**Market Segmentation Analysis Of**

**Electric Vehicle Market in India**

**Problem Statement:**

Task is to analyze the Electric Vehicles Market in India using Segmentation analysis and come up with a feasible strategy to enter the market, targeting the segments most likely to use their product in terms of Geographic, Demographic, Psychographic, and Behavioral. In this report we analyze the Electric Vehicles Market in India using segments such as region, price, charging facility, type of vehicles (e.g., 2 wheelers, 3 wheelers, 4 wheelers etc.), retail outlets, manufacturers, body type (e.g., Hatchback, Sedan, SUV, Autorickshaw etc.), safety, plug types and much more.

**Fermi Estimation:**

**Wild Guess:** Around 8-10% people will have electric vehicles by the end of 2023 in India.

**Educated Guess:** Employment rate = it is the ratio of number of available labor force to the population of People in the working age.

We think there are about 1.5 billion Indians in the world. Let's assume the only people over18 and under 60 works, assuming that they account for around 60% of the population then that would make 0.9 billion Indians in the working class.

Out of the 0.9 billion people not all are employed, assuming only 2023 had 45%employment rate that would bring the number around 405 million. Since, not everyone can afford an electric vehicle, let’s assume only people above middle class can afford an electric vehicle, that would be 40 million. Not everyone buysan electric vehicle. Let’s assume out of these 40 million only 10 million are willing to buy an electric vehicle. Variables and Formulas: Let E(x) be the employment rate of the year x (in %).

Let P(x) be the population of the year x.

Let A(x) be the number of available Labor in the year x.

Let r be the ratio of Indians between the age of 18 and 60 to the total population of India.

E(x) = (A(x)\*100)/(P(x)\*r) This formula will formulate the Employment ratio for the year x.

**Gathering More Information:** Estimation for the population of the year 2022 can be obtained by the increase in population each year P (2019) = 1.3676 billion P (2020) = 1.3786 billion P (2021) = 1.39199 billion P (2020)-P (2019) = 11million P (2021)-P (2020) = 13.39 million the mean would be 12.195 million thus P (2022) = 1.44185 billion assuming A(x) is constant every year= 471,688,990r=0.6C=0.75 E (2022) = (471,688,990/(1,441,850,000\*0.6))\*0.75E (2022) = 42% Conclusion: By this analysis, we conclude that by the end of the year 2024 there would a Employment rate of 42%.

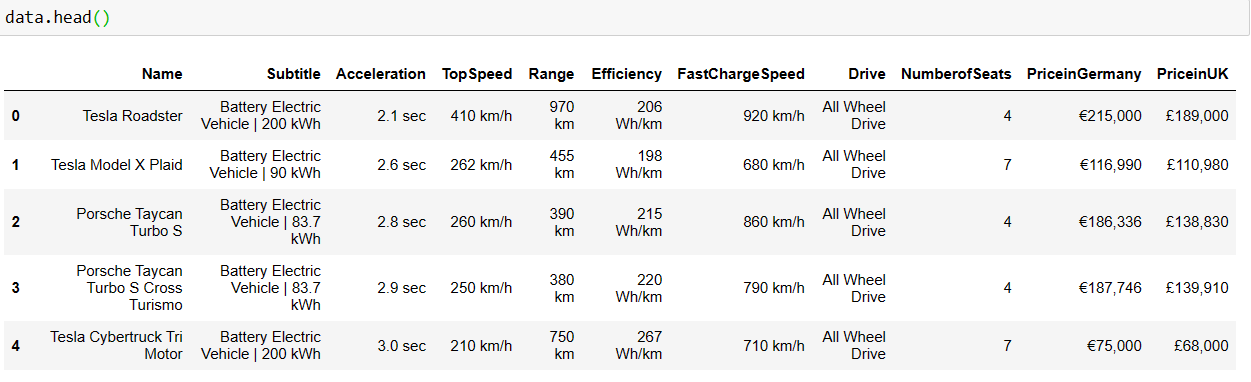
That would make 42% of 405 million i.e., 170 million. Out of these 170 million only 10% afford EV'S. So around 17 million people will have EV's by the end of 2024".

**Data Pre-processing:**

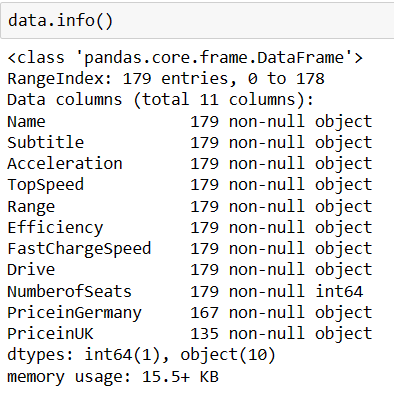
In this step we are going to convert the raw data into a meaningful data which will be used for clustering.

