MARKET SEGMENTATION – ELECTRIC VEHICLES

BY

YASHVEER KOTHARI.A

MARKET SEGMENTATION CASE STUDY - ABSTRACT

This report explores various aspects of market segmentation, focusing on the process of identifying, analysing, and targeting distinct consumer groups to enhance marketing efficiency and effectiveness. It underscores the significance of strategic planning, robust data analysis, and effective profiling in market segmentation. By combining traditional statistical methods with modern visualization techniques, organizations can create well-defined market segments that inform successful marketing strategies, aligning consumer needs with the organizational value proposition.

DATA PRE-PROCESSING

Required libraries

To perform EDA and clustering on the collected data, the following Python libraries are used:

1. Pandas: for data handling/manipulation
2. Matplotlib and Seaborn: for data visualization
3. Scikit-learn: for the k-means clustering algorithm and some other algorithms.

```
# importing the dependencies
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
```

LOADING THREE DATASETS

- **DATASET 1:**

	Region	2W	3W	4W	Bus	Chargers
0	Uttar Pradesh	9852	42881	458	197	207
1	Maharastra	38558	893	1895	186	317
2	Karnataka	32844	568	589	57	172
3	Tamil Nadu	25642	396	426	0	256
4	Gujarat	22359	254	423	22	228

- **DATASET 2:**

	State/UT	EV Charging Facility
0	Andhra Pradesh	65
1	Arunachal Pradesh	4
2	Assam	19
3	Bihar	26
4	Chandigarh	4

- **DATASET 3:**

[4]:

	Brand	Model	AccelSec	TopSpeed_KmH	Range_Km	Efficiency_WhKm	FastCharge_KmH	RapidCharge	PowerTrain
0	Tesla	Model 3 Long Range Dual Motor	4.6	233	450	161	940	Yes	AWD
1	Volkswagen	ID.3 Pure	10.0	160	270	167	250	No	RWD
2	Polestar	2	4.7	210	400	181	620	Yes	AWD
3	BMW	iX3	6.8	180	360	206	560	Yes	RWD
4	Honda	e	9.5	145	170	168	190	Yes	RWD

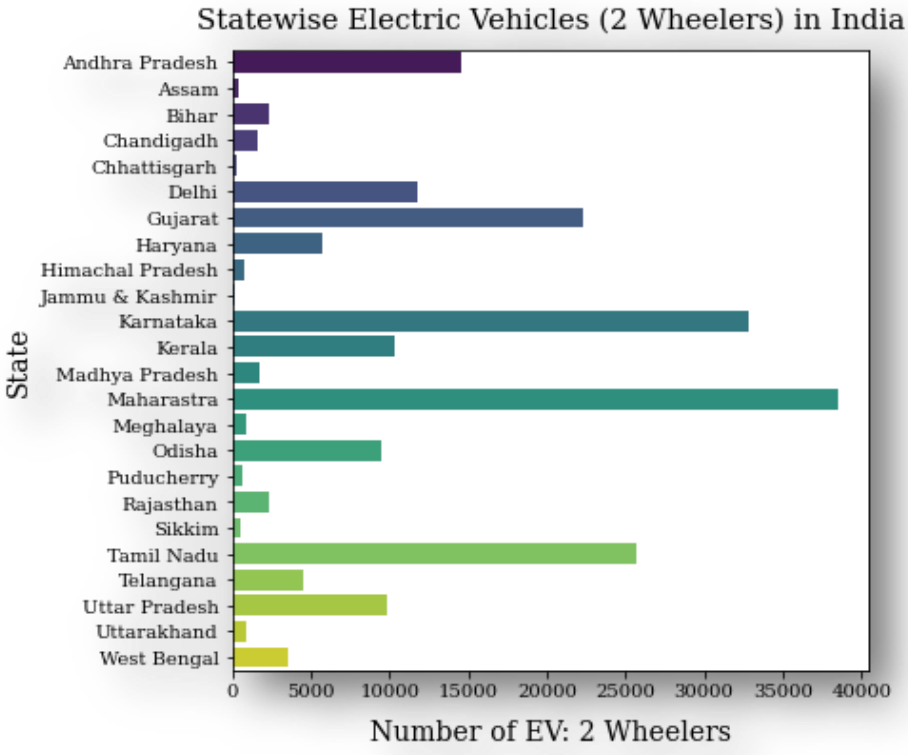
EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis, popularly abbreviated as EDA, is one of the most important steps in the data science pipeline. It is the process of gaining the information present inside the data with the help of summary statistics and visual representations.

