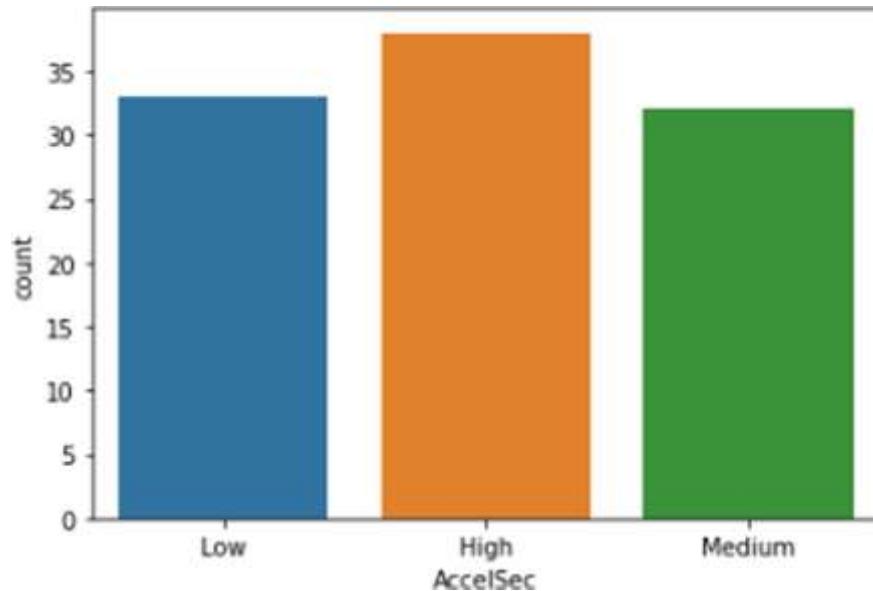
The bar plot for Range_Km is (less than 250: Low; between 250 and 400: Medium; greater than 400: High):



The bar plot for Efficiency_Whkm is (less than 170: Low; between 170 and 200: Medium; greater than 200: High):

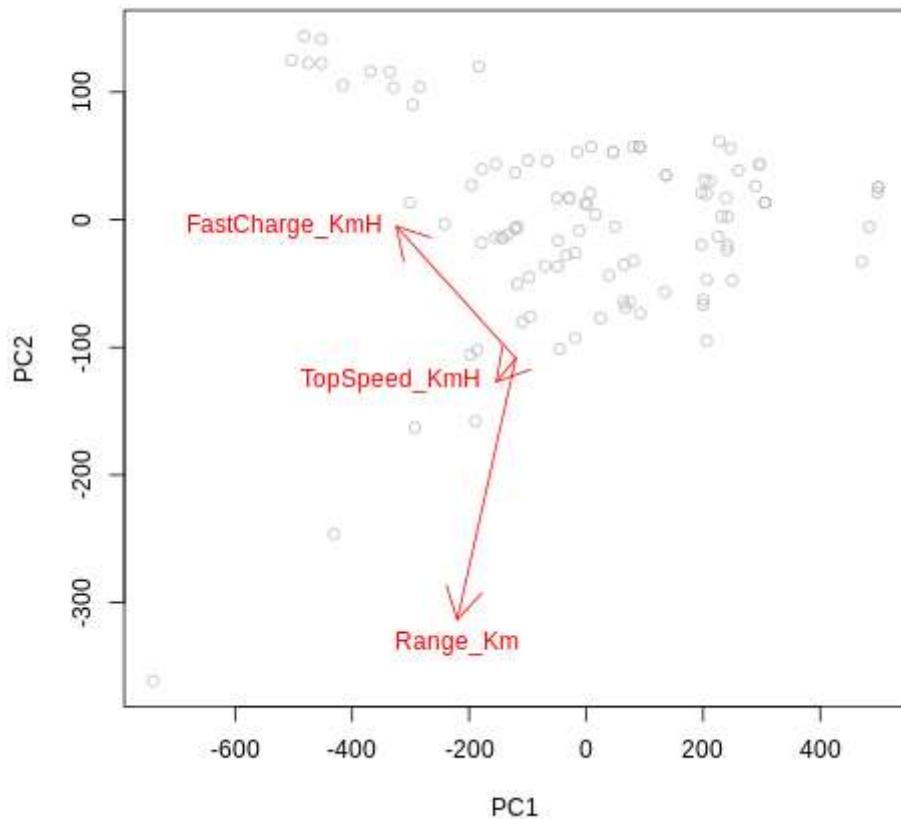


The bar plot for AccelSec is (less than 6: Low; between 6 and 8: Medium; greater than 8: High):



