Electric Vehicle Market in India

Market Segmentation

Spandana

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', '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- makers to encourage EVs adoption in the backdrop of emerging sustainable transport market.

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

Electric Vehicles (EV) Market Segmentation by Product Type

The market is segregated by product into BEV, FHEV, MHEV (48V), PHEV, EREV, FCEV, and PFCEV segments. The BEV segment, which holds the largest market share is projected to capture more than 60% of the market by 2031. BEVs emit zero tailpipe emissions, rendering them a cleaner and more environmentally sustainable option than traditional internal combustion engine vehicles. To bolster their adoption, governments worldwide are implementing more stringent emissions regulations and providing incentives, accelerating the growth of BEVs.

As concerns about climate change and air pollution increase, governments worldwide are progressively enacting bans on sales of new internal combustion engine (ICE) vehicles. These bans, designed to mitigate greenhouse gas emissions and enhance air quality, are compelling automakers to expedite their shift towards electric mobility, propelling the demand for non-emission transportation options, such as BEV.

Electric Vehicles (EV) Market Segmentation by End-Use Type

Electric Vehicles are primarily offered in two end-use types, personal and commercial. The personal vehicle segment currently dominates the market, holding over 95% of the market share. Consumers are increasingly exhibiting awareness regarding the deteriorating air quality, particularly in urban areas, due to the emissions from ICE vehicles. Consequently, consumers are seeking greener transportation alternatives such as personal EVs.

Governments and local authorities are pivotal in encouraging the adoption of personal EVs among urban consumers. Numerous cities are enacting policies to stimulate EV adoption, including offering subsidies and tax incentives and establishing dedicated EV infrastructure. These initiatives not only reduce the cost of ownership but also improve the overall convenience of owning and operating an EV in urban regions.

Electric Vehicles (EV) Market Analysis by Region

Asia Pacific's EV market volume sales were the highest in 2023, capturing over 50% of the overall market size. China led the regional market, accounting for over 75% of the EV market in 2023. The dominance of the Chinese EV industry can be attributed to its strategic combination of investments, infrastructure development, and technological capabilities. Moreover, in June 2023, the Chinese government extended its tax exemption policy on new electric vehicles from its original expiration date in 2023 to 2027. This extension is aimed at boosting domestic sales and enhancing the appeal of foreign investment opportunities in the electric vehicles market.

The key economies in the region, such as China, Japan, South Korea, and India, have established ambitious goals to decrease carbon emissions. This commitment has resulted in the implementation of supportive policies and funding to accelerate the adoption of electric transportation in the region, contributing to the overall market growth momentum.

Leading Companies in the Electric Vehicles (EV) Market

- Toyota Motor Corp.
- BYD Auto Co. Ltd.
- Tesla Inc.
- Volkswagen AG
- Hyundai Motor Co.
- Geely Automobile Holdings Ltd.
- Mercedes-Benz Group AG
- Stellantis N.V.
- Bayerische Motoren Werke AG (BMW)
- SAIC Motor Corp. Ltd.

Data Collection:-

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

1. https://www.kaggle.com/datasets
2. https://data.gov.in/
3. https://www.data.gov/

Market Segmentation

Target Market:

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

Behavioral Segmentation: Behavioral segmentation is the process of sorting and grouping customers based on the behaviors.
Example: prior experience with the product, amount spent on the purchase, etc.

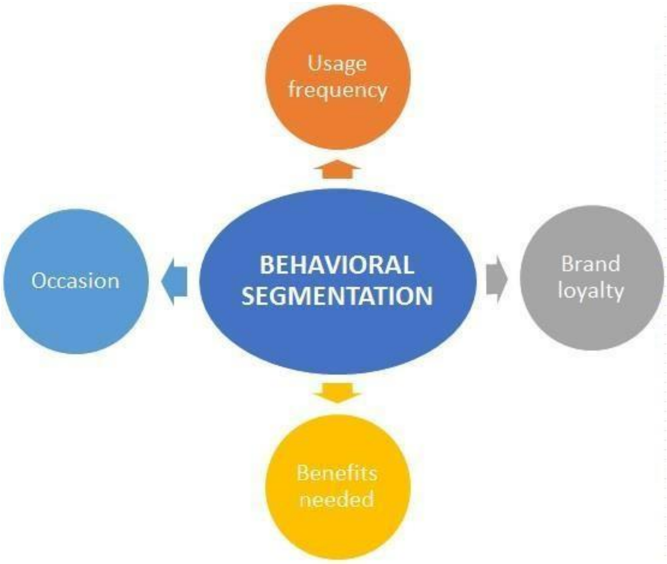


Fig 1. Behavioral Segmentation

Advantage:

1. It enables companies to accurately identify and target customers.
2. It helps organisations to change marketing campaigns based on customers' consistent behaviour or even erratic behaviour.
3. It helps establish brand loyalty further, built upon customers who have shown a strong affinity towards the brand.

Disadvantage: 1.not always readily available.

2. Behavioural segmentation is mainly performed with qualitative data rather than quantitative data. Thus, making forecasts, budgets, expenses all depends on specific assumptions and parameters.

Psychographic Segmentation: Psychographic segmentation is a market research method used to divide a market or customer group into segments based on their beliefs, values, lifestyle, social status, activities, interests and opinions and other psychological criteria.

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

Disadvantage: increased complexity of determining segment memberships for consumers.

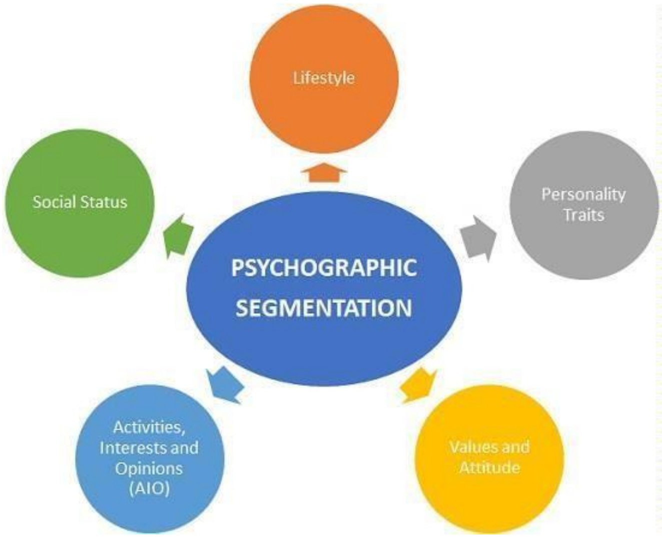


Fig 2: Psychographic Segmentation

Socio-Demographic Segmentation:

Demographic segmentation is a precise form of audience identification based on data points like age, gender, marital status, family size, income, education, race, occupation, nationality, and/or religion. It's among the four main types of marketing segmentation and perhaps the most commonly used method.