IMPLEMENTING EDA ON DATASETS

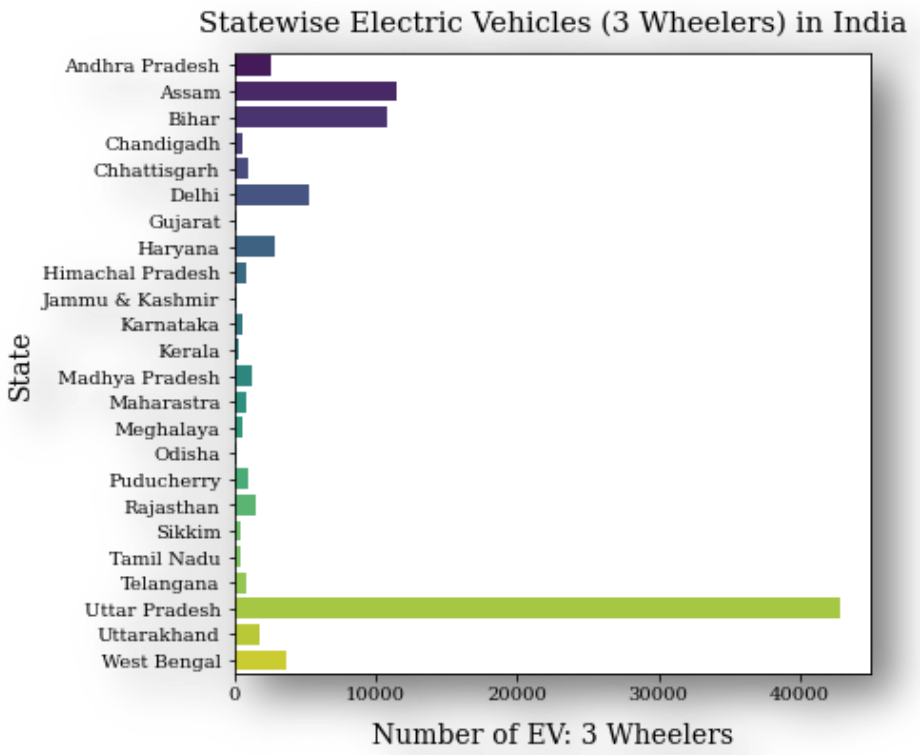
- **ANALYSIS OF 2-WHEELER EVs:**

  Observation: Maharashtra, Gujarat, Tamil Nadu, Karnataka, and Andhra Pradesh are among the top states with most EV 2-wheelers while Assam, Himachal Pradesh, Sikkim, J&K with the least.



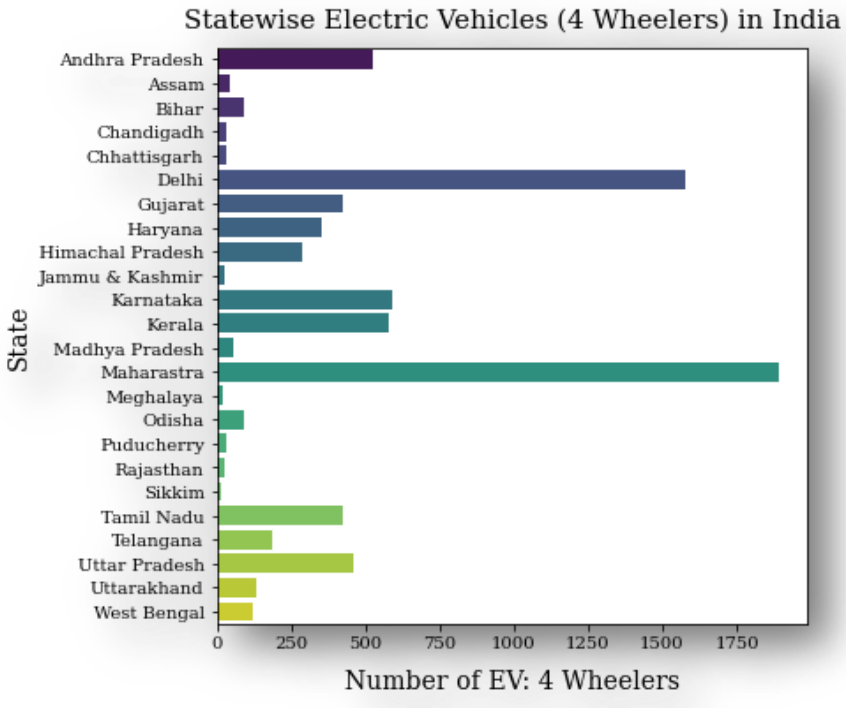
- **ANALYSIS OF 3-WHEELER EVs:**

  Observation: Uttar Pradesh, Assam, and Bihar are among the top states with most EV 3-wheelers while the remaining states do not seem to depend on the same.



