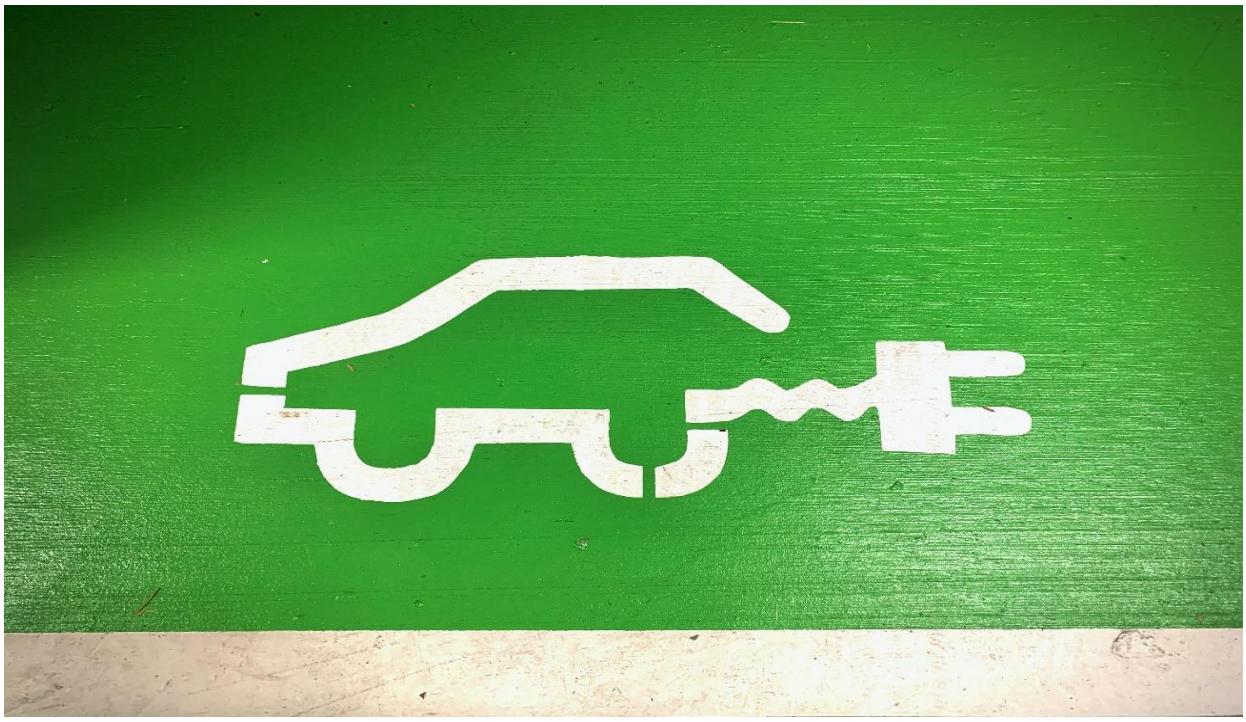
Market Segmentation Analysis of Electric Vehicles Market in India:



Introduction :

Electric vehicles (EVs) have been gaining popularity in India in recent years due to a number of factors, including government policies, rising fuel prices, and increasing environmental awareness among consumers. The Indian government has set an ambitious goal of having 30% of all vehicles on the road be electric by 2030, which has led to a number of incentives and initiatives to promote EV adoption.

The Task is to analyse the Electric Vehicles Market in India using Segmentation analysis which involves dividing the market into distinct groups based on certain characteristics. One way to segment the market is by vehicle type, which includes cars, two-wheelers, and commercial vehicles.

In this report we analyse the Electric Vehicles Market in India using segments which includes Descriptive Statistics of the dataset and analysis that include questions such as:

I. Which Car Has a Top speed?

II. Which car has fastest acceleration ? etc.

Data collection:

Segmenting:

Image segmentation can be used as a technique for customer segmentation, although it is not a commonly used method in the field of marketing. Typically, customer segmentation is done based on demographic, geographic, behavioural and psychographic characteristics of customers. However, image segmentation can be used in conjunction with other segmentation techniques to gain a deeper understanding of customer behaviour.

For example, image segmentation can be used to analyse images of customers' social media profiles or their interactions with a company's products or services. By segmenting these images into different regions, such as facial features or objects being used, it may be possible to identify patterns and preferences among different customer groups. This information can then be used to tailor marketing strategies and product offerings to specific customer segments.

Implementation:

Data Preprocessing

Data Cleaning:

The process of data collection involves gathering information from various sources and organizing it in a meaningful way. In this case, the collected data is compact and serves two main purposes: visualization and clustering. To achieve

these objectives, Python libraries such as NumPy, Pandas, Scikit -Learn, and SciPy are used in the workflow. These libraries provide essential tools for data manipulation, analysis, and visualization. Additionally, the results obtained from the workflow are ensured to be reproducible, which means that they can be replicated with the same level of accuracy and precision in the future. This ensures the reliability and credibility of the data analysis process.

