

# DATA SCIENCE ASSIGNMENT 03

## Introduction.

This assignment is about the clustering algorithm, which is very useful in data science when you make a group of data for a grouping in a data set. In this assignment, we use data from the World Bank that is related export goods. This time, we make some additional libraries, such as the Python third, which is built in for the word related data science, similar to SK. Learn a few things from that library, such as preprocessing for data normalization 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 may also utilize certain built-in libraries that we used in prior assignments. Can we check one by one and the various findings that we share with you in this poster? We not only obtained the results, but we also got it with the assistance of data visualization, which we used to construct the data visualization for more precise learning. We will now check how many clusters we can generate from this dataset using simply this clustering technique. And we describe data accordingly on the left side. This data is about the 2019 and 2020 and description about the whole dataset.

	2019	2020
count	209.000000	209.000000
mean	39.663855	35.908145
std	28.713277	28.693743
min	0.630574	0.435765
25%	23.061582	20.015022
50%	31.592055	27.873644
75%	48.131642	43.416895
max	205.482078	204.690741

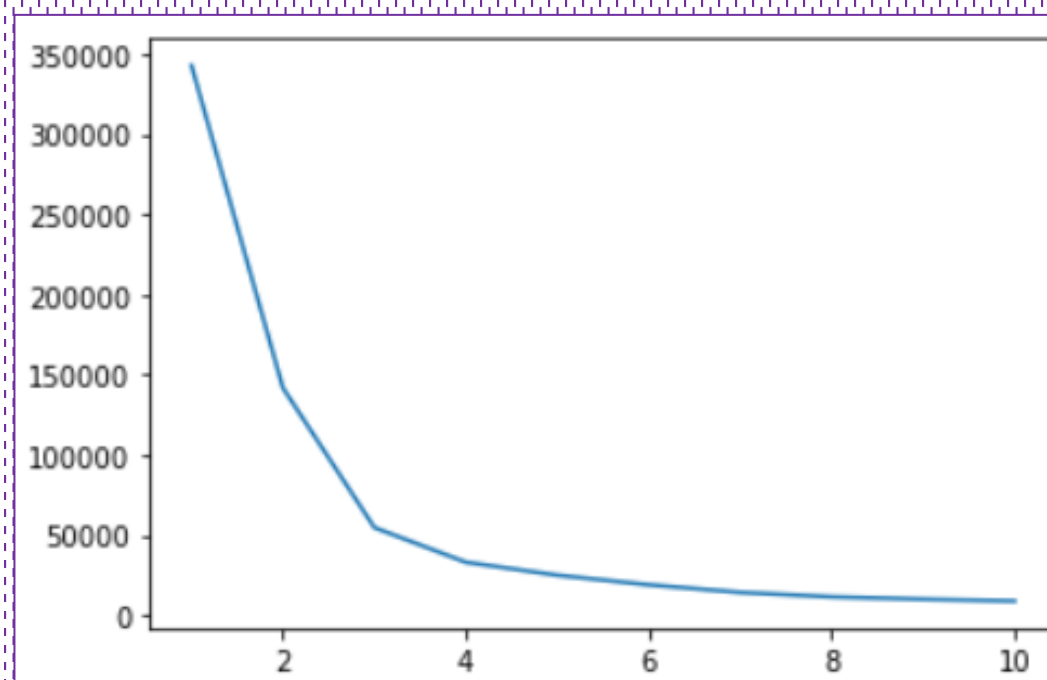
	2019	2020
0	0.114749	0.108612
1	0.102394	0.080894
2	0.188982	0.182871
3	0.149752	0.111015
4	0.083306	0.079097

## Normalization.

Normalization is the process of converting real-valued numeric properties into a 0 to 1 range. Data normalization is used in machine learning to make model training less sensitive to feature scale. As a result, our model can converge to more correct weights, making it more accurate. The normalize function for data normalization is part of the Python preprocessing library. It takes an array as input and turns the values into a 0 to 1 range. After subtracting the minimum value from each item, we divide the result by the range. Range is the difference between the maximum and minimum value.

## Elbow Curve

Before clustering, it's critical to determine how many clusters are possible in a particular data set by applying the elbow curve approach to the library SK. Learn that this is a method of inertia by default and append it to the list, then plot the entire list data using matplotlib and claim that where the line becomes straight until that point, note the value of x and assume that number of clusters.

