**Electric Vehicle Market Segmentation**

Market segmentation is an important strategy for emerging markets to study and deploy growing transportation technologies such as electric vehicles (EVs) to achieve widespread acceptance. EV usage is predicted to skyrocket shortly as a low-emission and low-cost vehicle, sparking a significant amount of future academic study interest. The primary goal of this study is to investigate and identify unique sets of possible EV buyer groups based on psychographic, behavioral, and socioeconomic characteristics, using an integrated research framework of 'perceived benefits-attitude-intention'.

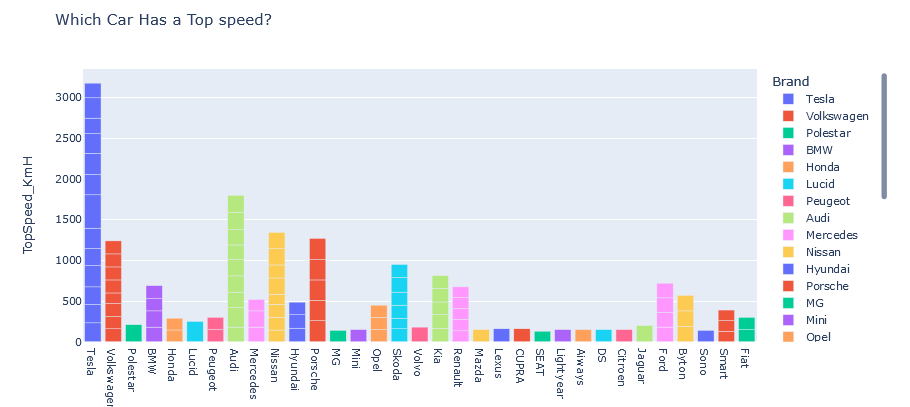
Libraries:

* [**SKLearn**](https://scikit-learn.org/stable/)**:** Simple and efficient tools for predictive data analysis
* [**Seaborn**](https://seaborn.pydata.org/)**:** Seaborn is a Python data visualization library based on Matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
* [**Plotly**](https://plotly.com/python/getting-started/)**:** The Plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering various statistical, financial, geographic, scientific, and 3-dimensional use cases.
* [**Matplotlib**](https://matplotlib.org/)**:** Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.
* [**Numpy**](https://numpy.org/)**:** Caffe-based Single Shot-Multibox Detector (SSD) model used to detect faces
* [**Pandas**](https://pandas.pydata.org/)**:** 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.
* Statsmodels: Python library that lets you analyze data, estimate statistical models, and run statistical tests. There is a comprehensive set of descriptive statistics, statistical tests, charting routines, and outcome statistics for each type of data and estimator.

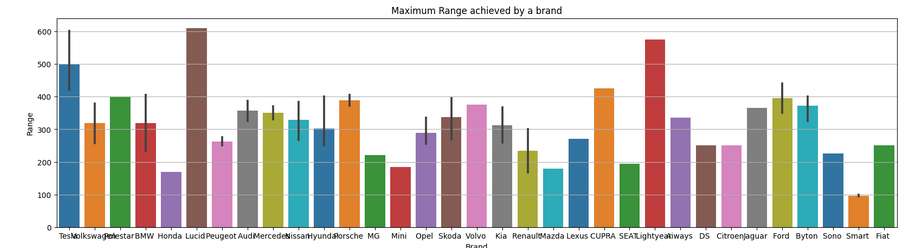
Data Preprocessing: Data pre-processing is a critical stage in the development of a machine learning model. Initially, data may not be clean or in the appropriate format for the model, resulting in misleading results. Data pre-processing involves transforming data into the format that we require. It is used to deal with noise, duplication, and missing values in a dataset. Data pre-processing operations include importing, dividing, scaling, and scaling attributes. Preprocessing of data is essential to improve model accuracy.