In [ ]:
```
import pandas as pd
data = pd.read_csv('/content/drive/MyDrive/indian-auto-mpg.csv')
```

In [ ]:
```
data.head()
```

Out[ ]:

	Unnamed: 0	Name	Manufacturer	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Engine CC	Power	Seats	Mileage Km/L	Price
0	0	Maruti Wagon R LXI CNG	Maruti	0	2010	72000	CNG	Manual	First	998	58.16	5	26.60	1.75
1	1	Hyundai Creta 1.6 CRDi SX Option	Hyundai	1	2015	41000	Diesel	Manual	First	1582	126.20	5	19.67	12.50
2	2	Honda Jazz V	Honda	2	2011	46000	Petrol	Manual	First	1199	88.70	5	18.20	4.50
3	3	Maruti Ertiga VDI	Maruti	2	2012	87000	Diesel	Manual	First	1248	88.76	7	20.77	6.00
4	4	Audi A4 New 2.0 TDI Multitronic	Audi	Coimbatore	2013	40670	Diesel	Automatic	Second	1968	140.80	5	15.20	17.74

In [ ]:
```
data.describe()
```

Out[ ]:

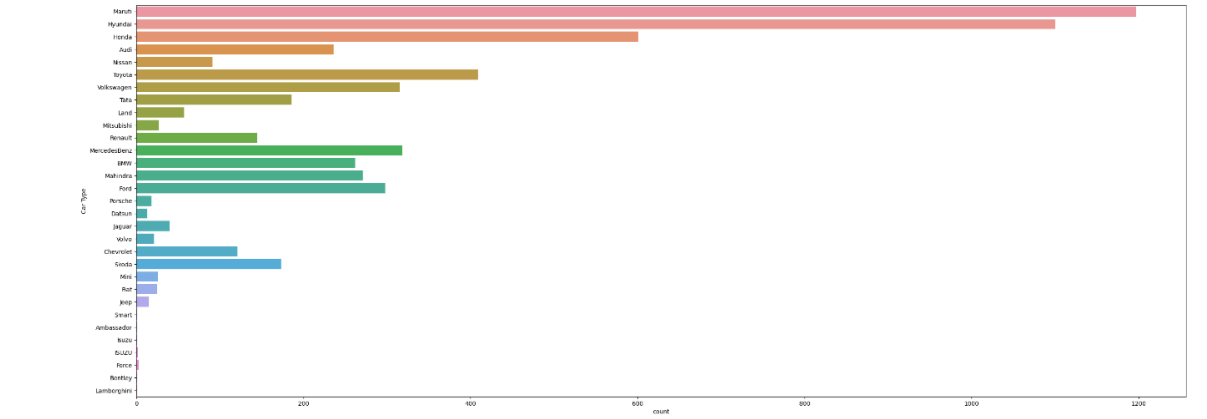
	Unnamed: 0	Year	Kilometers_Driven	Engine CC	Power	Seats	Mileage Km/L	Price
count	5975.00000	5975.000000	5.975000e+03	5975.000000	5975.000000	5975.000000	5975.000000	5975.000000
mean	3008.80887	2013.386778	5.867431e+04	1621.606695	112.599819	5.278828	18.179408	9.501647
std	1739.30056	3.247238	9.155851e+04	601.036987	53.659495	0.808959	4.521801	11.205736
min	0.00000	1998.000000	1.710000e+02	624.000000	34.200000	0.000000	0.000000	0.440000
25%	1502.50000	2012.000000	3.390800e+04	1198.000000	74.000000	5.000000	15.200000	3.500000
50%	3010.00000	2014.000000	5.300000e+04	1493.000000	92.700000	5.000000	18.160000	5.650000

Exploratory Data Analysis:

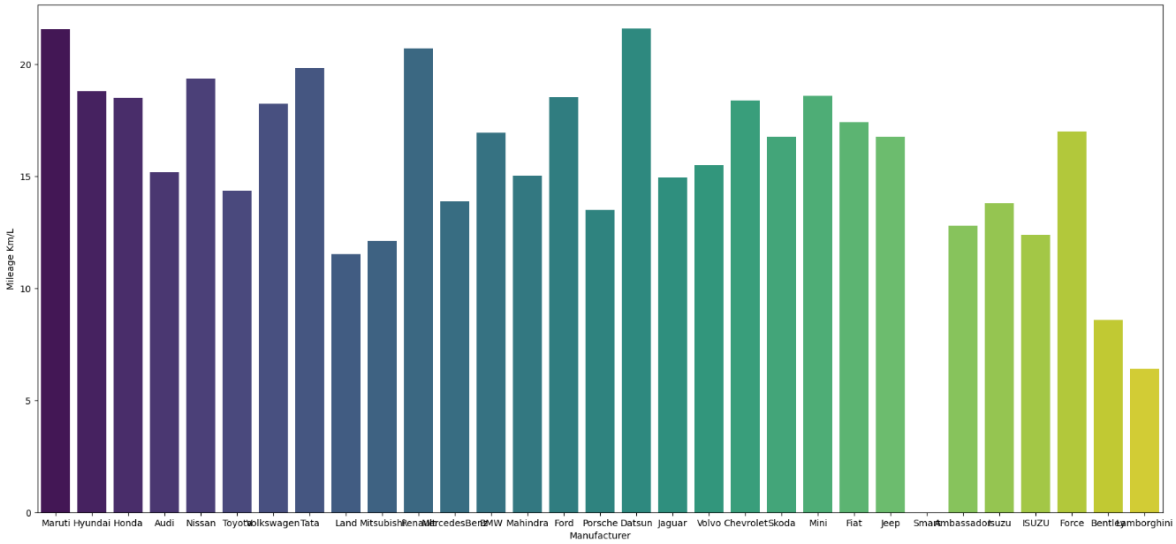
It mainly focuses on the exploratory data analysis of a dataset using principal component analysis (PCA). The dataset consists of correlated features, and PCA is used to transform the data into linearly uncorrelated features. This allows for the reduction of dimensionality, which can make classification, regression, or any other form of machine learning more cost-effective. The paper starts by analysing the data without PCA and then proceeds to use PCA to obtain the principal components. The paper concludes that PCA is an effective method for exploratory data analysis as it allows for a better understanding of the dataset and its underlying structure. Additionally, it provides a useful tool for data pre-processing, which can improve the performance of machine learning models.

