

## *Project – 2: Market Segmentation Analysis of Electric Vehicle Market in India*



### **Team – Afrin**

#### Team Members

Afrin Siddiqui  
Ajay Nirmal  
Saksham Singh  
Shaik Sirajuddin

*Batch – 17-22-S-B-FLI*  
*FEYNN LABS*

# Index

<b>1. Abstract</b>	<b>3</b>
<b>2. Problem Statement</b>	<b>4</b>
<b>3. Data Collection and Preprocessing</b>	<b>7</b>
<b>4. Segment Extraction</b>	<b>15</b>
<b>5. Profiling and describing potential segments</b>	<b>16</b>
<b>6. Selection of Target Segment</b>	<b>17</b>
<b>7. Customizing the Market Mix</b>	<b>19</b>
<b>8. Potential Customer base in the early market</b>	<b>20</b>
<b>9. Optimal Market Segments</b>	<b>21</b>
<b>10. Code Implementation</b>	<b>22</b>

## 1. Abstract

Customer segmentation in machine learning can help you save money on marketing initiatives by reducing waste. We can target campaigns to the proper people if we know which consumers are similar to each other.

Marketing managers utilize market segmentation as a decision-making process to select a target market for a particular product and develop an effective marketing strategy.

In this essay, the Indian market for electric vehicles has been studied using the concepts and methods of market segmentation analysis.

Electric vehicle adoption varies significantly by region, country and market, but it's all driven by a particular profile of a consumer.

Market segmentation becomes a crucial tool for evolving transportation technology such as electric vehicles (EVs) in emerging markets to explore and implement for extensive 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

## 2. Problem Statement formulation using Fermi Estimation

To analyze the Electric Vehicle market in India using Segmentation Analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use Electric Vehicles.

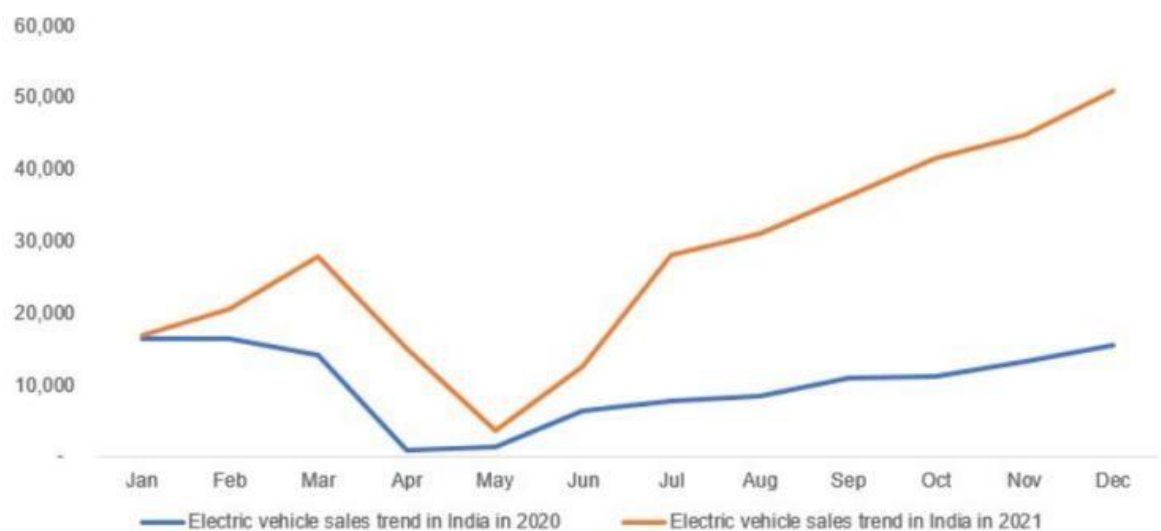


### Background

- **The Indian EVMarket**

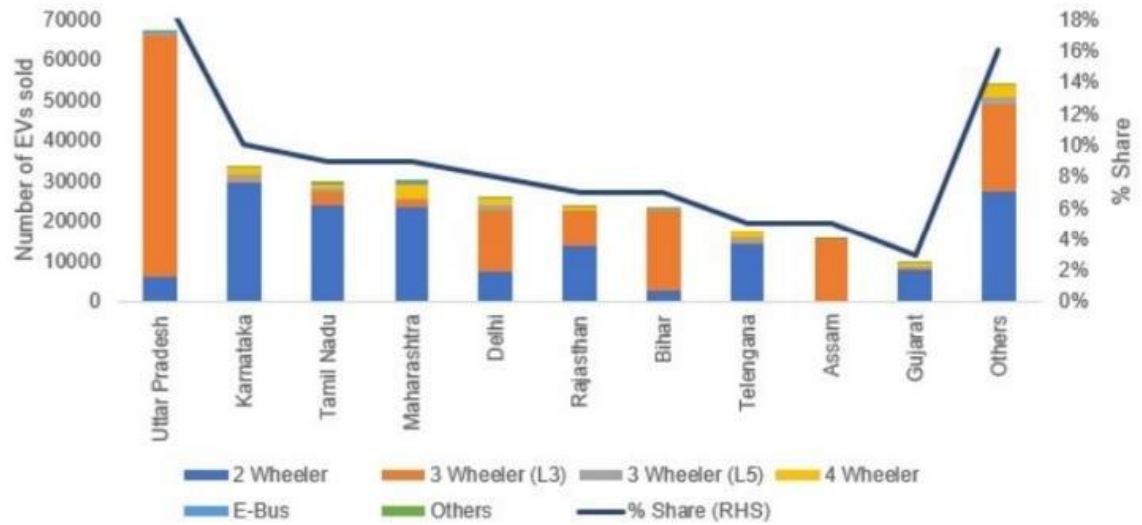
India's EV drive opens up a variety of economic prospects in the infrastructure, energy, and mobility sectors. These potentials include those in the battery, among others.

#### Electric Vehicle Sales Trend in India (2020-21)



By 2030, the EV industry could add 10 million direct jobs, which would result in the creation of 50 million indirect jobs, according to the Ministry of Skill Development and Entrepreneurship (MSDE).