Dataset Origin :

<https://www.kaggle.com/datasets/kkhandekar/quickest-electric-cars-ev-database>

Initially the dataset looks like,

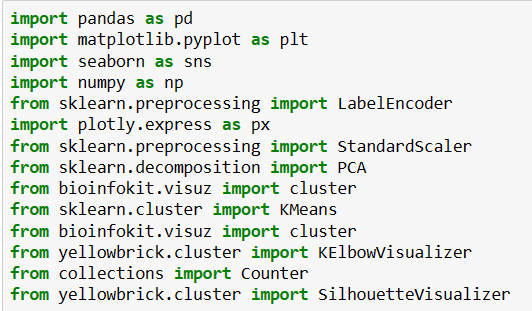
The information about the dataset,



The above information seems that even though the data is numerical but the type of the data is object. So we need to pre-process the data.

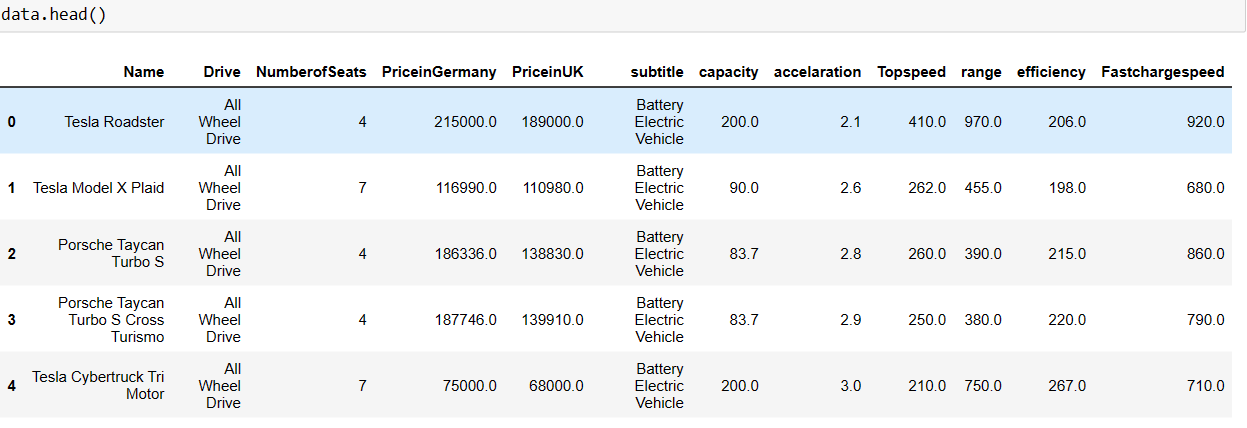
**Required Libraries:**

In order to perform EDA and Clustering on collected data, the following libraries are used.



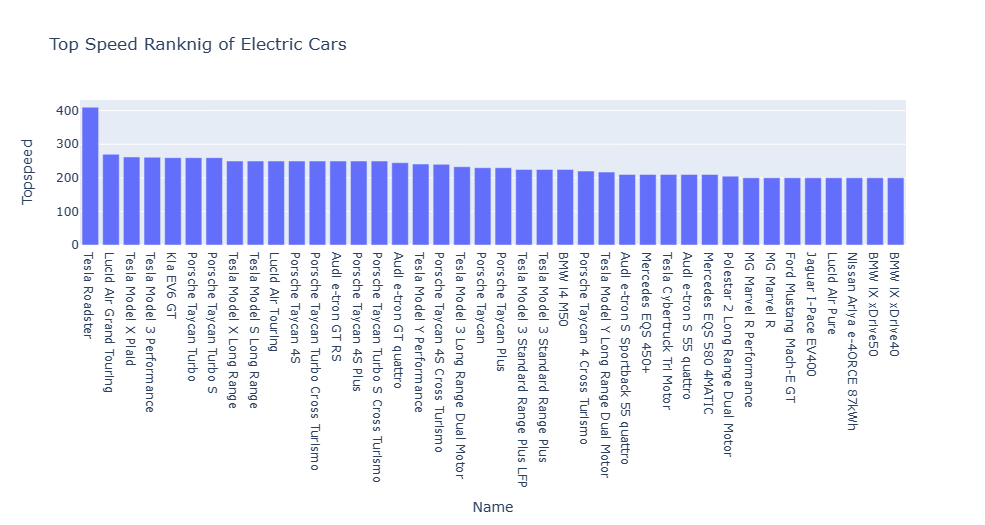
We are going to pre-process the data by using the below steps,