Comparison of cars in our data:

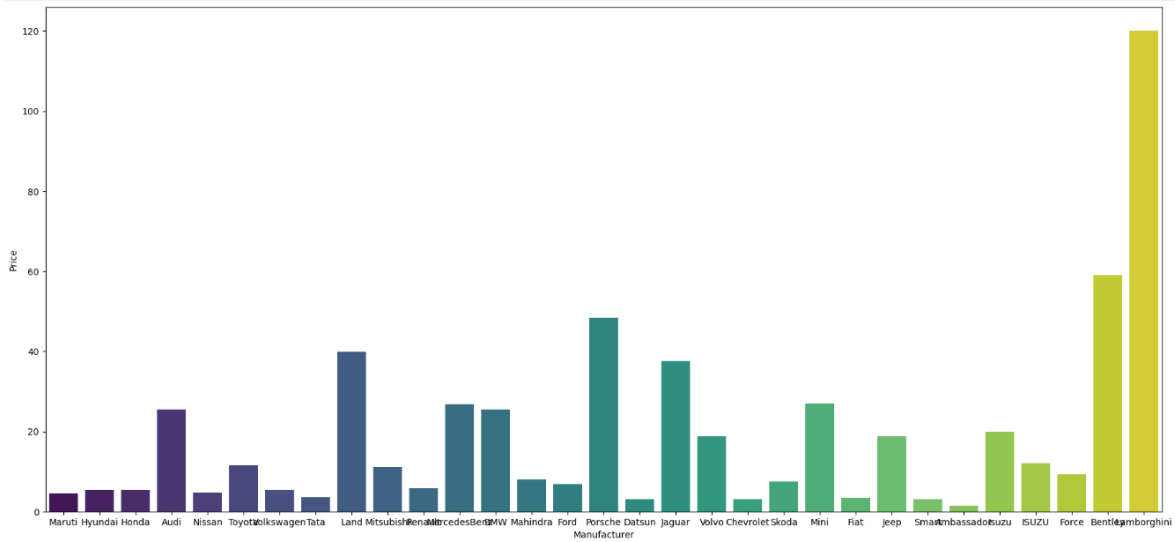
Which car is mostly used?

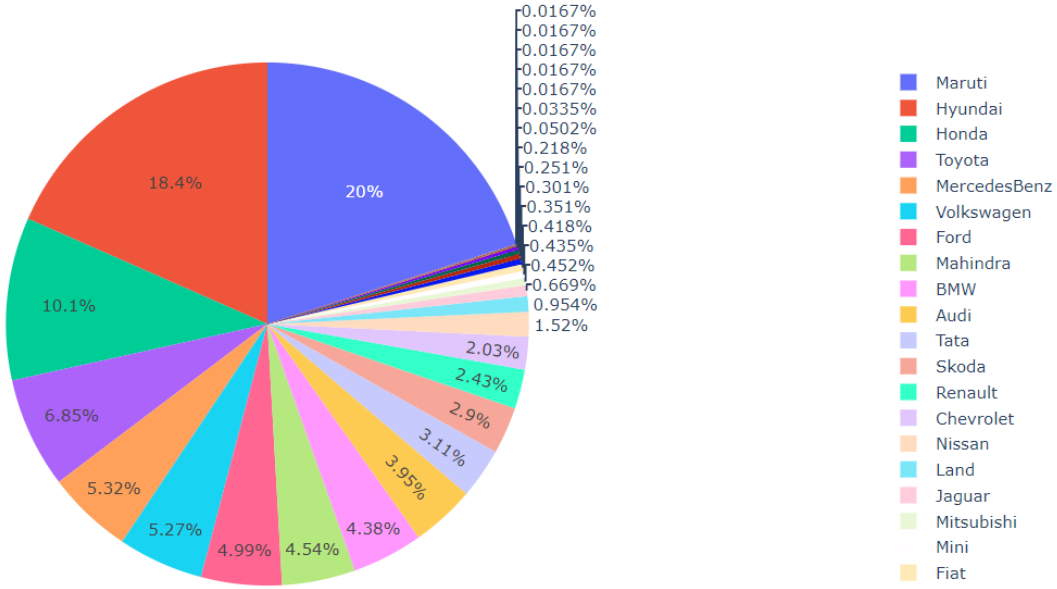


Mileage of different cars:

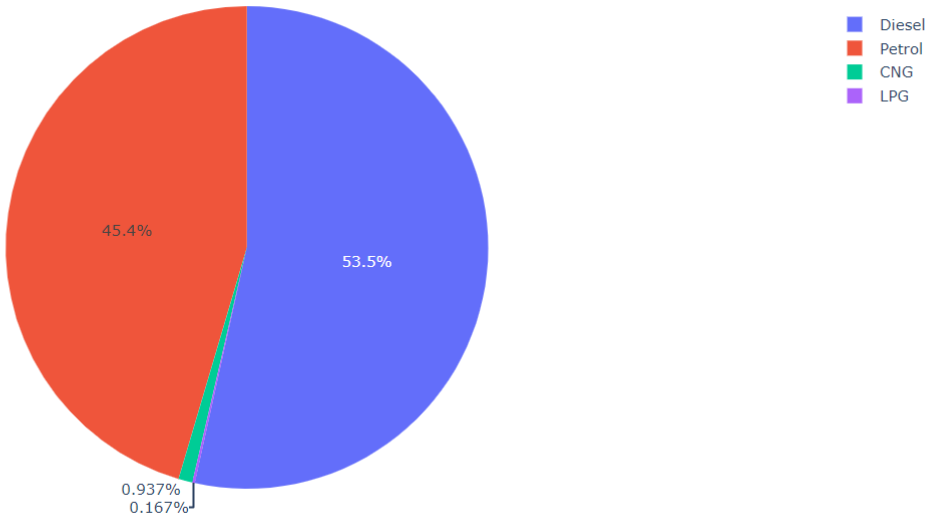


Prices of different cars:a

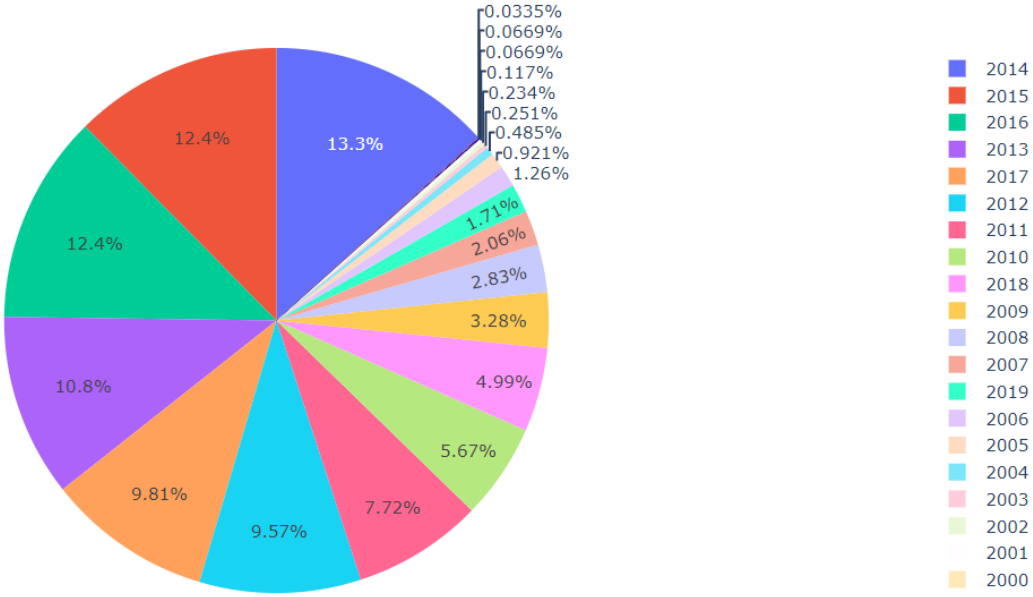




Types of cars used:



Number of cars purchased corresponding to year:

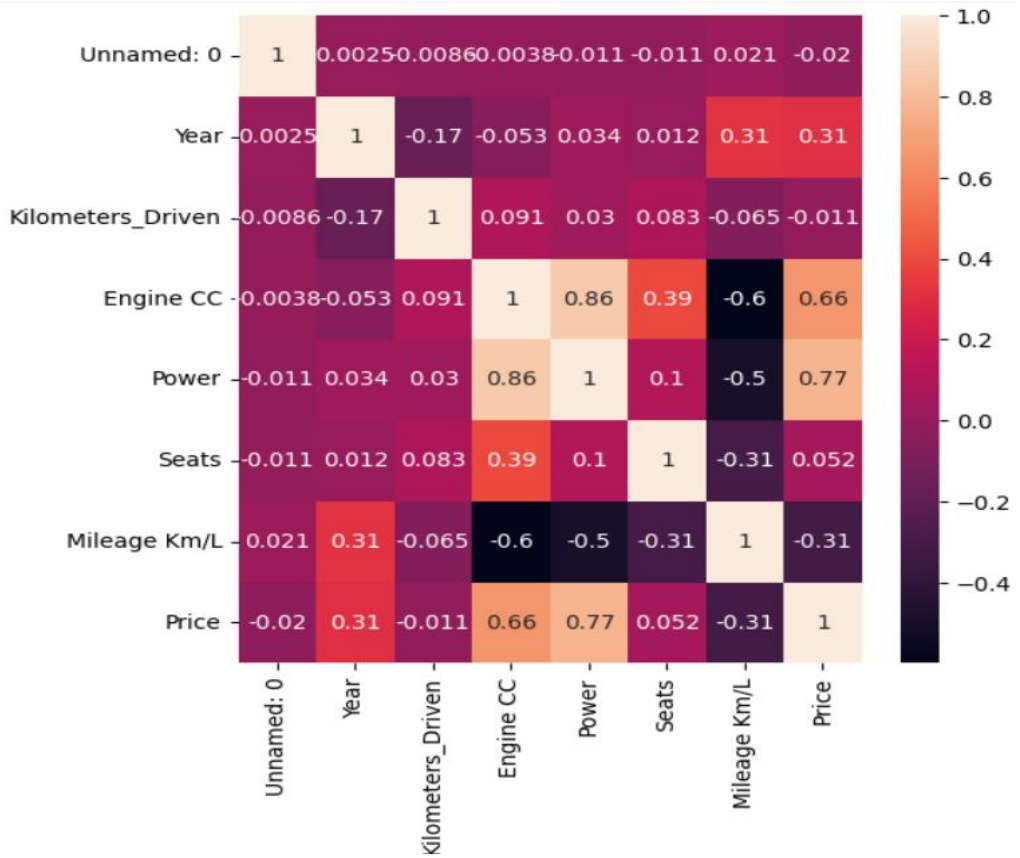


Tools of PCA used are:

1)Correlation Matrix:

- A correlation matrix is a statistical tool used to measure the strength and direction of the linear relationship between two or more variables in a dataset.

- It is a square matrix where each element in the matrix represents the correlation coefficient between two variables.
- The diagonal of the matrix represents the correlation of each variable with itself, which is always 1.
- A positive correlation coefficient indicates that the variables are positively related, while a negative coefficient indicates a negative relationship.
- The magnitude of the coefficient indicates the strength of the relationship, with values close to 1 or -1 indicating a strong correlation, and values close to 0 indicating a weak correlation.
- Correlation matrices are useful in exploratory data analysis as they allow for the identification of patterns and relationships within the dataset, which can inform further analysis and modeling. They can also be used in feature selection, where variables with high correlation coefficients can be removed to reduce multicollinearity and improve model performance.



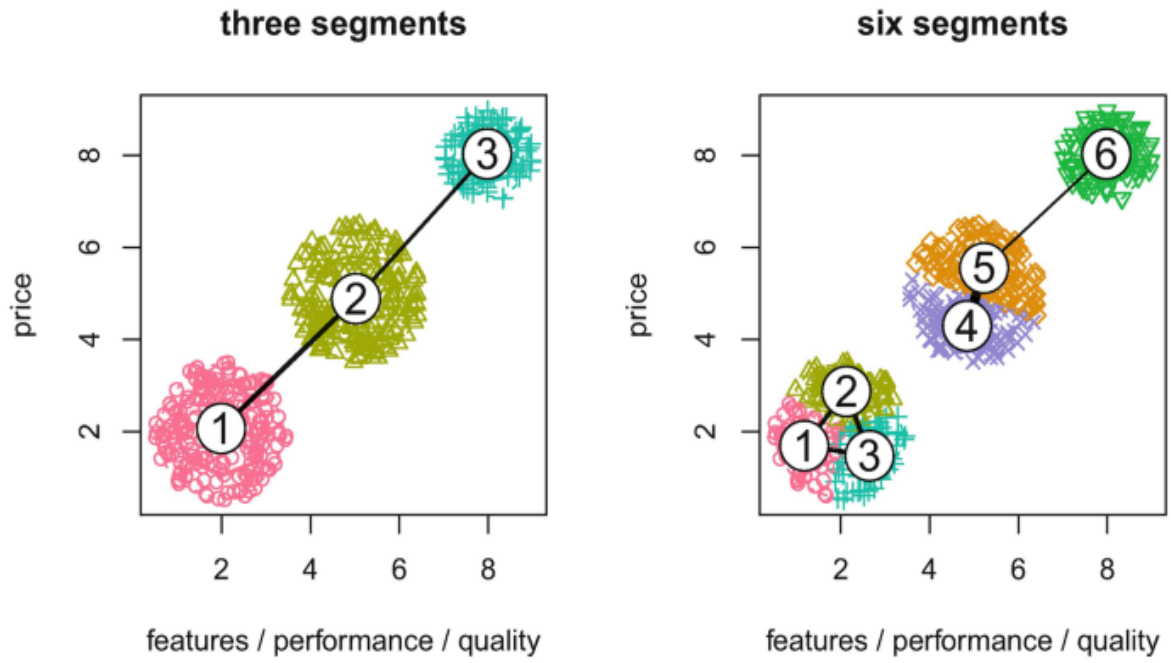
2)Screen Plot:

- A screen plot is a graphical tool used in principal component analysis (PCA) to visualize the relative importance of each principal component (PC) in the dataset.
- It is a plot of the eigenvalues of the principal components against their corresponding component number. The eigenvalues represent the amount of variance explained by each PC, and the screen plot allows for the identification of the number of significant PCs in the dataset.
- The plot typically shows a steep drop in eigenvalues, with the first few components explaining most of the variance in the data, followed by a flat line representing the insignificant components.
- The screen plot is a useful tool in exploratory data analysis as it helps in determining the optimal number of principal components to retain for subsequent analysis. This can prevent overfitting in the model and improve its interpretability.
- In addition, the screen plot can be used to compare the relative importance of different variables in the dataset, which can inform data pre-processing and feature selection.

