Abstract

Market segmentation becomes a crucial tool for evolving transportation technology such as electric vehicles (EVs) in emerging markets to explore and implement for ex- tensive adoption. EVs adoption is expected to grow phenomenally in near future as low emission and low operating cost vehicle, and thus, it drives a considerable amount of forthcoming academic research curiosity. The main aim of this study is to explore and identify distinct sets of potential buyer segments for EVs 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 findings posit that the three distinct sets of young consumer groups have been identified and labelled as 'Conservatives', 'Indiffer- ents', and 'Enthusiasts' which are deemed to be buddying EV buyers The implications are recommended, which may offer some pertinent guidance for scholars and policy- makers to encourage EVs adoption in the backdrop of emerging sustainable transport market.

In this report we are going to analyse the data and solve the problem using Fermi Estimation by breaking down the problem.

KeyWords : Electric vehicles, Market segmentation, Cluster analysis, Attitude to- wards electric vehicles, Subjective norms, Adoption intention, Sustainable transporta- tion.

Data Collection

Provided by the organization.

Market Segmentation

Target Market:

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

Behavioral Segmentation: searches directly for similarities in behavior or reported behavior.

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

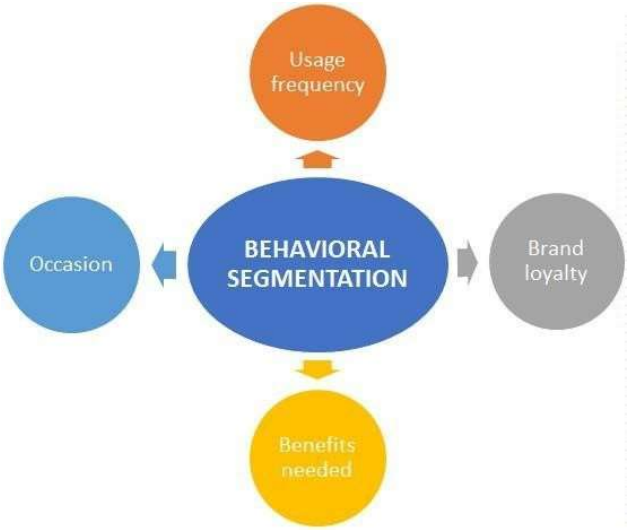


Figure 1: Behavioral Segmentation

Advantage: uses the very behavior of interest is used as the basis of segment extrac- tion.

Disadvantage: not always readily available.

Psychographic Segmentation: grouped based on beliefs, interests, preferences, aspi- rations, or benefits sought when purchasing a product. Suitable for lifestyle segmenta- tion. Involves many segmentation variables.

Advantage: generally more reflective of the underlying reasons for differences in con- sumer behavior.

Disadvantage: increased complexity of determining segment memberships for con- sumers.

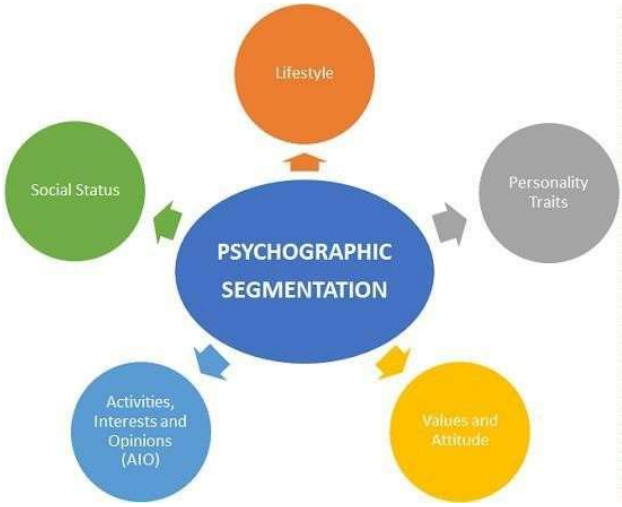


Figure 2: Psychographic Segmentation

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

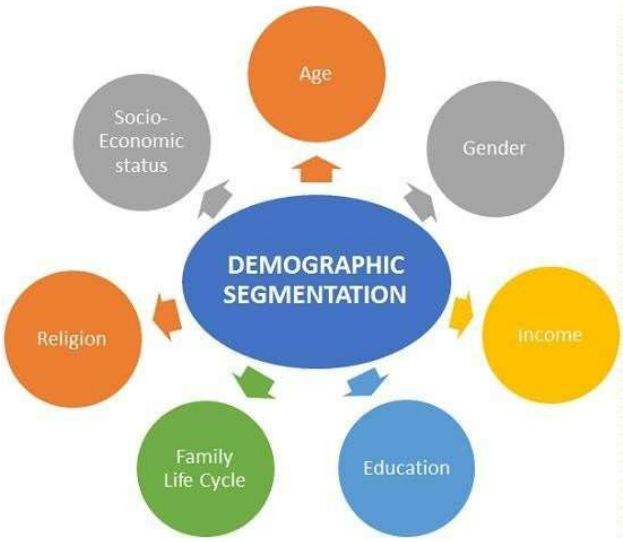


Figure 3: Behavioral Segmentation

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, homoge- nous 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 pro- filing 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.