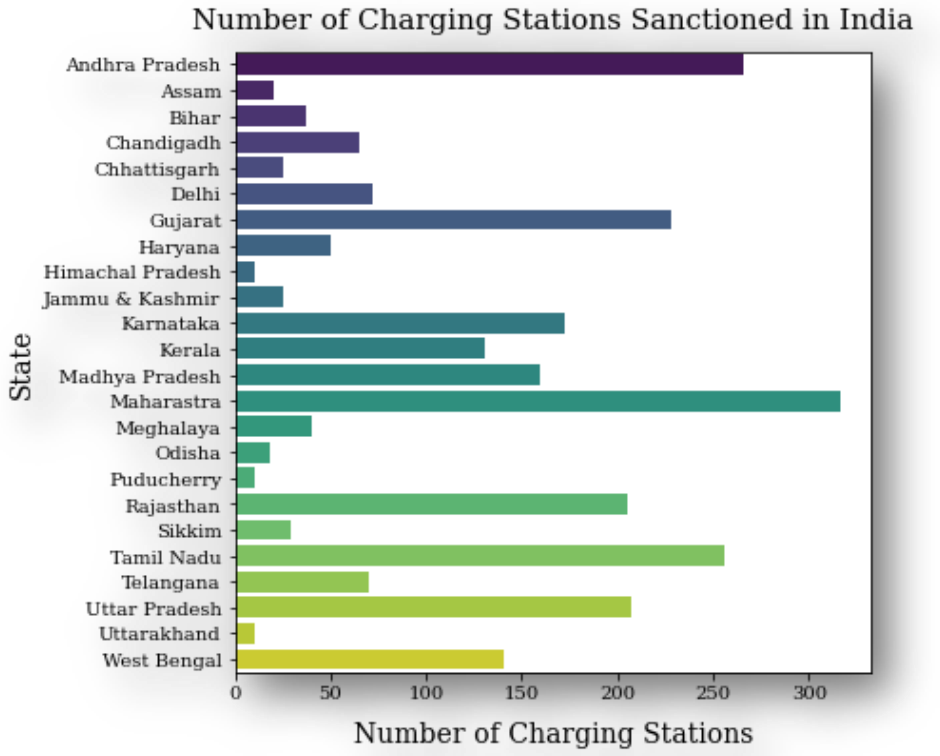
- **<u>ANALYSIS 4-WHEELER EVs:</u>**

  Observation: Maharashtra, Delhi, Karnataka, Kerala and Andhra Pradesh are among the top states with the majority of EV 4-wheelers while the remaining states have less number of EV 4-wheelers.



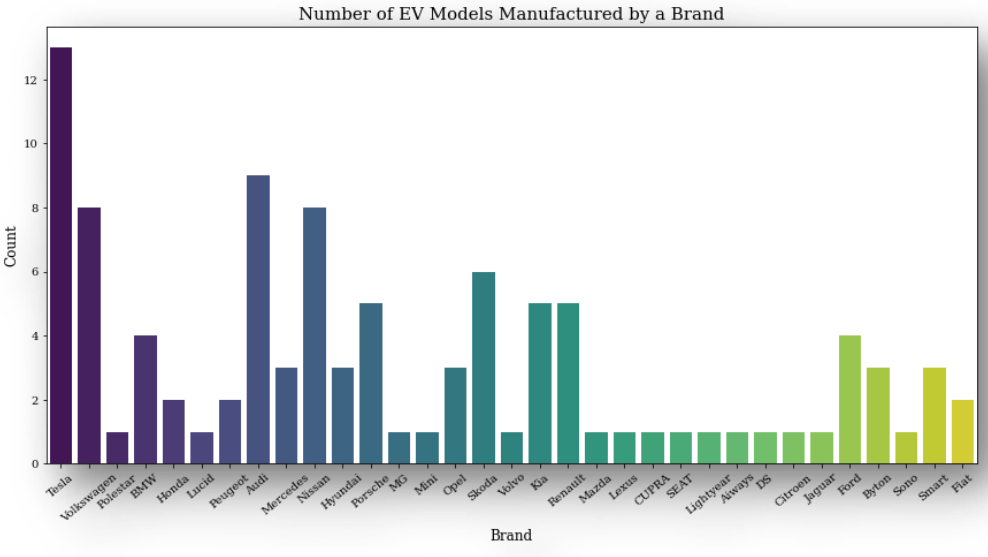
- **ANALYSIS OF CHARGING STATIONS SANCTIONED:**

  Observation: Maharashtra, Gujarat, Karnataka, Kerala, Uttar Pradesh, Rajasthan, and Andhra Pradesh are among the top states with the majority of EV charging stations sanctioned while the remaining states have a smaller number of the same.