The Final Preprocessed Data looks like below as shown in figure

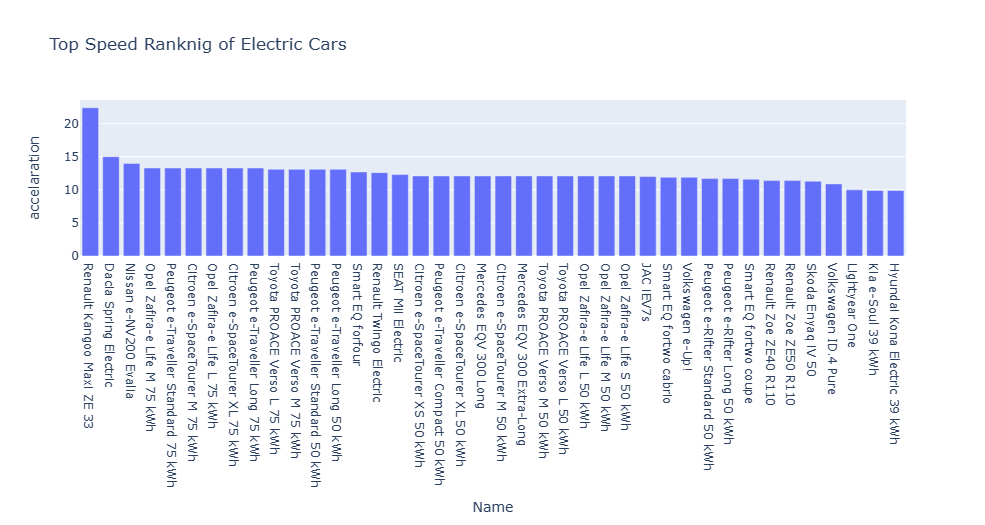
**Exploratory Data Analysis:**

In data mining, Exploratory Data Analysis (EDA) is an approach to analyzing datasets to summarize their main characteristics, often with visual methods. EDA is used for seeing what the data can tell us before the modelling task.

*Top Speed Ranking of Electric cars:*

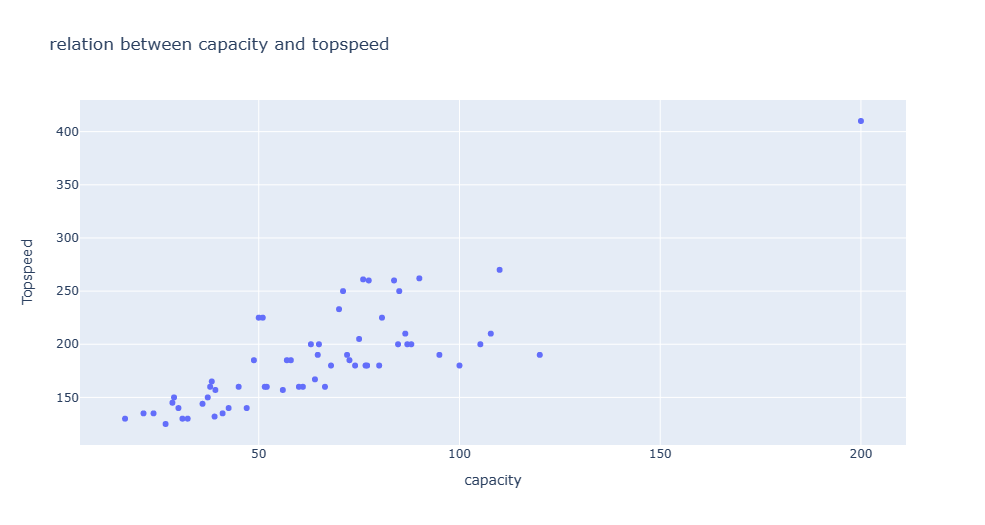
**Observations from above Graph:**

The above graph shows that the highest speed ranking of top 40 electric cars. Among them the top 5 ranked cars are *Tesla Roadster, lucid air grand touring, Tesla Model x Plaid, Tesla Model 3 performance, Kia EV6 GT* .