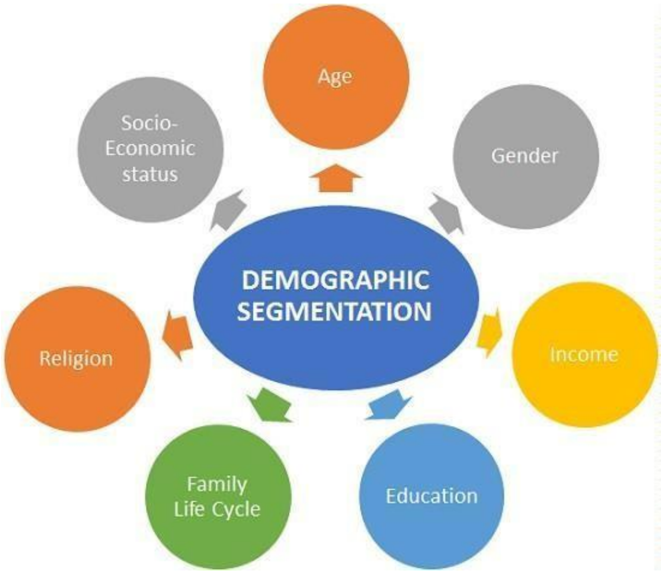


Fig3: Demographic Segmentation

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.

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 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.

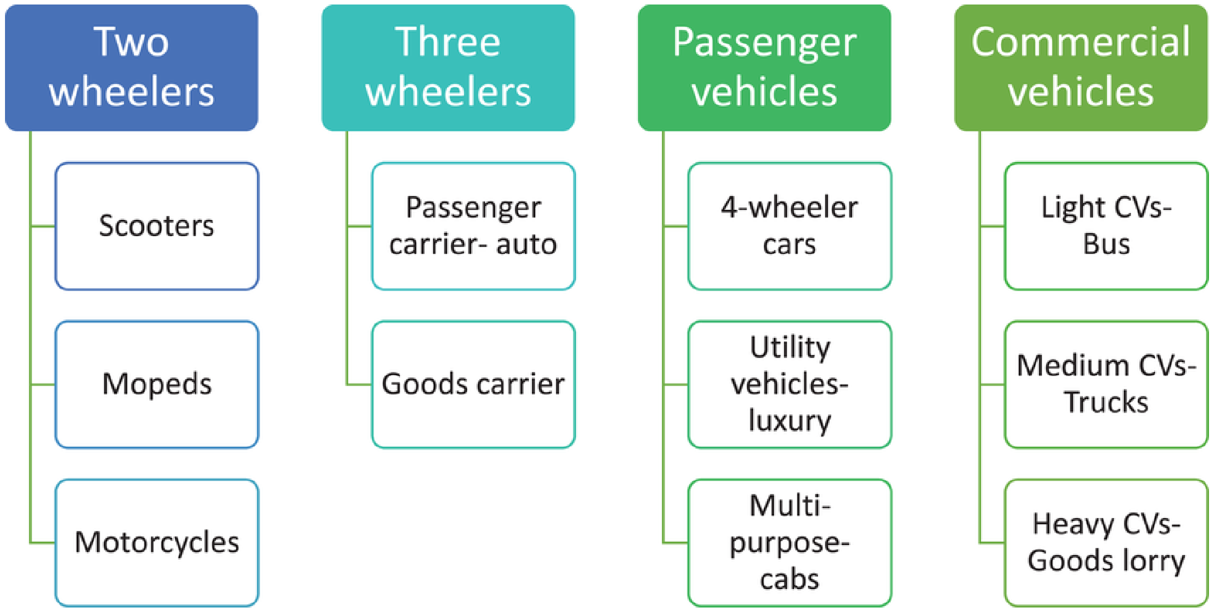


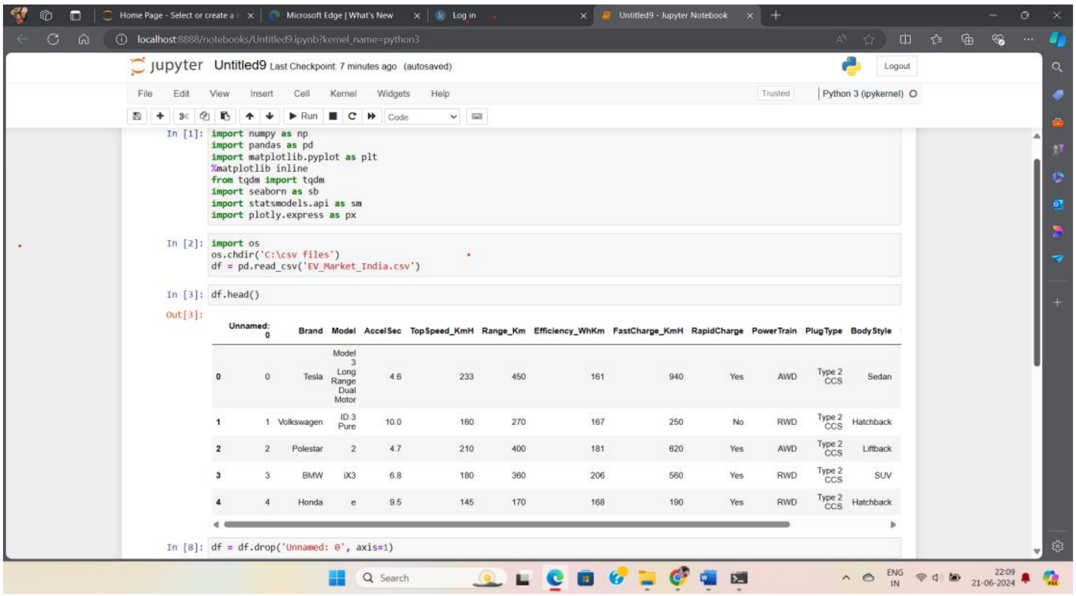
Fig 4 : Electric vehicle market segmentation

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.

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.