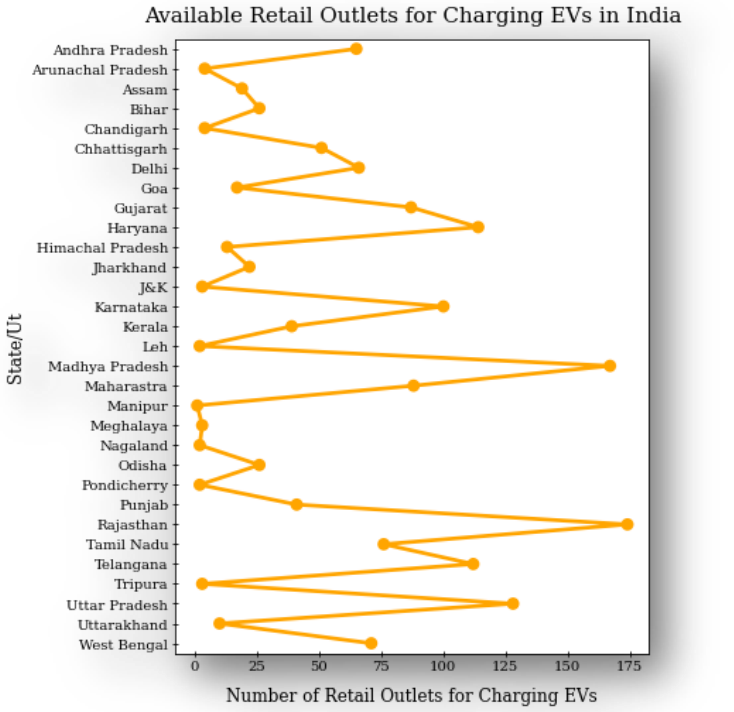
- **ANALYSIS OF EVs BASED ON BRANDS:**

  Observation: Tesla, Audi, Volkswagen, Nissan, Skoda tops the list of EVs with the maximum number of models in the Indian automobile market.



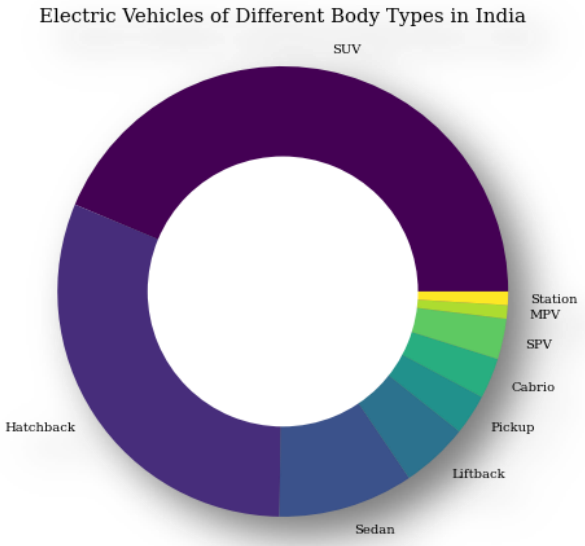
- **ANALYSIS OF RETAIL OUTLETS FOR EV CHARGING:**

  Observation: Rajasthan, Madhya Pradesh, Maharashtra, Karnataka, and Uttar Pradesh are among the top states with most retail outlets for EV charging while the remaining states have a smaller number of the same.



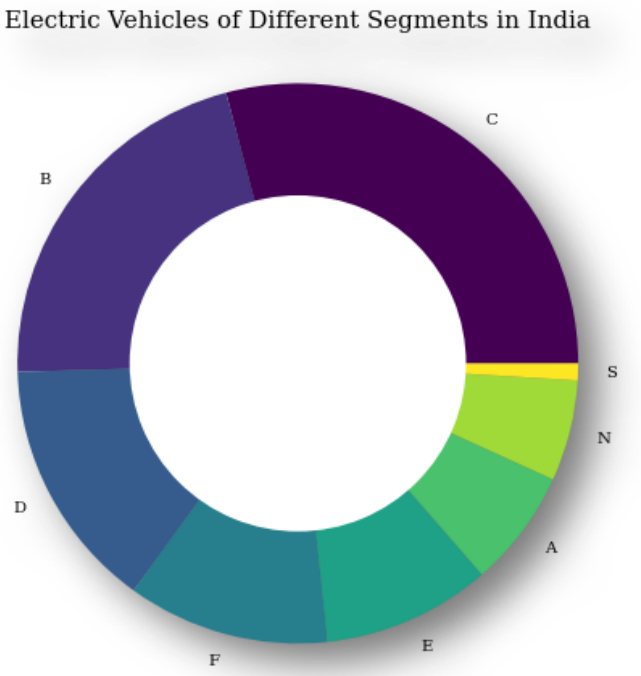
- **ANALYSIS OF DIFFERENT BODY TYPES OF EVs**

  Observation: SUV and Hatchback body types form the majority while Station and MPV the minority.