## State -Wise-EV Sales Trend in 2021

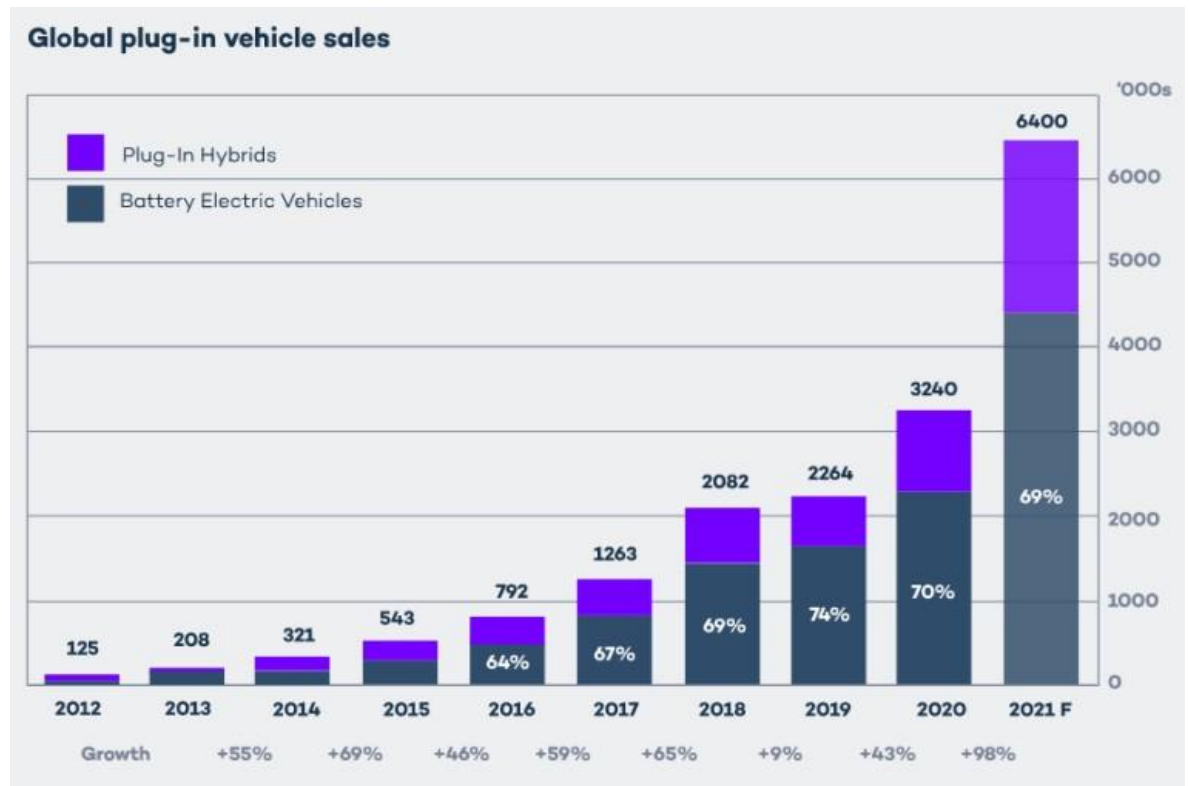


According to the table below, a number of automakers intend to enter the EV market:

Company	EV related plans
Kia	Kia plans to manufacture small SUV EVs in India for global markets in 2025.
Maruti Suzuki	Maruti Suzuki plans to launch its first EV model in India by 2025.
Tata Motors	Tata Motors bags an order worth US\$ 678 million (Rs 5,000 crore) order from the government for electric buses; it plans to launch 10 more EVs in India.
Hyundai	Hyundai plans to launch IONIQ 5 EV in India by the second half of 2022.
Hopcharge	Hopcharge, a Gurgaon- based start-up has created the world's first on-demand doorstep fast charge service.
MG Motors	MG Motors India has partnered with Bharath petroleum for expanding the EV charging infrastructure.
Mahindra & Mahindra	Mahindra and Mahindra targets to launch 16 EV models across its SUV and LCV categories by 2027.

- **Global Trends**

The industry is unmistakably preparing for the ambitious goal of zero emission targets set for 2050, which will be mostly driven by EVs. The trend of EV sales globally is displayed.



- The Indian government has created business opportunities and programs to promote the use of electric vehicles.
  - FAME India Scheme
  - PLI Scheme
  - Battery Swapping Policy
  - Tax exemption of up to Rs.1,50,000 (US\$ 1,960) under section 80EEB of income tax while purchasing an EV (2W or 4W) on loan.
  - Reduction of customs duty on nickel ore (key component of lithium-ion battery) from 5% to 0%.
  - State- wise reduction of road tax and other incentives

### 3. Data Collection and Pre-processing

#### Data Collection

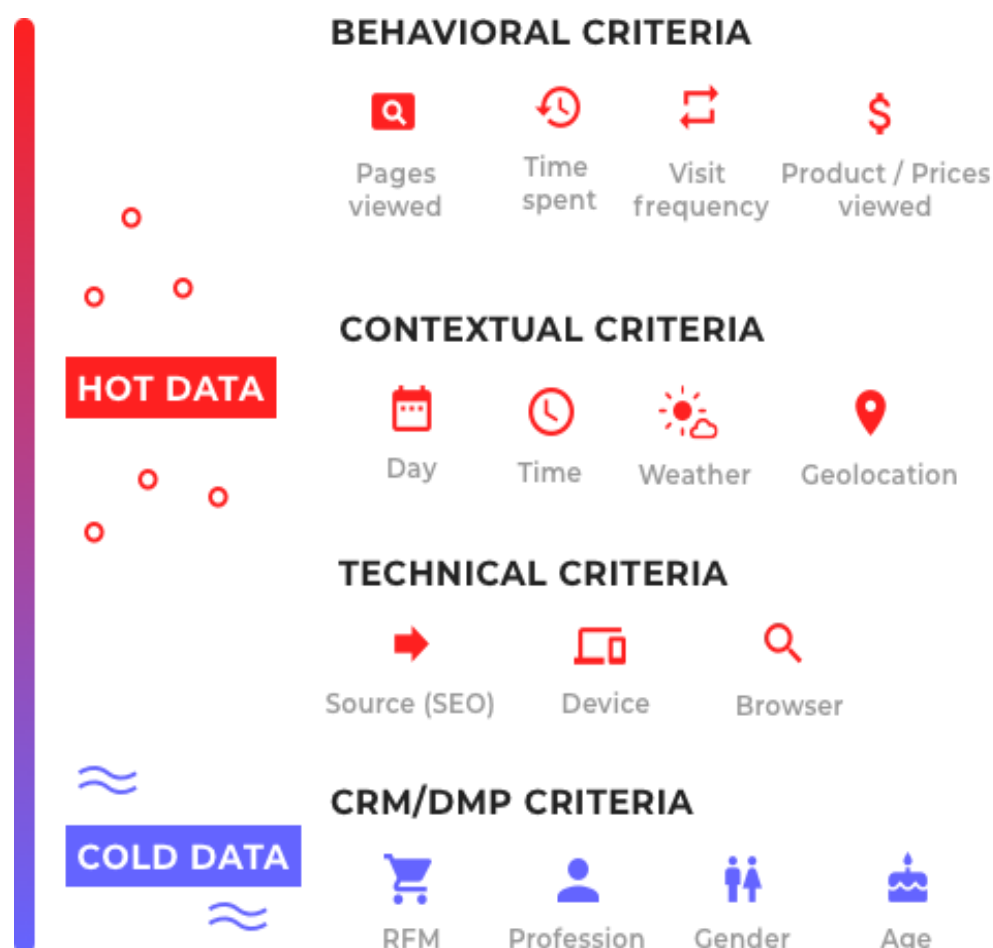
*THE DIFFERENT TYPES OF DATA TO SEGMENT AUDIENCE:*

Online, we distinguish between two types of data:

hot data which is generated by the visit, such as behavioral criteria (browsing journey, history), contextual criteria (geolocation, weather) or technical criteria (browser/device used)

cold data, or historical data, which you can find in your CRM system, such as demographic criteria (age, gender, socio-professional categories) or behavioral data from previous purchases (RFM).

To a large extent, this data overlaps with the segmentation criteria described earlier. Online, it's this information that you will use as segmentation criteria.





## Data Preprocessing

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data pre-processing has the activities like importing datasets, splitting datasets, attribute scaling, etc. Preprocessing of data is required for improving the accuracy of the model.

- **Frameworks and Libraries Used**

- **SKLearn**: Simple and efficient tools for predictive data analysis
- **Seaborn**: Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
- **Plotly**: The plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.
- **KElbowVisualizer**: The KElbowVisualizer implements the “elbow” method to help data scientists select the optimal number of clusters by fitting the model with a range of values for . If the line chart resembles an arm, then the “elbow” (the point of inflection on the curve) is a good indication that the underlying model fits best at that point. In the visualizer “elbow” will be annotated with a dashed line.
- **Matplotlib** : Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.
- **Numpy**: Caffe-based Single Shot-Multibox Detector (SSD) model used to detect faces
- **Pandas**: pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

The snapshot below shows that the dataset was checked for null values and its shape. No null values found. Hence directly moving on to EDA.

```
▼ Information of the type of data in seach column

[ ] df.info()

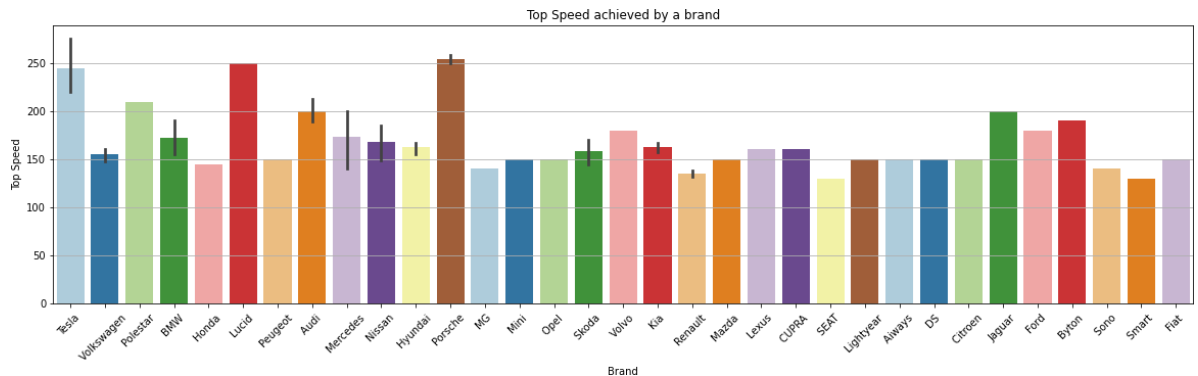
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103 entries, 0 to 102
Data columns (total 15 columns):
#   Column              Non-Null Count  Dtype  
---  -
0   Brand               103 non-null   object 
1   Model               103 non-null   object 
2   AccelSec            103 non-null   float64
3   TopSpeed_KmH        103 non-null   int64  
4   Range_Km            103 non-null   int64  
5   Efficiency_WhKm      103 non-null   int64  
6   FastCharge_KmH      103 non-null   int64  
7   RapidCharge         103 non-null   int64  
8   PowerTrain          103 non-null   object 
9   PlugType            103 non-null   object 
10  BodyStyle           103 non-null   object 
11  Segment             103 non-null   object 
12  Seats               103 non-null   int64  
13  PriceEuro           103 non-null   int64  
14  inr(10e3)           103 non-null   float64
dtypes: float64(2), int64(7), object(6)
memory usage: 12.2+ KB
```



- EDA

Exploratory Data Analysis provides visual clues about whether it is likely to form insightful clusters when combining multiple variables together. It is also an imperative step because choosing an appropriate clustering algorithm is reliant on the shape of the cluster. Some center-based algorithms (e.g. K Means) are more adaptable towards globular shapes clusters and they tend to break linear shaped clusters apart. While density-based algorithms (e.g. DBSCAN) are better at clusters with irregular shape and a lot of noise.

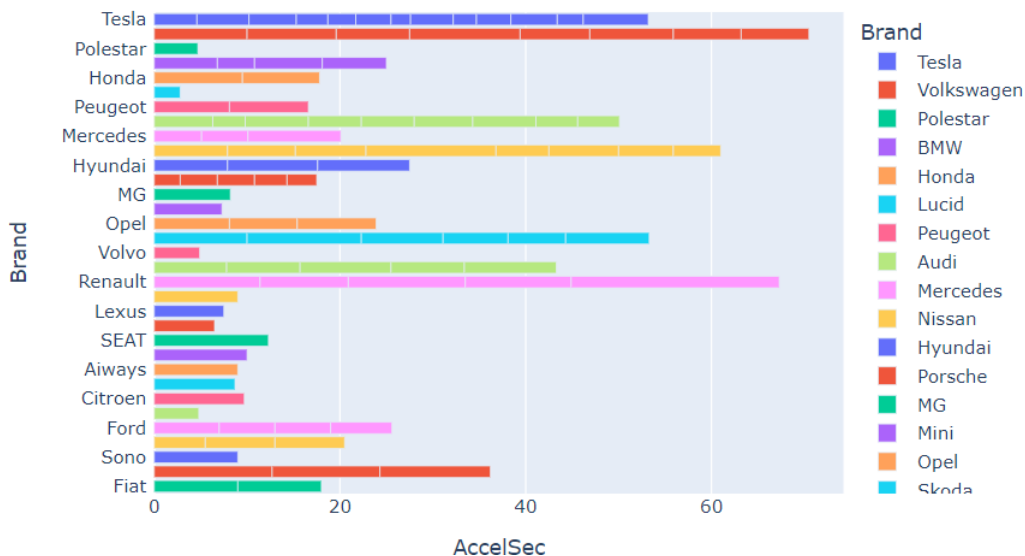
## 1. Top Speed



- Tesla Brand EVs have the top-most speed. The max-speed is 410kmph.