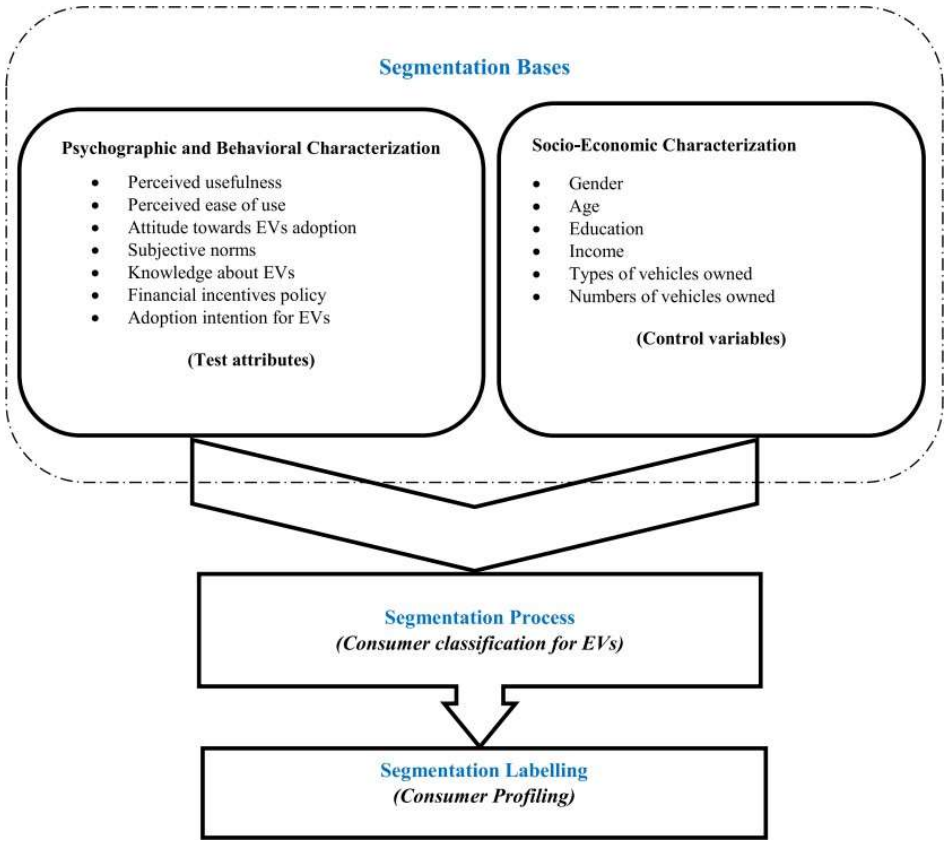


Figure 4: Market Segmentation Electric Vehicles

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.

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

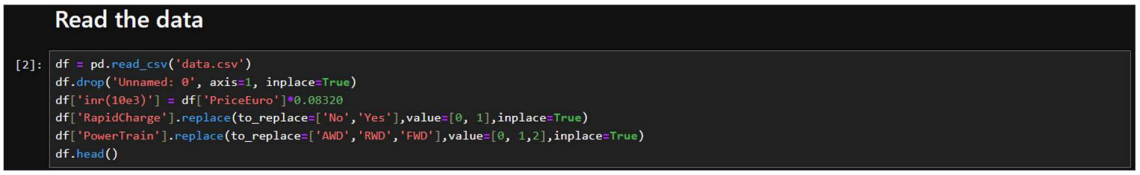
Packages/Tools used:

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

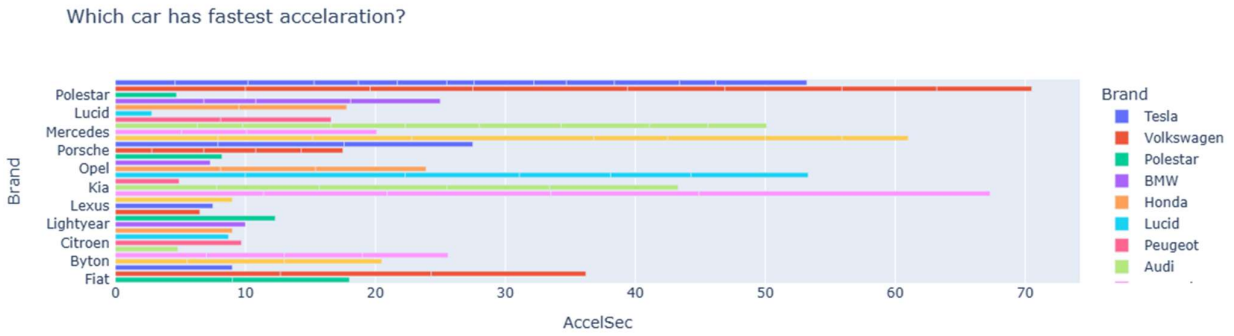
The data collected is compact and is partly used for visualization purposes and partly for clustering. Python libraries such as NumPy, Pandas, Scikit-Learn, and SciPy are used for the workflow, and the results obtained are ensured to be reproducible.

