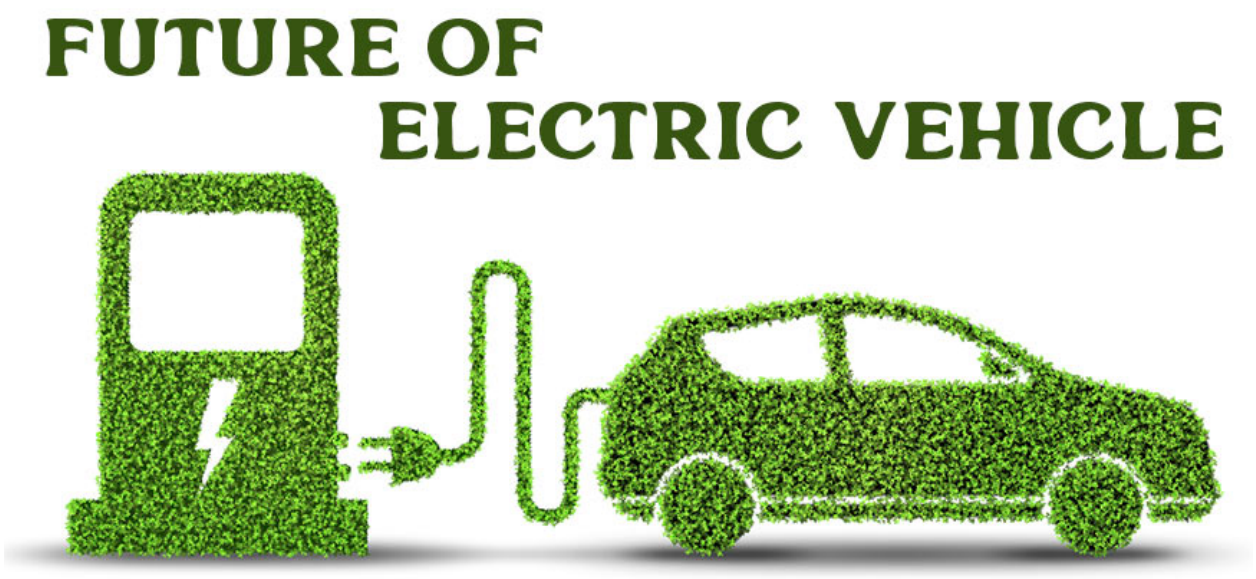
Problem Statement: Analysis of the Electric Vehicle (EV) Market in India

Jayanth Atipamula



The objective is to conduct a segmentation analysis of the EV market in India and determine the most viable market segments to target for entry into the market. This analysis will inform the development of a strategic plan for targeting the most promising market segments for the adoption and use of electric vehicles in India.

Data Collection Data was extracted from the various websites mentioned below for EV market segmentation.

Link for data :-
https://bit.ly/41MNhVg

Columns explanations:

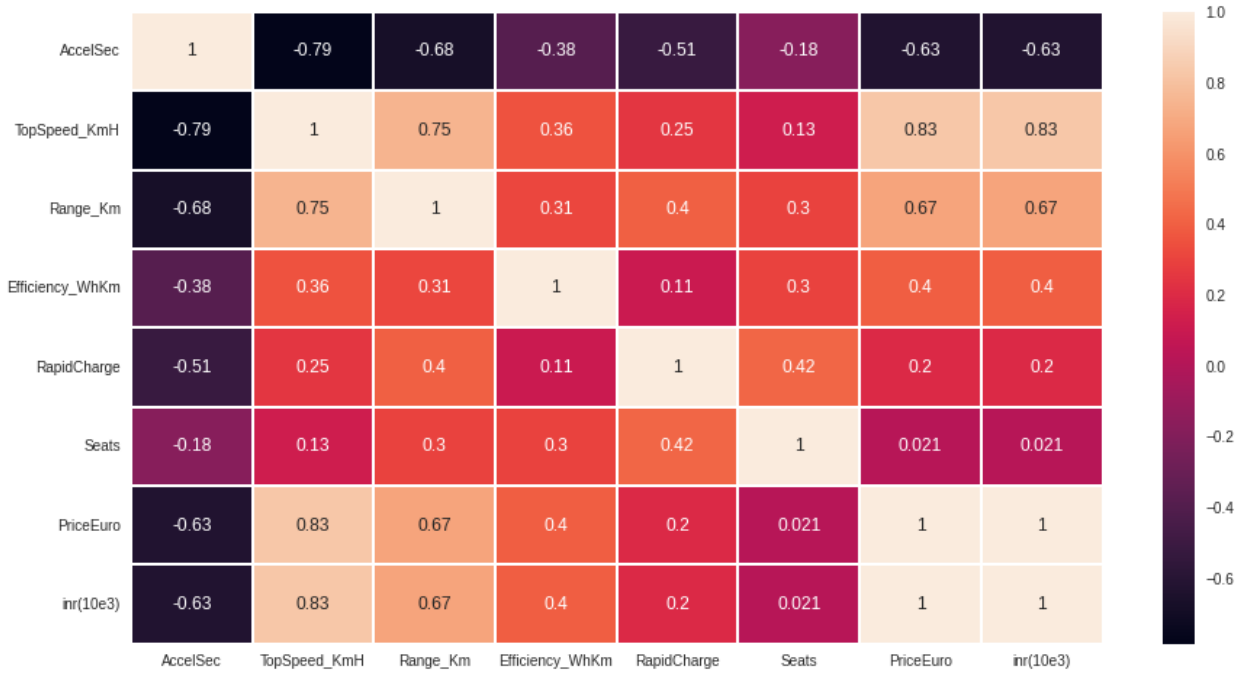
- **Brand: The brand or manufacturer of the electric vehicle.**
- **Model: The specific model name of the electric vehicle.**
- **AccelSec: The time it takes for the electric vehicle to accelerate from 0 to 100 km/h, in seconds.**