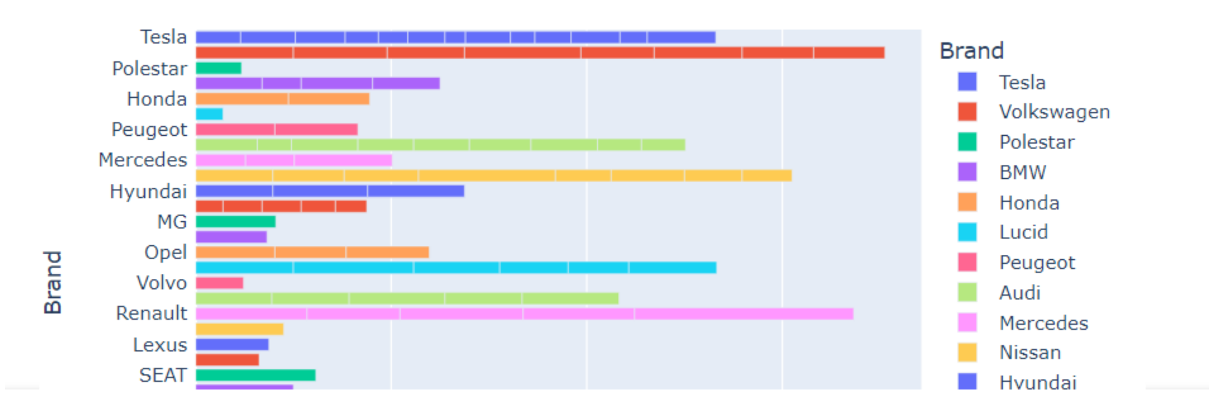


EDA

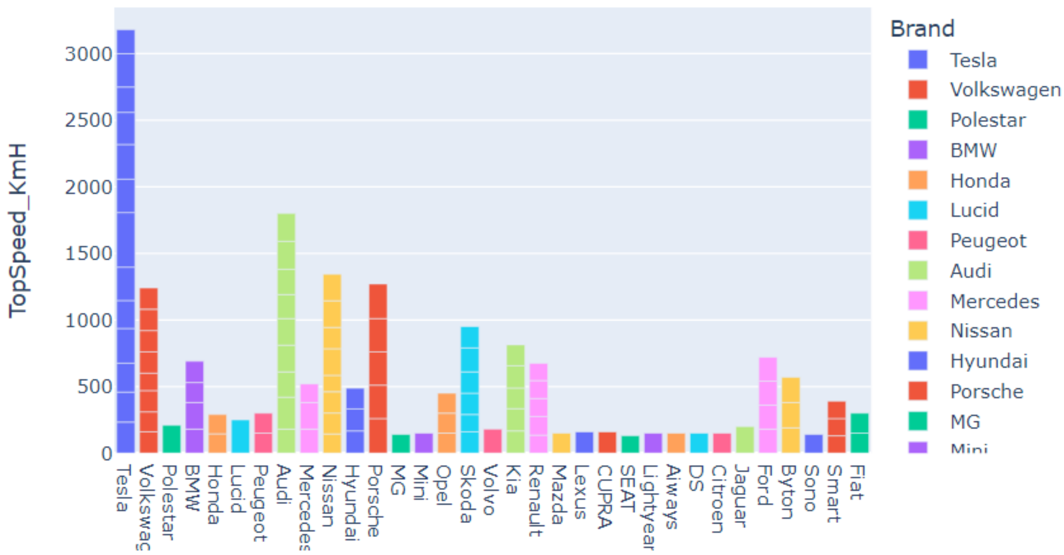
We start the Exploratory Data Analysis with some data Analysis drawn from the data without Principal Component Analysis and with some Principal Component Analysis in the dataset obtained from the combination of all the data we have. PCA is a statis- tical process that converts the observations of correlated 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.

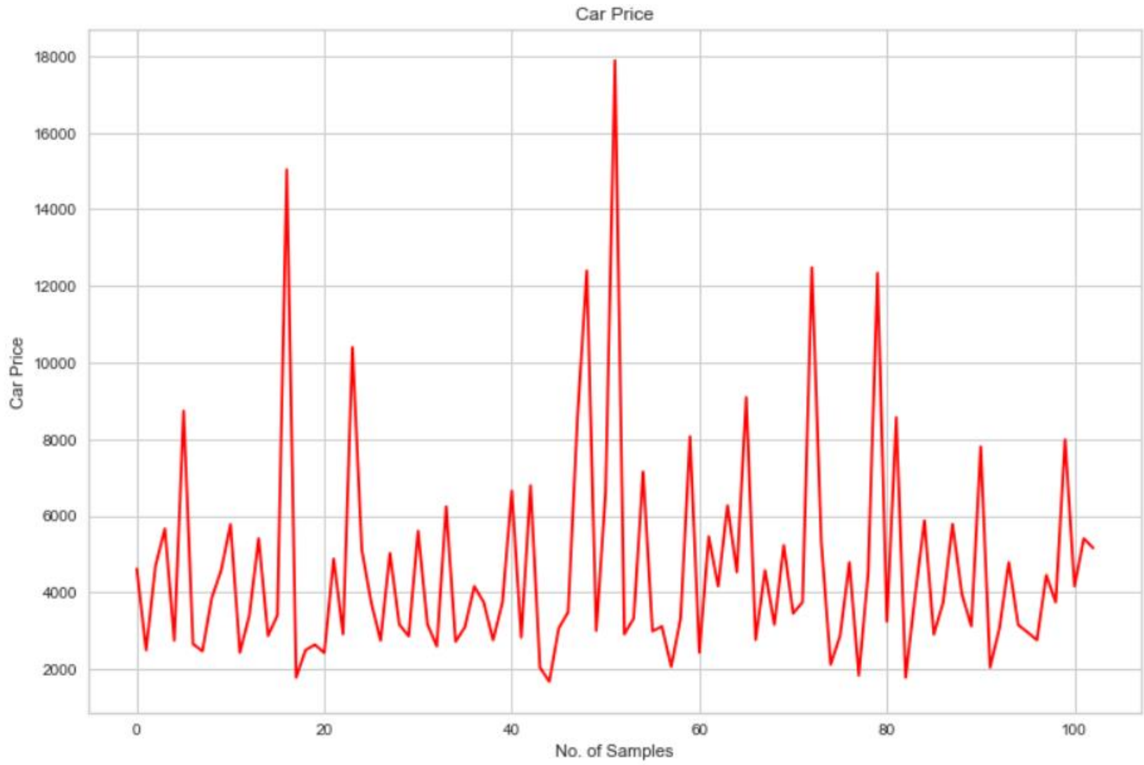
Comparision of cars in our data

Which car has fastest accelaration?



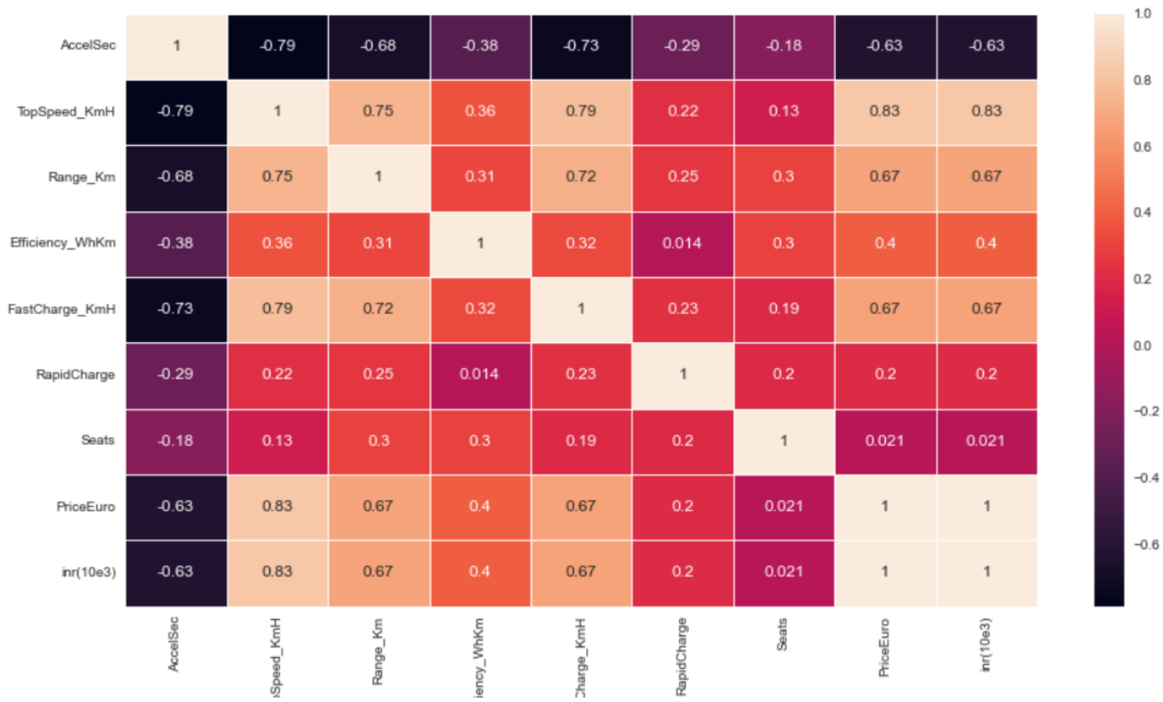
Which Car Has a Top speed?