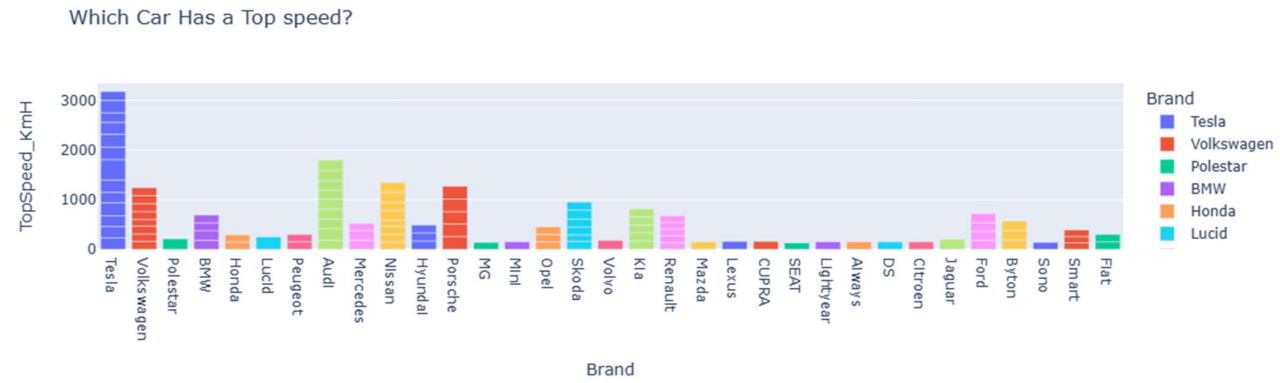


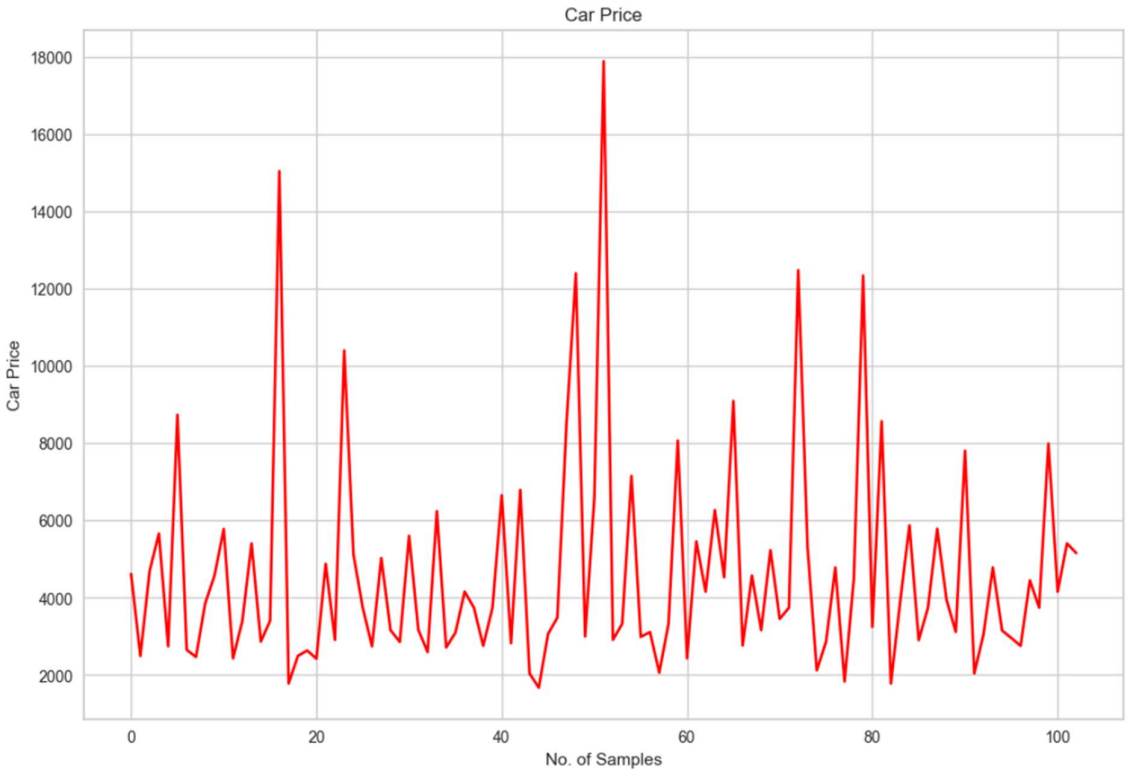
EDA

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.

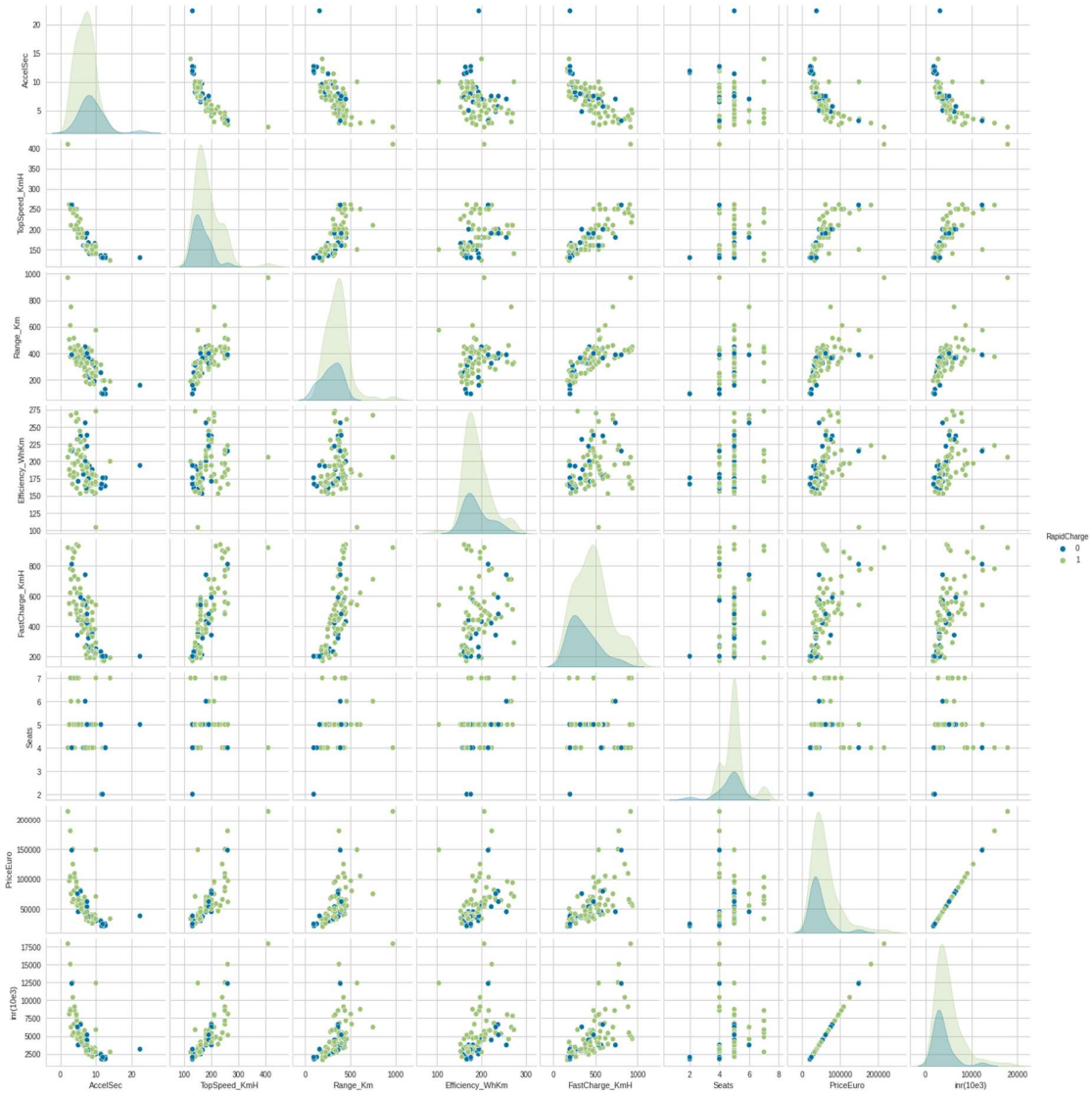
Comparision of cars in our data






Car Price
Car Price
No. of Samples
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
20
40
60
80
100

PairPlot:


AccelSec
TopSpeed_KmH
Range_Km
Efficiency_WhKm
FastCharge_KmH
Seats
PriceEuro
inr(10e3)
RapidCharge
0
1

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.