- **TopSpeed_KmH: The maximum speed that the electric vehicle can achieve, in kilometers per hour.**
- **Range_Km: The estimated range of the electric vehicle on a single charge, in kilometers.**
- **Efficiency_WhKm: The energy efficiency of the electric vehicle, measured in watt-hours per kilometer.**
- **FastCharge_KmH: The charging speed of the electric vehicle using a fast charger, in kilometers per hour.**
- **RapidCharge: A binary variable indicating whether the electric vehicle has the capability of using rapid charging technology.**
- **PowerTrain: The type of powertrain used in the electric vehicle (e.g. all-wheel drive, front-wheel drive, rear-wheel drive).**
- **PlugType: The type of plug used to charge the electric vehicle.**
- **BodyStyle: The body style of the electric vehicle (e.g. sedan, SUV, hatchback).**
- **Segment: The market segment that the electric vehicle is aimed at (e.g. luxury, mid-size, compact).**
- **Seats: The number of seats in the electric vehicle.**

Exploratory Data Analysis:

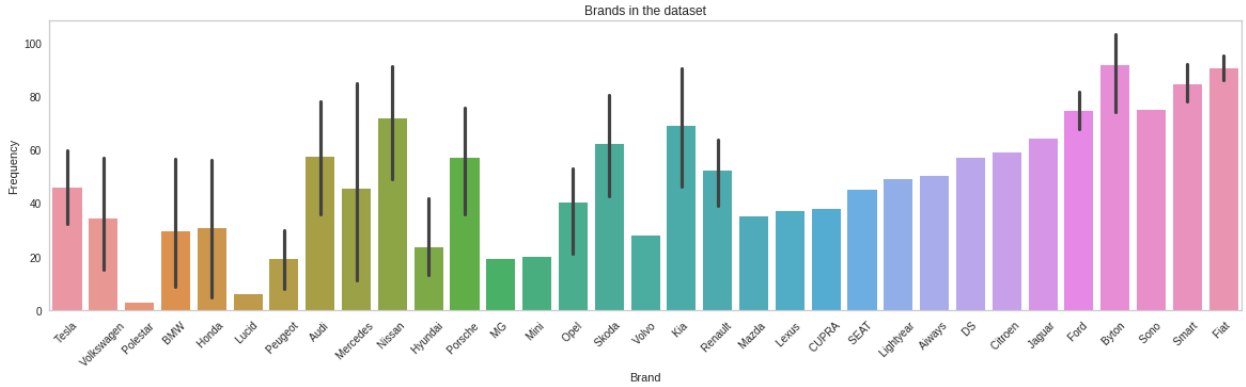
Exploratory Data Analysis (EDA) is a comprehensive investigation of a dataset to uncover its underlying structure. For a company, EDA is crucial as it reveals trends, patterns, and relationships that may not be immediately obvious. Our analysis of the dataset involved three types of analysis: univariate analysis (examining a single variable/column from the dataset), bivariate analysis (examining two variables/columns from the dataset together), and multivariate analysis (examining more than two variables/columns from the dataset together).

Heatmap to show correlation of data



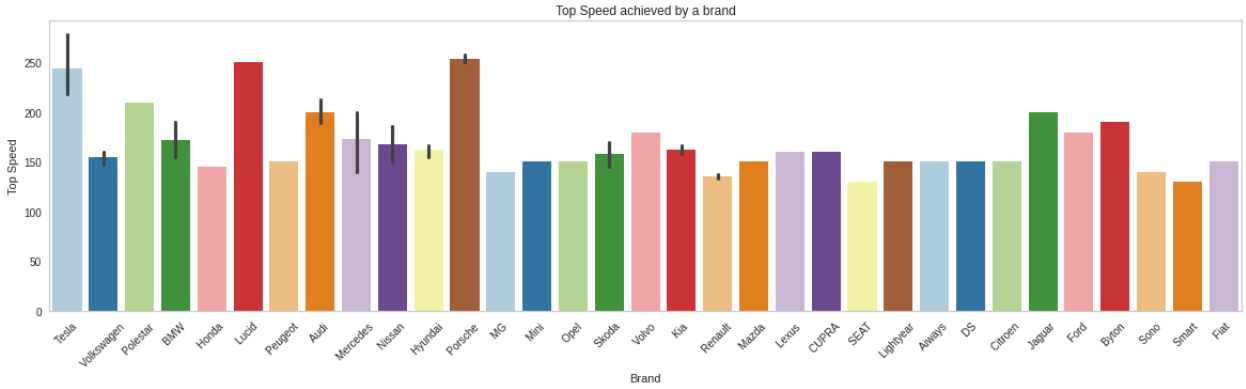
We have observed that two columns in our dataset are highly correlated. Firstly, there is a positive correlation of 0.83 between the 'price' and 'top speed' columns. This correlation is to be expected since a higher price often results in more advanced features. Secondly, there is a positive correlation of 0.75 between the 'top speed' and 'range' columns

Frequency of the brand



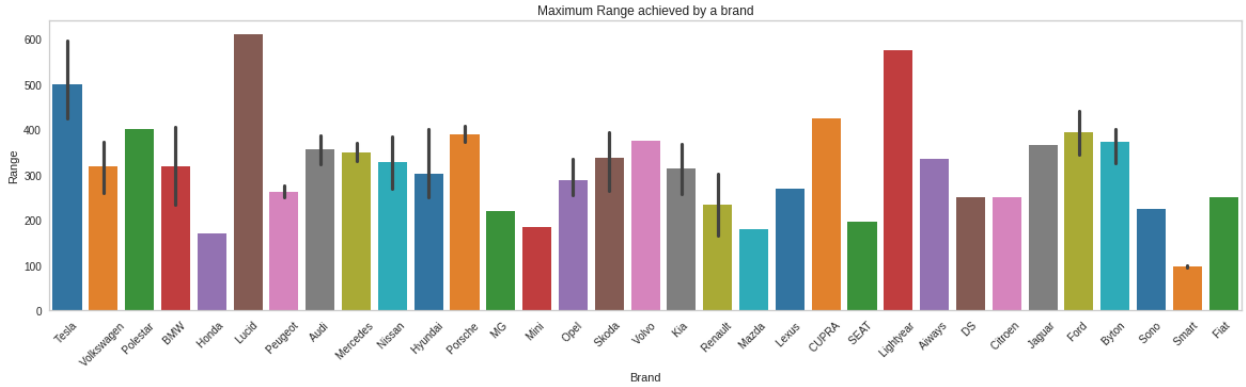
Top speed of the car by brands

The fastest cars are produced by Porsche, Lucid, and Tesla, while Smart produces the slowest cars in our dataset.