*Top Speed Ranking of Electric cars By Acceleration:*

**Observation from above Graph:**

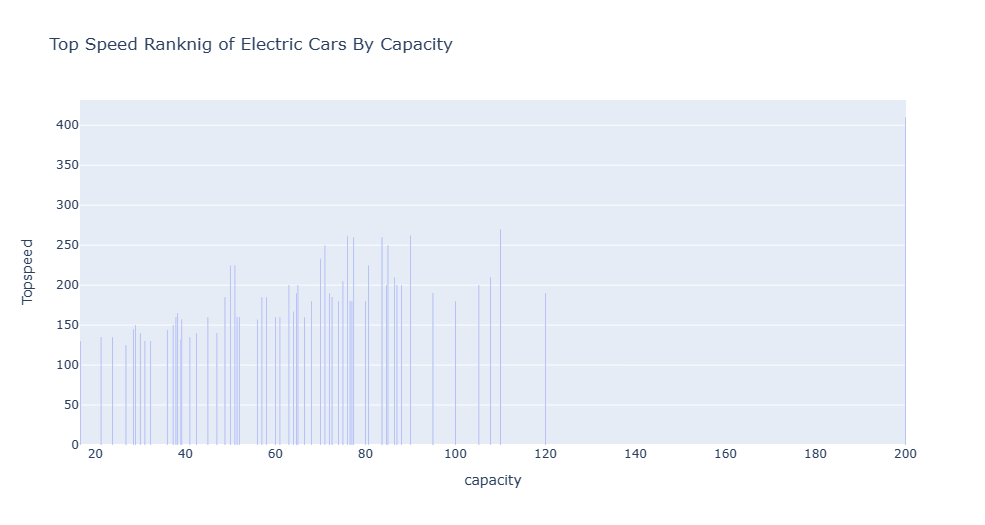
The above graph shows that the top speed ranking of electric vehicles by acceleration. Among them the top 5 electric vehicles are *Renault Kangoo Maxi ZE 33, Dacia Spring Electric, Nissan e-NV200 Evalia, Opel Zafira-e Life M 75 kWh, Peugeot e-Traveller Standard 75 kWh.*

*Relation Between Capacity And Top Speed:*

**Observations from above Graph:**

From the above graph there is a linear relation between TopSpeed and Capacity that describes If Capacity increases then the topspeed also increases.

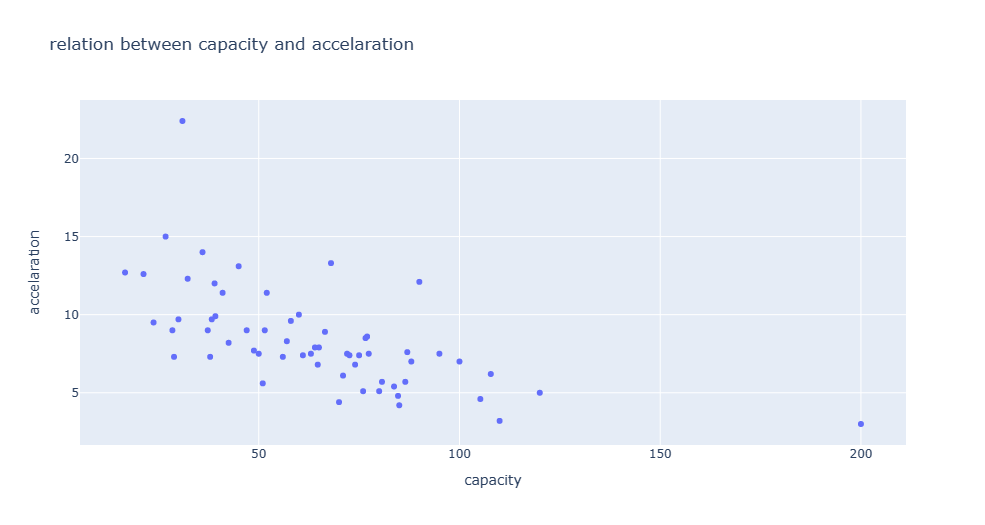
*Top Speed Ranking of Electric Cars By Capacity:*

**

**Observations from above Graph:**

The above graph shows that the Top Speed ranking of Electric Cars By Capacity which indicates that the there is a linear relationship between the speed and capacity and for the capacity of 200 has topspeed 420.