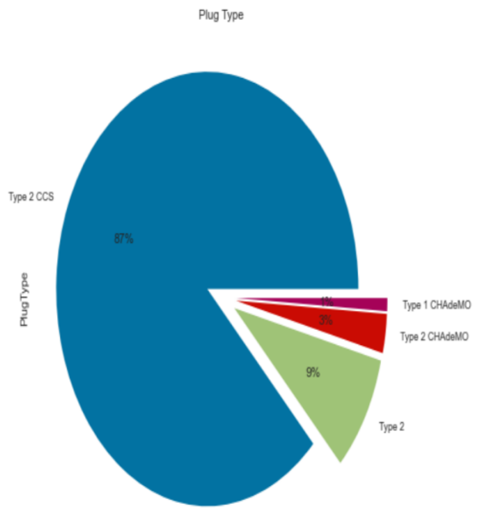


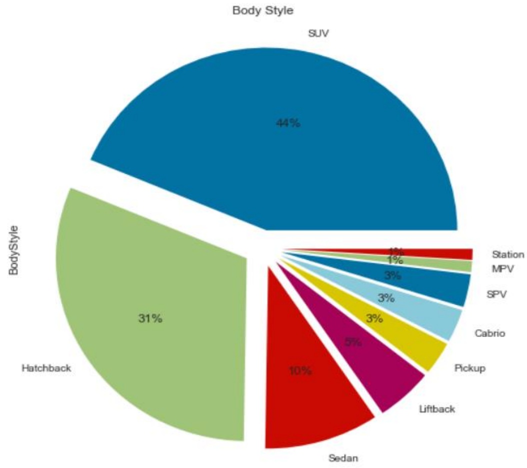


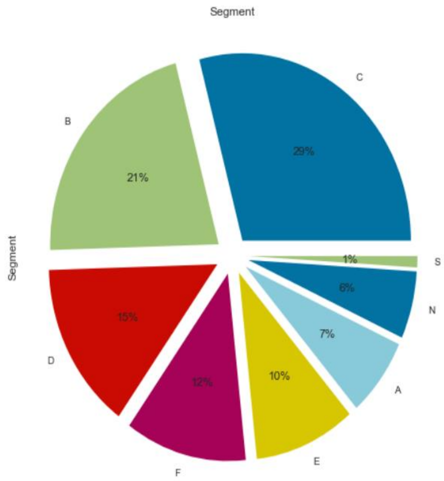
Correlation Matrix: A correlation matrix is simply a table that displays the correlation. It is best used in variables that demonstrate a linear relationship between each other. Coefficients for different variables. The matrix depicts the correlation be- tween all the possible pairs of values through 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.

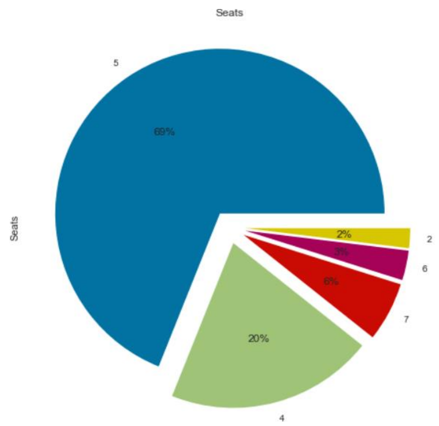
<AxesSubplot:>



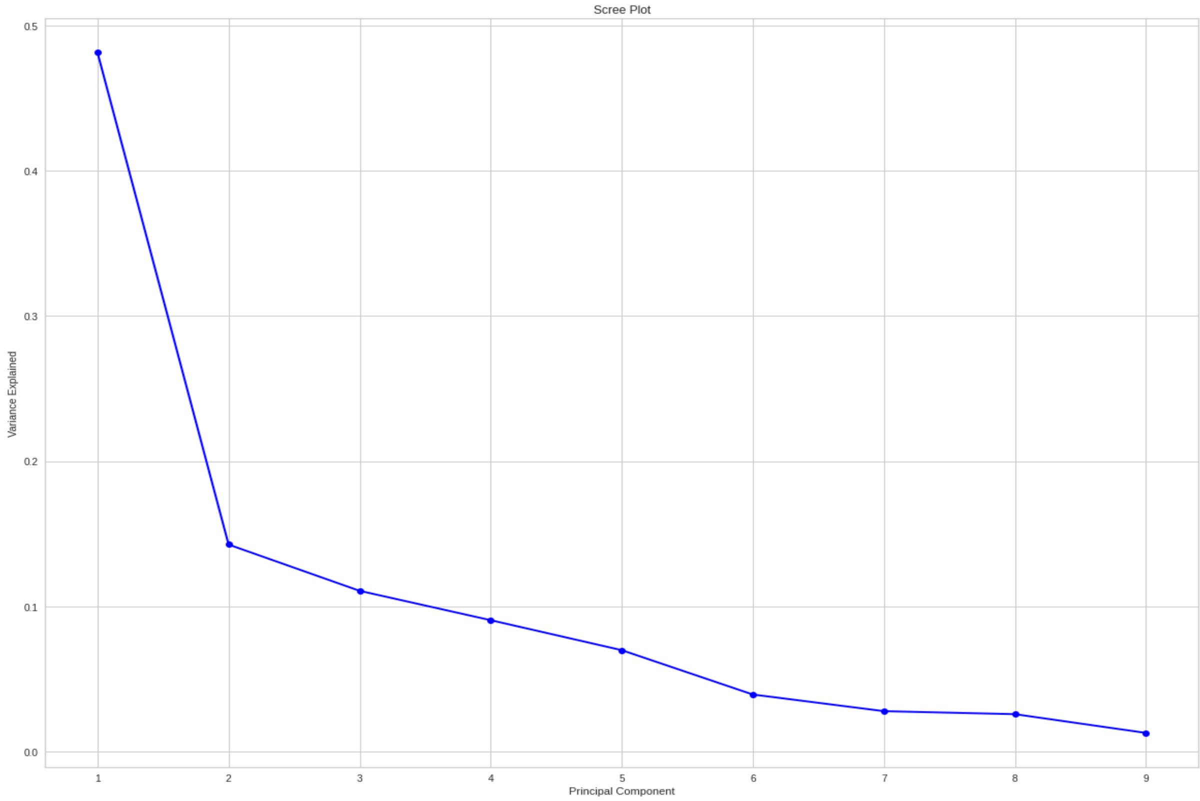








Scree Plot: is a common method for determining the number of PCs to be retained via graphical representation. It is a simple line segment plot that shows the eigenvalues for each individual PC. It shows the eigenvalues on the y-axis and the number of fac- tors on the x-axis. It always displays a downward 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.