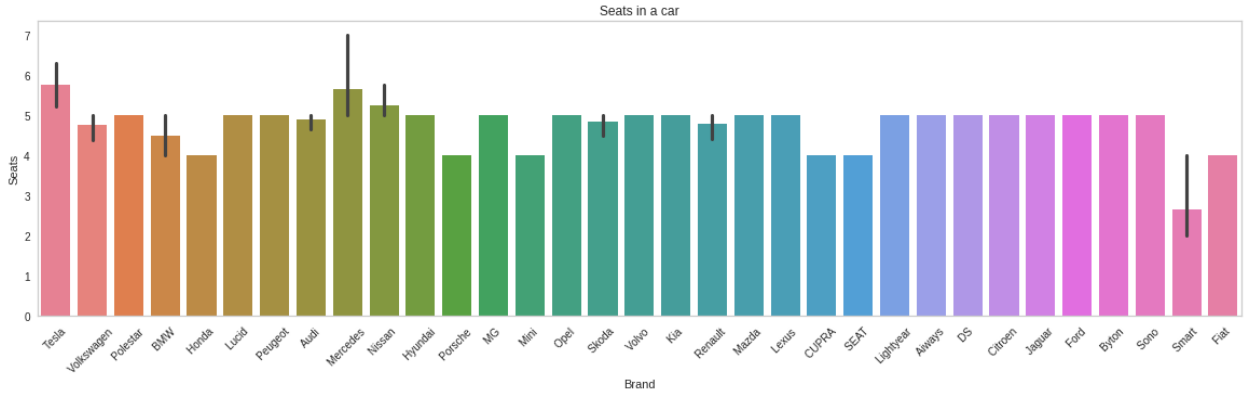


Range a car can achieve

Lucid, Lightyear and Tesla have the highest range and Smart the lowest

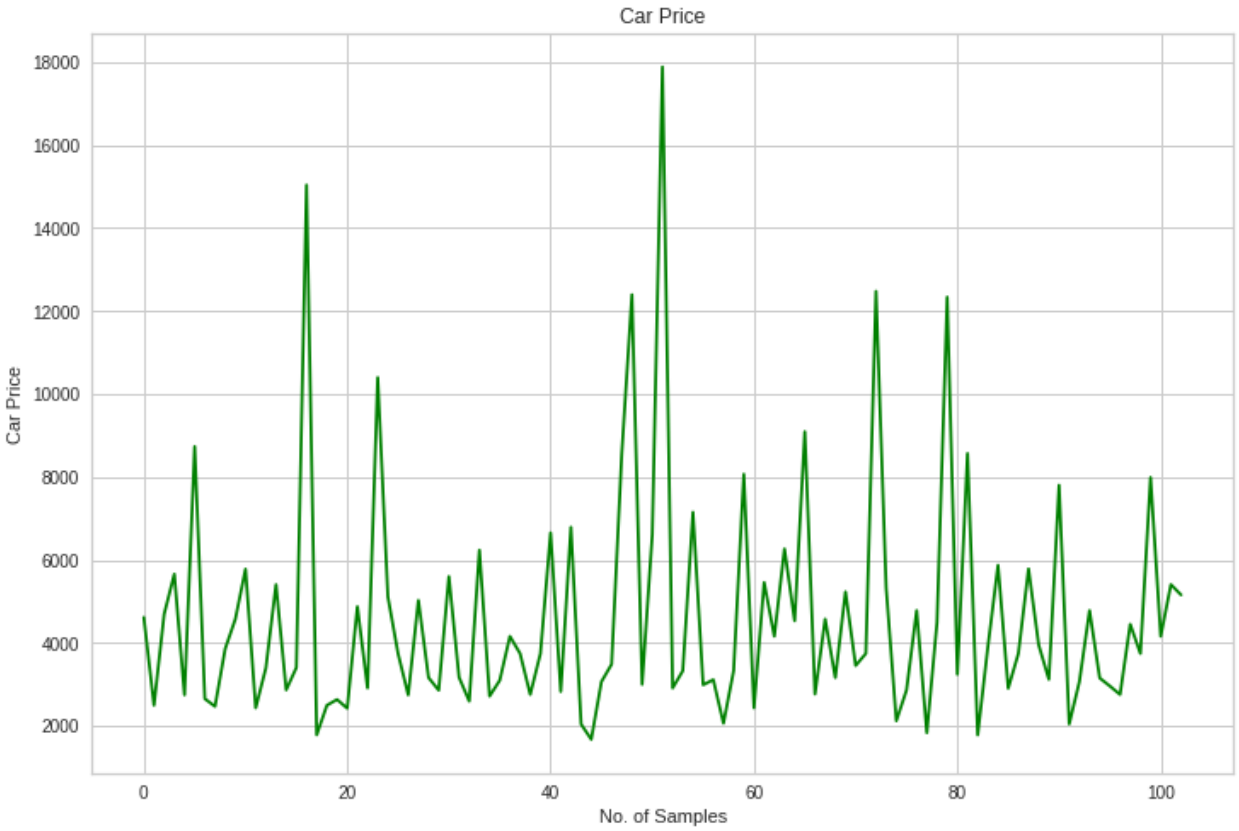


Number of seats in car



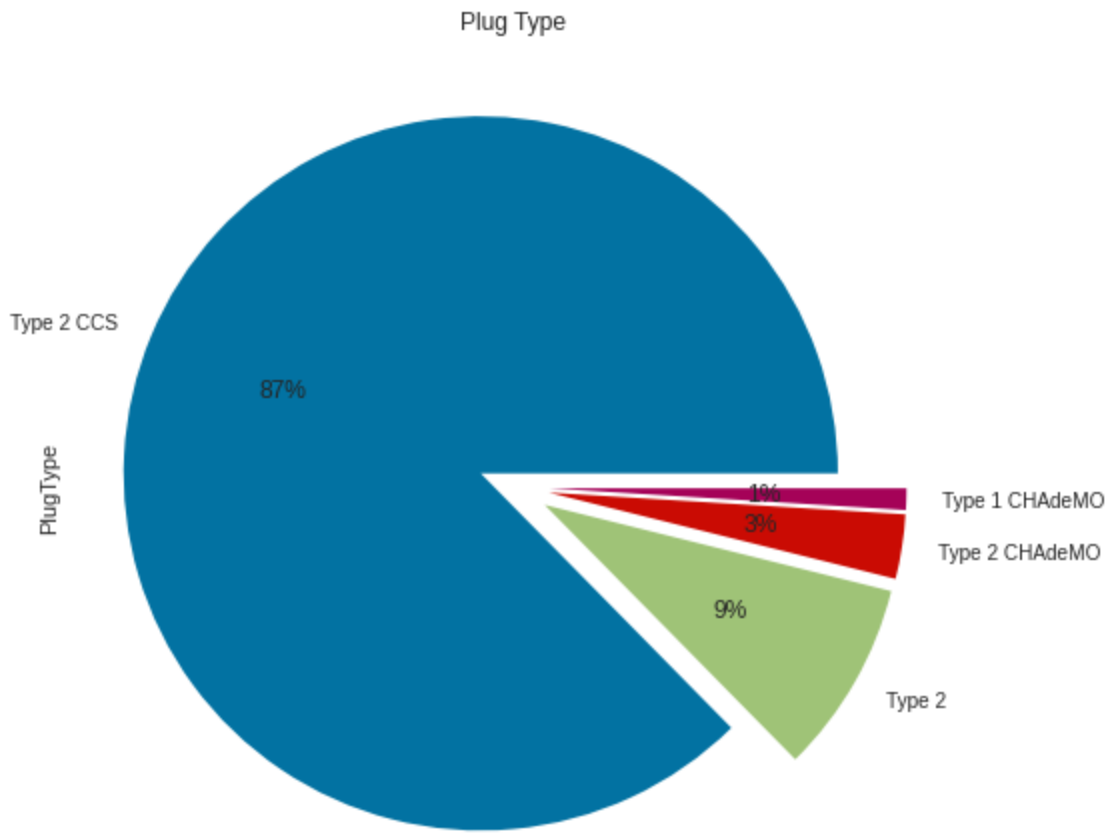
Mercedes, Tesla and Nissan have the highest number of seats and Smart the lowest

Price of cars (in Euro)



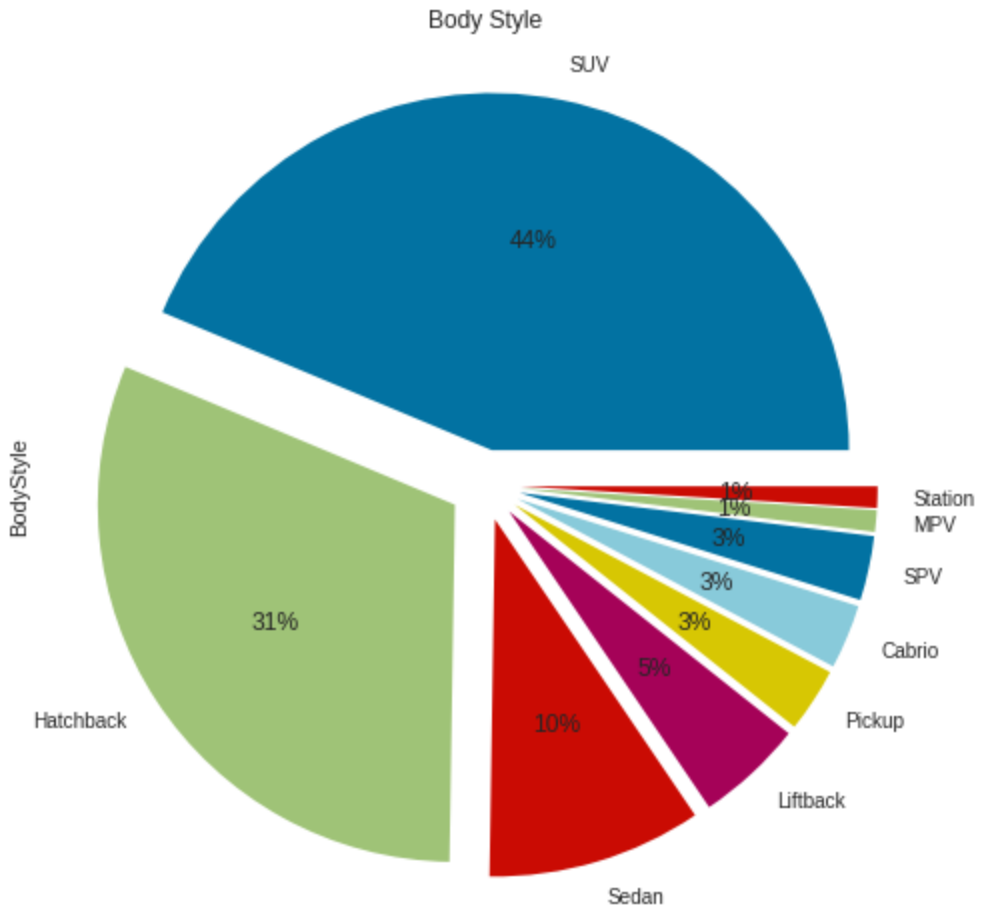
Lightyear, Porsche and Lucid are the most expensive and SEAT and Smart the least

Type of Plug used for charging



Cars and their Body style

Most cars are eiher SUV or Hatchback



Segment Extraction

Segment extraction is a crucial step in market analysis. One popular method for segmenting unlabeled data is K-Means Clustering, which groups data points into clusters based on their similarity in features and patterns. Clustering is an unsupervised machine learning technique that can identify common patterns within multivariate datasets without any external guidance.