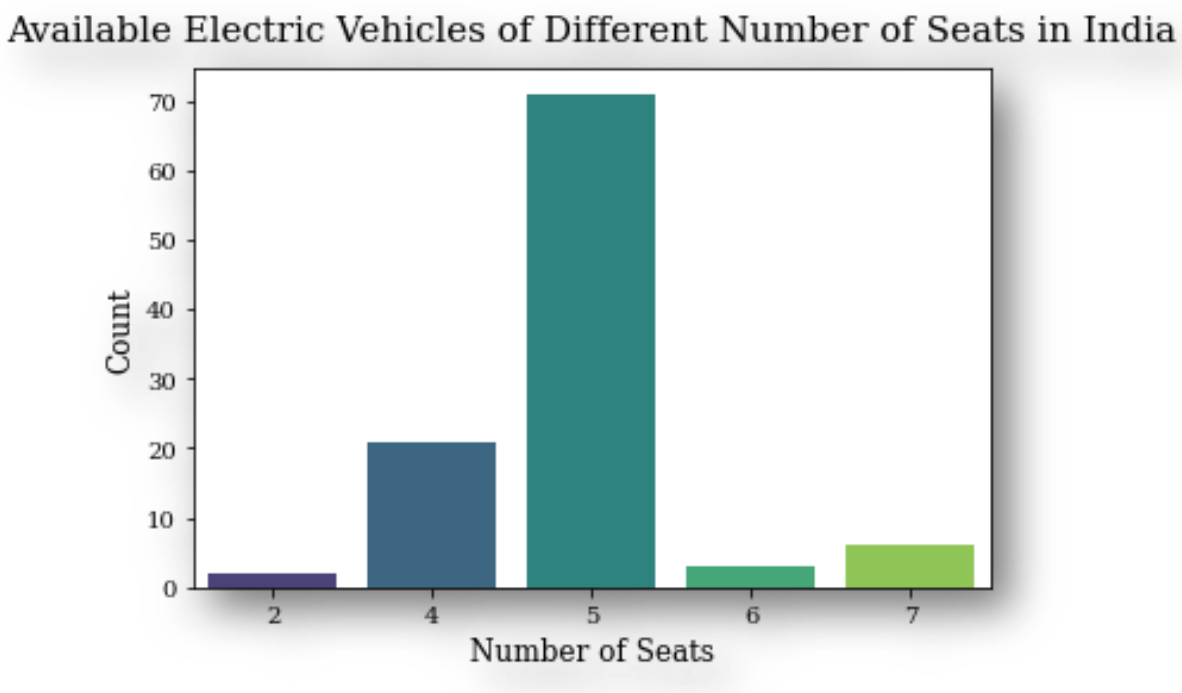
- **ANALYSIS OF DIFFERENT SEGMENTS OF EVS:**

  Observation: B and C body segments form the majority while S and A the minority.



