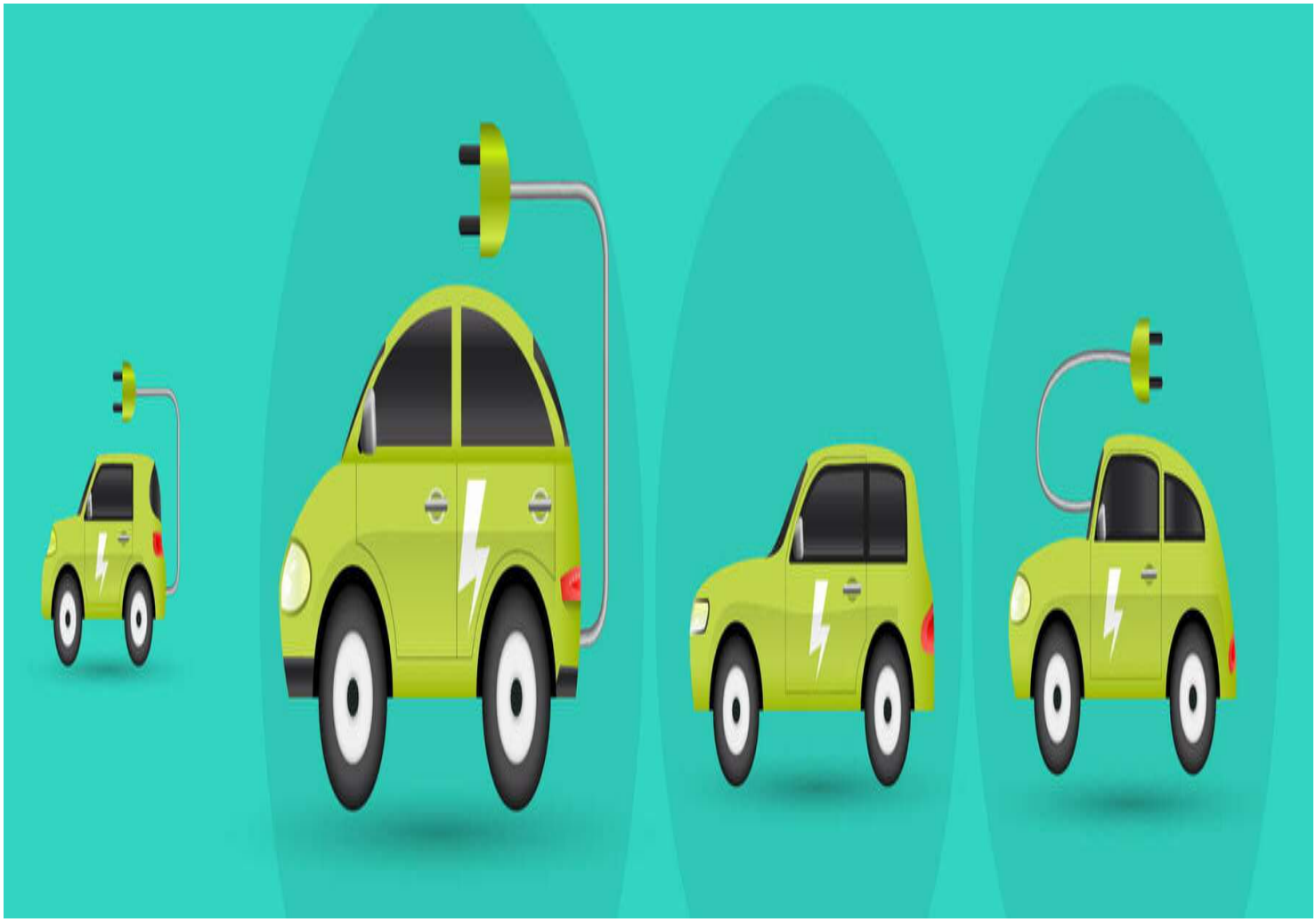
MARKET Segment Analysis of Electric vehicle market

By saikat Adhikary submitted to FeyNN Labs

- Github link-https://github.com/saikatpythondev/ml-internship



Introduction:

Electric cars run at least partially on electricity. Unlike conventional vehicles that use a gasoline or diesel-powered engine, electric cars and trucks use an electric motor powered by electricity from batteries or a fuel cell. In terms of air pollution and greenhouse gas emissions, electric cars and trucks are often cleaner than even the most efficient conventional vehicles. Exactly how clean depends on the type of vehicle and the source of the electricity. When battery electric EVs are powered by the cleanest electricity grids, greenhouse gas emissions from EVs are comparable to a car getting over 100 miles per gallon. When charged exclusively with renewable electricity like solar or wind, charging and operating an EV can be nearly emission free.

