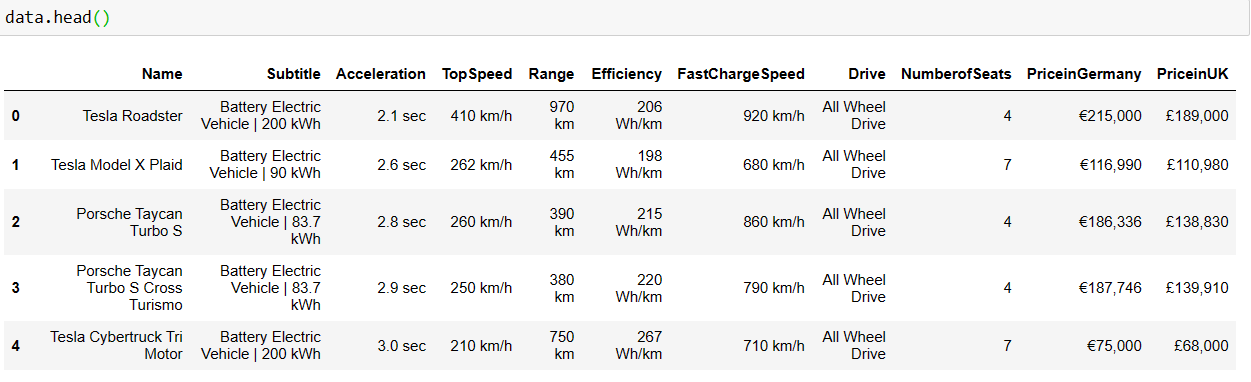
**Data Pre-processing:**

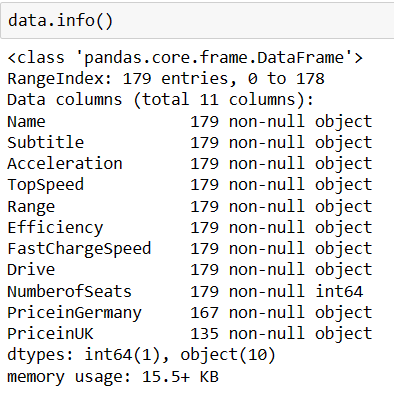
In this step we are going to convert the raw data into a meaningful data which will be used for clustering.

Dataset Origin :

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

Initially the dataset looks like,

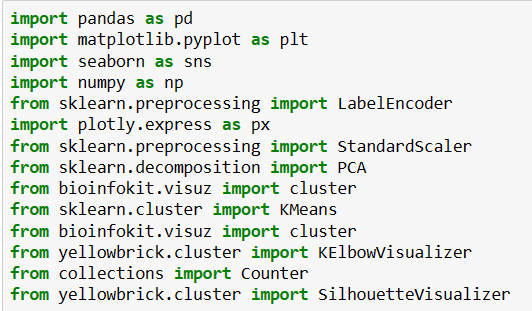
The information about the dataset,



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.