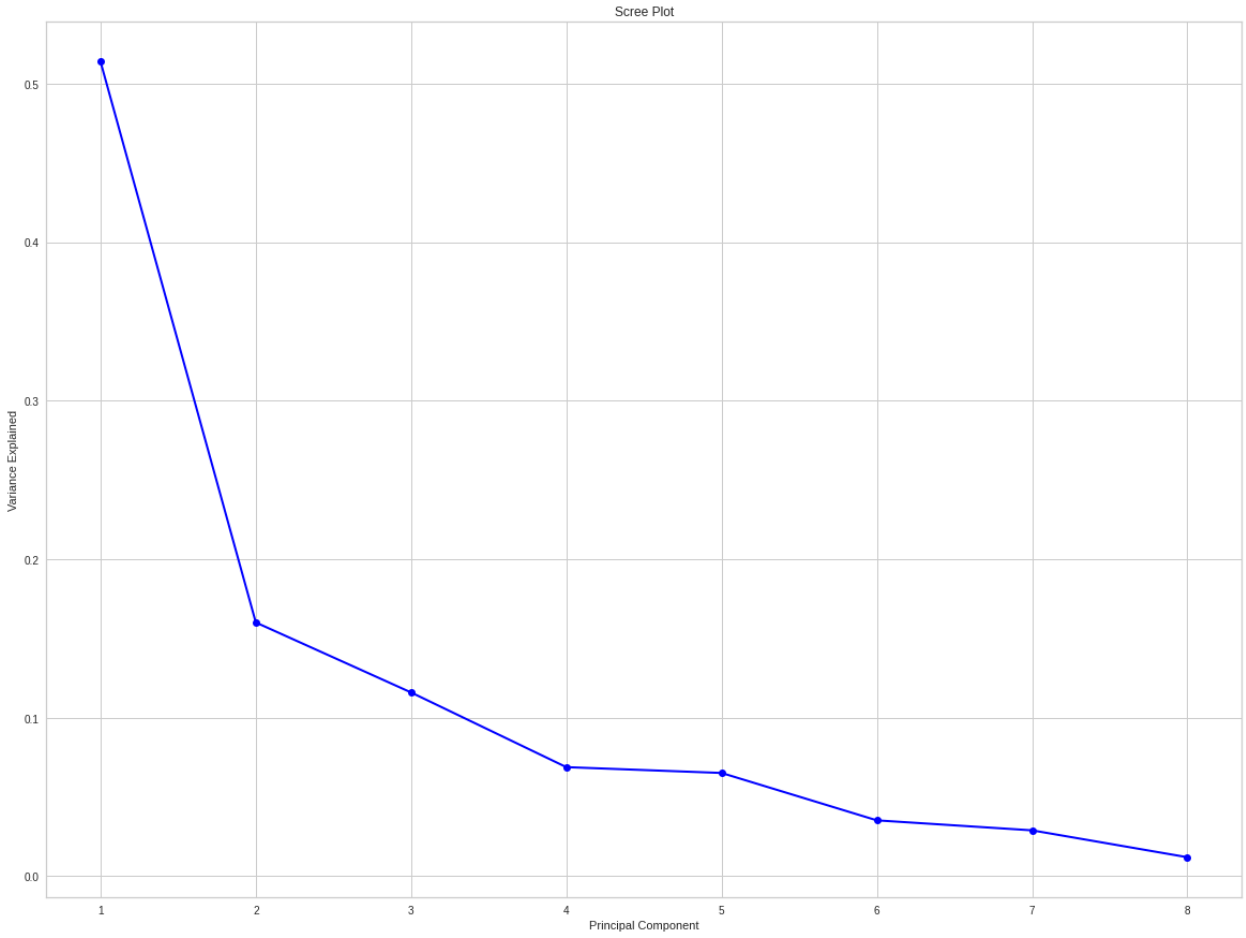
To determine the optimal number of clusters, we used the Elbow method. By varying the number of clusters from 1 to 10, we calculated the Within-Cluster Sum of Square (WCSS) for each value of K. WCSS is the sum of squared distances between each data point and the centroid in a cluster. The WCSS value is largest when K=1 and decreases as the number of clusters increases. When we plot the WCSS against the number of clusters, we observe an elbow shape. The optimal number of clusters is the K value corresponding to the elbow point, where the graph changes rapidly and then moves almost parallel to the X-axis

