K-MEAN CLUSTERING

By
SHAHZAIB NAZIR.
DATA SCIENCE ASSIGNMENT 03

Introduction.

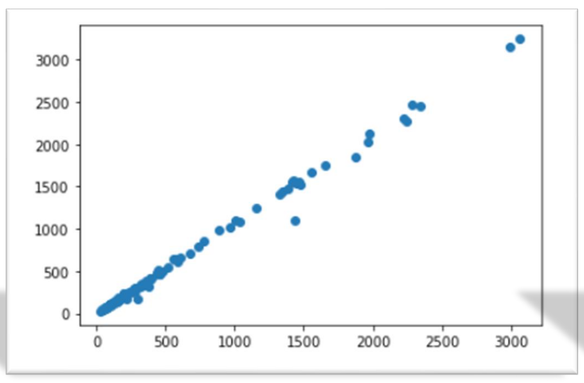
This assignment is about the clustering algorithm that is very meaningful in the data science when you make a group of data for a grouping in a data in this assignment we use data set of the World bank that is related to the GDP per capita this time we make some additional libraries how the Python third is built in for the word related data science just like SK learn from that library import few things like preprocessing for the normalization of the data and Science Pi we import curve fit method and from the same library as K learn we import the cluster where we use k mean cluster you also use some built-in libraries that we also used in the previous assignments like Pandas matplotlib, number etc. Can we check one by one and the different results that we share with you in this poster we not only e get the results but we also got it really with the help of data visualization we make the visualization of the data for more precise learning. Here we use only this clustering technique now we move toward the check how much cluster we make from this dataset.

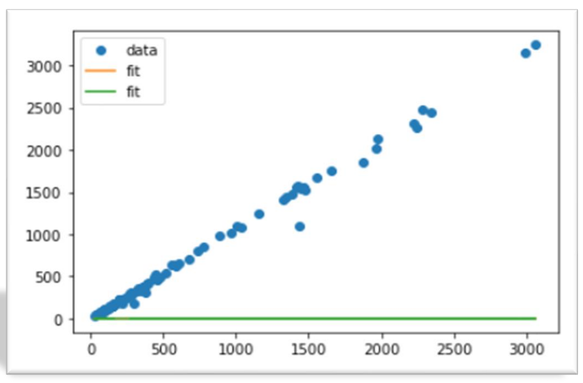
Clustering.

Clustering is a collection of strategies for grouping data into clusters. Clusters are roughly described as collections of data objects that are more comparable to each other than to data objects from other clusters. The unsupervised machine learning technique k-means clustering is used to find groupings of data objects in a dataset. There are a variety of clustering methods available, but k-means is one of the most popular and accessible.

Elbow Curve

Before clustering this is very important to check how many possibilities of clusters in a given data set by using a method of elbow curve adding to the given library SK learn this is by default a method of inertia and append this to the list and after that plot there is whole list data using matplotlib and we say that where the line become straight till that point notice the value of x and assume that number of clusters.



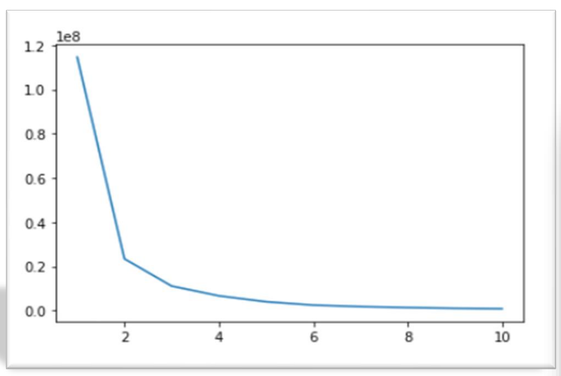


K mean cluster

This is the k-means cluster in which we observed clearly three different clusters in the scattered form of matplotlib graph according to elbow curve vi set number of cluster equals to 3 e and we observe three different clusters of k mean algorithm now we observe all these three plasters separately and check which clusters having more elements on left top figure is about whole dataset then is curve fit after it is elbow chart