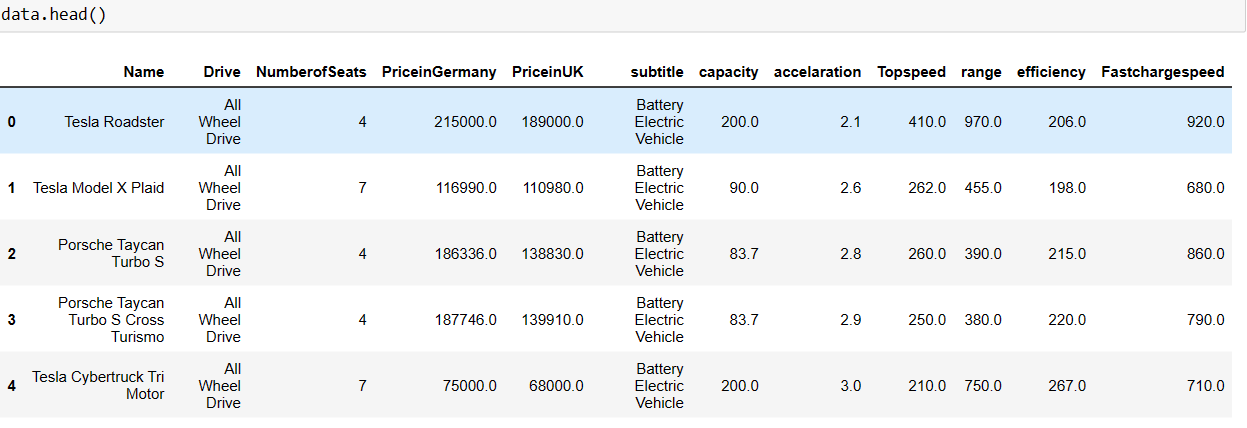
**Required Libraries:**

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



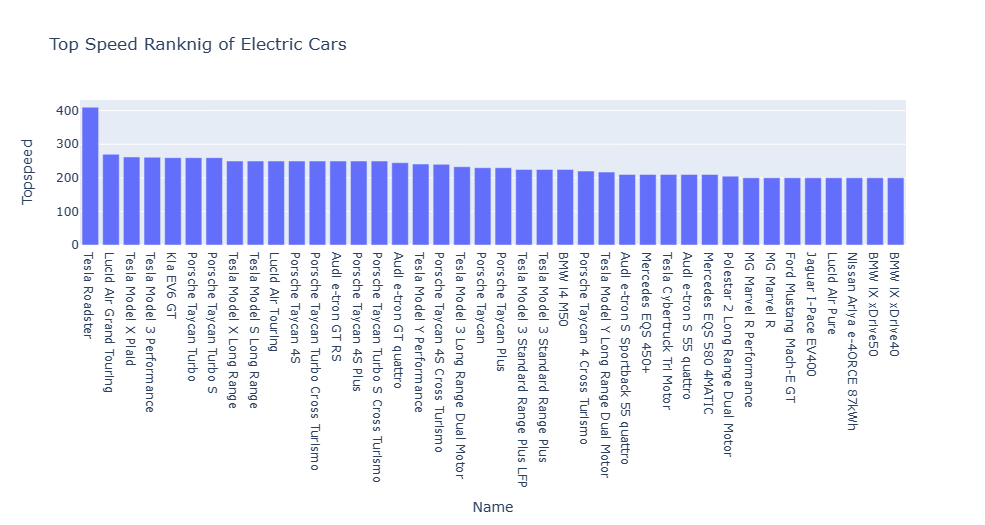
We are going to pre-process the data by using the below steps,