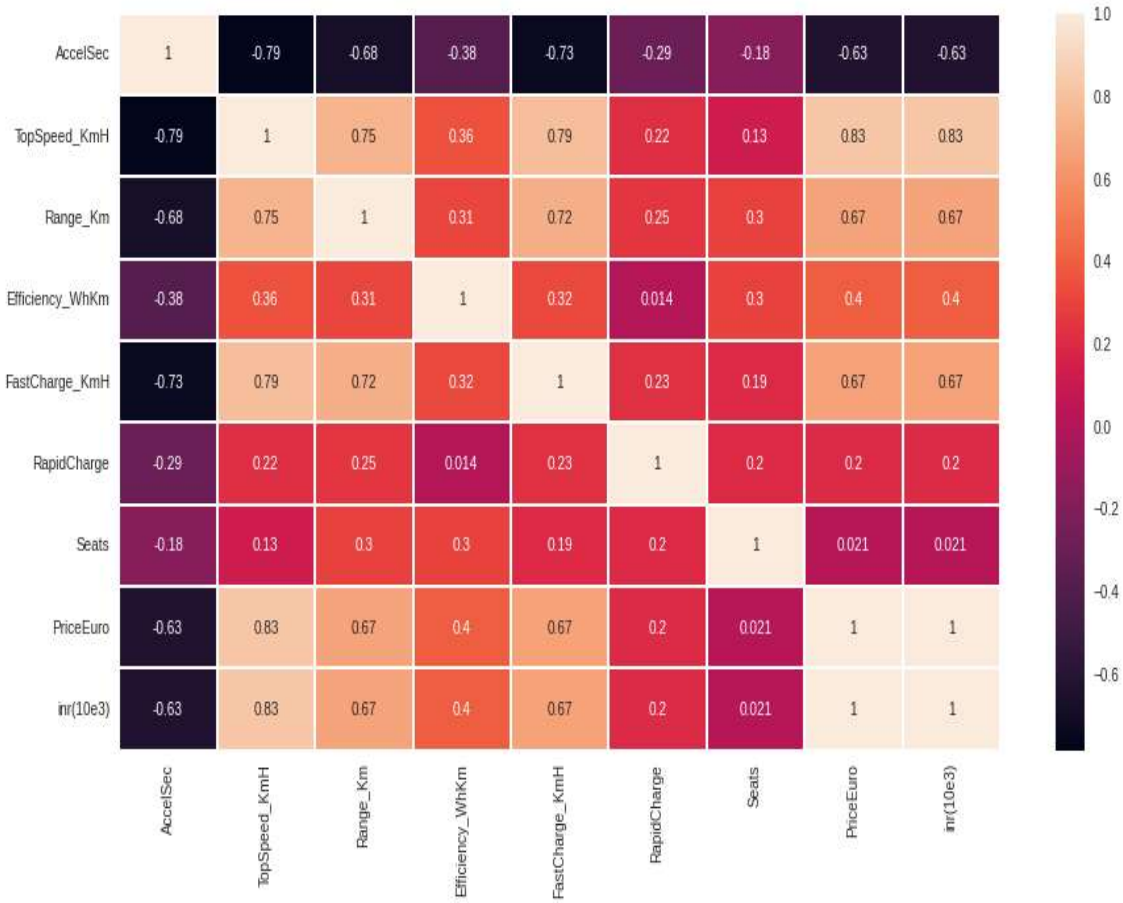
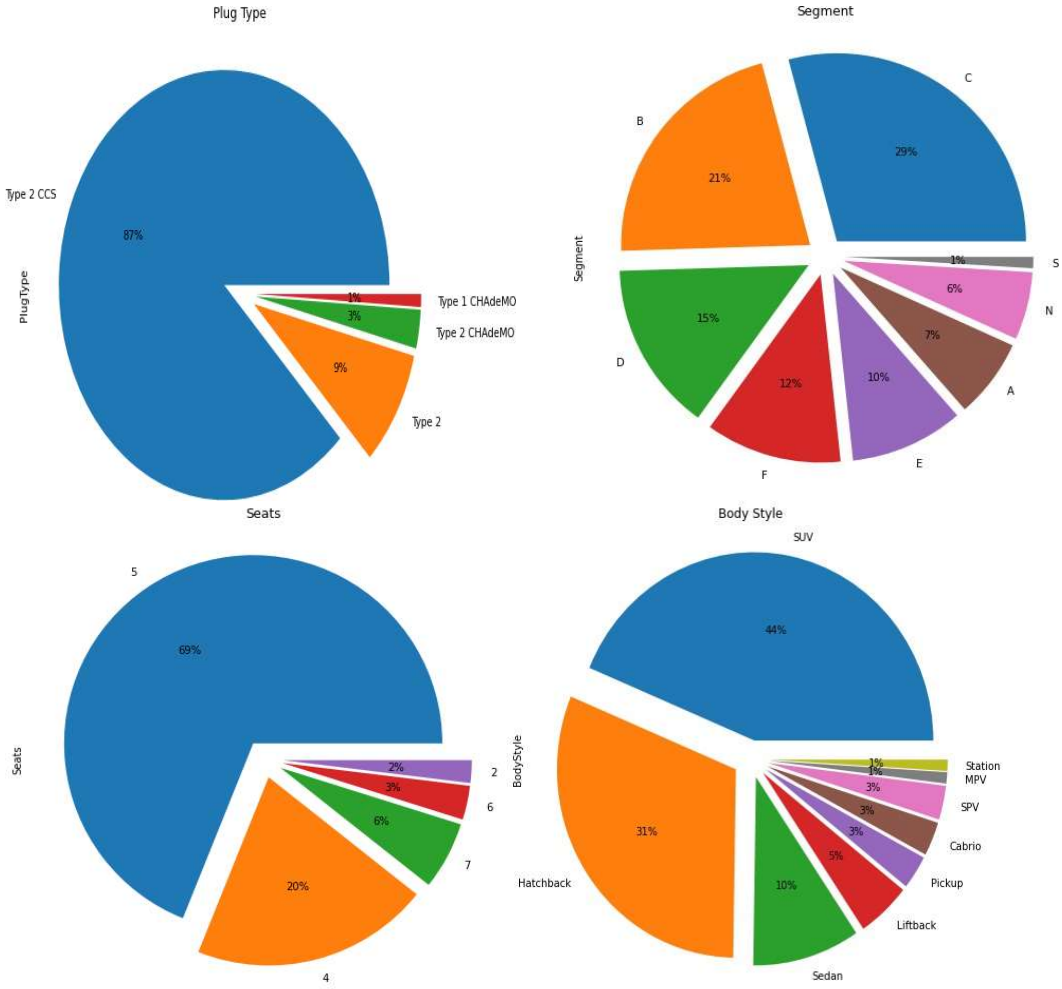