Segmentation Extraction

Inferences drawn from my Segmental Analysis and Targeting Segments:

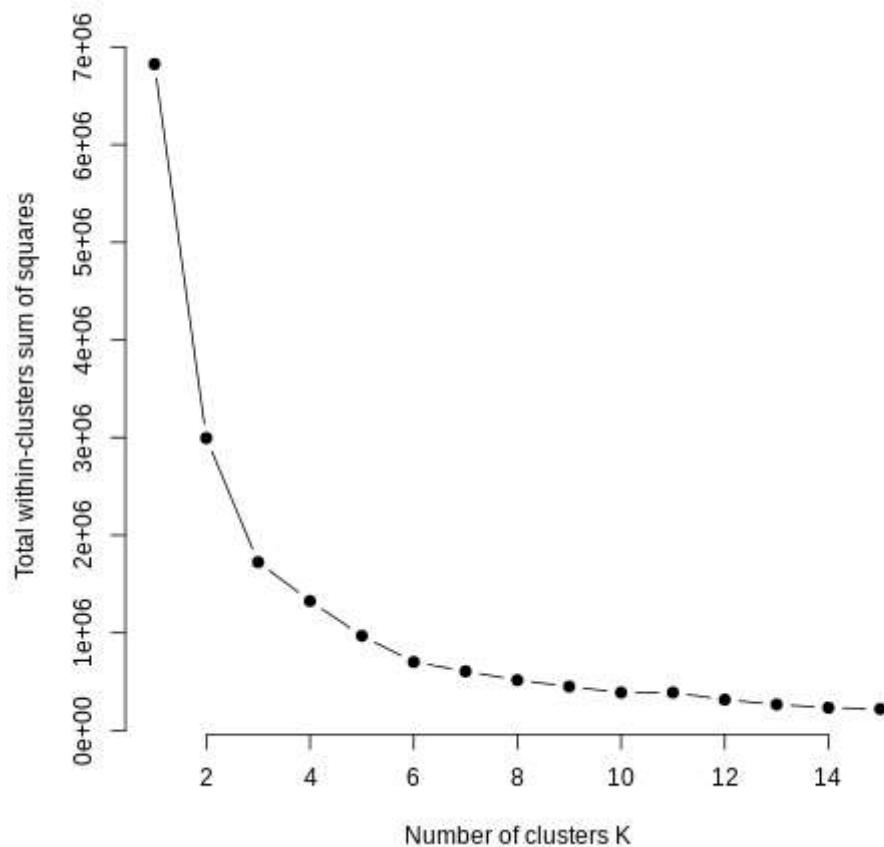


1) So from the Principle Component Analysis it can be clearly seen that the entire market depends mainly on the following 3 factors which are basically the Principle Components.