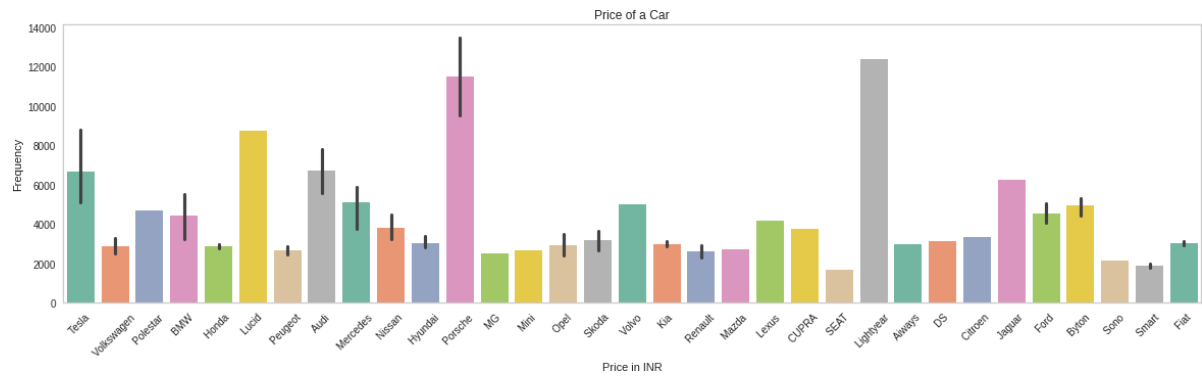
## 2. Acceleration

Which car has fastest acceleration?



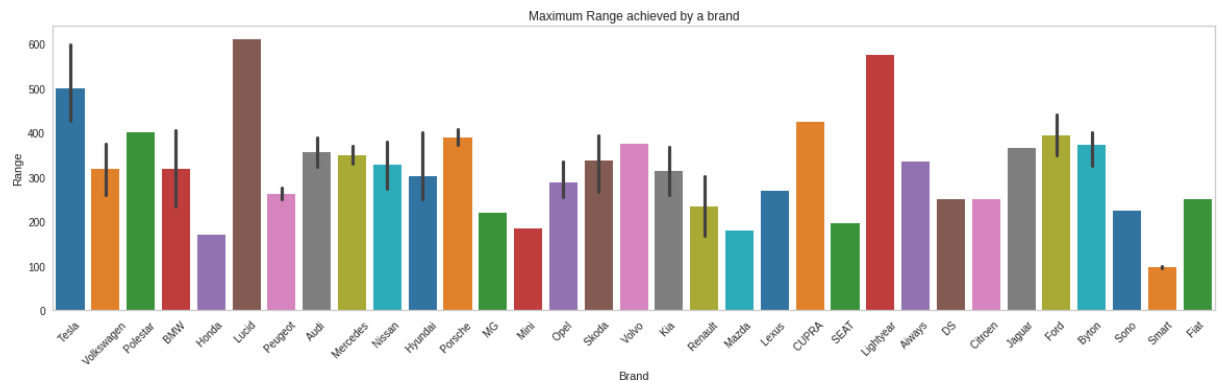
- 12.1 m/sec acceleration is the most common.
- Renault has the max acceleration which is 22.

### 3. Car Price (INR)



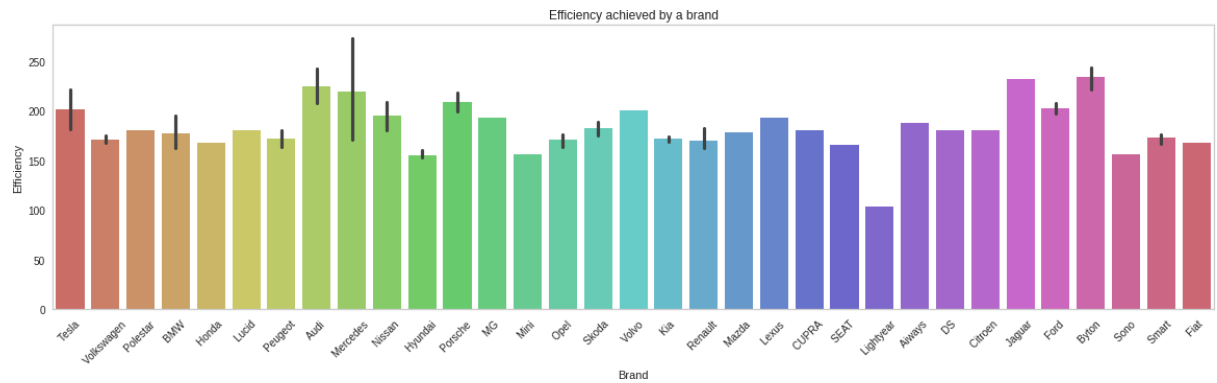
- **Lightyear, Porsche and Lucid are the most expensive.**
- **SEAT and Smart the least expensive.**

### 4. Range



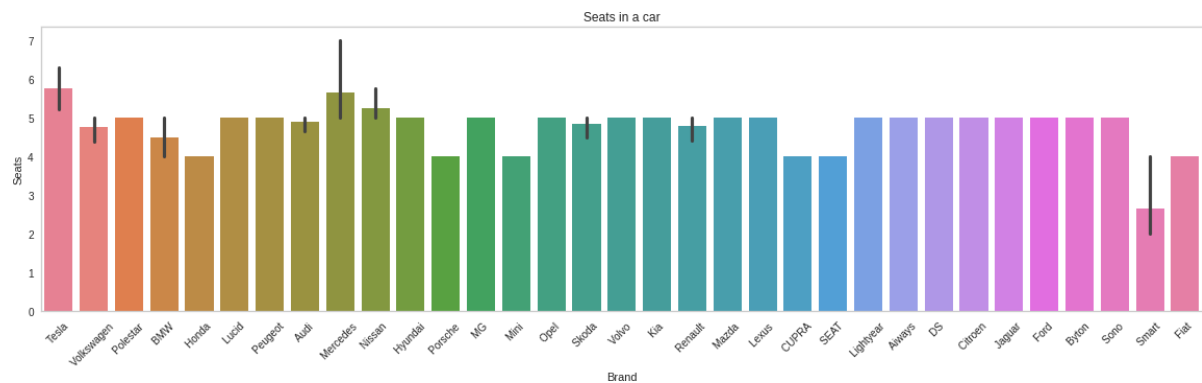
- **Tesla Brand EVs have the max range and Smart the lowest.**