Figure 5: Correlation Matrix for the dataset



Now we can see that the requirements of what type of cars are most needed for cus- tomers and from the past 10 years there is a rapid growth of Electric vehicles usage in India

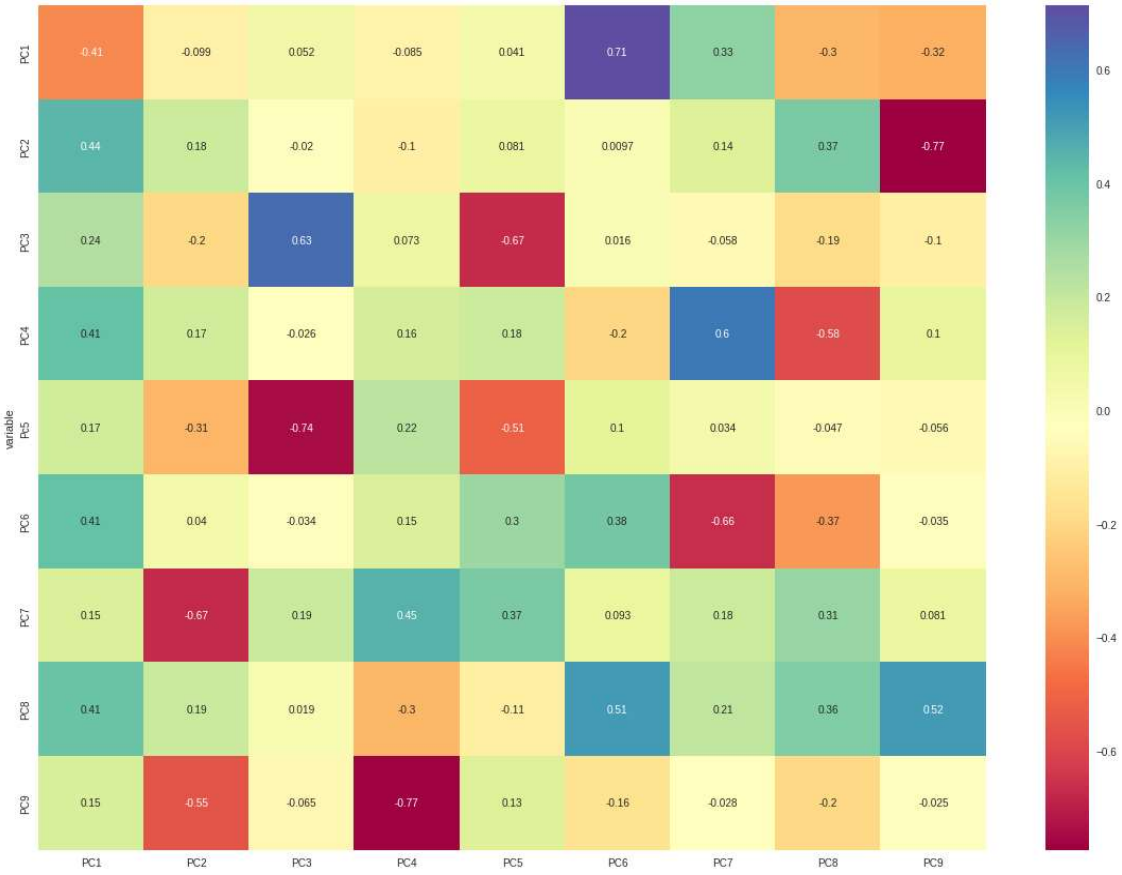


Figure 6: Correlation matrix plot for loadings

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 fac- tors 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 com- ponents 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 flat- tens out. The proportion of variance plot: The selected PCs should be able to describe at least 80% of the variance.