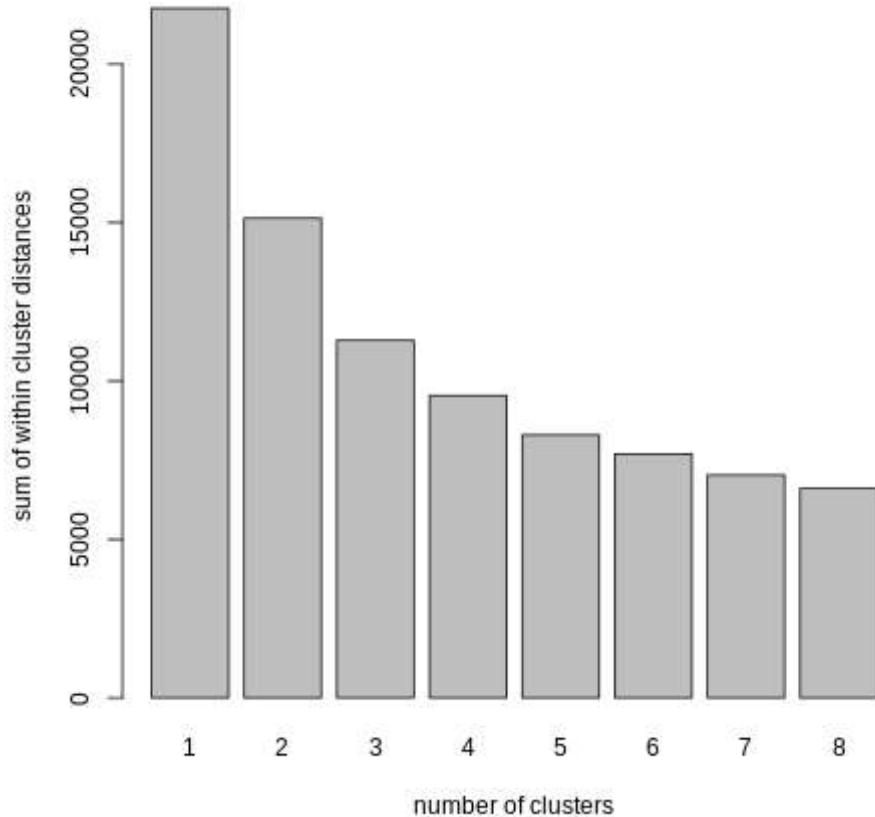
Namely-

- a) FastCharge_KmH
- b) TopSpeed_KmH
- c) Range_Km

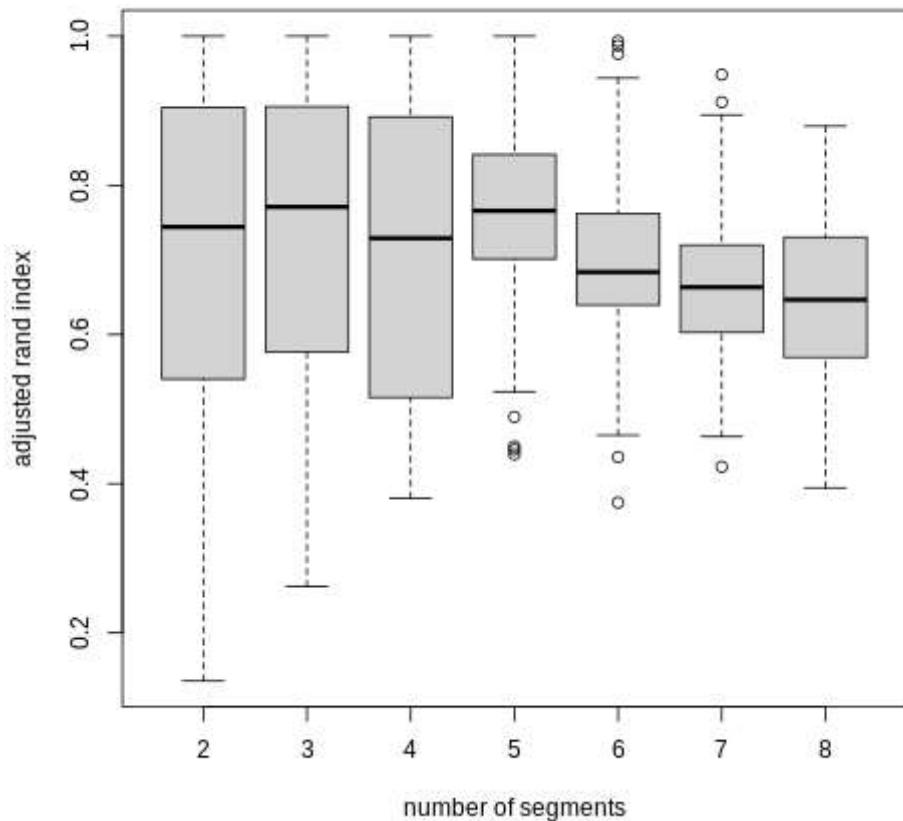
And all these Principle components are relatively independent of each other.



