**Market Segmentation**

**of**

**Electric Vehicle market**

**in**

**India**

Kushagra Bhatnagar

Electric Vehicle is an emerging market in India, new businesses open to investing in this market require to study the trends and needs of the market in India, thus market segmentation is an important strategy that needs to be implemented by the businesses.

Among the many type of segmentations, psychographic segmentation is needed to study the requirements, needs, and behaviour of the customers.

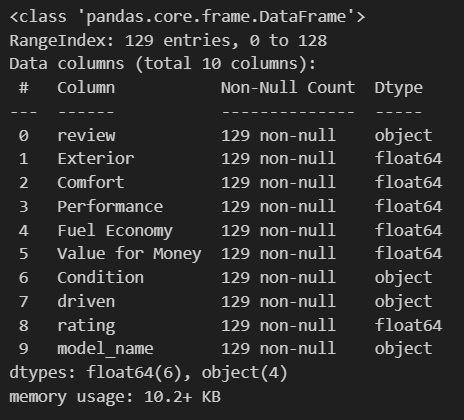
To perform psychographic segmentation, a dataset containing the customer reviews, their needs and expectations can be used. The dataset can be divided into segments based on the features that appeal to them the most.

**Dataset description:**

The dataset used in this project is taken from [Kaggle](https://www.kaggle.com/datasets/deadprstkrish/ev-cars-user-reviews-india).

The dataset contains the reviews of electric vehicle users obtained from carwale.com

The original structure of the dataset is as follow:



There are 3 car models in this dataset:

* Tata Nexon EV
* Hyundai Kona
* Tata Tigor EV

The features ‘exterior’, ‘comfort’, ‘performance’, ‘fuel economy’, ‘value for money’ and ‘rating’ contains rating from 0.0 to 5.0 with 5.0 being the best.

The ‘review’ column contains detailed reviews by the users in natural language.

The ‘condition’ column contains the condition of the vehicle as new, old and ‘not purchased’

The ‘driven’ column contains the amount that the vehicle was driven for. The values in this column are:

* Did a short drive once
* Few thousand kilometres
* Few hundred kilometres
* Haven't driven it
* It’s my mate since ages

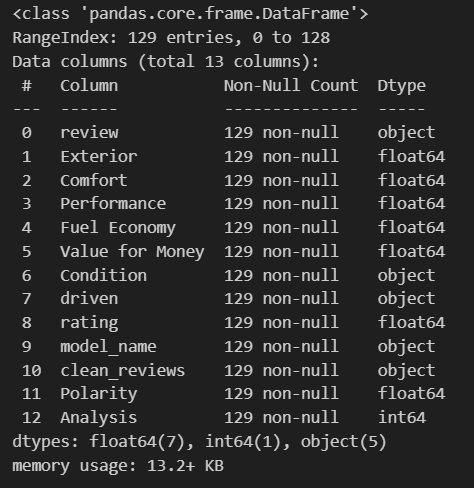
Text Analysis was performed on the reviews column to get the sentiment of the language used to describe the experience of different users.

The reviews were first cleaned using NLTK’s PorterStemmer and stopwords and regex libraries.

The cleaned reviews are then passed to the sentiment.polarity method from the TextBlob library to measure the polarity of the text in the reviews. If the polarity is between 0 and 1, the review is positive, if it is between -1 and 0, the review is negative, and if it is 0, the review is neutral.

Two more columns are added to the dataset named ‘polarity’ and ‘analyses.’ The ‘polarity’ column contains values from -1 to +1 depending on the sentiment of the reviews. The ‘analysis’ column contains distinct values of 1, 0 or -1 as per the polarity of the review.

The new structure of the dataset is as follow:



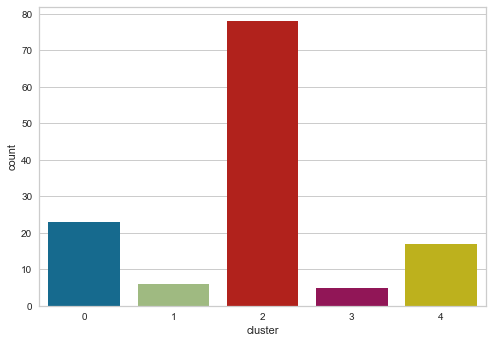
PCA is performed on the features containing numerical values in the dataset, the results of PCA are stored.

The dataset is divided into multiple set of clusters using K-means algorithm, and the within cluster distance is measured. The set of cluster with minimum within cluster distance is taken for segmentation of the dataset.