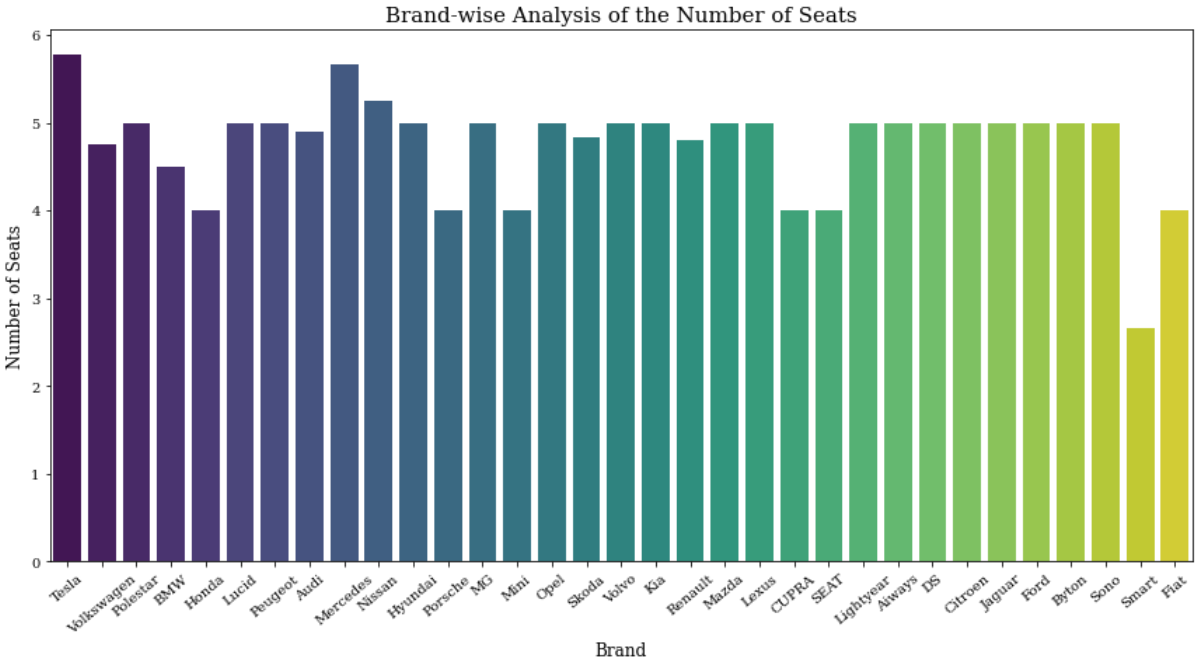
- **ANALYSIS OF EVS OF DIFFERENT NUMBER OF SEATS:**

  Observation: EVs with 5 sitters dominate the market while EVs with 2 sitters are less in number.



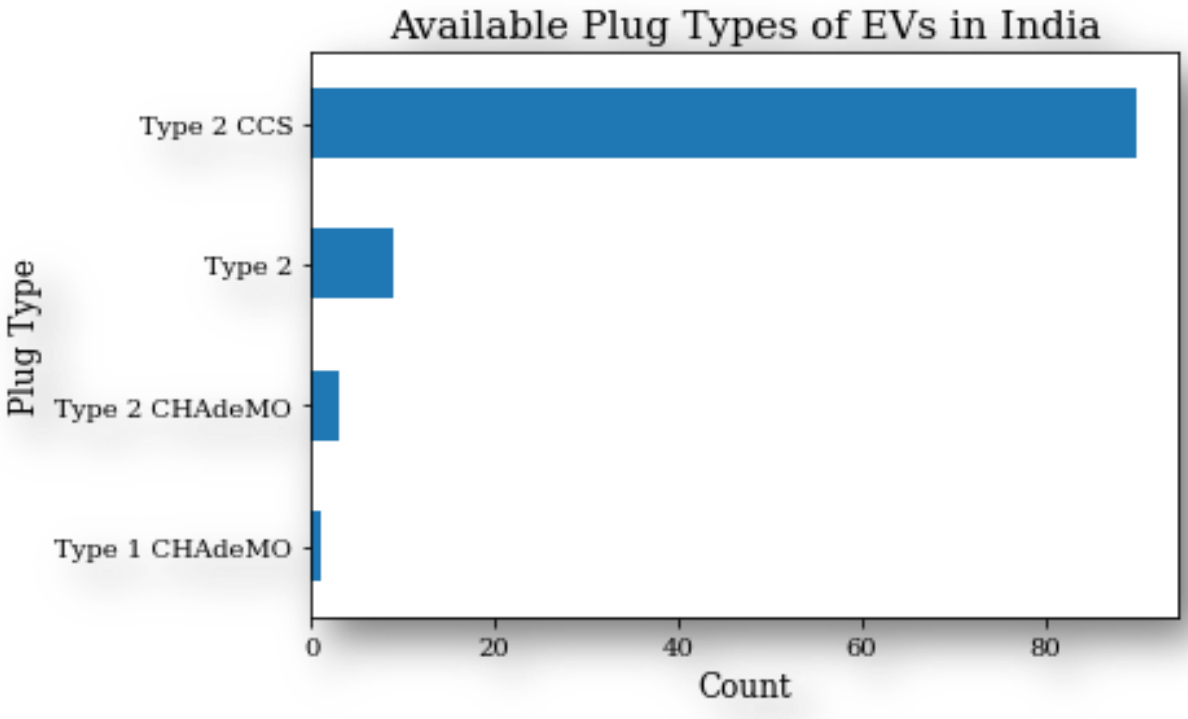
- **ANALYSIS OF THE NUMBER OF SEATS BY EACH BRAND:**

  Observation: Based on the number of seats, Tesla, Mercedes, and Nissan have the maximum number of seats and smart the minimum.