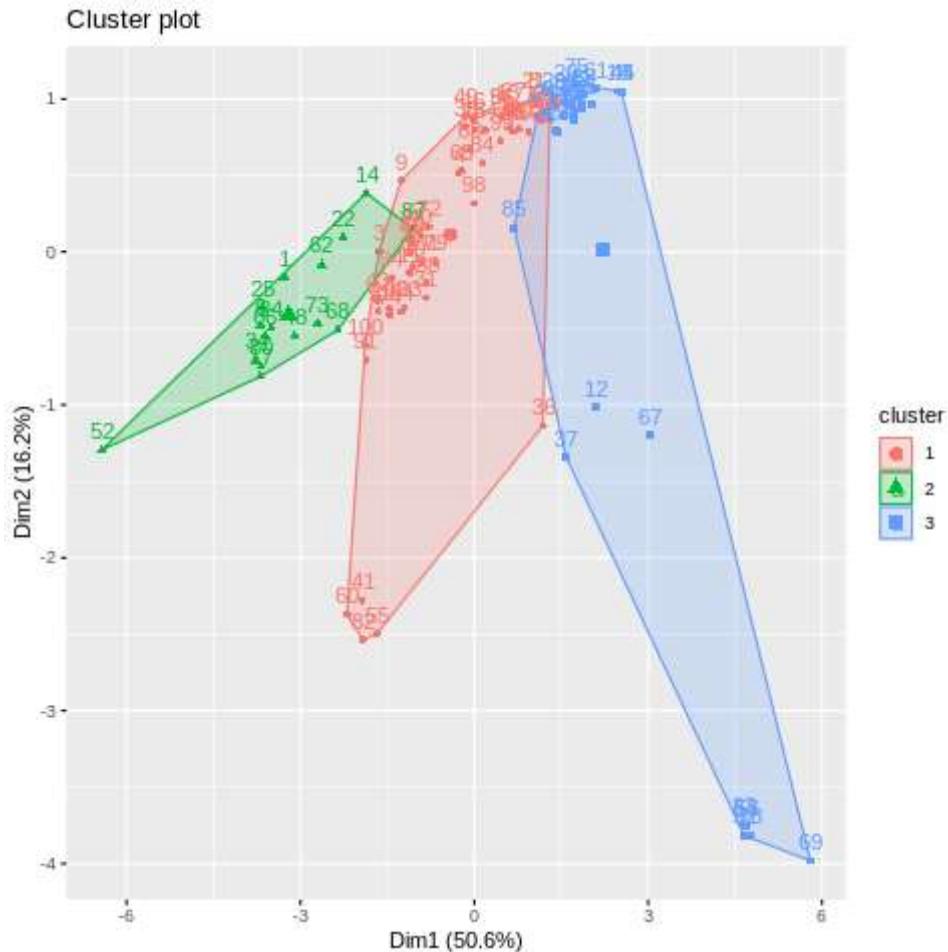
2)From the Kmean cluster analysis as well we see that the ideal market segment is somewhere between 2 and 4(i.e the elbow point)



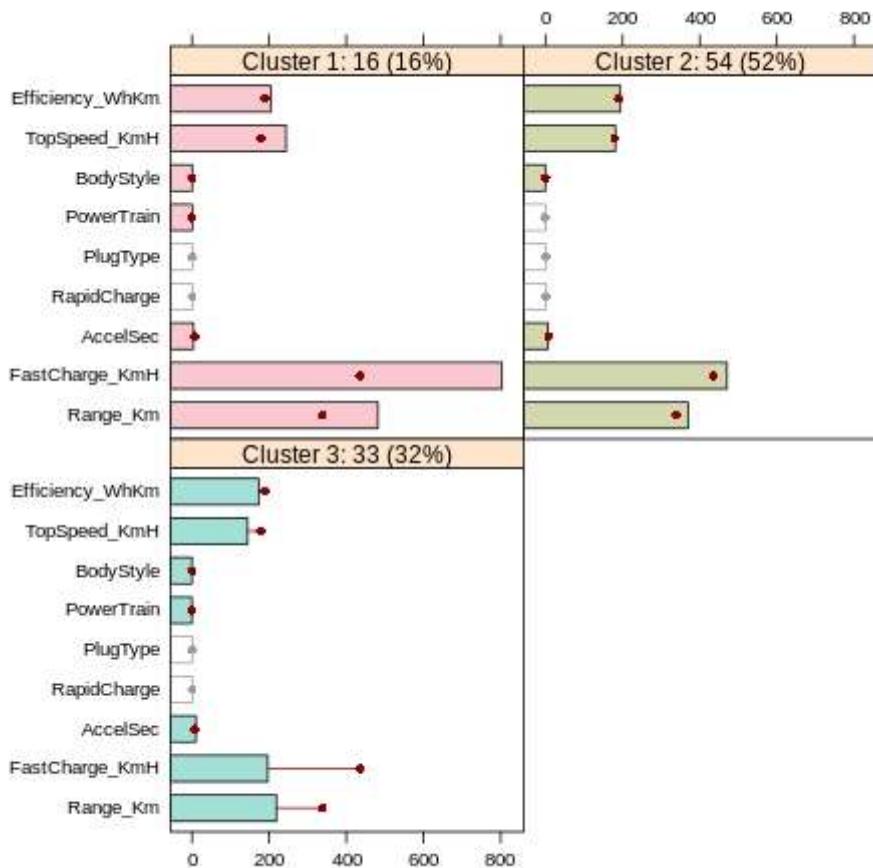
- 3) From the above scree plot deduced from step flexclust analysis we see that elbow point is around 3. But still we are not clear whether this is the best number of segments.