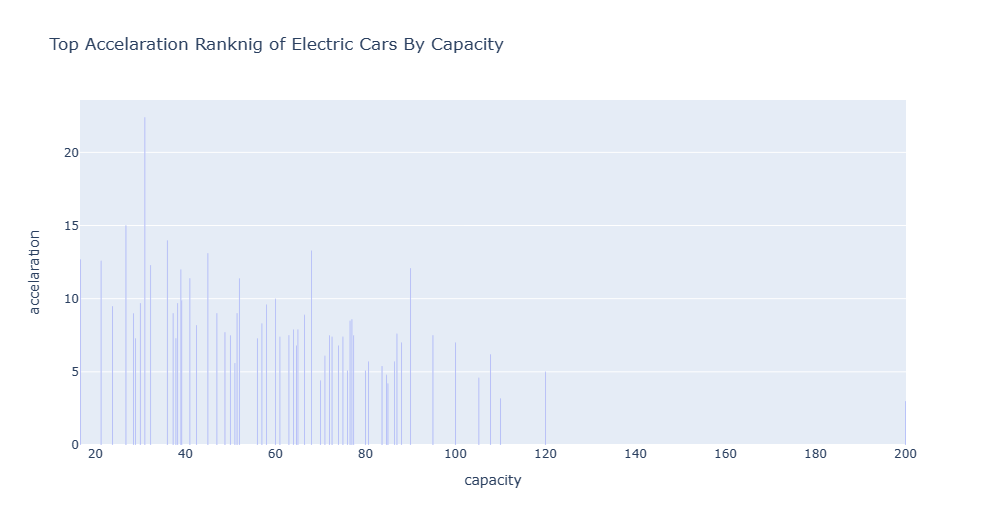
*Relation Between Capacity And Acceleration:*

**

**Observations from above Graph:**

From the above graph there is a linear relation between Acceleration and Capacity that describes If Capacity increases then the topspeed decreases. That it shows a negative correlation between acceleration and capacity.

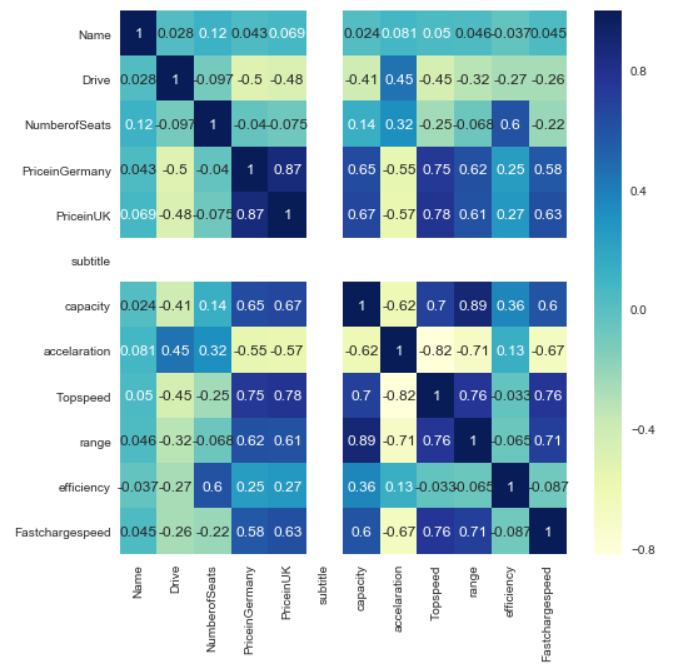
*Top Acceleration Ranking of Electric Cars By Capacity:*

**

**Observations from above Graph:**

The above graph shows that the Top acceleration ranking of Electric Cars By Capacity which indicates that the relation between the acceleration and the capacity and for the capacity of 20 has acceleration 12. There is a negative correlation between acceleration and the capacity.

*Correlation Matrix:*

**

The correlation matrix define that the relation among all the variables in the data.

**Segmentation Approaches:**

**Clustering:**

Clustering is a type of unsupervised learning method of machine learning. In the unsupervised learning method, the inferences are drawn from the data sets which do not contain labelled output variable. It is an exploratory data analysis technique that allows us to analyze the multivariate data sets.