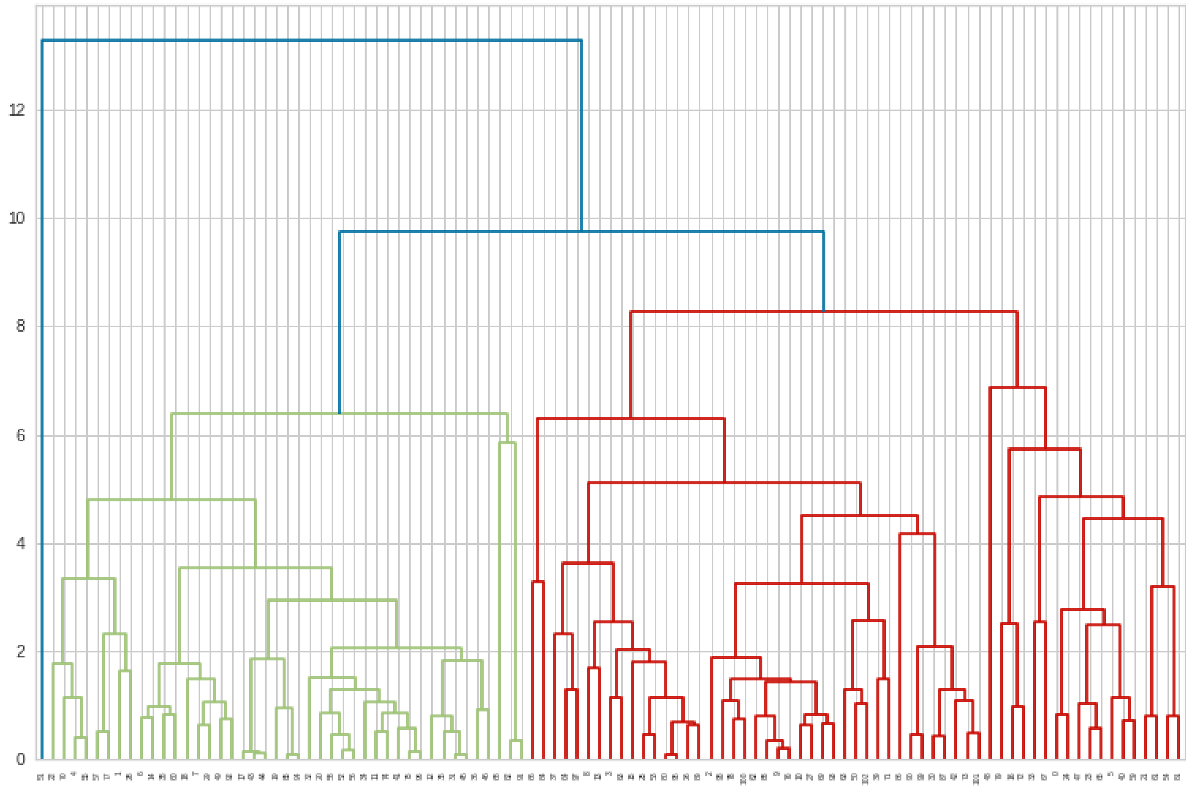


Scree plot

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 choose 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.