Though electric cars can be more expensive to purchase than their conventional counterparts, the higher upfront cost is often reduced through federal and state incentives. The cost to refuel an electric car can also be a fraction of the cost of gasoline, meaning that electric cars can have a lower total cost of ownership. Switching to an electric car can save on average over $700 a year in fueling costs, and over one thousand dollars a year in some cities.

Problem statement:

The global **Electric Vehicle Market** size is projected to grow from 8,151 thousand units in 2022 to 39,208 thousand units by 2030, at a CAGR of 21.7%. Factors such as growing demand for low emission commuting and governments supporting long range, zero emission vehicles through subsidies & tax rebates have compelled the manufacturers to provide electric vehicles around the world. This has led to a growing demand for electric vehicles in the market. Countries around the world have set up targets for emission reductions according to their own capacity.

Increasing investments by governments across the globe to develop EV charging stations and Hydrogen fueling stations along with incentives offered to buyers will create opportunities for OEMs to expand their revenue stream and geographical presence. The EV market in Asia Pacific is projected to experience steady growth owing to the high demand for lower cost efficient and low-emission vehicles, while the North American and European market are fast growing markets due to the government initiatives and growing high-performance Passenger vehicle segment. However, relatively less number of EV charging stations and hydrogen fuel stations, higher costs involved in initial investments, and performance constraints could hamper the growth of global electric vehicle market.

In this project I have shown the EV market analysis using unsupervised classification algorithm and come up with a segments most likely to use their product in terms of geographic ,demographic,psychographic,behavioral

Data collection:

- **The csv file with data is collected from kaggle.com**

  With total 14 columns represent:

  1)Brand:the name of the brand

  2) Model:name of the model

  3) AccelSec

  4) TopSpeed_KmH:the maximum speed of the vehicle

  5) Range_Km:

  6) Efficiency_WhKm:the efficiency of the motor

  7) FastCharge_KmH:fast charge in km/h

  8) RapidCharge:is the vehicle is rapid chargeable yes or no?

  9) PowerTrain

  10) PlugType:type of the plug

  11) BodyStyle:bodystyle

  12) Segment:segment

  13) Seats:no of seats

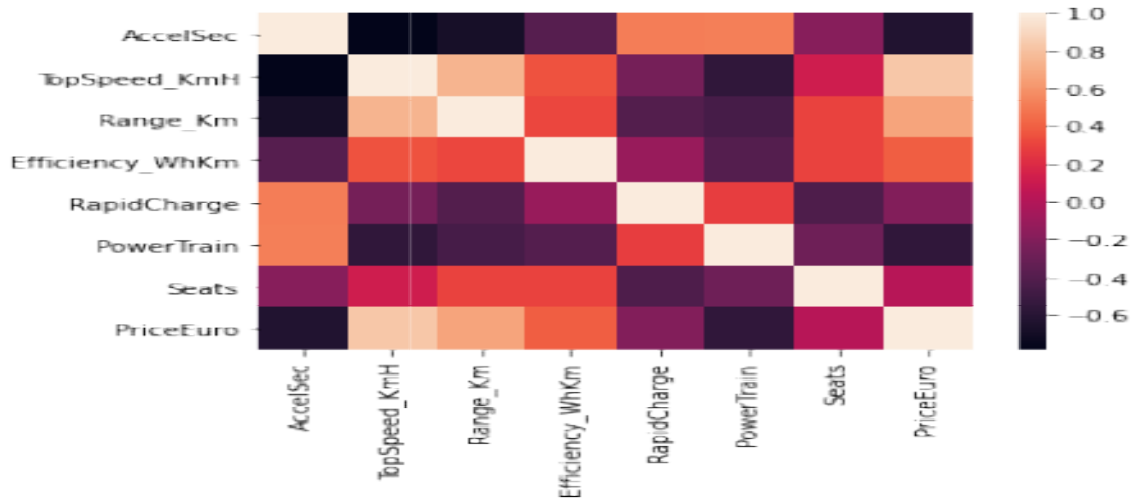
  14) PriceEuro:price in Euro

Preprocessing:

First we have to import the essential libraries of python after EDA analysis we have to convert the categorical variable to numerical variable a that we have to
Fade the data in kmeans clustering algorithm to get the segment of the market

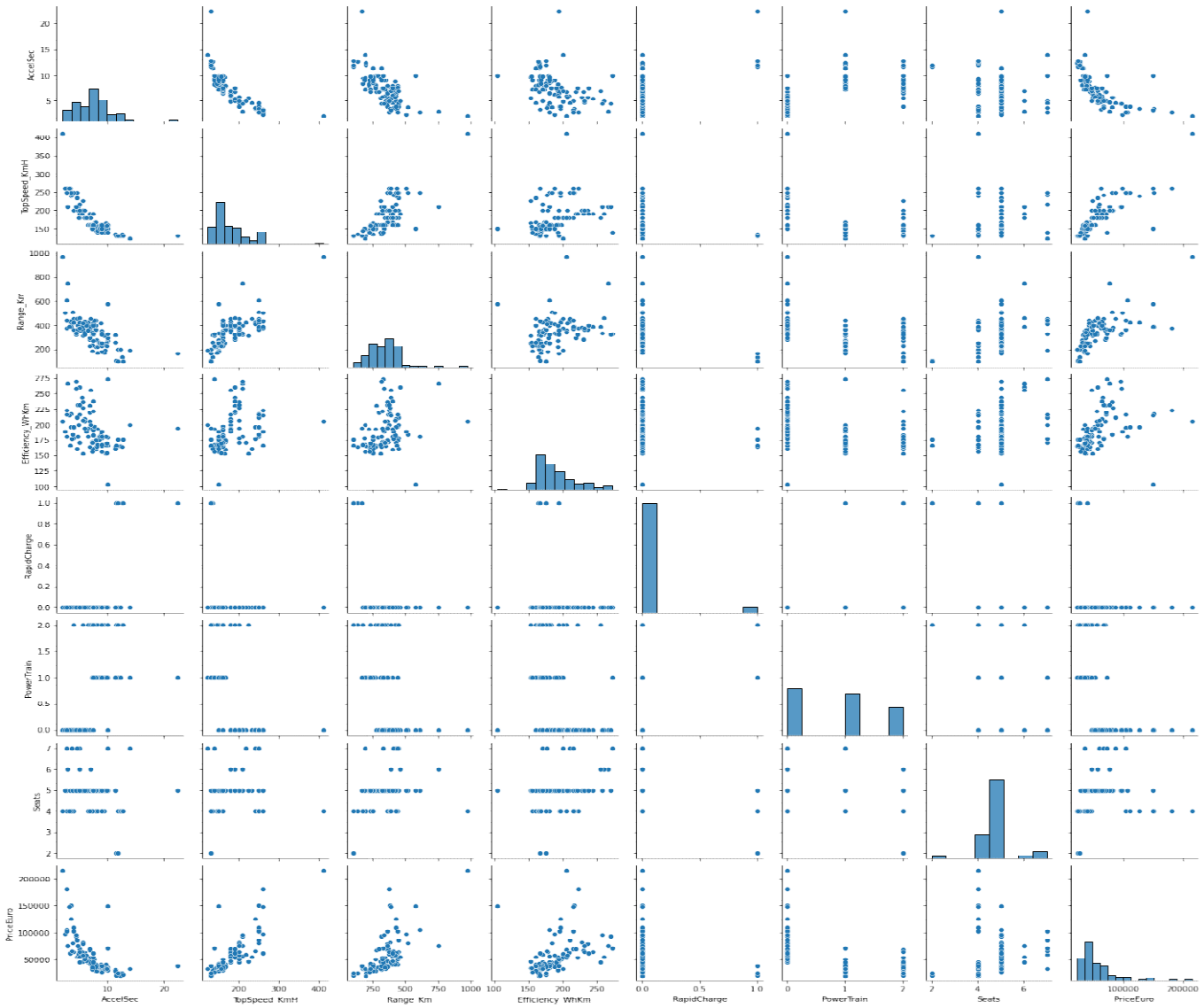
EDA:

1)The correlation of the features in the dataset shown in figure below:



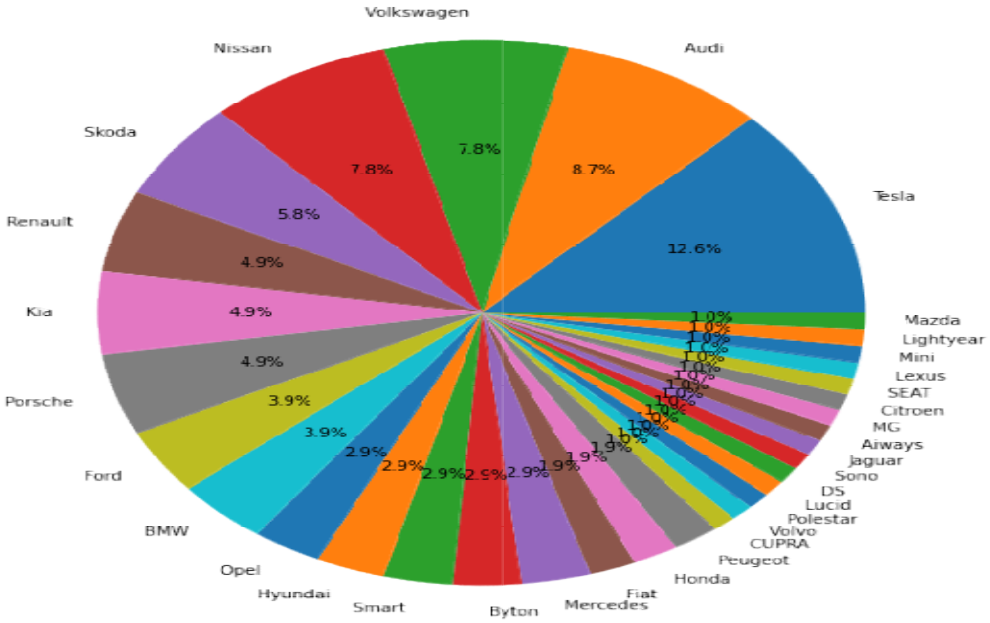
This is the seaborn heatmap of correlation matrix,the value of each element varies between -1 to +1,notice some values are highly correlated i.e,the absolute balue of the matrix is close to 1 they are removed to make the model is more better

2) the seaborn pairplot is shown in the figure



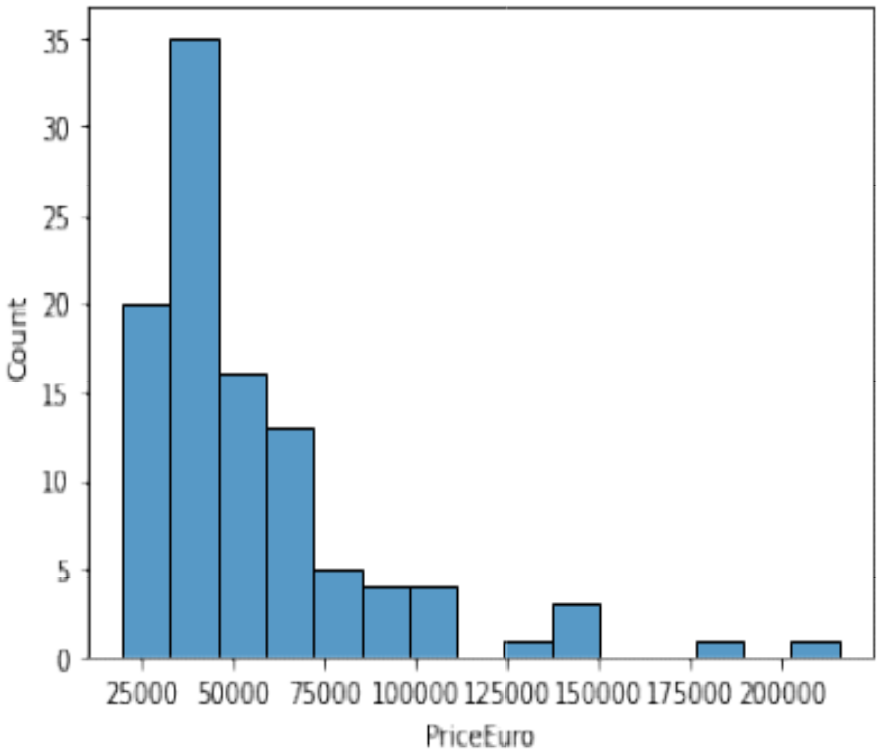
It represent the graph between each and every column in the dataset