- 4) Finally from the boot flexclust analysis it can be deduced that Segment 3 has the second highest stability, the highest being 5. But if the number of segments becomes 5, then each of the segments becomes too small to be invested upon. So the ideal and most viable number of segments for this EV market is 3.



- 5) If we look at each of the cluster 1, 2 and 3 , which we get on classifying the entire market into primarily 3 segments, then we find that cluster 2 and cluster 3 is largest, i.e. it capture more than 50%(precisely 84%) of the data present in the dataset.



6)

- a) If we look into cluster 2 we find that it includes cars with respective models which have high FastCharge_KmH and Range_Km.
- b) Cluster 3 includes cars and their respective models which at par values of Efficiency_WhKm, TopSpeed_KmH, FastCharge_KmH and Range_Km.

Conclusions:-

The marketing managers should focus their attention on the Cluster 2 and Cluster 3 group of cars, as they are highest in Numbers. Factories who are manufacturing EV should make sure that they make cars which have high-

.

High Efficiency_WhKm, TopSpeed_KmH, FastCharge_kmH and Range_Km

Profiling and Describing Potential Segments

The goal of Profiling is to be able to differentiate between the groups with the proper distinction that means each group has some unique characteristics about it and then that cluster is the ideal solution to segment the values.

The libraries used for profiling are:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
```

The segments that are used for profiling are:

1. TopSpeed_KmH
2. Range_Km
3. FastCharge_KmH

The steps to perform profiling are as follows:

1. Find the count of each segment which gives how many observations or records are present in each of the segments.
2. Find the overall average and the individual segment-wise average for each of the attributes.

Step 1: Firstly, we find the total size of the cluster and then find the break up of the values in each of the respective segments.

So, we obtained how many observations are present in each of the segments. Post which, we sort the segments based on the labels and not the highest value within the segments:

```
df.cluster1.size
```

```
df.cluster1.value_counts()
```

Now, we repeat the above two steps for other clusters, in our case for cluster2 and cluster3, and then combine the result for each cluster K value into one single array:

```
size=pd.concat([pd.Series(df.cluster1.size),pd.Series.sort_index(df.cluster1.value_counts()), pd.Series.sort_index(df.cluster2.value_counts()), pd.Series.sort_index(df.cluster3.value_counts())])
```

We store the Segment Size and the segment distribution in the respective data frames and concatenate the two outputs:

```
Seg_size=pd.DataFrame(size, columns=['Seg_size'])
```

```
Seg_Pct = pd.DataFrame(size/df.cluster1.size, columns=['Seg_Pct'])
```

Step 2:

A good indication of the distribution of data is the mean value, hence will find the average value for each of the variables and in each cluster.

Firstly, finding the overall average for each variable:

```
df.apply(np.mean).T
```

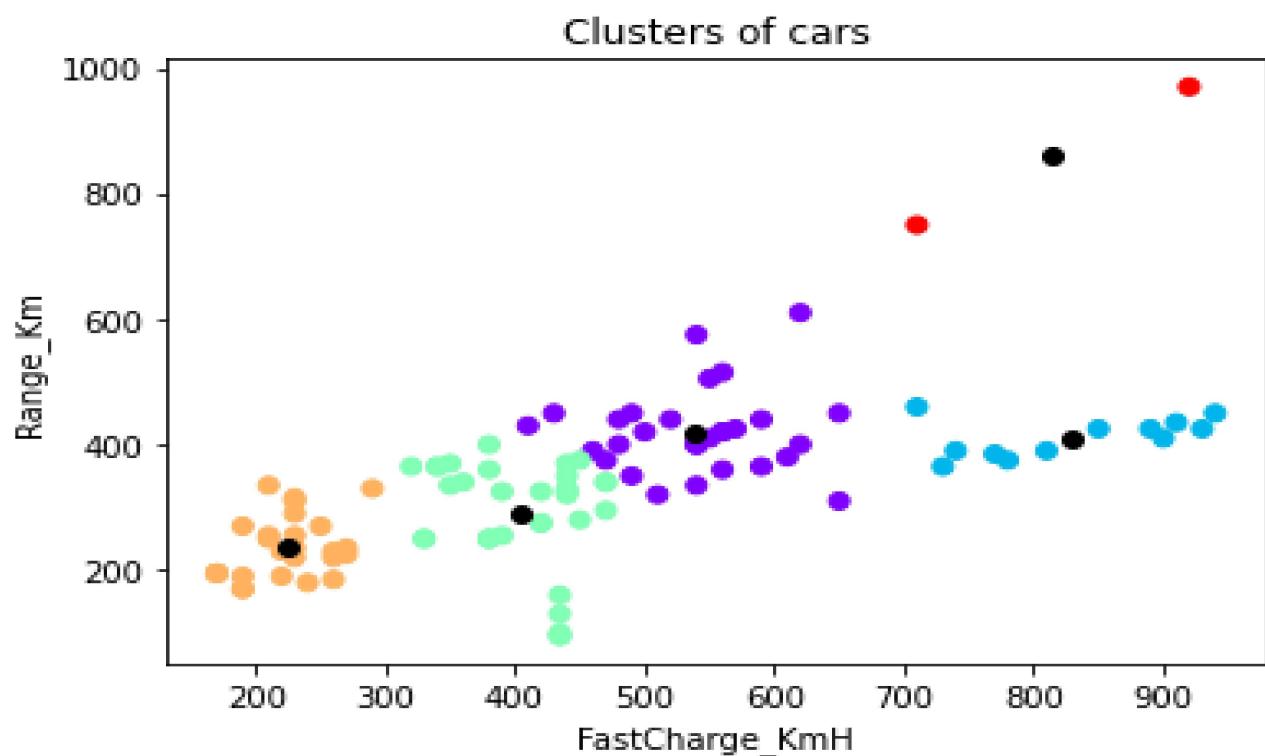
By grouping over each cluster, we find the Segment wise average for each variable:

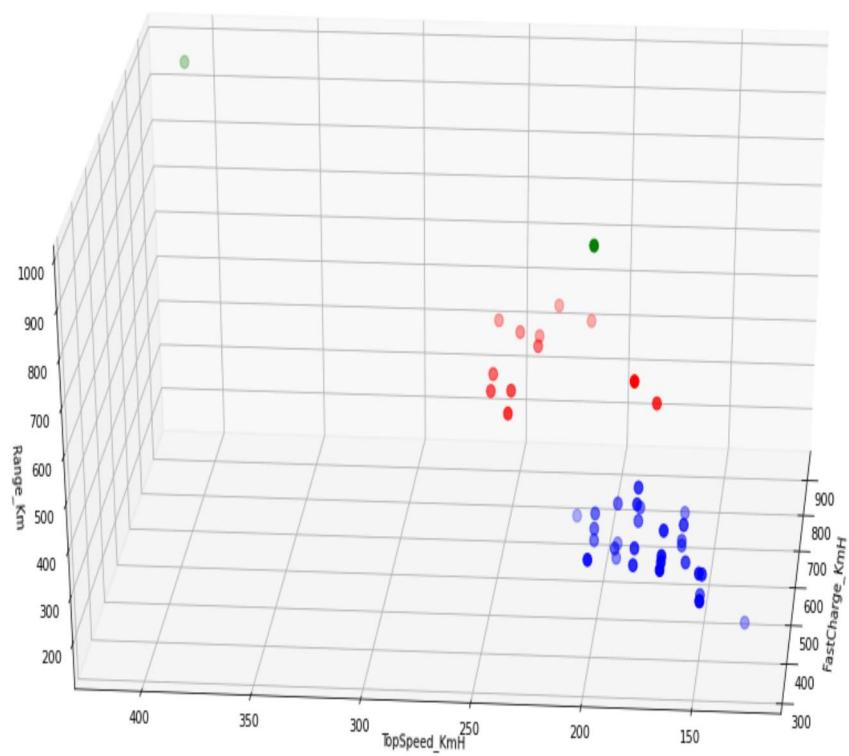