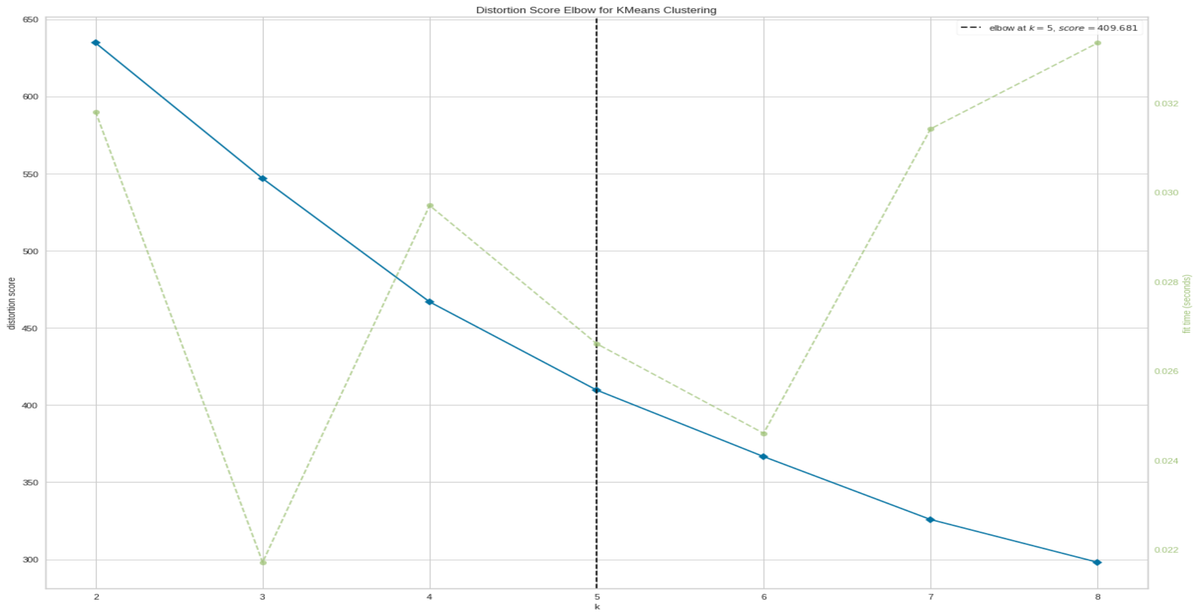


: *Dendrogram Plot for our Dataset*

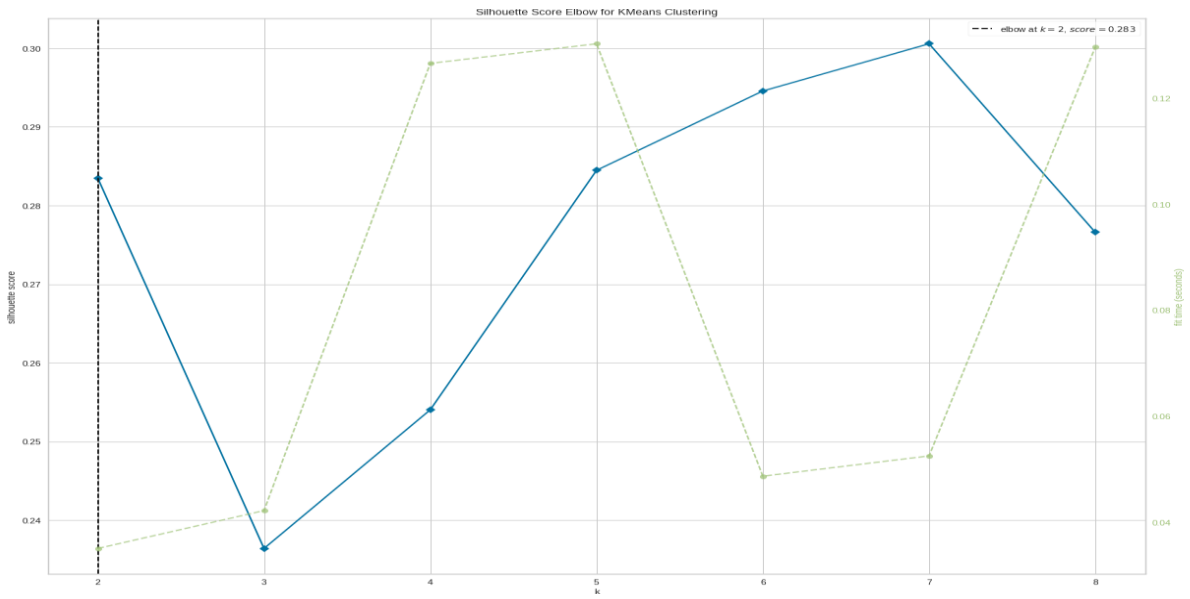
Elbow Method

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.

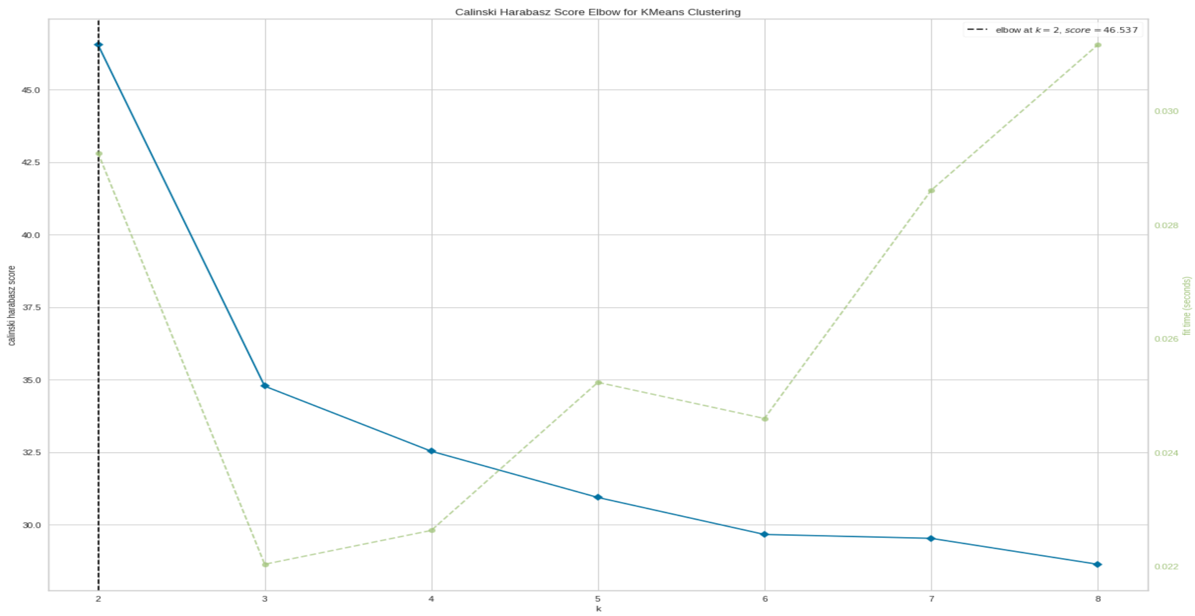
The KElbowVisualizer function fits the KMeans model for a range of clusters values between 2 to 8. As shown in Figure, the elbow point is achieved which is highlighted by the function itself. The function also informs us about how much time was needed to plot models for various numbers of clusters through the green line.



Evaluating the clusters using Distortion



Evaluating the clusters using silhouette



Evaluating the clusters

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.

<u>K-Means Algorithm</u>

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:

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

And M-step is :

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

$$\Rightarrow \mu_k = \frac{\sum_{i=1}^{m} w_{ik} x^i}{\sum_{i=1}^{m} w_{ik}}$$

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:

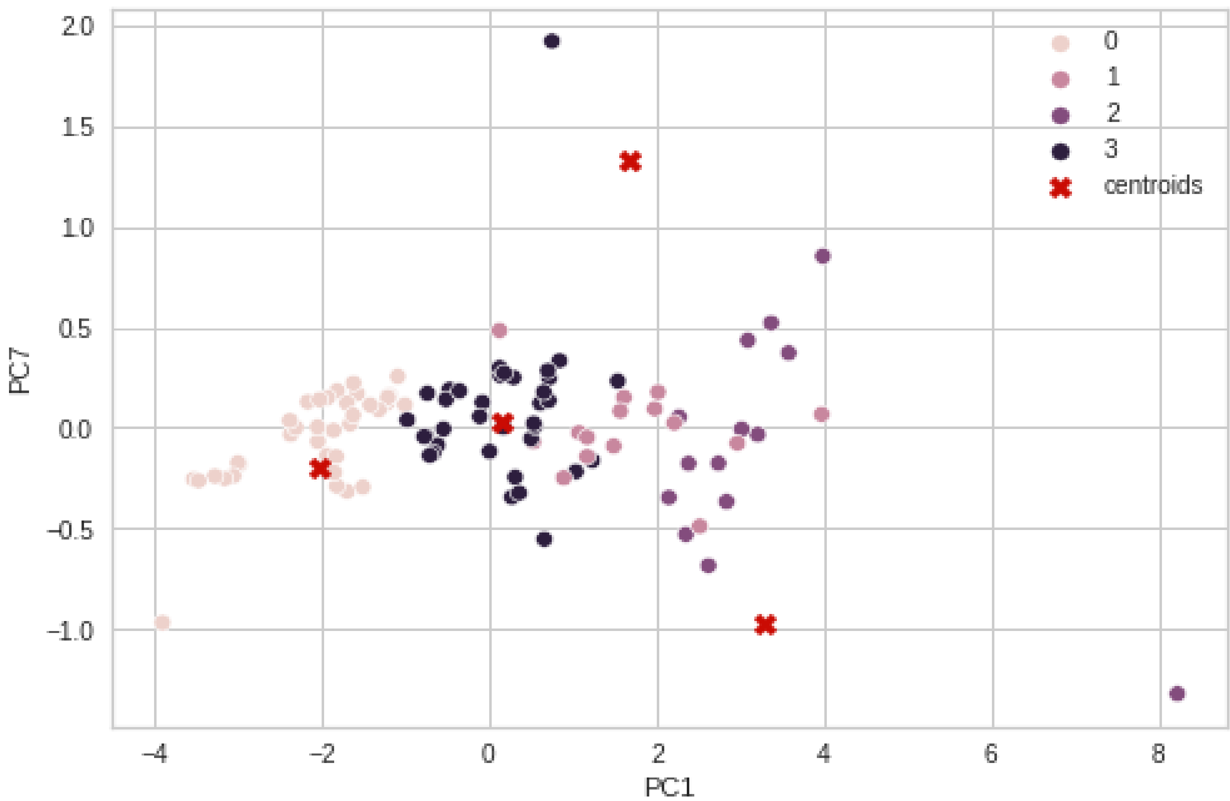
1. Get a meaningful intuition of the structure of the data we're dealing with.
2. Cluster-then-predict where different models will be built for different subgroups if we believe there is a wide variation in the behaviors of different subgroups.