In this project, the minimum within cluster distance was a result of a set of cluster containing 5 clusters, therefore the dataset was divided into 5 segments.

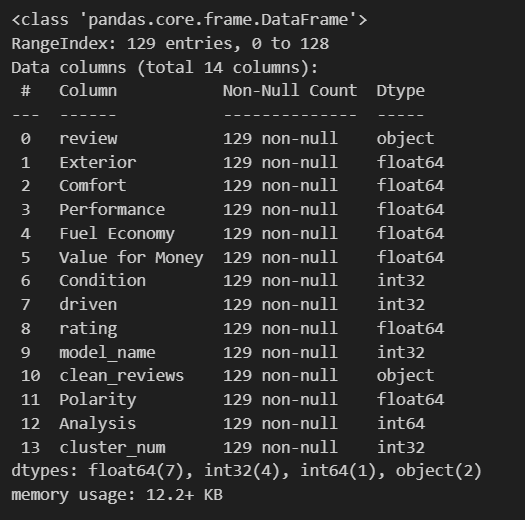
The dataset was divided into 5 segments, with the number of customers per cluster as follow:

* Cluster 2 78
* Cluster 0 23
* Cluster 4 17
* Cluster 1 6
* Cluster 3 5



The ‘driven’, ‘condition’ and ‘model\_name’ columns were converted to numerical type in order to fit the decision tree algorithm that was later used to determine the importance of various features in the segmentation process of the dataset.

After the segmentation, and changing the type of the 3 features to int, the structure of the dataset is:

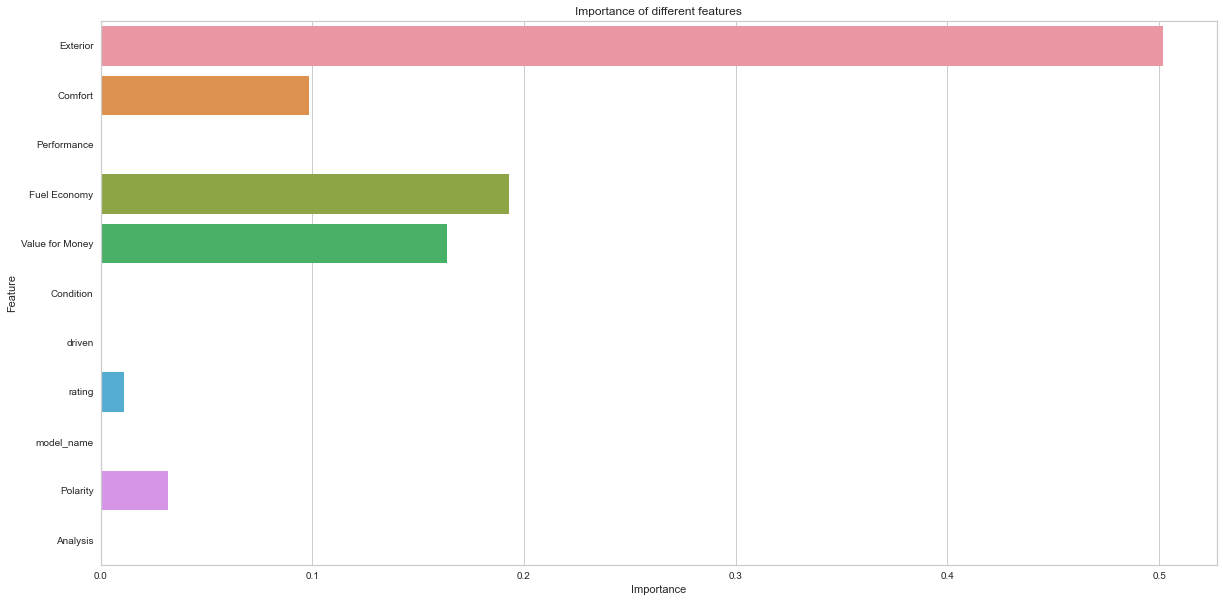


The dataset was split into 70:30 ratio for training and testing the decision tree algorithm.

The decision tree could predict the segment of the user based on their review with 92.3076% accuracy.

The most feature playing the most important role in the segmentation was the Exterior of the car, followed by value for money and comfort. The fuel economy also played an important role in the segmentation.

The importance of various features is shown in the bar graph below:



**GitHub:**

<https://github.com/Bhatnagar621/MS_EV.git>