3)the pie plot of brand shown in figure :

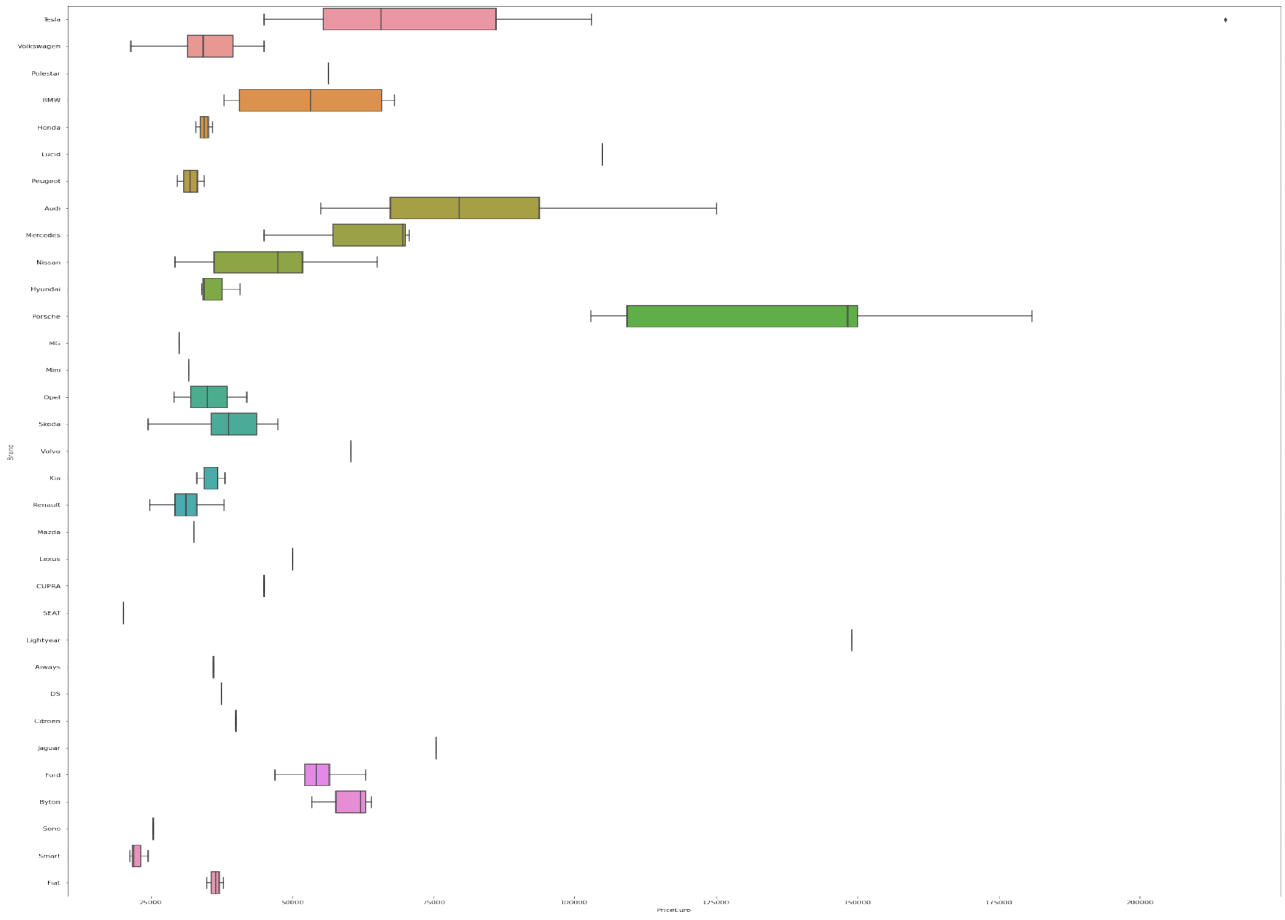


it represent Tesla and Audi has highest 12.6% and 8.7% market share and lowest market share is Mazda 1%

4)



the histogram of price range is shown in the figure ,it shows the histogram is right-screwed and the price range of each brand is given in boxplot



Unsupervised ml algorithm:

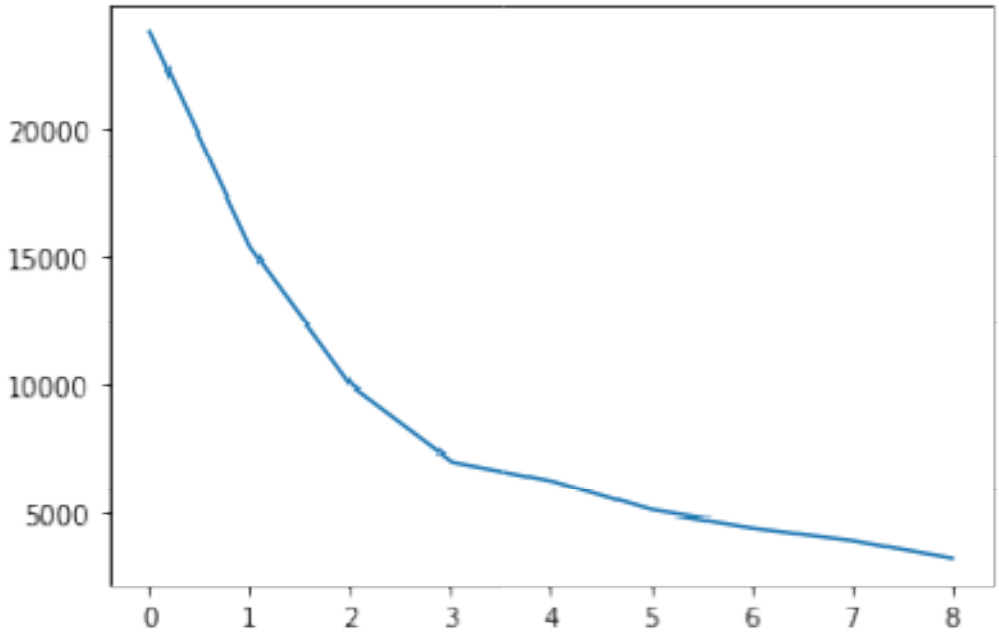
K-means Clustering:

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training. It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters. The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters.

The value of k should be predetermined in this algorithm.

We start by pre-processing the data and cleaning it. This essentially involves null-handling and label encoding the ordinal parameters of the data. The data is then passed into the Scikit-Learn K-Means Clustering model to obtain the elbow curve for the ideal number of clusters. Using the

"elbow" or "knee of a curve" as a cutoff point is a common heuristic in mathematical optimization to choose a point where diminishing returns are no longer worth the additional cost. In clustering, this means one should choose a few clusters so that adding another cluster doesn't give much better modeling of the data. The intuition is that increasing the number of clusters will naturally improve the fit (explain more of the variation), since there are more parameters (more clusters) to use, but that at some point this is over-fitting, and the elbow reflects this.



Acknowledgment

We would like to thank the administration of FeyNN Labs for providing us with the opportunity of carrying out such a detailed analysis and providing us with the exposure to industry business analytics.

We would also like to thank our parents, friends and associates for their constant support, aid, and advice.

The project has required no form of funding and is absolute independent research by the

Machine learning Interns of FeyNN.AI