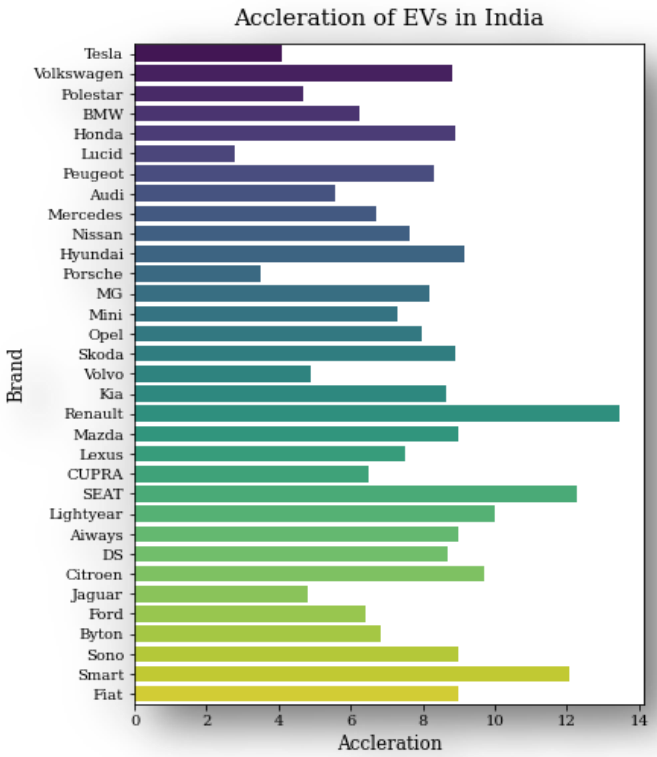
- **Analysis of different plug types:**

  Observation: EVs with plus type of 'Type 2 CCS' seem to dominate the market.



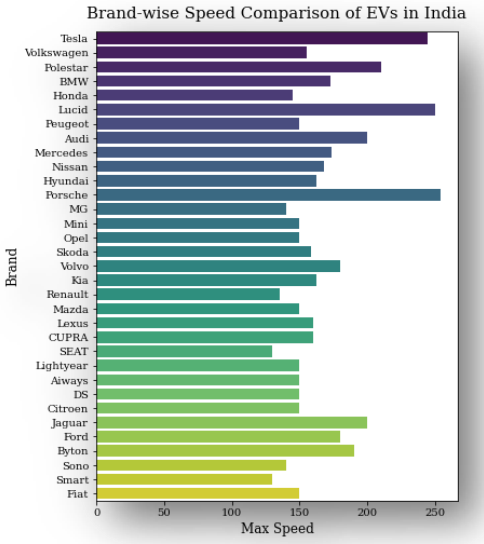
- **ANALYSIS OF EVS BASED ON ACCELERATION**:

  Observation: Based on acceleration, EVs from Renault, Seat and Smart are the top performers while Tesla, Lucid and Porsche do not make it to the same.



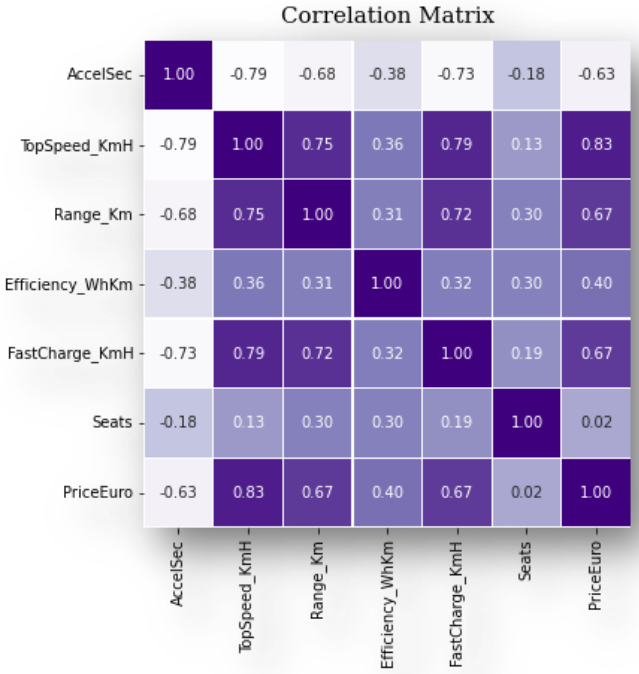
- **ANALYSIS OF EVs BASED ON SPEED:**

  Observation: Based on speed parameter, EVs from Tesla, Lucid and Porsche are the top performers while Renault, Smart and SEAT do not make it to the same.