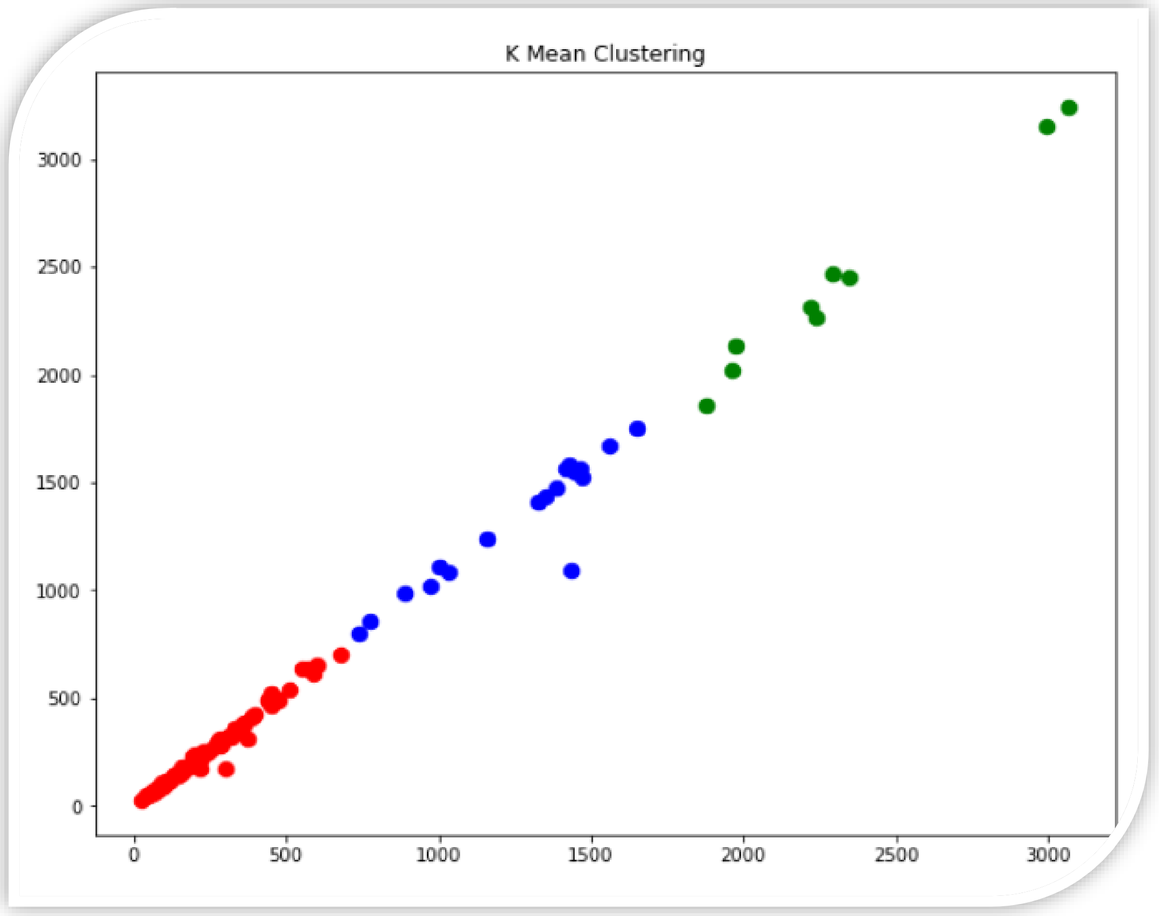
Normalization.

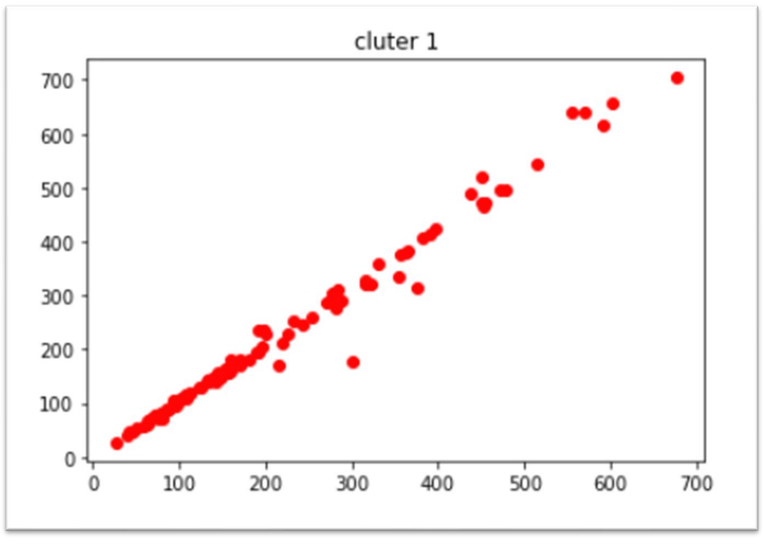
Rescaling real-valued numeric properties into a 0 to 1 range is known as normalization. In machine learning, data normalization is used to make model training less sensitive to feature scale. As a result, our model can converge to better weights, resulting in a more accurate model. Python includes a preprocessing library that includes the normalize function for data normalization. It accepts an array as input and converts its values to a range of 0 to 1. We divide the result by the range after subtracting the minimum value from each entry. The difference between the maximum and least value is called range.