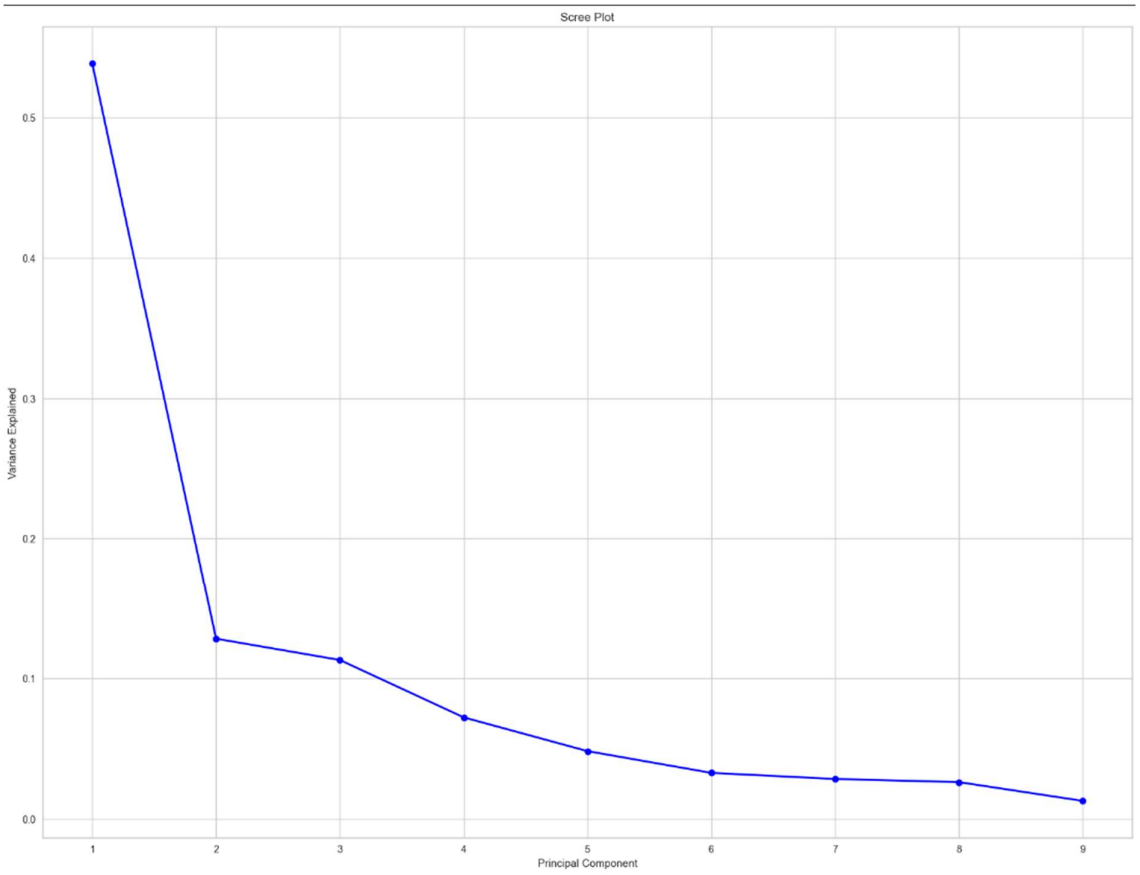


Figure 7: Scree Plot for our Dataset

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.

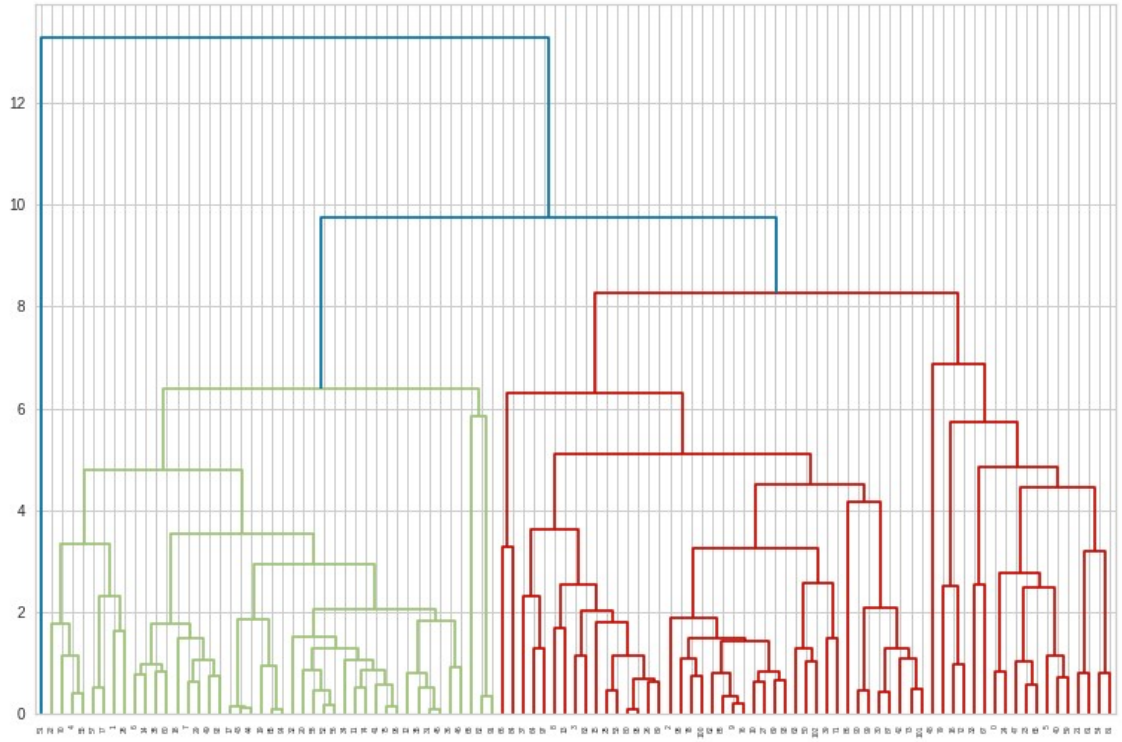


Figure 8: Dendrogram Plot for our Dataset

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 KElbowVisualizer function fits the KMeans 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.