Extracting Segments:

Extracting segments refers to the process of identifying and separating specific sections or parts from a larger whole. This process can be applied to a variety of contexts, including data analysis, text mining, and media editing. The purpose of extracting segments is to isolate and focus on the relevant information or content within the larger whole.

To extract segments, it's important to first identify the purpose and criteria for the extraction. This involves determining which parts of the larger whole are relevant and which can be disregarded. Once the criteria are established, tools such as algorithms, software, or manual selection can be used to extract the desired segments.

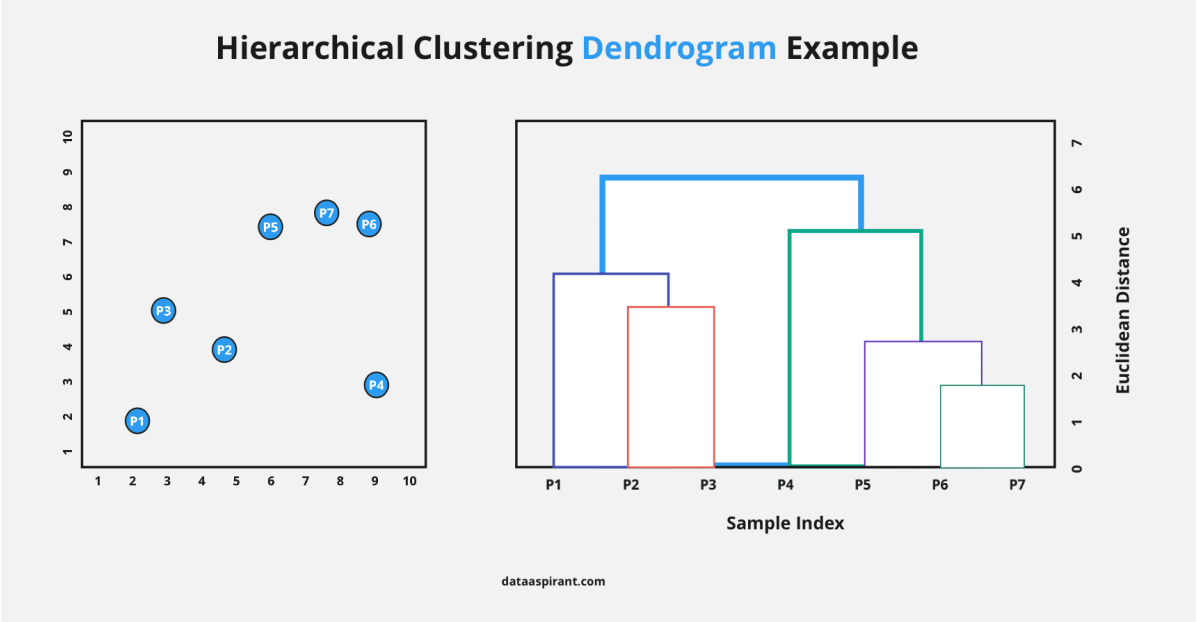


In this task we used mainly two methods namely:

1) Dendrogram:

It is a hierarchical clustering algorithm used in data analysis to group similar objects into clusters. The method constructs a tree-like diagram called a dendrogram that shows the relationships between the clusters.

It is built by starting with each object in its own cluster and then iteratively merging clusters based on their similarity. The similarity between two clusters is typically measured using a distance metric, such as Euclidean distance or correlation coefficient. The algorithm proceeds by merging the two closest clusters at each step until all objects are in a single cluster.



Advantages of Dendrogram:

i. It can handle large datasets and is not sensitive to the initial choice of clusters.
ii. It can provide insights into the underlying structure of the data and help identify subgroups or patterns that may not be apparent from other methods

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.

2)Elbow Method:

The elbow method is a technique used in data analysis to determine the optimal number of clusters in a dataset. It involves plotting the within-cluster sum of squares (WSS) against the number of clusters, and identifying the "elbow" point in the graph where adding more clusters does not significantly reduce the WSS. The optimal number of clusters is then chosen at this elbow point.