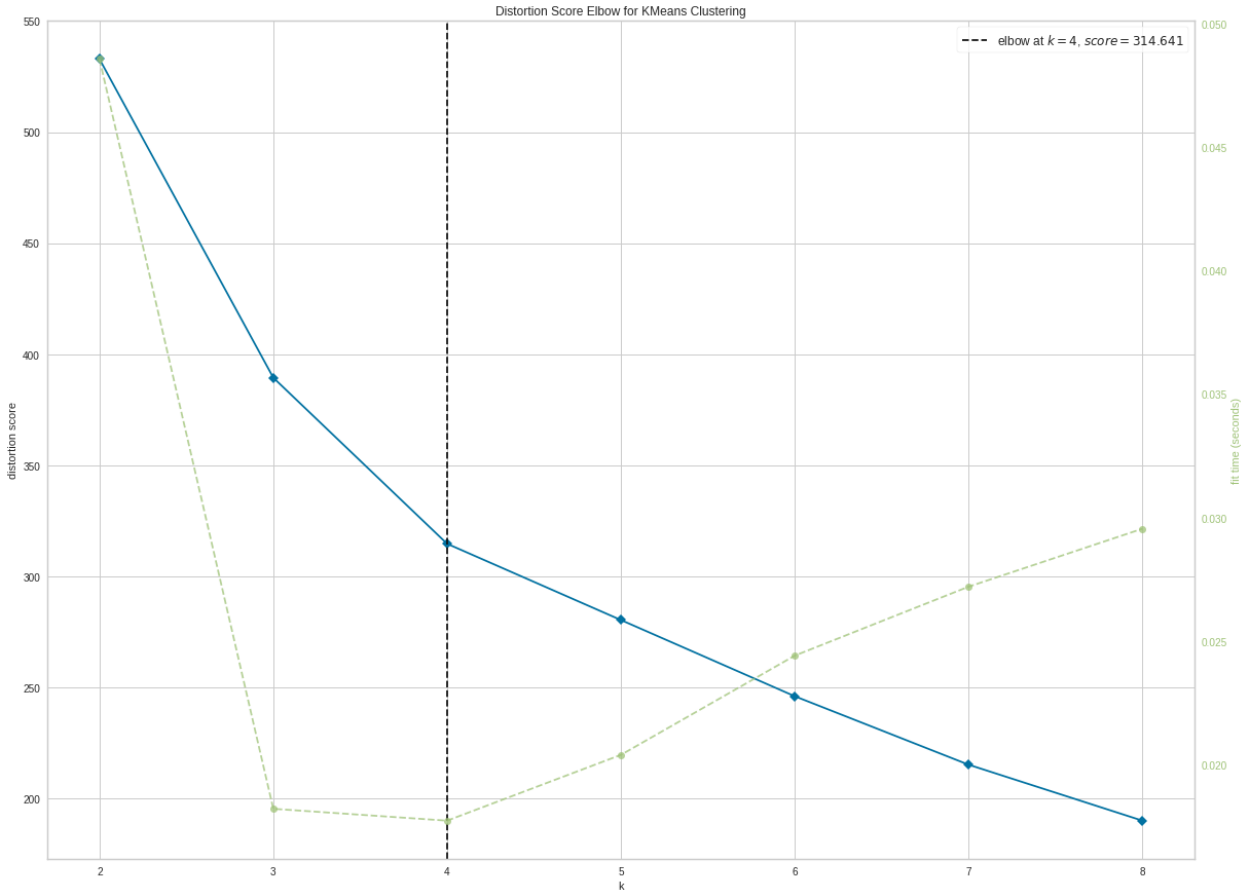
Potential Improvements to Segmentation:

The following approaches could lead to improved results in the segmentation process:

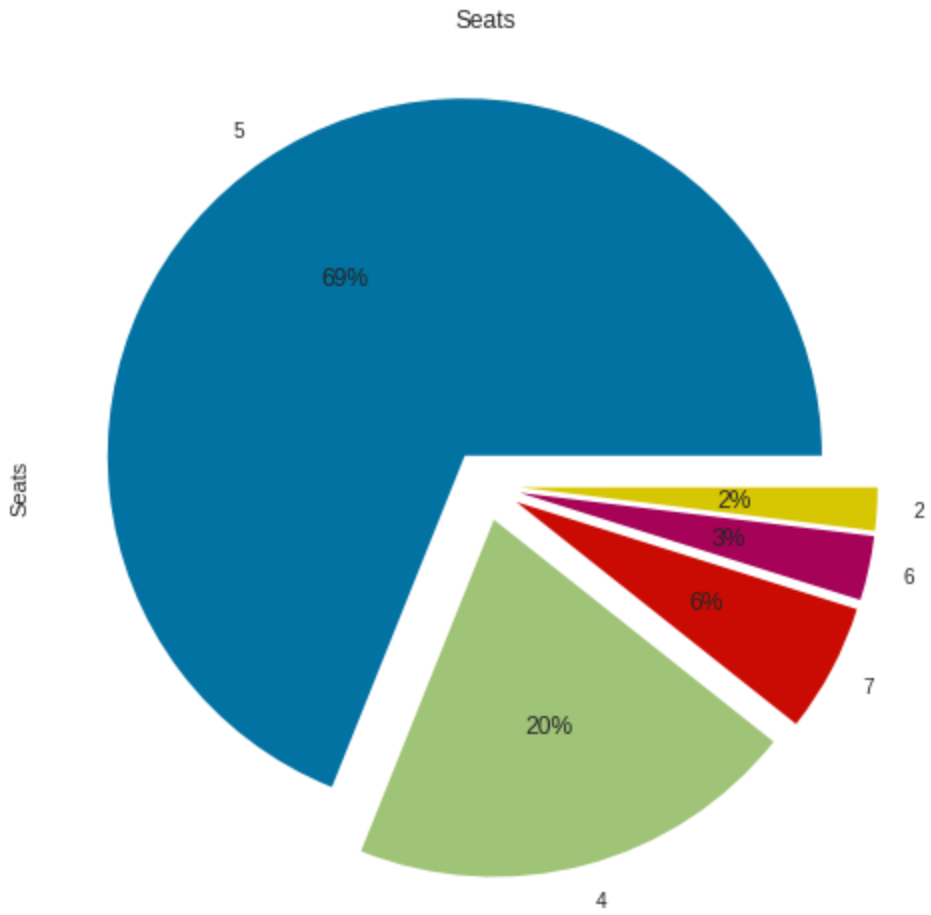
- We could try applying multiple segmentation algorithms on the dataset to identify the optimal result, though this could be a time-consuming process.