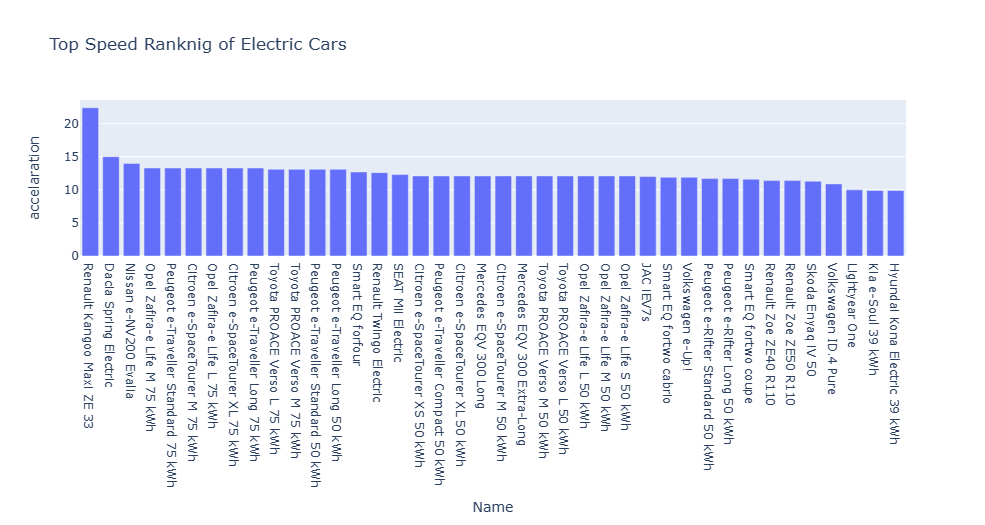


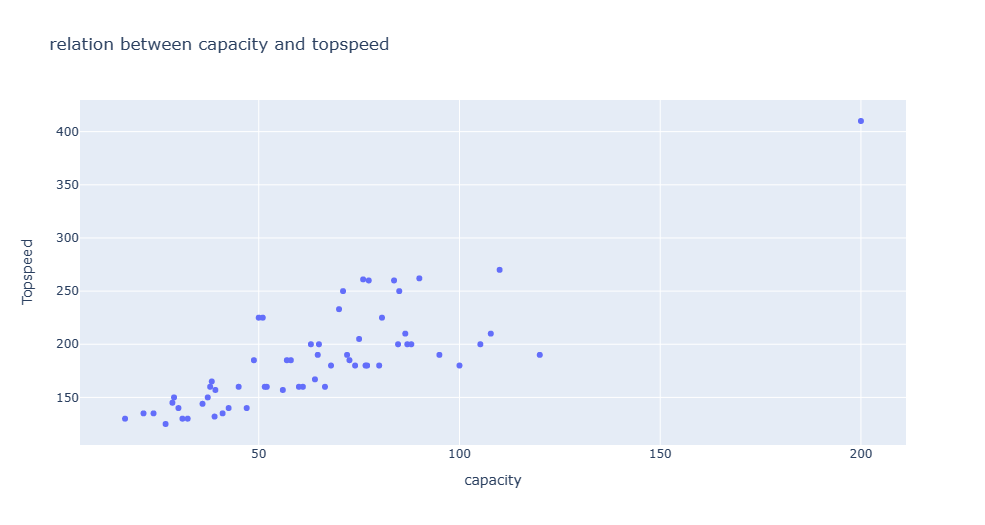
The Final Preprocessed Data looks like below as shown in figure

**Exploratory Data Analysis:**

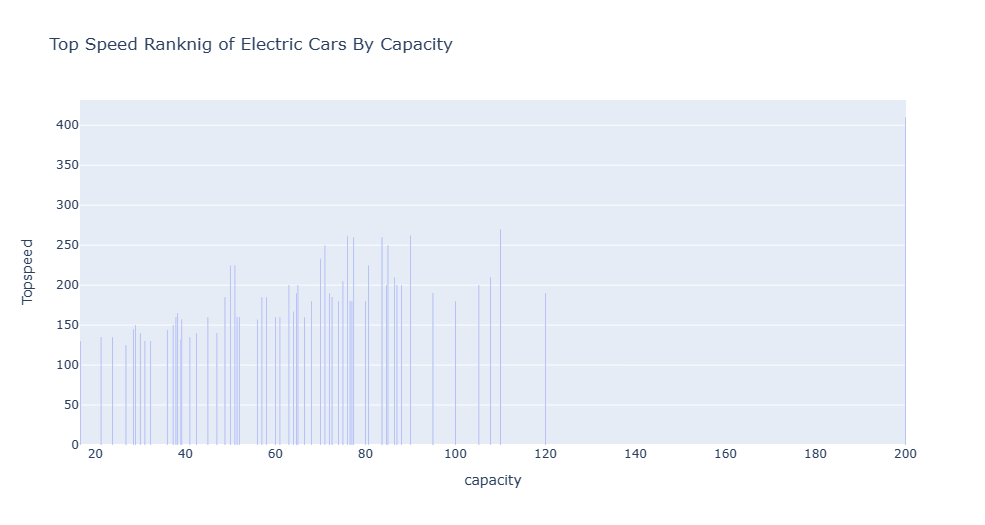
In data mining, Exploratory Data Analysis (EDA) is an approach to analyzing datasets to summarize their main characteristics, often with visual methods. EDA is used for seeing what the data can tell us before the modelling task.