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

```
#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 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.

Linear Regression().fit($X_{train,ytrain}$) 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.

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

```
]: 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)
```

```
]: print(lm.intercept_)

   4643.522050485437
```

```
]: lm.coef_

]: array([ 1101.58721,  -741.20904,   208.53617,   508.32246,   122.3533 ,
           1579.00686,   333.61147, -1079.99512,  1461.72269])
```

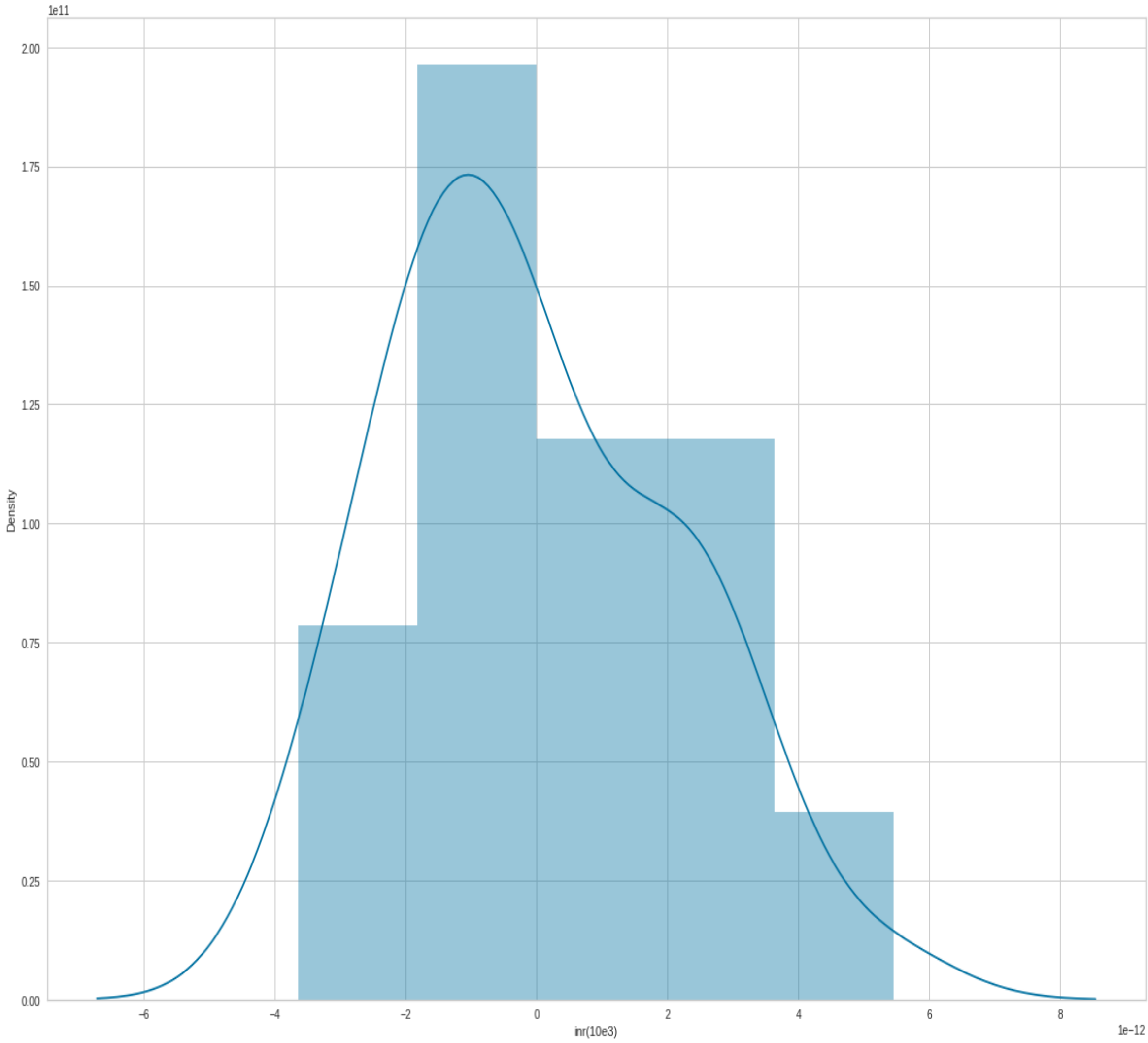
```
]: X_train.columns

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

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

]:

	Coeff
PC1	1101.5872
PC2	-741.2090
PC3	208.5362

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:

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

```
MAE: 1.5049971850766312e-12
MSE: 4.908923039972432e-24
RMSE: 2.215608954660644e-12
```

```python
metrics.mean_absolute_error(y_test,predictions)
```

```
1.5049971850766312e-12
```

```python
metrics.mean_squared_error(y_test,predictions)
```

```
4.908923039972432e-24
```

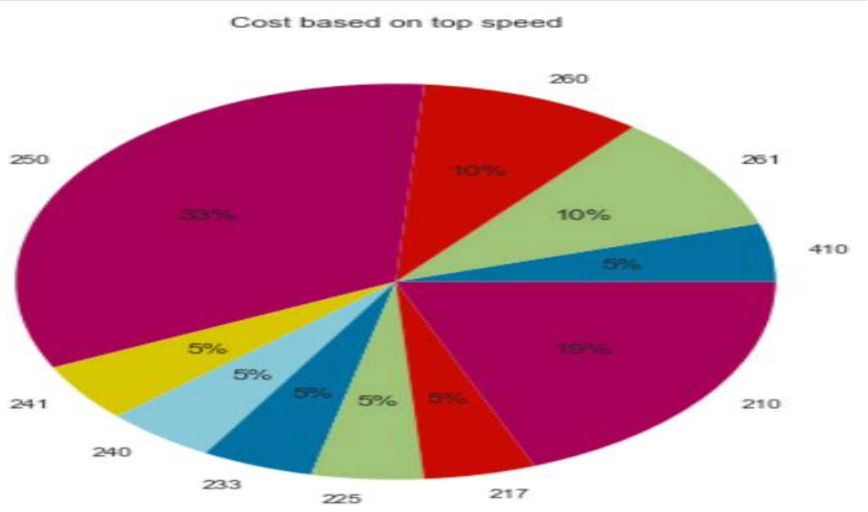
```python
np.sqrt(metrics.mean_squared_error(y_test,predictions))
```