- Data collection is a critical factor in obtaining accurate results, and visiting each region to collect high-quality, precise data may yield better results. Additionally,

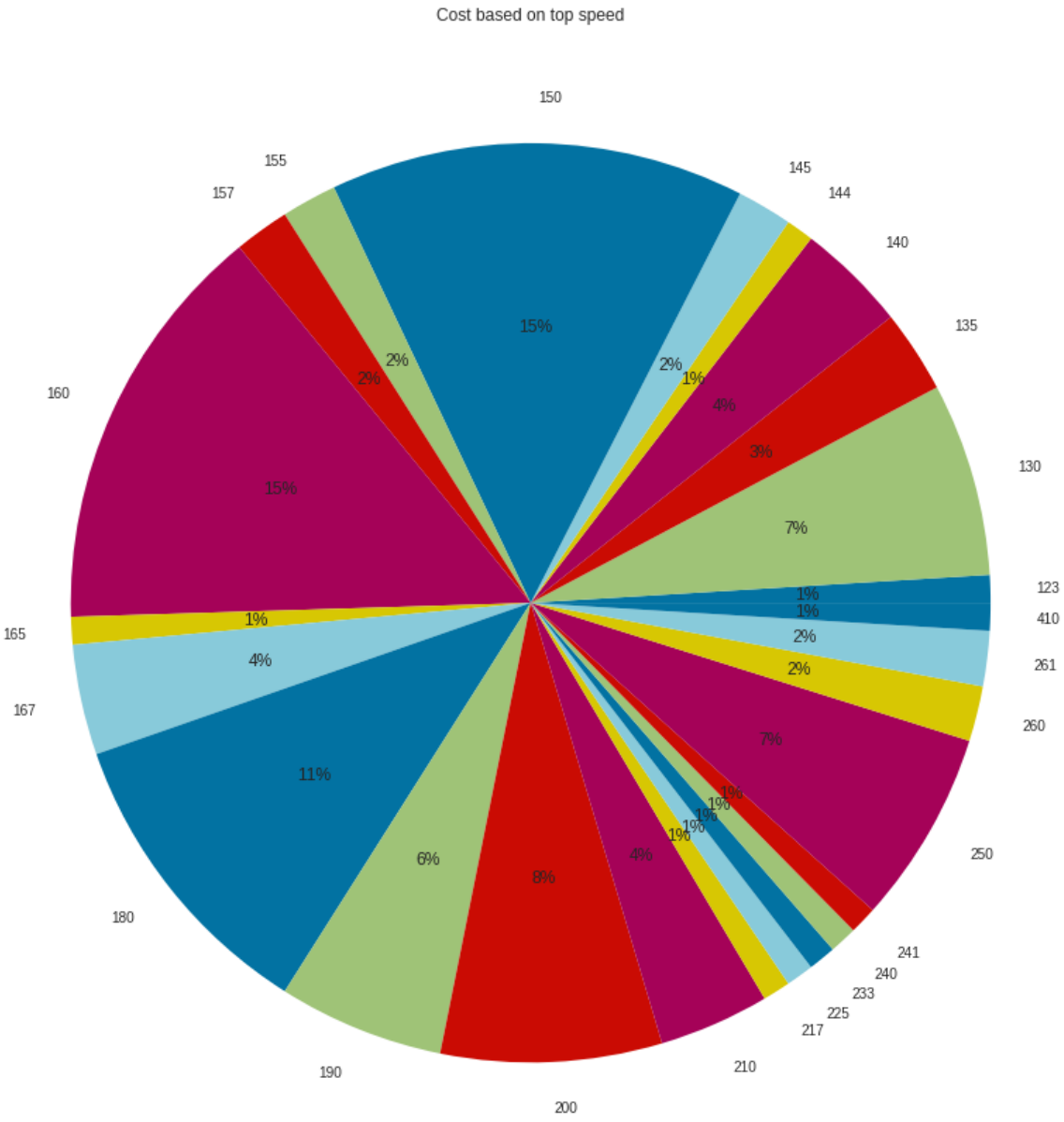



Distortion Score Elbow for KMeans Clustering
elbow at $k = 4$, $score = 314.641$
distortion score
fit time (seconds)
k
550
500
450
400
350
300
250
200
0.050
0.045
0.040
0.035
0.030
0.025
0.020
2
3
4
5
6
7
8