The WSS is calculated as the sum of the squared distances between each point and the centroid of its assigned cluster. The idea behind the elbow method is

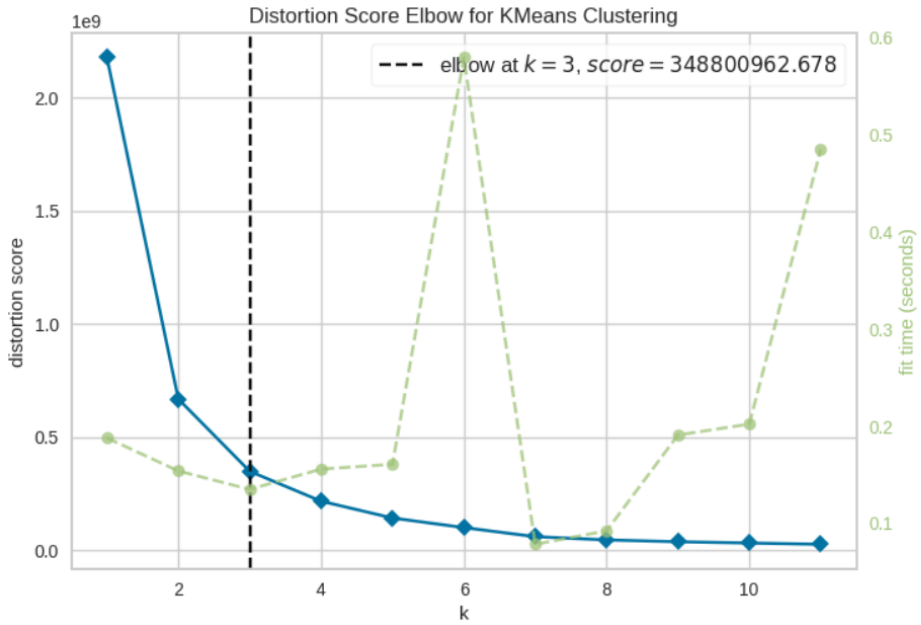
that as the number of clusters increases, the WSS generally decreases, since each point is closer to its assigned centroid. However, beyond a certain number of clusters, the reduction in WSS becomes less significant, and adding more clusters may lead to overfitting or a loss of interpretability.

Advantages of Elbow Method:

i. The elbow method is easy to understand and implement, making it accessible to researchers and practitioners with little expertise in clustering or data analysis.

ii. It provides a quantitative measure of the optimal number of clusters based on the within-cluster sum of squares (WSS), which can be calculated automatically using software tools.

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 need to plot models for various numbers of clusters through the green line.

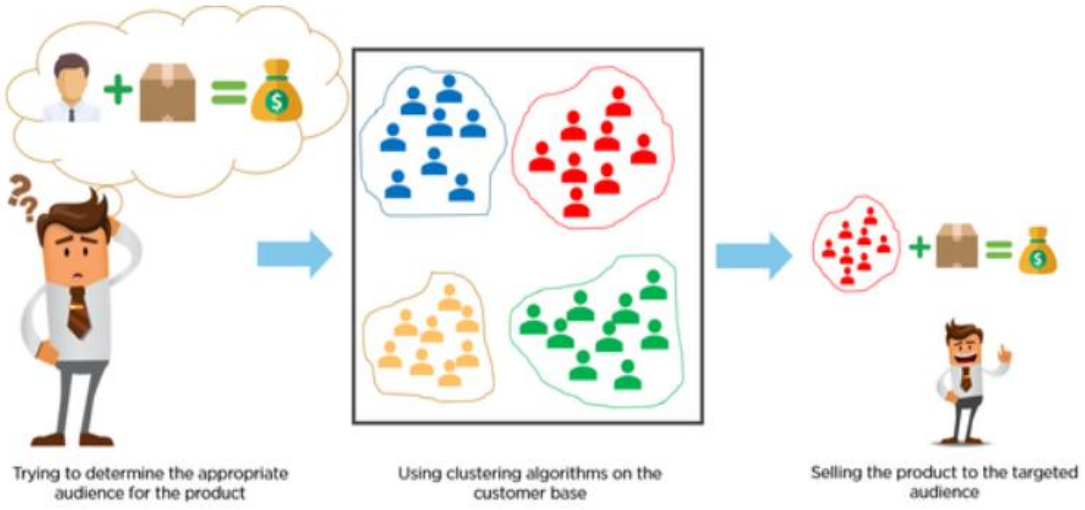


Analysis and Approaches used for Segmentation:

Clustering:

Clustering in segmentation can help organizations understand their customer base or target audience better, and create more effective marketing and communication strategies. By dividing customers into segments, organizations can tailor their products, services, and messaging to the unique needs and preferences of each segment, thereby improving customer satisfaction and loyalty.

It typically involves collecting data on customers or objects, pre-processing the data to remove anomalies, selecting a clustering algorithm appropriate for the dataset, selecting relevant features, clustering the data to create segments, evaluating the quality of the clustering results, and interpreting the results to formulate marketing or business strategies for each segment.