*Top Speed Ranking of Electric cars:*

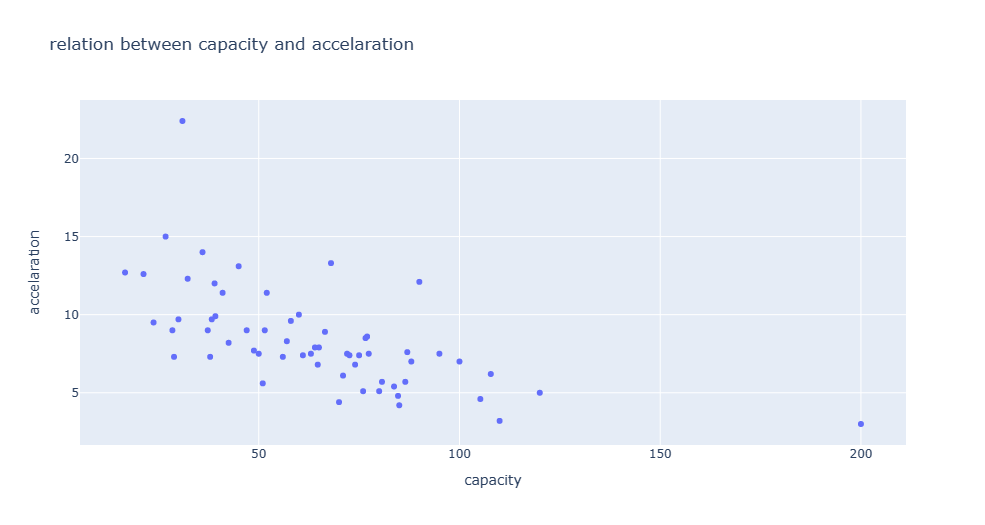
*Top Speed Ranking of Electric cars By Acceleration:*

*Relation Between Capacity And Top Speed:*

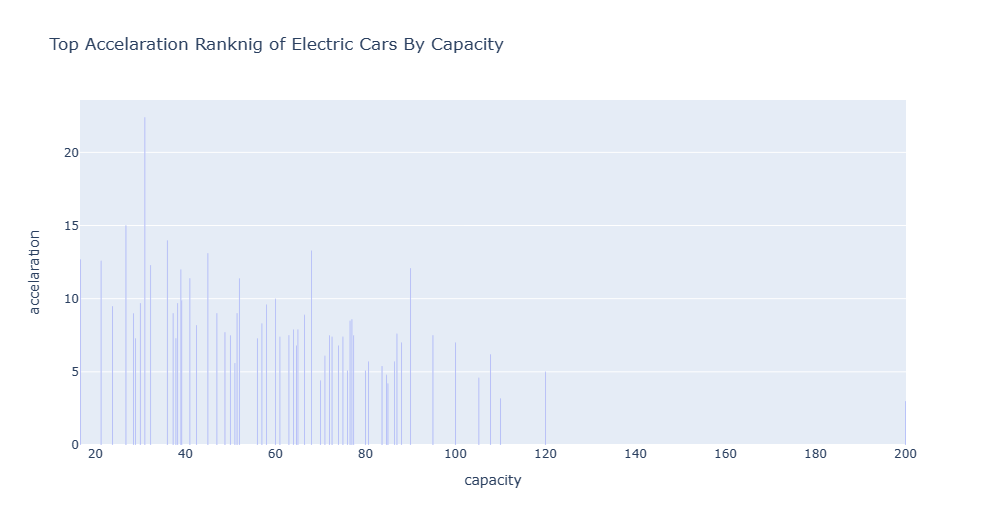
*Top Speed Ranking of Electric Cars By Capacity:*

**

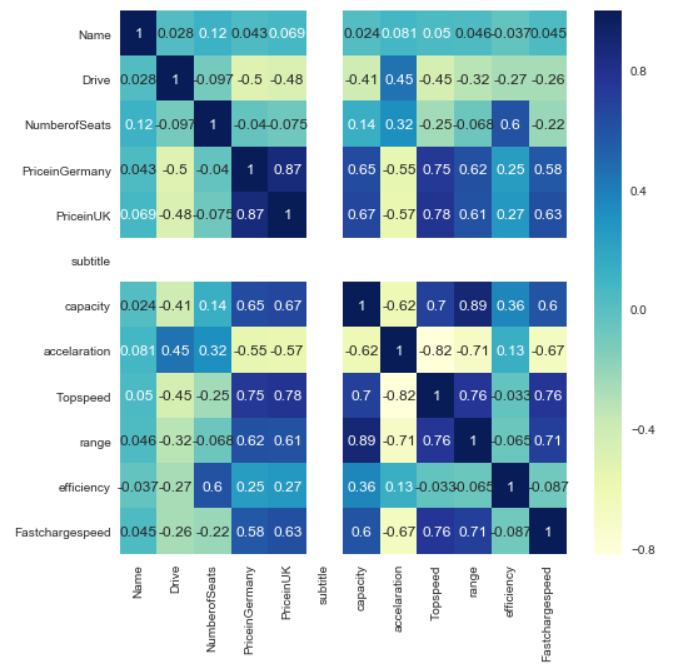
*Relation Between Capacity And Acceleration:*