## Download: dataset [here](file:///E:\ElectricCar.csv)

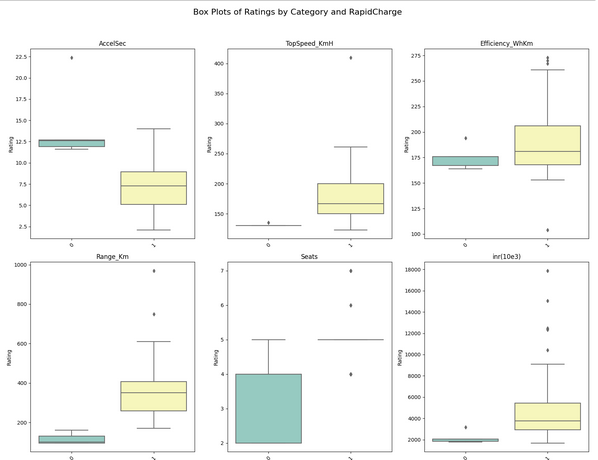
Results:



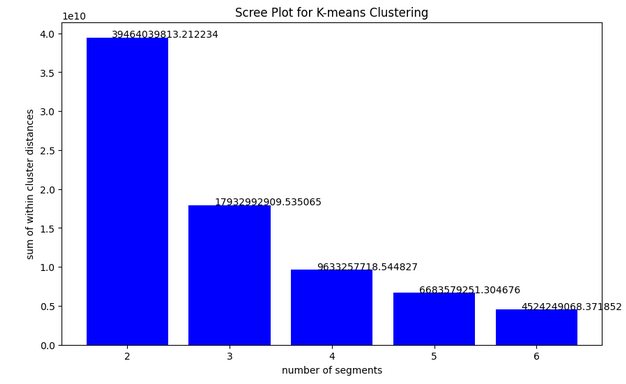
There are different types of brands and models in the dataset, the above visual helps us in finding the top speed with respect to the brand to give appropriate output. In the above chart, Tesla is on the top with high speed following with Audi and Volkswagen. This brand in market has huge demand for its speed.

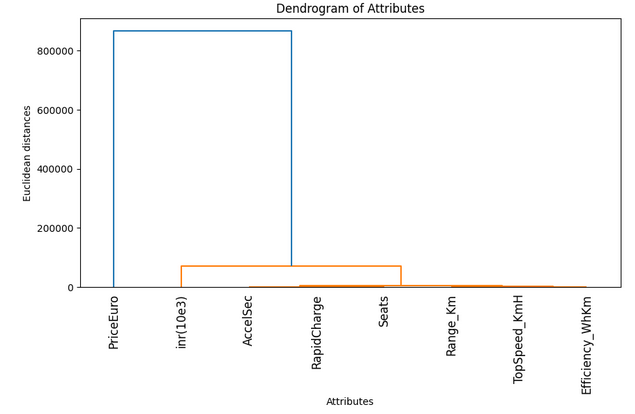


Each brand vehicle has some range, based on the mileage of the brand an electric vehicle can be sold. So, here the Lucid has around 600+ range followed by Light year which is 590, and then Tesla. Finding the maximum range of the brands of electric vehicles will know the high mileage for future data to calculate which brand vehicle is better.

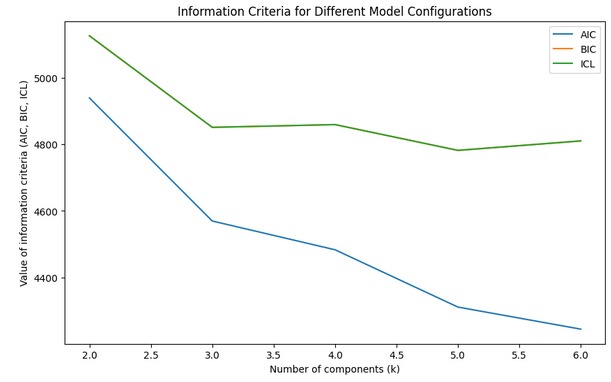


Based on Rapid Charge the following columns show the performance, the columns are AccelSec, TopSpeed, Efficiency, Range, Seats, and Price. As observed the Rapid Charge is high for most of the following data but least at seats.