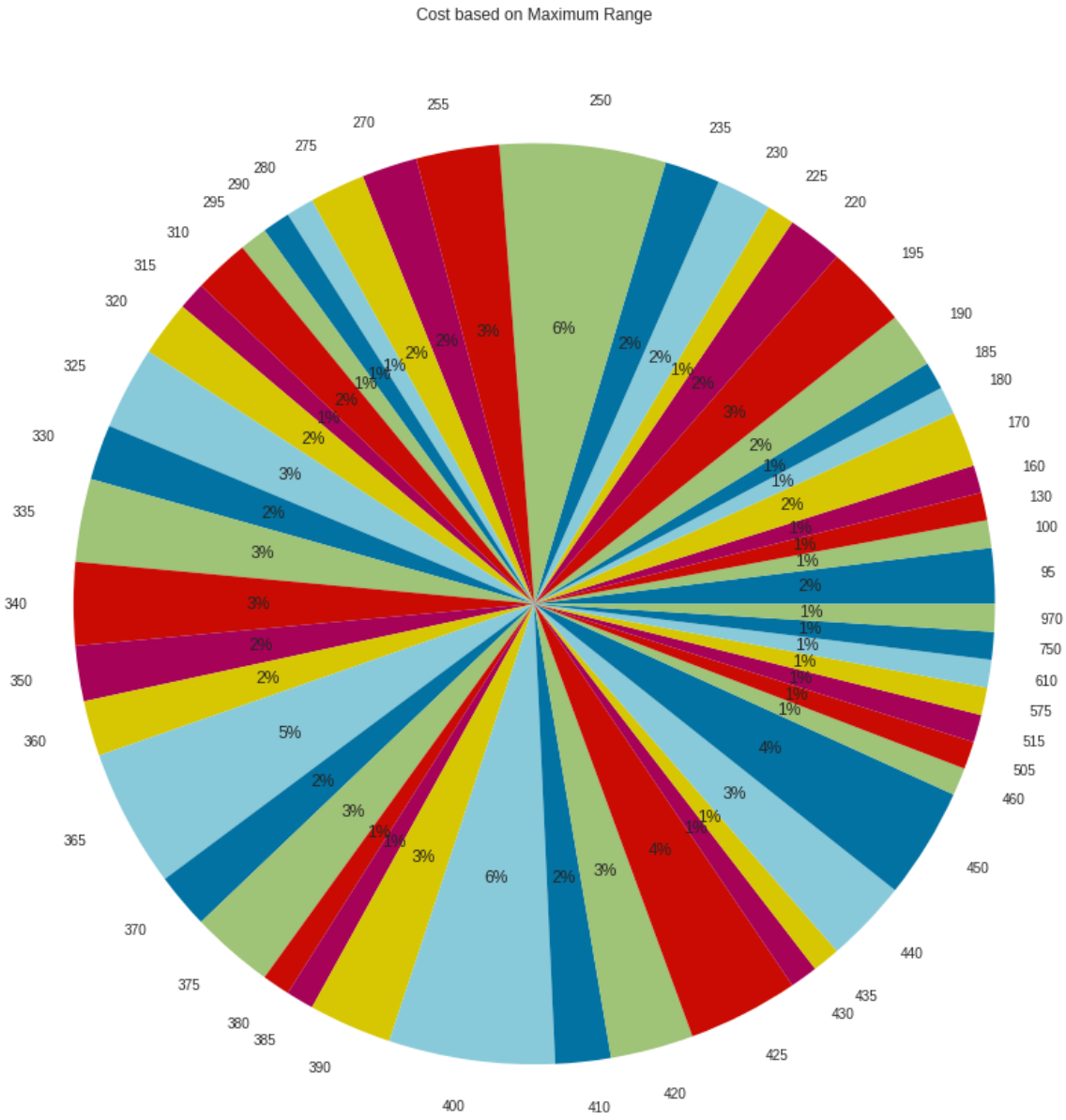
Seats in car



TOP SPEED OF CARS

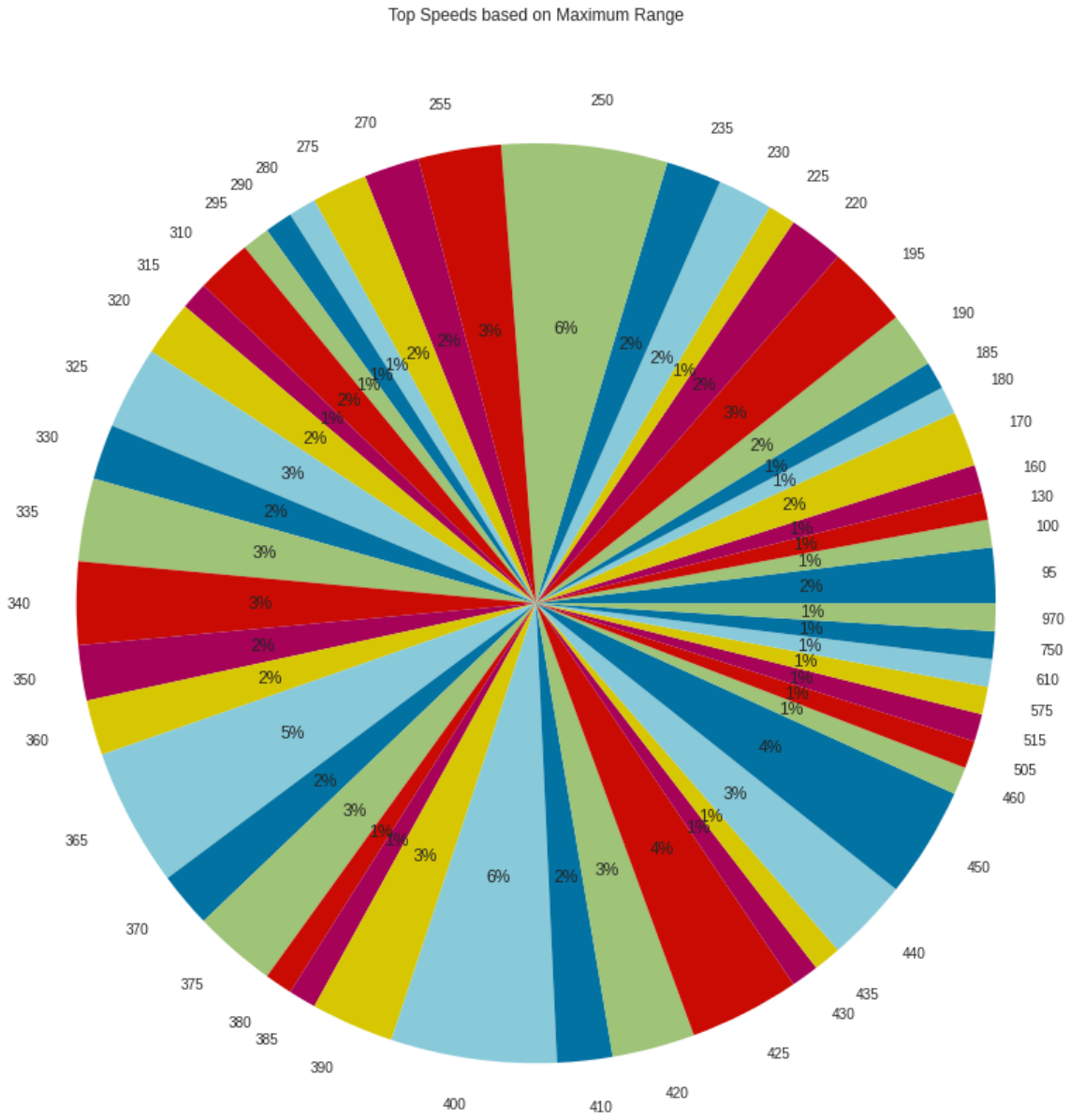


COST BASED ON MINIMUM RANGE



TOP SPEED BY MAXIMUM RANGE

- purchasing a paid dataset could save considerable effort in data collection



GITHUB LINK https://github.com/JayanthAtipamula/EV_MARKET_SEGMENTATION