**

*Top Acceleration Ranking of Electric Cars By Capacity:*

**

*Correlation Matrix:*

**

**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.

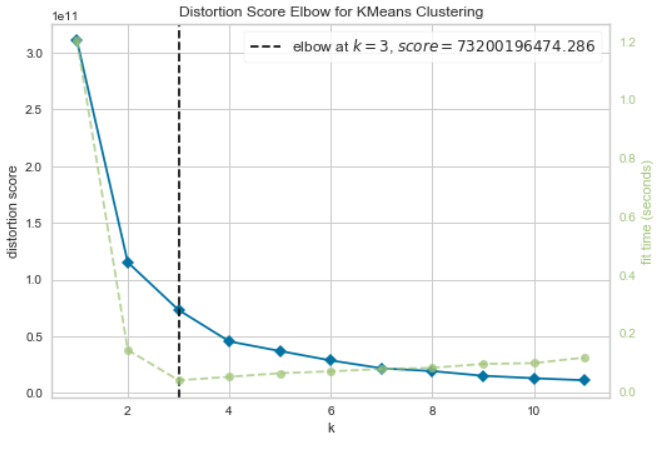
**Step-6:** If any reassignment occurs, then go to step-4 else go to FINISH.

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

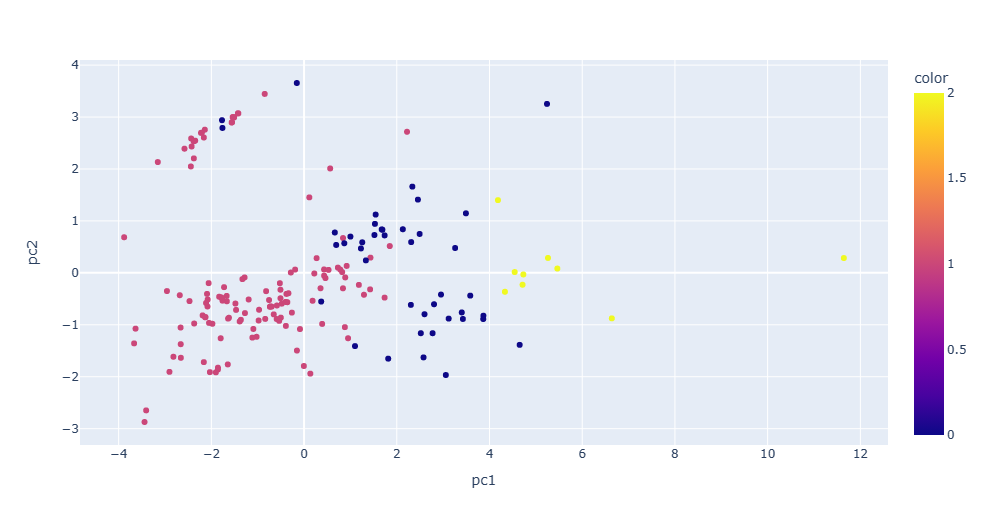
**Elbow Method:**

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:

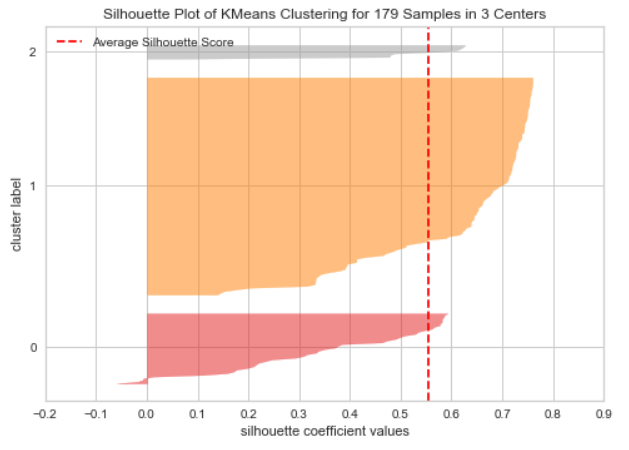


The clusters can be visualized as shown in below,



**Obtaining the Stability of the Clusters:**

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.