## 5. Efficiency



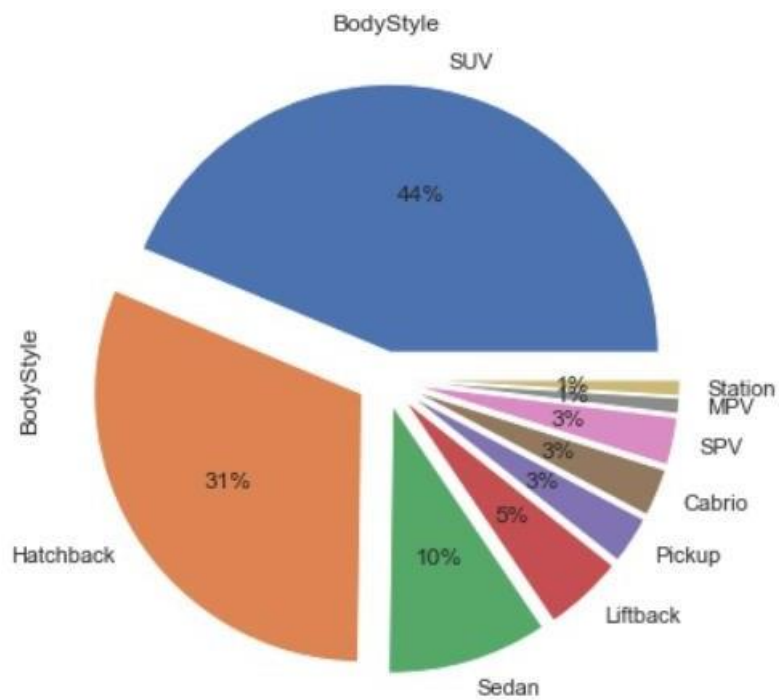
- Byton , Jaguar and Audi have maximum efficiencies and Lightyear has the least.

## 6. No. of seats



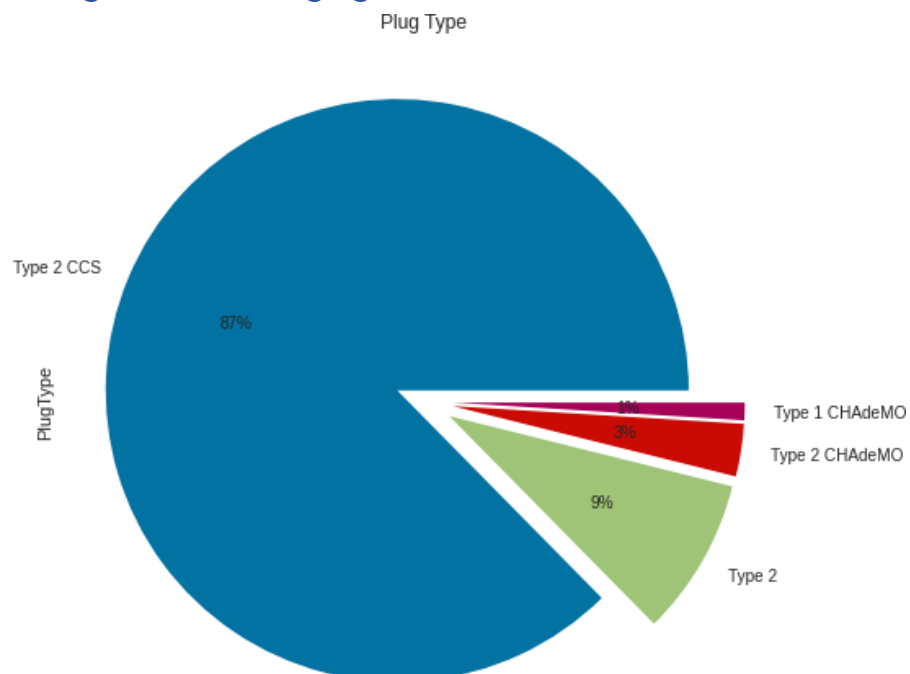
- 5 seat EV is most commonly used.
- Mercedes, Tesla and Nissan have the highest number of seats and Smart the lowest.

## 7. Body Style



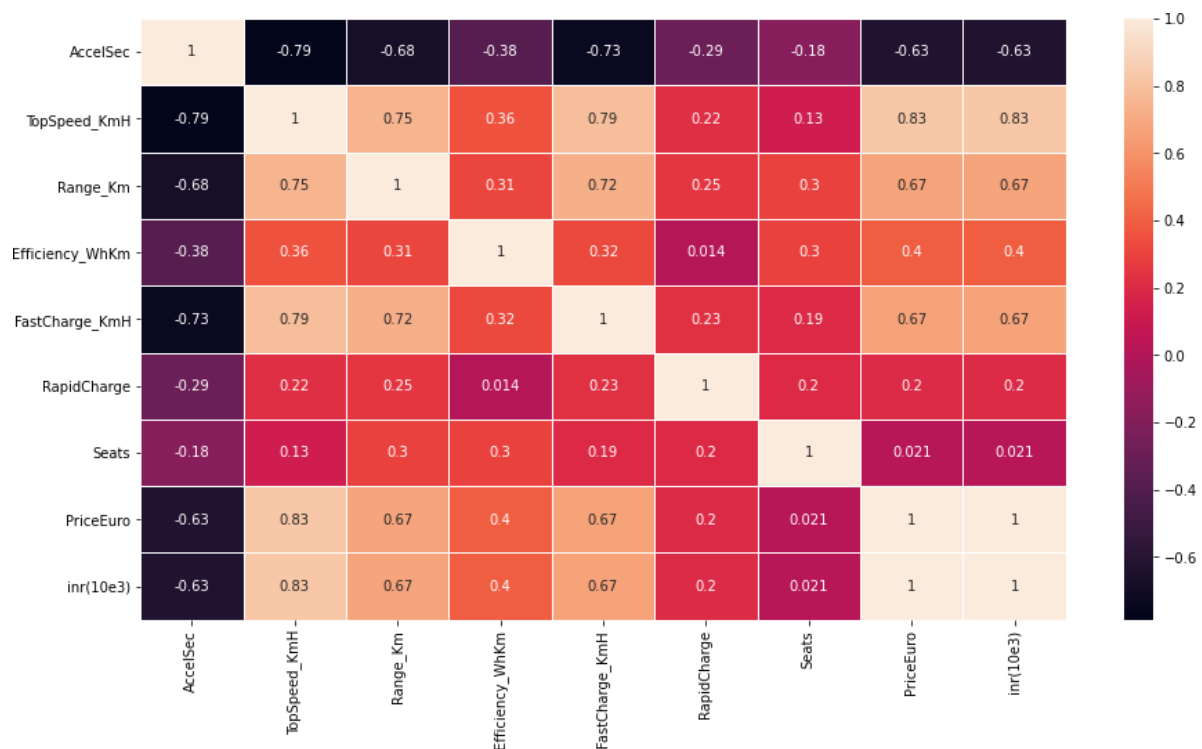
- **Maximum EV's are either SUV or Hatchbacks.**