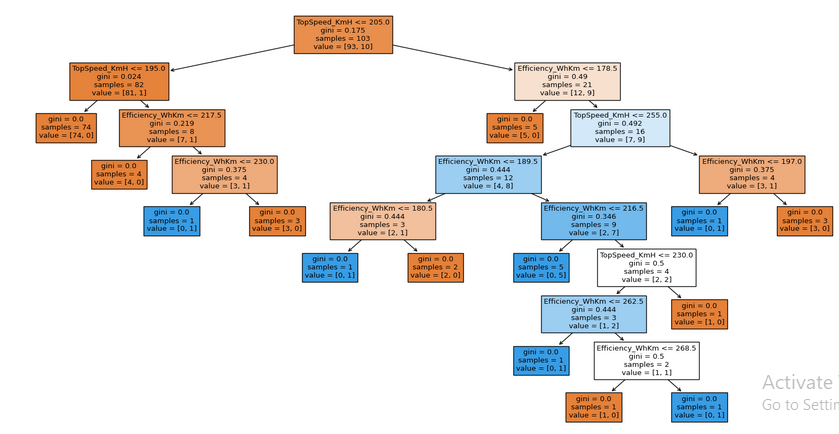


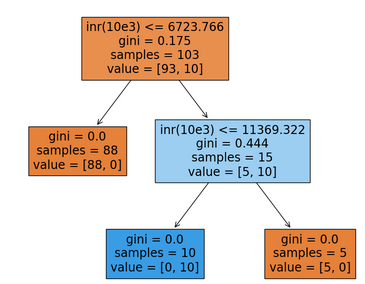
KMeans clustering on a dataset, assesses the stability of the clustering, and analyzes the resulting clusters. The results are stored in various data structures for further analysis and comparison. The scree plot helps identify the "elbow" or the point where the decrease in inertia becomes less significant, indicating that adding more clusters may not significantly improve the clustering quality.

In the above visual we see there are two major clusters they are PriceEuro and other involves AccelSec, RapidCharge, Seats, Range\_km, Top Speed\_kmH,Efficieny\_whkm.The attributes in x-axis and Euclidean distance in y-axis, where the major part is divided by inr and seats, further seats are divided into 5 subdivisions they are accel sec, rapid-charge, range\_km,topspeed\_kmh, and efficiency. The distinct attribute is the price of Euro as it's just a conversion of price into euro. This shows how the Dendrogram has created a relationship between clusters and attributes.



Information criteria for the mixture models of binary distributions with 2 to 6 components(segments) for the Electric Vehicle data set. The information criteria values AIC, BIC and ICL on the *y*-axis for the different number of components (segments) on the *x*-axis. As can be seen, the values of all information criteria decrease quite dramatically until three components (market segments) are reached. If the information criteria are strictly applied based on statistical inference theory, the ICL recommends – by a small margin – the extraction of six market segments. The BIC also points to six market segments. The AIC values decrease beyond six market segments, indicating that at least seven components are required to suitably fit the data. Three market segments might be a good solution if a more pragmatic point of view is taken; this is the point at which the decrease in the information criteria flattens visibly.





**Conclusion**:  
The first division in the decision tree, "rapid charge," is the most significant predictor for data splitting.   
• After 'seats' is the second most used element in decision-making within the tree, indicating its significant impact on outcomes.   
• 'Range\_Km' is a splitting property that appears after ‘rapid charge ‘ and 'driven', showing its relevance in finer categorizations.  
The highest speed is Tesla (410), high-frequency Byton. Maximum Range Lucid and light-year, Plug type 2 CCS, body shape is hatchback and SUV, preferred seats are 5, the rapid charge is excellent in efficiency. The clusters are separated into six segments. The optimal number of clusters in terms of stability: 2. Stability score.   
Decision tree This tree classifies samples based on a few features:  
• Top speed (km/h) • Range (km) • Efficiency (Wh/km) • Price (Euro) • INR (10e3)  
The tree measures the quality of the splits using the Gini impurity index; a Gini index of 0 denotes a complete separation of classes.

• The feature 'TopSpeed\_KmH' is the most important factor in deciding sample categorization in this decision tree model, indicating that it may be the most relevant predictor.

• The first division in the decision tree, "rapid charge," is the most significant predictor for data splitting.   
• After 'seats' is the second most used element in decision-making within the tree, indicating its significant impact on outcomes.   
• 'Range\_Km' is a splitting property that appears after ‘rapid charge ‘ and 'driven', showing its relevance in finer categorizations.