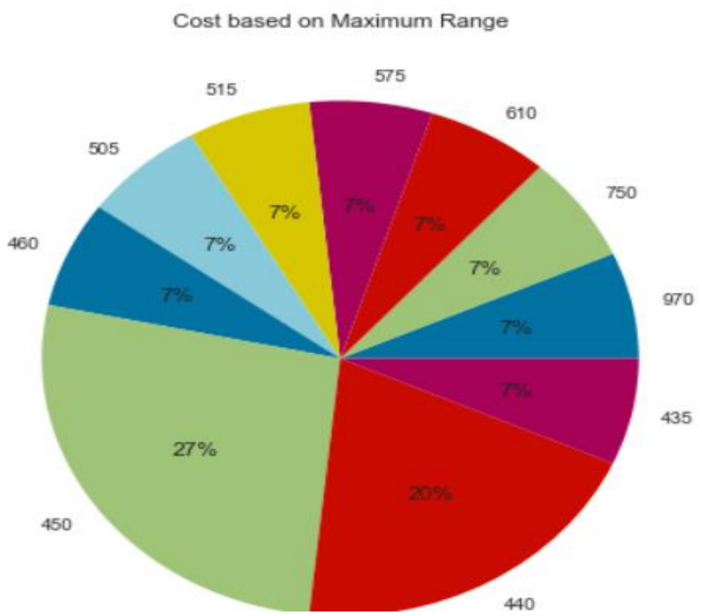
```
2.215608954660644e-12
```

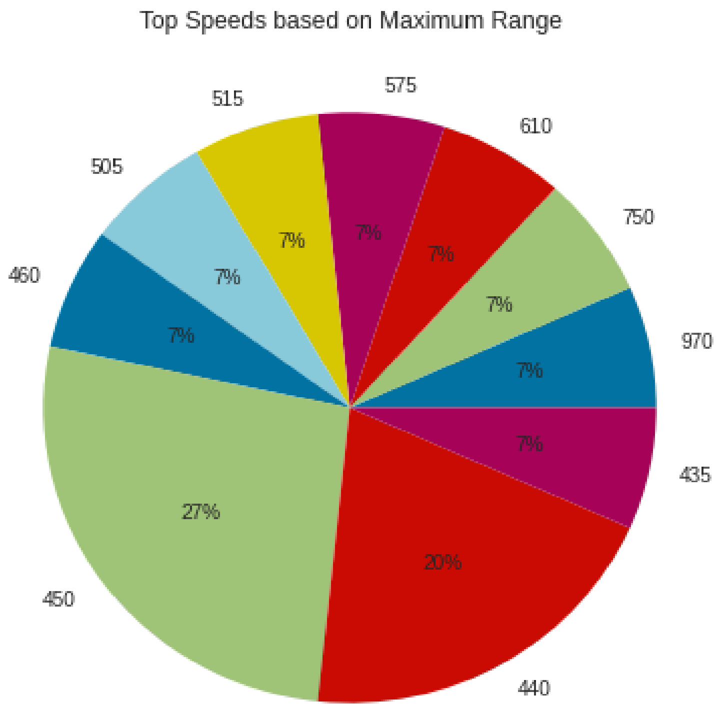
Profiling and Describing the Segments

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

Pie Chart:







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**.

Customizing the Marketing Mix

In our electric vehicle market strategy, customizing the marketing mix is paramount for appealing to Segment 1 and Segment 2, our identified target segments. For Product Customization, we plan to enhance features tailored to the specific desires of each segment. Addressing dissatisfaction points, such as improving performance and service experience for Segment 1, and emphasizing visual appeal and value for money for Segment 2, is central to

product refinement. Diverse offerings within each segment ensure a broad spectrum of choices, aligning with varied tastes and budgets. Price Customization involves setting competitive and flexible pricing structures. Segment 1 will benefit from affordable options, while Segment 2 might accept a slightly higher price point for value-added features. Promotion Customization demands targeted advertising,

focusing on reliability and service improvements for Segment 1, and aesthetics and affordability for Segment 2. Tailored promotional events and online campaigns further engage these segments effectively. In terms of Place Customization, we'll establish accessible distribution channels in urban areas for Segment 1 and in suburban and semi-urban regions for Segment 2. Strengthening our online presence ensures seamless online purchasing experiences, emphasizing virtual showrooms and customer support platforms. Additionally, People and Process Customization involves training customer service representatives to address segment-specific concerns empathetically. Efficient processes, streamlined for customization requests and service appointments, enhance customer satisfaction and brand loyalty. This tailored approach ensures our electric vehicles resonate with the distinct needs of Segment 1 and Segment 2, fostering market relevance and customer preference.

Potential Early Market Customer Base

In the analysis of the potential early market customer base, two primary segments emerge: Segment 1, encompassing 330 members (39% of consumers), and Segment 2, comprising 277 members (33% of consumers). Analyzing the price range data, the logical target price for Segment 1 falls between ₹51,094 and ₹1,67,844, while for Segment 2, it ranges from ₹51,094 to ₹1,37,890.

Calculating the potential sales (profit) in this early market scenario involves multiplying the number of potential customers in each segment by our targeted price range. For instance, if our target price for Segment 1 is set at ₹1,20,000 the potential profit from this segment alone would amount to ₹39.60 crores. Similarly, for Segment 2 with a target price of ₹1,10,000, the potential profit would be ₹30.47 crores.

Segment 1 demonstrates the larger potential, with a significantly higher market share and a broader customer base, making it a primary focus for our early market penetration efforts.

These calculated potential profits underscore the substantial market opportunity within these segments, guiding our strategic decisions effectively.

Most Optimal Market Segments

In the context of selecting the most optimal market segment for our electric two-wheeler vehicles, thorough analysis and evaluation have pointed to Segment 1 as the ideal choice. Representing 39% of consumers, this segment boasts significant opportunities and a large customer base, making it a strategic target for market penetration. Its substantial market potential, coupled with its balanced blend of technical specifications and price range,
positions it as the most promising market segment for our electric vehicles.

The recommended technical specification range for Segment 1, presented in Table 10.1, ensuring alignment with the diverse needs and preferences of the market:

Table 10.1 Technical specification of electric vehicle two-wheeler for segment 1

Specification	Recommended Range (in INR)
Price	70,688 – 1,29,063
Riding range	89 - 180 km

Top speed	58 - 116 kmph
Weight	76 - 120 kg
Battery charging time	3 - 5 hours
Rated power	1200 - 5500 W

This comprehensive analysis ensures our market entry strategy is finely tuned to cater to the demands and expectations of the chosen segment, setting the stage for a successful and sustainable venture into the electric vehicle market.

References

[1] Deepak Jaiswal,Arun Kumar Deshmukh(2022) *Who will adopt electric vehicles? Segmenting and exemplifying potential buyer heterogeneity and forthcoming re- search*, Journal of Retailing and Consumer Services
.

[2] Dolnicar, S., Grun Bettina, amp; Leisch, F. (2019). *Market segmentation analysis understanding it, doing it and making it useful*. Springer Nature.

[3] McDonald, M., amp; Dunbar, I. (2003). /emphMarket segmentation. Butterworth- Heinemann

[4] Society of Manufactures of Electric Vehicles, (2023, October), https://www.smev.in/ev- industry

[5] S. Dolnicar, B. Grun, F. Leisch, Market Segmentation Analysis, Springer Open, Management for Professionals, 2018