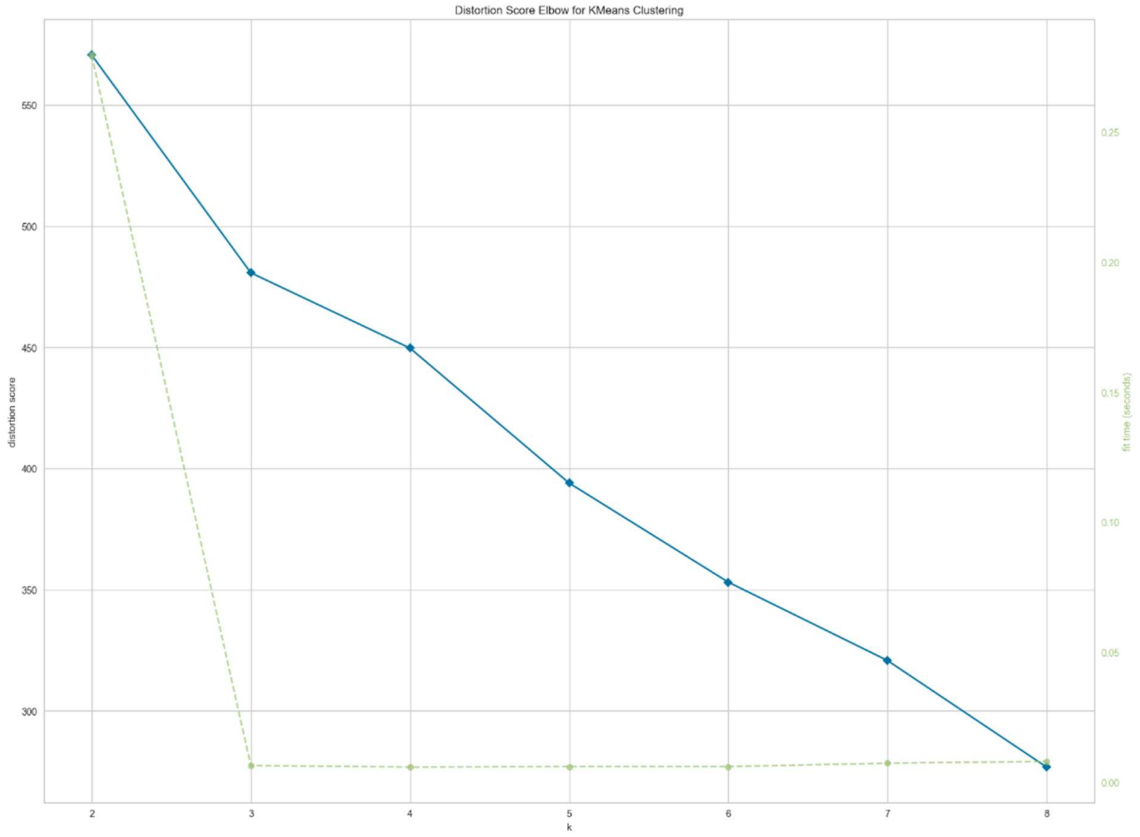


Figure 9: Evaluating the cluters using Distortion

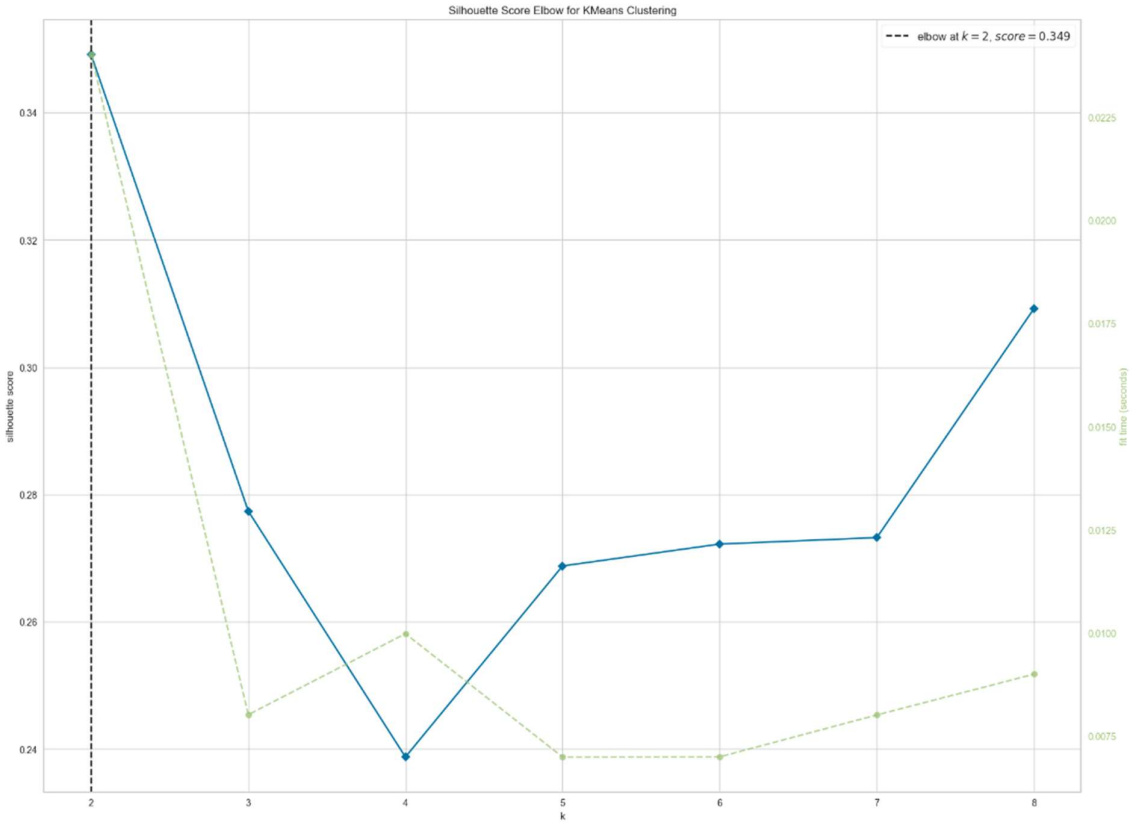


Figure 10: Evaluating the cluters using silhouette

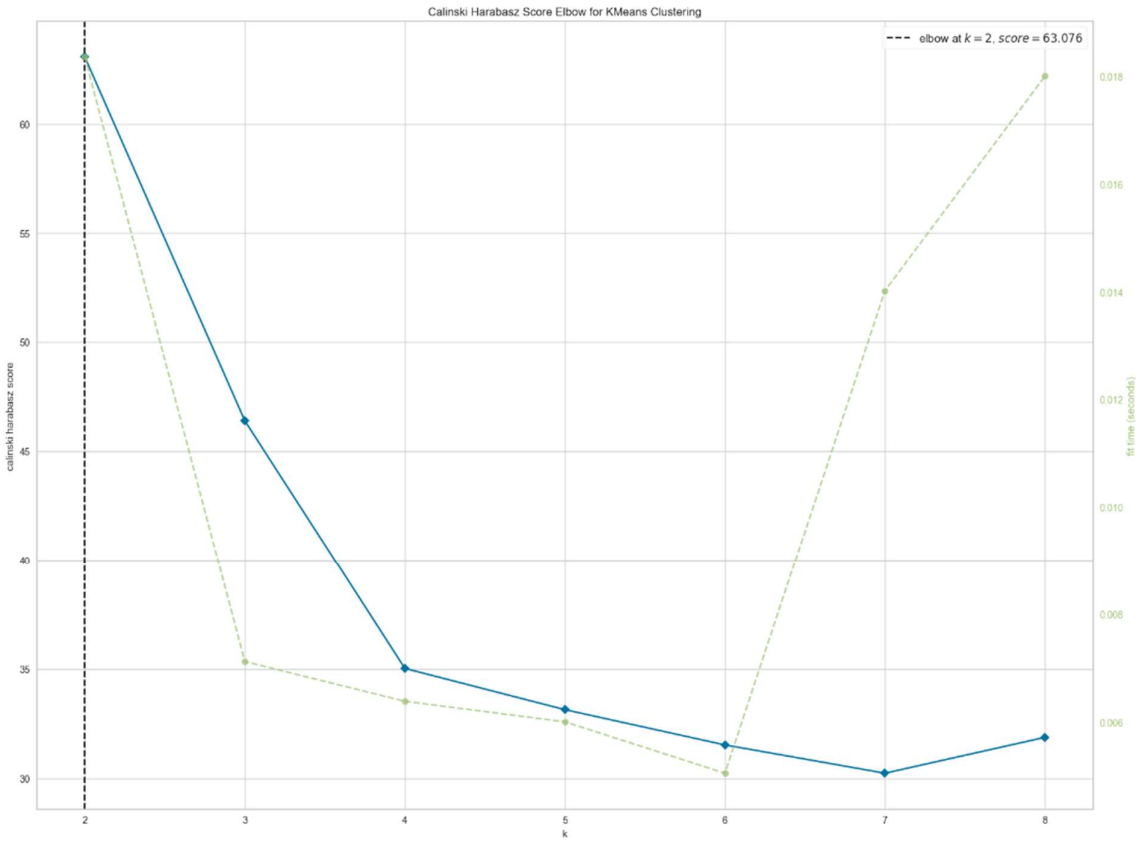


Figure 11: Evaluating the cluters using calinskiharabasz

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 iden- tifying 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 homo- geneous (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.