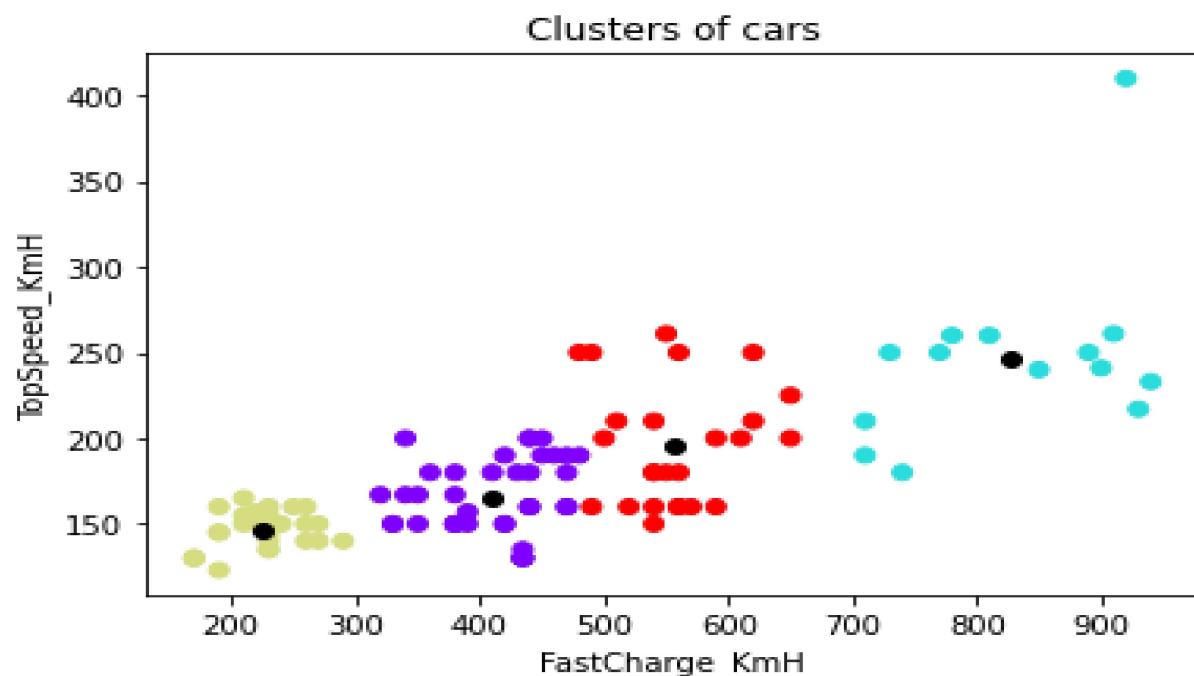
```
df.groupby('cluster1').apply(np.mean).T
```

Concatenating the above two averages overall average and individual-segment wise average:

```
Profiling_output = pd.concat([df.apply(lambda x: x.mean()).T,
                             df.groupby('cluster1').apply(lambda x: x.mean()).T,
                             df.groupby('cluster2').apply(lambda x: x.mean()).T,
                             df.groupby('cluster3').apply(lambda x: x.mean()).T], axis=1)
```

The graph of three clusters which is used for profiling are as follow:





Analysing Market Segmentation

Customising Market Mix

There are basically 8 main points that market mix should contain while performing market segmentation and these are as followed:-

1. **Product Planning** -> Based on the dataset found we will be managing the product distribution of the EV Vehicles and providing the right vehicle market to the right customer at the right location. In order to achieve so a head office for the company must be set up in state from where all import and exports of the car can be done that are not yet manufactured in the country and to control the sales operation example – (Maharashtra or Gujarat) that are near to coastal region, other than that the central offices and showrooms for sales must be established in metro cities along with in the state capital where EV cars are more likely to be bought. To expand the market according to the market needs the showrooms and offices can be expanded to various cities across the country only if the need of the product reaches a certain criteria and profit could be made.

2. Packaging -> To move the product across the country from ports or manufacturing plants road transport will be used which includes trucks which can be either of 18.75 metres or preferably 22.5 metres which will transport the desired vehicles to particular state as per demand or railway transport can also be used to transport cars which will save a fortune. For this company must contact third party companies which are professional movers and packers who can handle their jobs smoothly.

3. Distribution Channels-> For the product to reach their destination to the customer the product after manufacturing from the manufacturing plant must be loaded for transportation which will be then delivered to the showroom or the warehouse of the company in the city. After all the paperwork and payment to be done by the customer the authorised company people can hand over the car to the customer. This distribution channel will not be affected by the mediocre person who takes more money for delivering the product and companies who do not provide any guarantee for their delivery of the product if the deal is being closed by a non-authorized dealer to the customer.

4. Pricing-> The pricing of the product will be divided into certain categories which will depend on the geographic region the product is being delivered. The ex-showroom price will be the same in the entire country along with that price will include RTO registration, road tax, insurance and other logistic charges. The charges for the accessories depend on customer to customer as per their needs.

5. Branding-> Branding of the product must be done by pasting the company's specific logo on the product which will be done through a consistent theme. Branding would also be done through billboards and mass media communication and social media. The colour theme of the logo will be

a combination of yellow-green to which human eyes are more sensitive and using a light shade with them. The branding would definitely be done by the customer who is using the product and of course the best branding is the mouth publicity that will surely be done by the customer.

6. Advertising-> Advertising of the company and its product will be done through 2 ways first by social media which is very powerful way of advertising and marketing these days targeting the youth and secondly by flex printing and in newspaper through which persons who are above 50's can be targeted as they are not active on social media but have assets of purchasing the product.

7. Servicing-> The cities in which the showrooms will be opened along with that the service centre for the vehicles will also be opened so proper and time to time maintenance could be done and all faulty behaviour in the vehicles can be rectified and customer satisfaction can be achieved.

8. Finding and analysis-> The marketing team, sales team and customer support team will be handling all customer queries regarding the product and their problems faced. They would also be monitoring the sales in every region and they will be working on how to increase the sales in the particular region along with providing the good customer service because at last a happy customer means a good running company.

Potential Customer Base & Segmentation

To find the right customer base and customer retention are the two important things which require more marketing efforts. Growing a customer base will be beneficial for the brand's success. Retaining those customers would

increase the possibility of becoming repeated users around 60% - 70%. The company provides great customer support and meets the requirements of customers on time in order to grab loyalty and keep good relationships with customers. Apparently there is a 5-20% chance that a new customer buys the product. The company as its last step should enquire how the customers got to know about the product and ask for their feedback regarding the current services offered by the company.

For the product to be successful ,it should be launched in 3 phases. In the first phase two showrooms will be opened in each metro cities of Mumbai and Delhi along with service centres. After analysing customer response to the product and how the market reacts to it, the company will proceed to the second phase. This includes opening showrooms within the cities where the product would be comparatively more demanded and further expand the market among other cities as well. The last phase includes opening the showrooms in metro cities more with traffic centres. The future plan is to reach out to the outskirts of current cities as well as the other cities in different states.

Geographical Segmentation: - The product will be displayed based on the geographic region and adaptable road for the corresponding car segments. The A segment cars will be preferred in populated cities where traffic is more. The B segment cars can be preferred for both city travel with a long distance. The J segment cars will be preferred in hilly regions with tough terrain where roads are muddy, snowy, having grassy patches. Ground clearance is a major thing but some customers would prefer money over functionalities. The C segment car will be preferred for those who need comfort during their travelling, also called family cars. The D segments cars will be preferred by the customers who are in need of large boot space along with a good cabin space. The F segment cars will be preferred majorly by business owners who live their lives in luxury and want to reflect that, these products are higher in cost as compared to others and maintenance is also high ,despite this ,it won't affect the person who owns it.

Demographic Segmentation: - This type of segmentation helps to narrow down the market and it will depend on according to the colour, fuel efficiency, power figures, efficiency, range and acceleration per second in the given dataset. These types of buyers can be found in any geographic region and it all depends upon their willingness to buy a satisfiable product with required features. Those requirements need to be arranged within a limited time for building a strong customer base.

Behavioural Segmentation: - This type of segmentation includes the behaviour of customers towards products also a good way of understanding about the customers. These customers are important for an organisation for the successful journey since they can be repeated users as well. To hold these customers ,companies must implement new strategies among existing customers and provide them with offers and discounts to retain them.

Most Optimal Market Segment

After the thorough research and analysis performed on the given dataset and from the obtained output it can be clearly seen that most optimal market segments that need to be targeted are top speed, fast charging and range. In the current scenario when people want to switch from a combustion engine to an electric car, they would think about the availability of charging stations. Petrol pumps are large in number but it is difficult to find charging stations ,as the electric vehicles are just starting to become familiar among the Indian population. Customers are more concerned about the range of the vehicle. More R&D work and update will occur parallelly in this area. As the customers would like to spend less time on their vehicle charging , the company targets these behavioural customers and market targeting would be carried out based on those specific customers. Generally car enthusiasts think about how much power a car produces over the other features. These types of customers will be targeted by using top speed criteria of the vehicles.

To start a new venture in electric vehicles the company needs to target top speed, range and fast charging as their key selling points to the customer which in turn create a strong grip of the company as a start-up in the market.

The link to find the code for all the analysis done above is attached herewith-

https://colab.research.google.com/drive/1OoHL4OuAW_B_amTL0_SCKWFYQQfw_UA?usp=sharing