## 8. Type of Plug used for charging



- **Most companies use Type 2 CCS and Type 1 CHAdeMo the least.**

## Heatmap for the dataset



**Correlation Matrix:** A correlation matrix is a table containing correlation coefficients between variables. Each cell in the table represents the correlation between two variables. The value lies between -1 and 1. A correlation matrix is used to summarize data, as a diagnostic for advanced analyses and as an input into a more advanced analysis.

The value of the correlation coefficient can take any values from -1 to 1.

**If the value is 1,** it is said to be a **positive correlation** between two variables. This means that when one variable increases, the other variable also increases.

**If the value is -1,** it is said to be a **negative correlation** between the two variables. This means that when one variable increases, the other variable decreases.

**If the value is 0,** there is no correlation between the two variables. This means that the variables changes in a random manner with respect to each other.

### **Inference:**

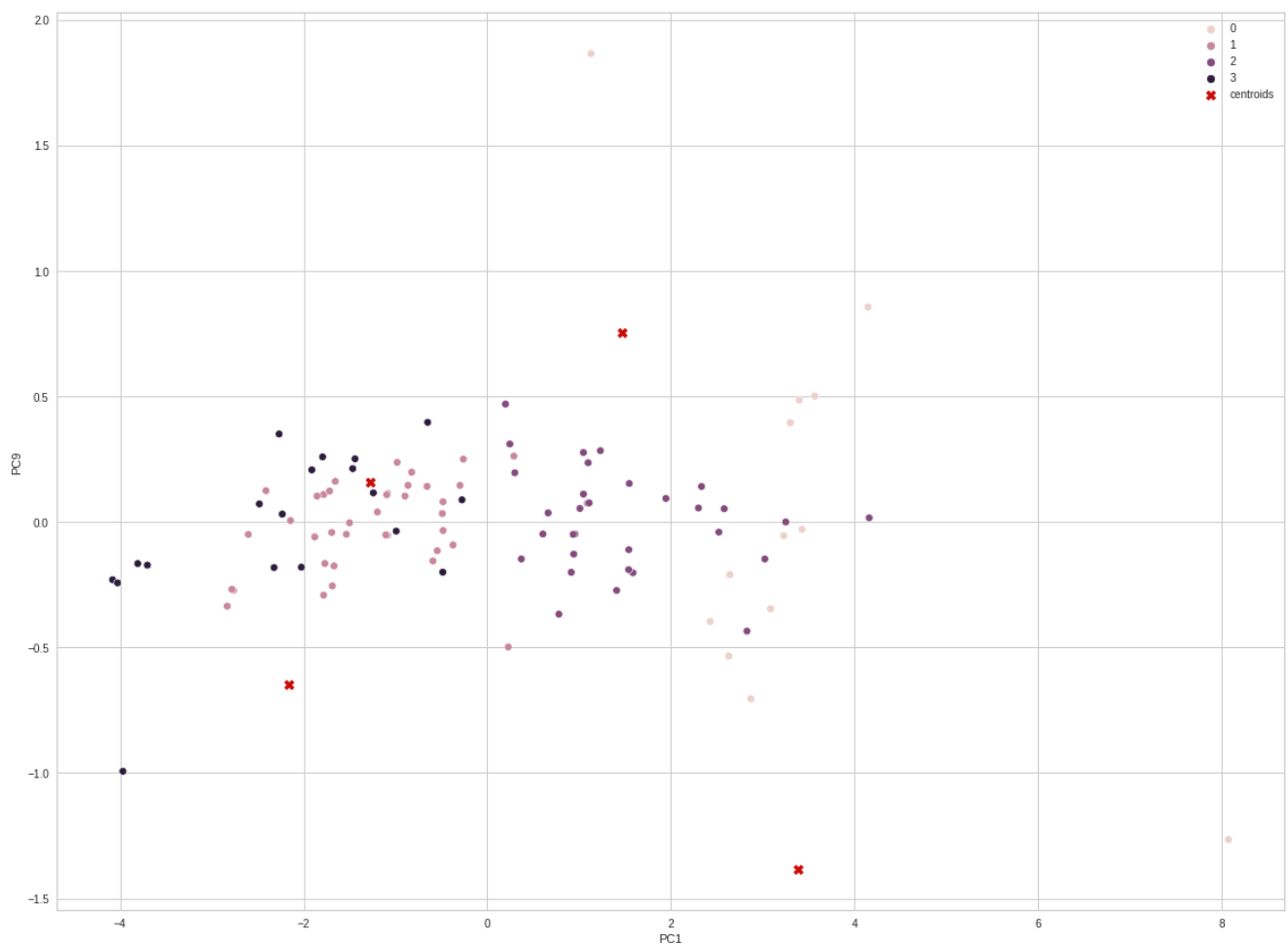
- PriceEuro, Inr(10e3) are highly correlated with TopSpeed\_KmH
- TopSpeed\_KmH, Range\_Km are also correlated to feature RapidCharge

## K - Means Clustering Algorithm

K Means Clustering is a centre-based clustering algorithm, which means that it assigns data points to clusters based on closeness or distance, following these procedures:

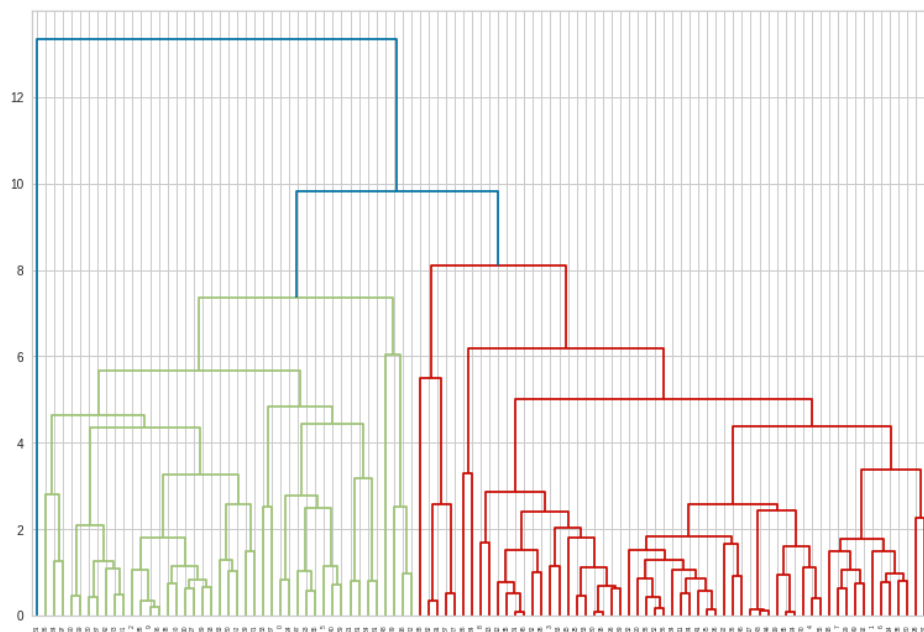
- i. Specify the number of clusters “K”.
- ii. Initiate K random centroids and assign each cluster a centroid: Centroid is the center of each cluster. There are k data points randomly selected as the centroids at the beginning, and the cluster label of other data points are later defined relatively to them. Consequently, different initial centroid assignments may lead to different cluster formations.
- iii. Form K clusters by assigning data points to the closest centroid: The closest centroid is usually defined by the smallest Euclidean distance but it can also be correlation or cosine depends on the use cases.
- iv. Recompute the centroid of each cluster : After all data points have been assigned to a cluster, for each cluster we recalculate the mean of all data points belonging to that cluster and define it as the new centroid.
- v. Reach convergence when centroids no longer change: Iterate step 2–4 until reaching the stopping criteria that either the centroids no longer change or the maximum number of iterations are reached.

This procedure also determines that K Means algorithm has the limitation of clustering points into circular shapes with similar size and it is also heavily reliant on the predefined number K.



## 4. Segment Extraction

**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

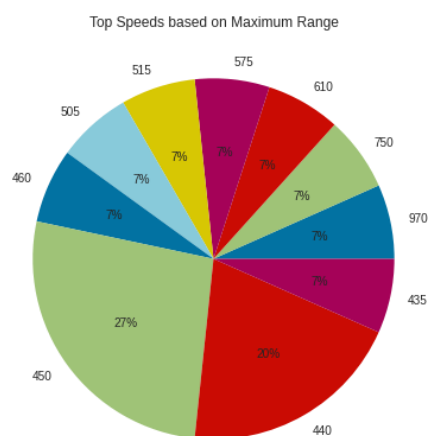
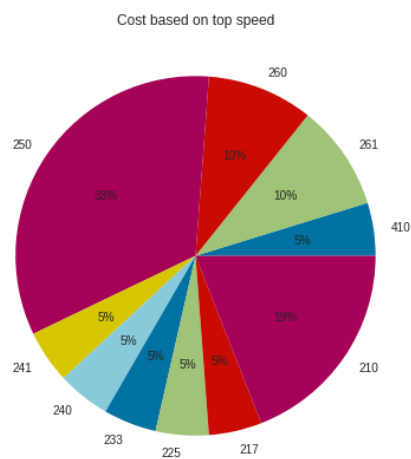
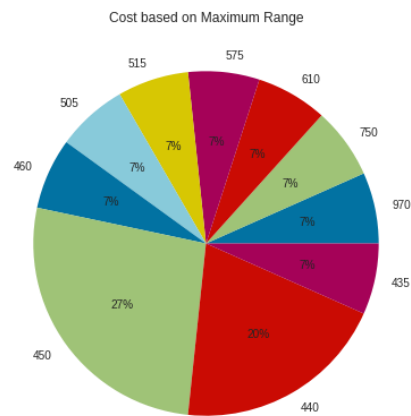


The clades are arranged according to how similar (or dissimilar) they are. Clades that are close to the same height are similar to each other; clades with different heights are dissimilar — the greater the difference in height, the more dissimilarity.



## 5. Profiling and describing potential segments

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

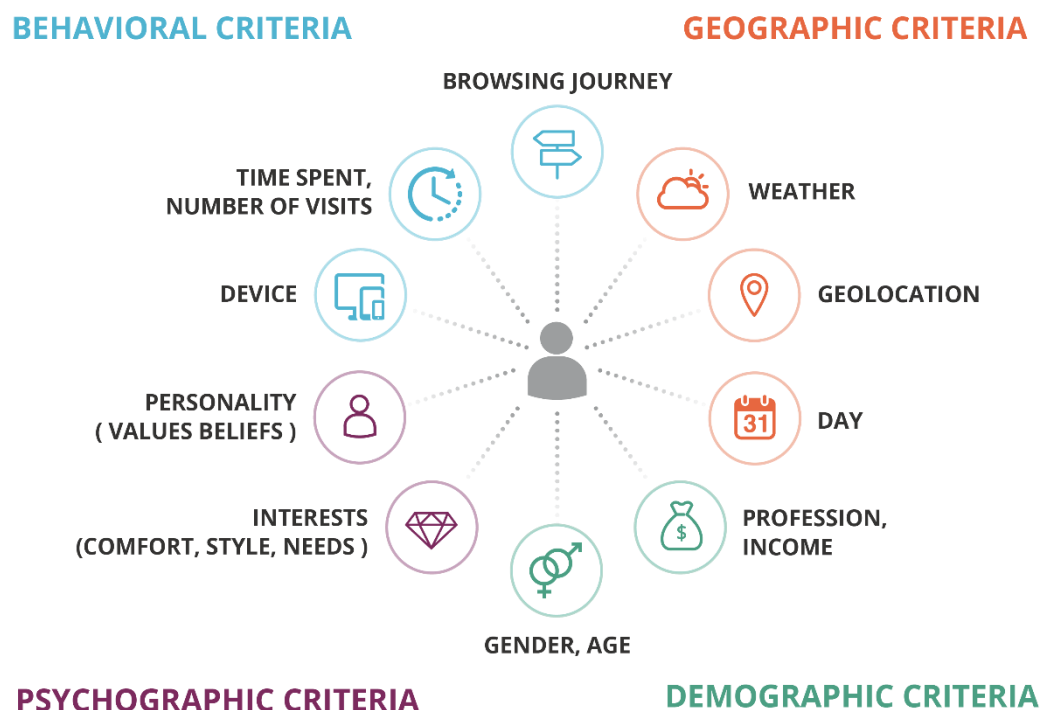


## 6. Selection of Target Segment

Segmentation involves creating homogenous groups made up of individuals with identifiable common characteristics. These might be place of residence, age, lifestyle or even how they behave on your website: these are what we call segmentation criteria.