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 behaviors 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 KMeans model. The derived clusters are shown in the following figure

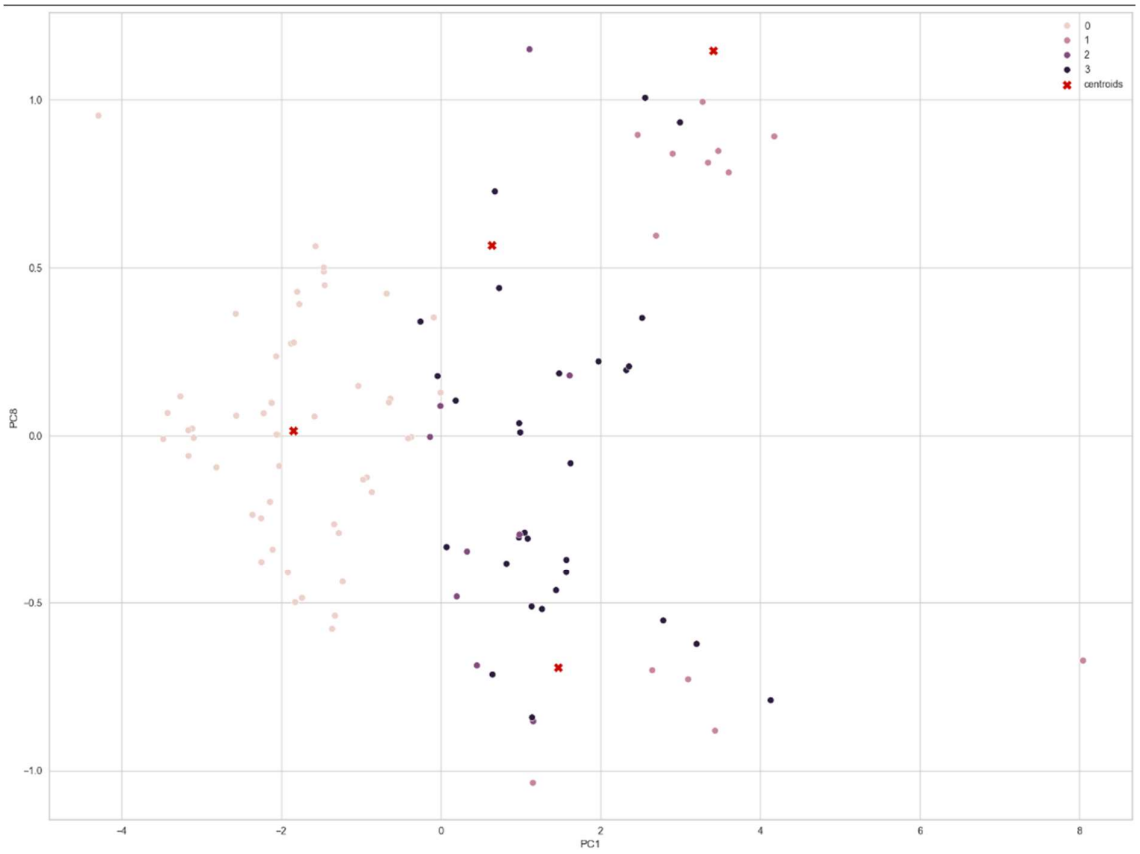
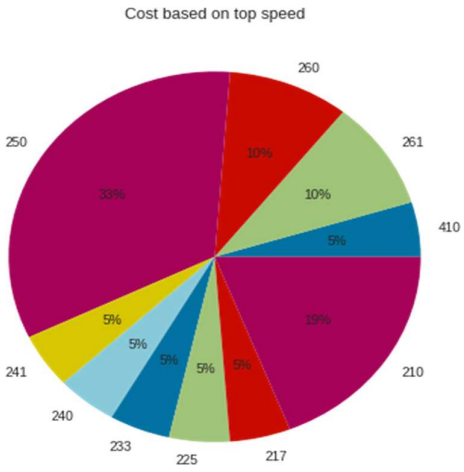


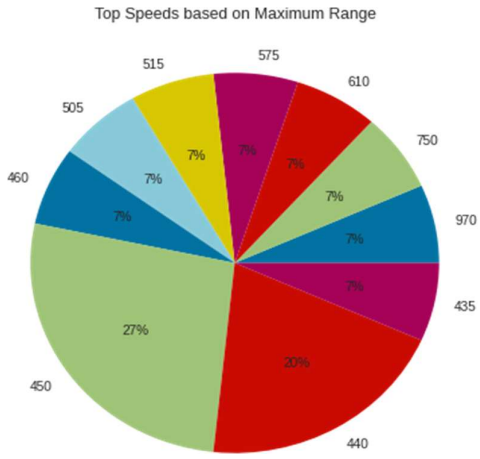
Figure 12: Segmented Clusters

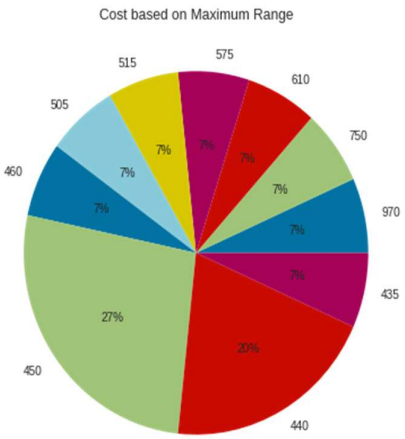
Profiling and Describing the Segments

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

Pie Chart:







Target Segments:

So from the analysis we can see that the optimum targeted segment should be belong- ing 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.