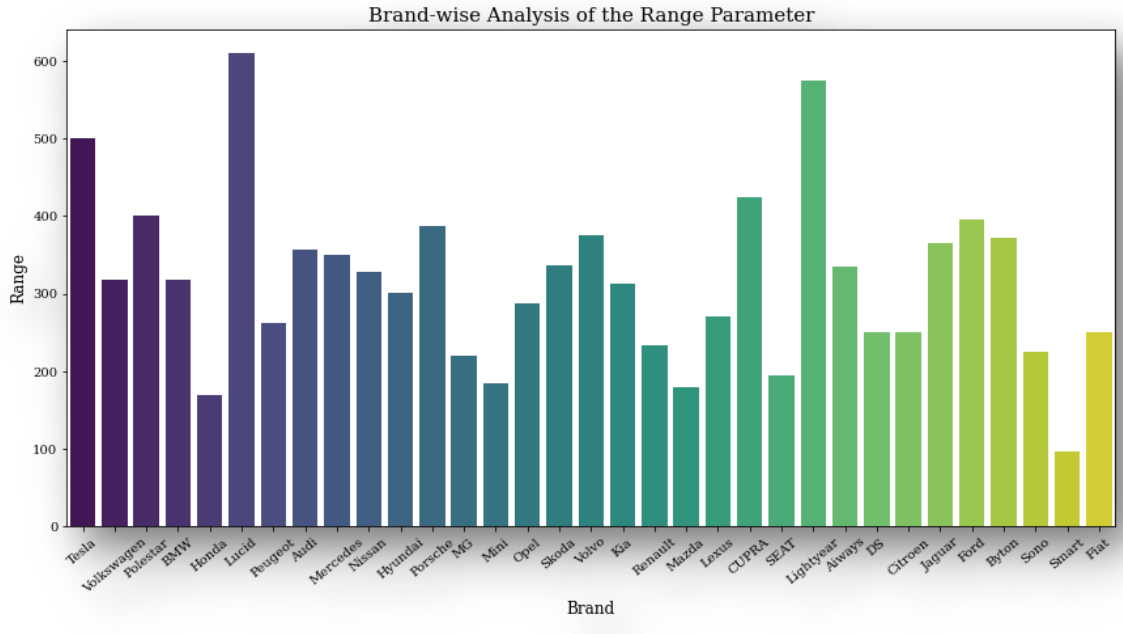
- **CORRELATION MATRIX:**

  To understand the correlation between different variables of the dataset. Top speed kmh has the highest correlation with price in Euro followed by Fast charge range and then range km.



- **<u>Analysis of EVs based on the range parameter:</u>**

  Observation: Based on range (Km), Lucid, Lightyear and Tesla have the highest range and smart the lowest.



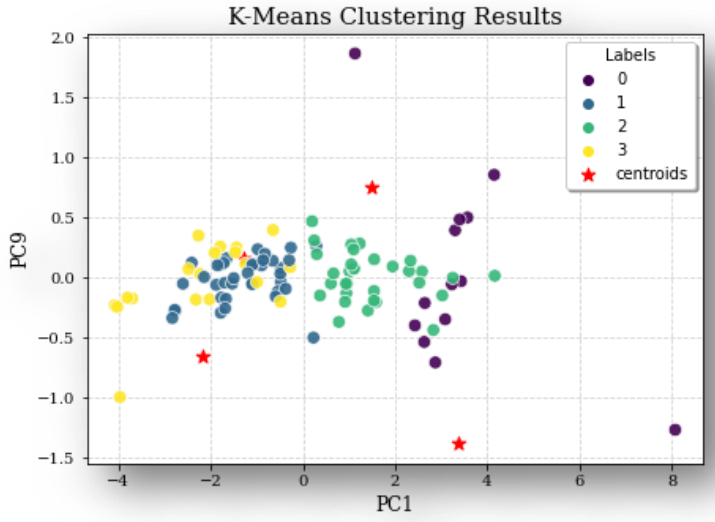
SEGMENTATION APPROACHES: KMEANS

- **CLUSTERING :**

  Clustering is an unsupervised machine learning technique of grouping similar data points into clusters. The sole objective of this technique is to segregate datapoints with similar traits and place them into different clusters. There are several algorithms to perform clustering on data such as k-means clustering, hierarchical clustering, density-based clustering etc.

- **K-MEANS CLUSTERING:**

  K-Means Clustering is an unsupervised learning algorithm whose job is to group the unlabelled dataset into different clusters where each datapoint belongs to only one cluster. Here, K is the number of clusters that need to be created in the process. The algorithm finds its applicability into a variety of use cases including market segmentation, image segmentation, image compression, document clustering etc. The below image is the results of clustering on one of our datasets.



The K-Means Algorithm works the following way:

1. Specify the number of clusters, i.e. K

2. Select K random points in the dataset. These points will be the centroids (centres) of each of the K clusters.

3. Assign each data point in the dataset to one of the K centroids, based on its distance from each of the centroids.

4. Consider this clustering to be correct and reassign the Centroids to the mean of these clusters.

5. Repeat Step 3. If any of the points change clusters, Go to step 4. Else Go to step 6.

6. Calculate the variance of each of the clusters.

7. Repeat this clustering 'n' number of times until the sum of variance of each cluster is minimum.

PRINCIPLE COMPONENT ANALYSIS:

Principal component analysis (PCA) is a linear dimensionality-reduction technique that is used to reduce the dimensionality of large data sets by transforming a large set of variables into a smaller one while preserving most of the information present in the large set.

ELBOW METHOD:

The Elbow method is a way of determining the optimal number of clusters (k) in K-Means Clustering. It is based on calculating the Within Cluster Sum of Squared Errors (WCSS) for a different number of clusters (k) and selecting the k for which change in WCSS first starts to diminish. When you plot its graph, at one point the line starts to run parallel to the X-axis and that point, known as the Elbow Point, is considered as the best value for the k (as 4 in the below figure).