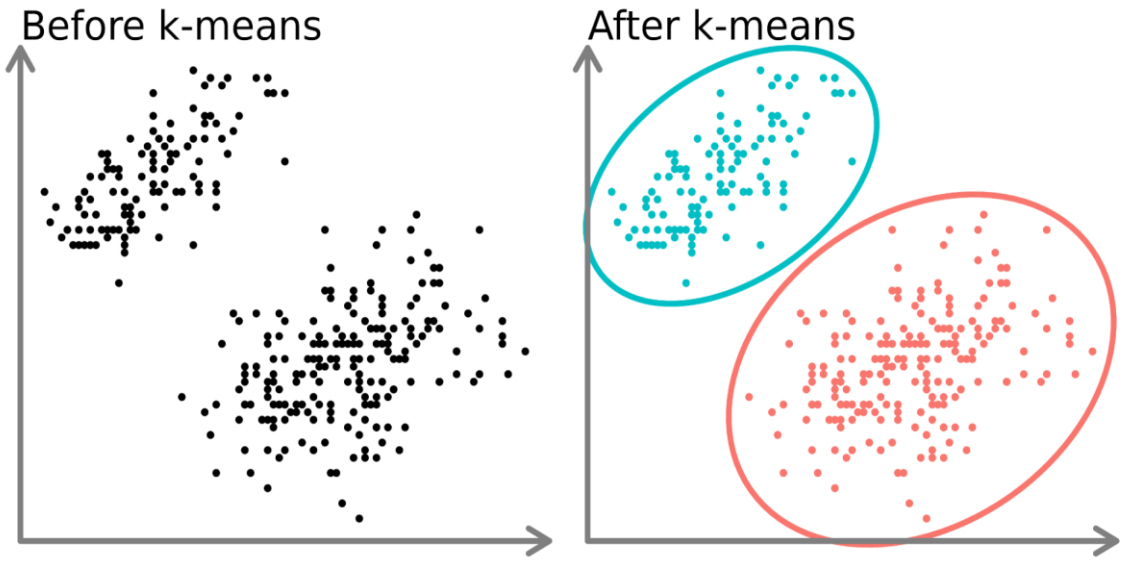


K-Means Algorithm:

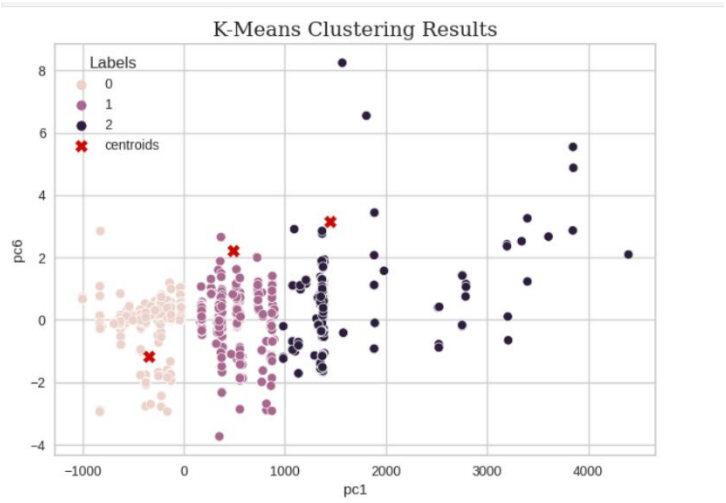
K-means is a popular clustering algorithm used in data analysis and machine learning to group similar data points together. The algorithm works by iteratively assigning each data point to the nearest centroid and recalculating the centroid of each cluster until convergence.

Here are the main steps of the K-means algorithm:

i. **Initialize**: Choose the number of clusters K and randomly initialize K centroids.
ii. **Assign**: Assign each data point to the nearest centroid.
iii. **Update**: Recalculate the centroid of each cluster based on the assigned data points.
iv. **Iterate**: Repeat steps 2 and 3 until convergence or a maximum number of iterations is reached.
v. **Terminate**: Stop the algorithm when the centroids no longer change significantly, or when a maximum number of iterations is reached.



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 below:



Target segments

Target segments are groups of customers or markets that a business aims to attract or serve with its products or services. Based on the analysis of customer data, certain characteristics are identified that define the optimal target segments for the business.

In the context of electric cars, the analysis suggests that the optimum target segment should belong to the following categories:

- Behavioural: Customers who prefer cars with 5 seats

Demographic:

- Top speed & range: Customers who are looking for cars with high top speeds and maximum range. These are important factors that influence the cost of the car, and hence the target market is likely to be large.
- Efficiency: Customers who value efficiency the most, and are looking for cars that are highly efficient.

By identifying these target segments, businesses can tailor their marketing and product development strategies to appeal to these customers. This can help improve sales and customer satisfaction, and ultimately lead to greater profitability for the business.

Psychographic:

Psychographic factors are characteristics of a target market that relate to their personality, lifestyle, values, and attitudes. Based on the analysis of customer data, certain psychographic factors can be identified that are important in determining the target market for a product or service.

In the context of electric cars, the analysis indicates that price is a key psychographic factor. The price range for electric cars is found to be between **16,00,000 to 1,80,00,000**. This information can be used by businesses to determine the target market for their electric cars, and to tailor their marketing strategies and product development efforts accordingly. For example,

businesses can focus on marketing their products to customers who prioritize value for money, or to those who are willing to pay a premium for high-end features and luxury. By understanding the psychographic factors that drive customer preferences, businesses can better meet the needs of their target market and achieve greater success in the marketplace.

In conclusion, the target segment that businesses should focus on is comprised of cars that prioritize efficiency, have high top speeds, and are priced between 16 to 180 lakhs. Furthermore, these cars should predominantly have 5 seats.

GitHub Link: https://github.com/adepurithesh/Market-segmentation-on-EV-s