The individuals within a same segment are supposed to have the same expectations and should react in a similar way to an offer, type of content or a specific message.

By cross-referencing different types of data, you can obtain a more in-depth analysis of your customers. Let's review the most common segmentation criteria.



### *i. GEOGRAPHIC CRITERIA*

This type of segmentation is based on the geolocation of your visitors and is one of the simplest criteria. It enables you to target your marketing actions based on where your visitors are or on weather conditions.

### *ii. DEMOGRAPHIC CRITERIA*

Demographic segmentation is the most commonly used criteria, since it requires information that can be collected easily and that enable you to quickly target a potential market. These criteria include gender, age, nationality, education, profession, income or family situation.

### *iii. PSYCHOGRAPHIC CRITERIA*

Psychographic segmentation focuses on the lifestyle of visitors: their interests, personalities, values, beliefs and opinions. To obtain this kind of information, you'll usually need to have your visitors complete questionnaires or surveys.

#### **iv. BEHAVIORAL CRITERIA**

Behavioral segmentation relies on the way visitors interact with the website. Some data depends on their immediate online behavior (online) while other data depends on their past offline behavior (offline) when dealing with the brand.

- Online: time spent on the website, pages visited, point of exit, purchase opportunity (urgent or not), purchase attitude, brand loyalty (registering for newsletters), search engine and device used, traffic source, etc.
- Offline: number of visits, purchase history, date and amount of latest orders (RFM).

## Target Segments

As the analysis suggests 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.***

## 7. Customizing the Market Mix



### Product

Specifying the product in light of client wants is one of the important choices a company must make when creating the product dimension of the marketing mix. Frequently, this involves tweaking an existing product rather than developing a brand-new one. In addition to branding and packaging, the product's packaging, whether or not warranties are offered, and whether or not after-sale support services are provided are also marketing mix considerations that fall under the product dimension.

### Price

When constructing the price dimension of the marketing mix, a firm must frequently decide how much to charge for a product and whether to offer discounts.

### Place

The key choice connected to the place dimension of the marketing mix is how to distribute the product to customers. This includes selecting whether the creator should sell directly to clients, whether a wholesaler, retailer, or both should be used, as well as whether the product should be offered exclusively online, exclusively offline, or both.

### Promotion

Making an advertising message that will appeal to the target market and choosing the best way to deliver it are common promotion decisions made while establishing a marketing mix. Personal selling, sponsorship, and public relations are all examples of marketing mix promotion tools.

- ***For our market strategy we can select location like Bangalore and other metropolitan cities where we find a variety of well – informed and young consumers interested in the EV market.***

## 8. Potential Customer Base in Early Market

Predictive targeting involves using a machine learning algorithm to analyze website visitor data, enabling brands to identify which visitors belong to a particular target segment. By observing visitor behavior and the correlations between them, the algorithm learns how to predict the behavior of each visitor. Eventually, it determines – with increasing accuracy over time – which visitor belongs to which segment.

Through machine learning, the predictive approach offers superior segmentation, since it is capable of learning from the behavior of visitors to offer them ever more relevant experiences depending on their browsing journey. And it can do this even if it's their first visit to the website.

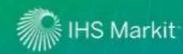
Predictive targeting is used to identify segments that cannot be found using the a priori approach.

It is essential when: the criteria that define the segment are vague, ambiguous or changeable (“I want to identify undecided visitors”, “I want to know which of my visitors will be influenced by scarcity effects”) the criteria that define the segment are too numerous and dissimilar, and impossible to model by hand.

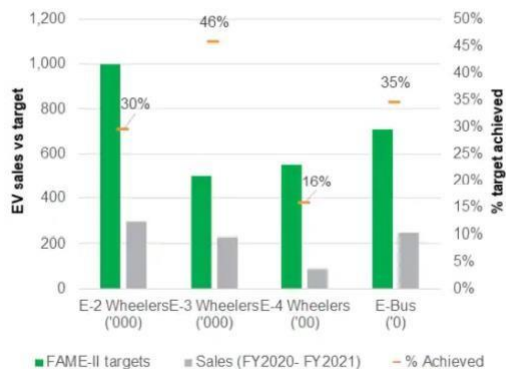
In these situations, predictive targeting enables you to identify, with certainty, a greater number of visitors belonging to a segment that you're seeking to target.

## 9. Optimal Market Segments

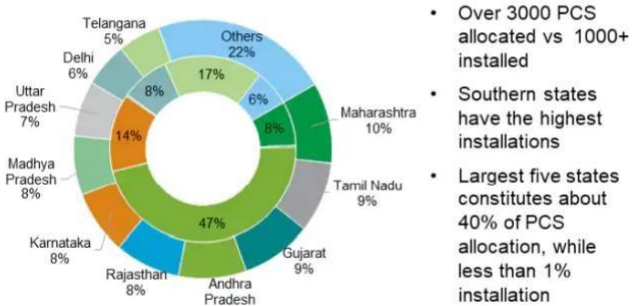
### India's preparedness for large-scale EV adoption



EV sales versus FAME-II target



PCS allocated versus installed



- Over 3000 PCS allocated vs 1000+ installed
- Southern states have the highest installations
- Largest five states constitutes about 40% of PCS allocation, while less than 1% installation

Note: FAME - Faster Adoption and Manufacturing of Electric (and Hybrid) Vehicles; PCS - Public Charging Stations  
Outer circle in the doughnut chart represents the state-wise distribution of allocated PCS, and the inner circle state-wise distribution of installed PCS  
Source: IHS Markit

© 2021 IHS Markit

Information contained in this infographic is part of IHS Markit Asia-Pacific Climate and Sustainability Capability

- Maharashtra and Andhra Pradesh have the highest number of EV charging stations sanctioned, but Andhra Pradesh has the most ambitious EV plan and Punjab the most balanced.
- Focusing on domestic manufacturing and research in battery and EV components will bring investments, jobs, and opportunities to serve demands.
- Tesla, Audi, Mercedes have max efficiencies and SUV, Pickup and SPV body styles have max efficiencies.

**Considering that 21 of the top 30 most polluted cities are in India, as well as the skyrocketing demand and price for oil, EVs are likely to become a need rather than a choice in the years to come.**

## 10. Code Implementation

<https://github.com/afrin7/Market-Segmentation-Analysis-of-EVs-Market-in-India.git>

\*\*\*