**KMeans Clustering:**

**Step-1:** Select the number K to decide the number of clusters.

**Step-2:** Select random K points or centroids. (It can be other from the input dataset).

**Step-3:** Assign each data point to their closest centroid, which will form the predefined K clusters.

**Step-4:** Calculate the variance and place a new centroid of each cluster.

**Step-5:** Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.

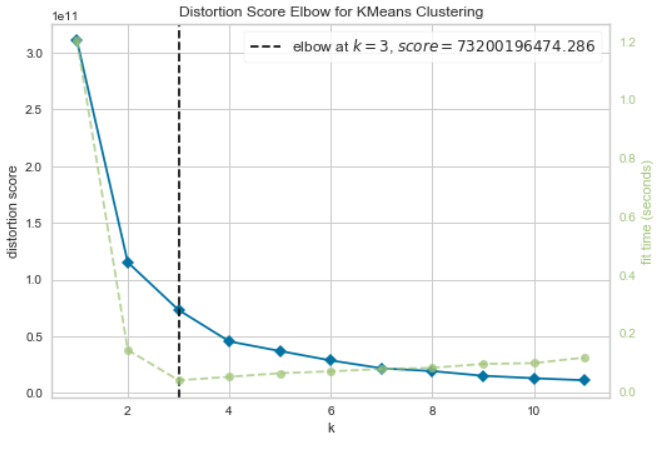
**Step-6:** If any reassignment occurs, then go to step-4 else go to FINISH.

**Step-7**: The model is ready.

**Elbow Method:**

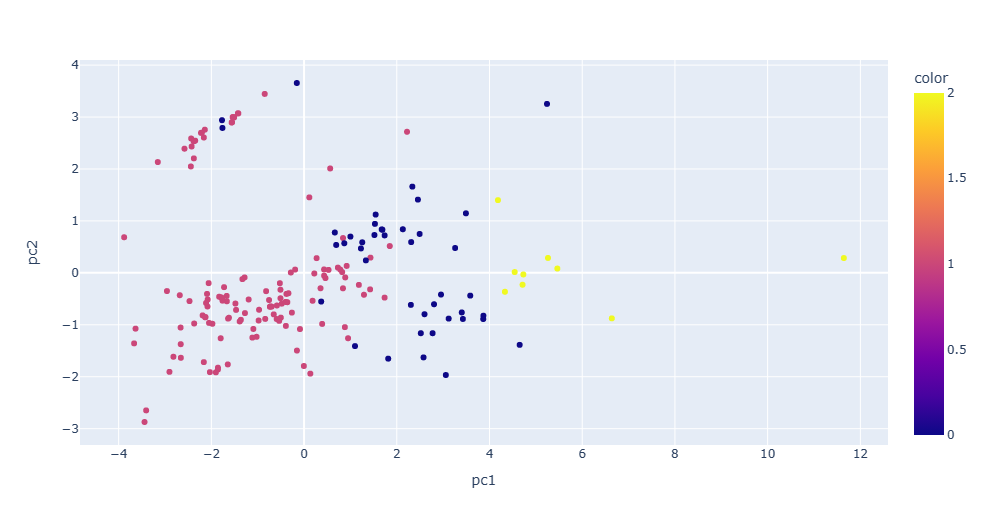
The elbow method is a graphical representation of finding the optimal 'K' in a K-means clustering. It works by finding WCSS (Within-Cluster Sum of Square) i.e. the sum of the square distance between points in a cluster and the cluster centroid.

The optimal clusters for our case study is 3, as shown below:



From the above graph we conclude that the data can be clustered into three clusters.

The clusters can be visualized as shown in below,

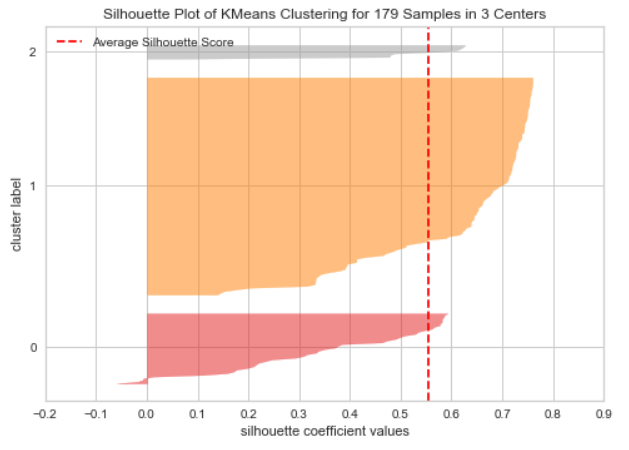


After completion of clustering of data into 3 clusters, that can be visualized as shown.

**Obtaining the Stability of the Clusters:**

Grouping items into clusters is a complex problem in unsupervised learning with inherent uncertainty. Stability is a measurement that characterizes the strength and reproducibility of a cluster and an items membership to a cluster.

The stability of the clusters is obtained from the below figure,



From the above figure the cluster 1 and 2 are perfectly stabled but not the cluster 0.