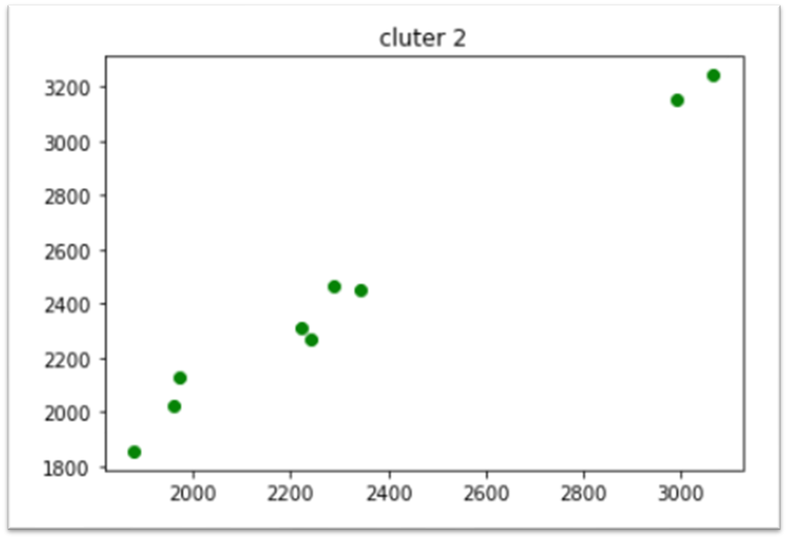


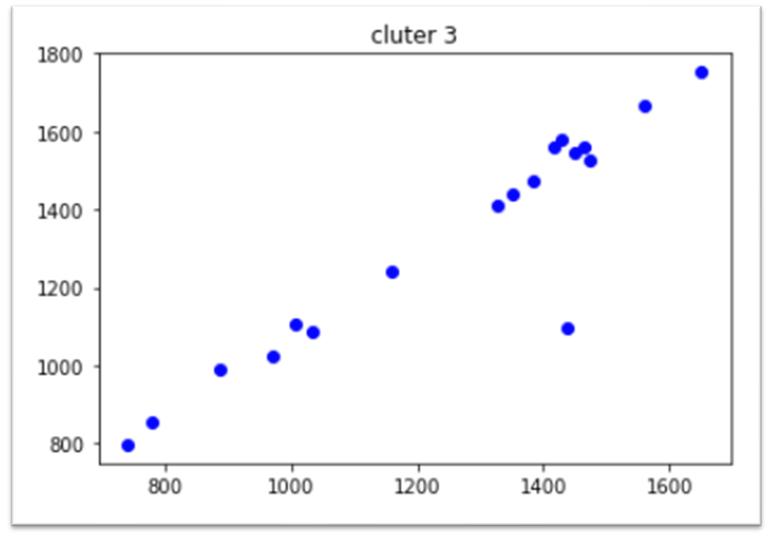
Curve_fit()

Curve fitting is a sort of optimization that involves determining the best set of parameters for a specific function that best fits a set of data. Curve fitting, unlike supervised learning, necessitates the definition of the function that maps examples of inputs to outputs. We also use built-in libraries for that to make curve fit method to fit the dataset.









Interpretation of Results.

according to this results we totally satisfied because y_kmeans clearly define that starting records is belong to 0 cluster most of the records presents in this cluter and afte that we saw 1 and 2 accordingly after that we saw using matplotlib visualization that we have these clusters in also visualization.
red denote 0
green denote 1
blue denote 2

Github link.
https://github.com/ShahzaibNazir/data-science-assignment-3