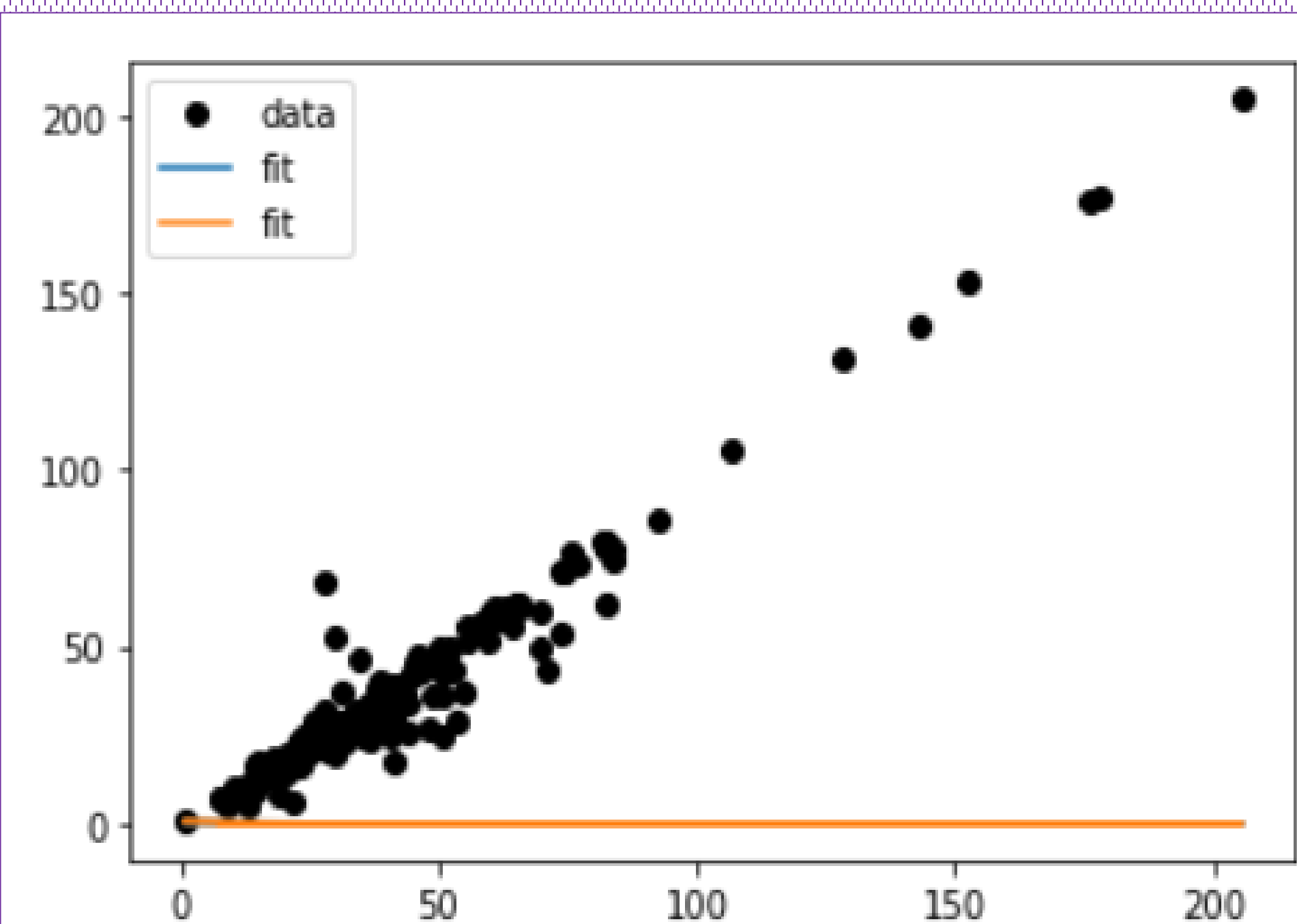


## Clustering.

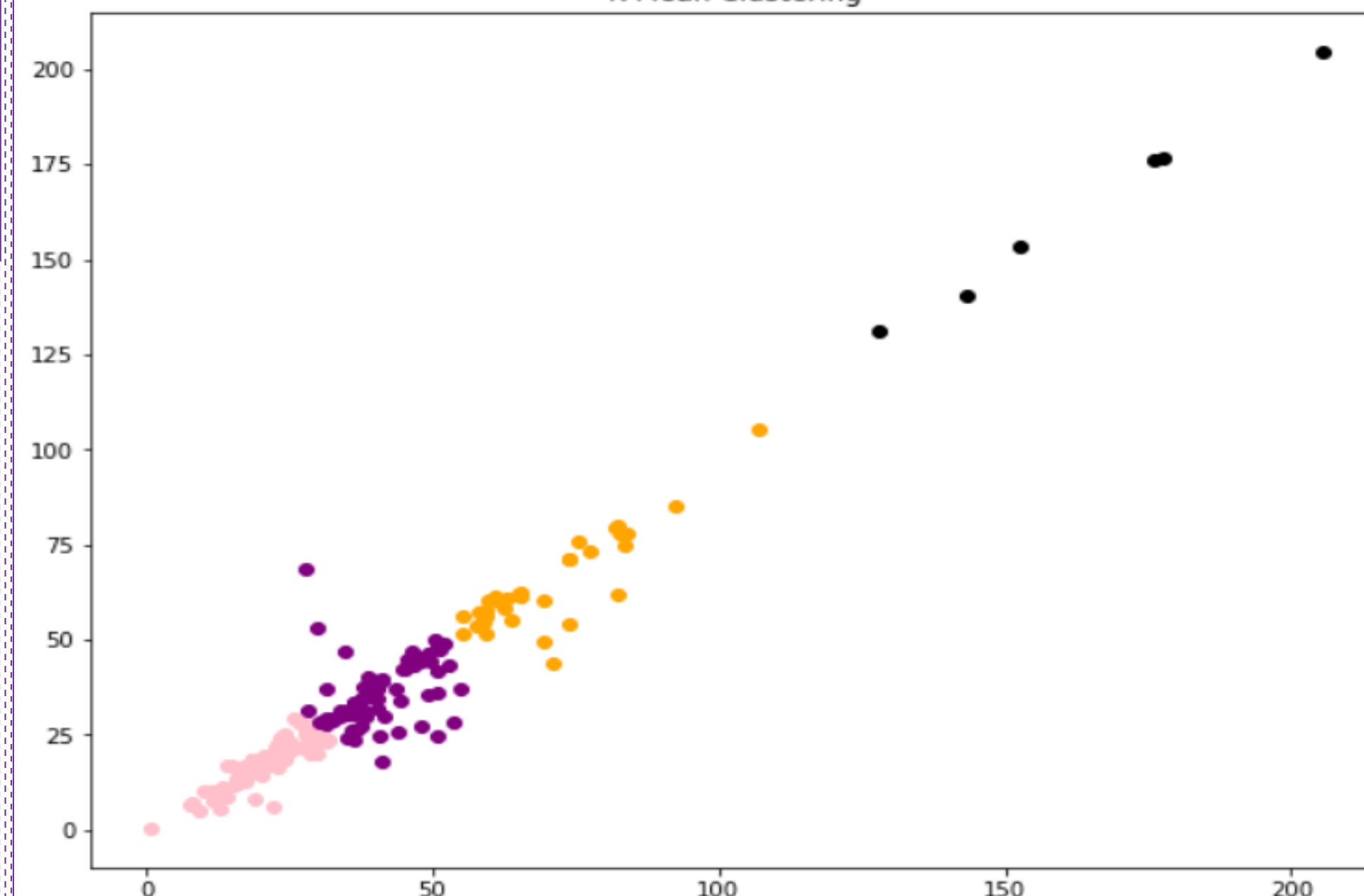
Clustering is a set of techniques for organizing data into groups. Clusters are groups of data objects that are more similar to one another than they are to data objects from other clusters. To locate groupings of data objects in a dataset, the unsupervised machine learning technique k-means clustering is utilized. There are several clustering methods to choose from, but k-means is one of the most popular and accessible.

## Curve fit method

Curve fitting is a type of optimization in which the optimal set of parameters for a certain function that best fits a set of data is determined. Unlike supervised learning, curve fitting requires the specification of the function that translates input examples to outputs. We also use built-in libraries to fit the dataset using the curve fit method.



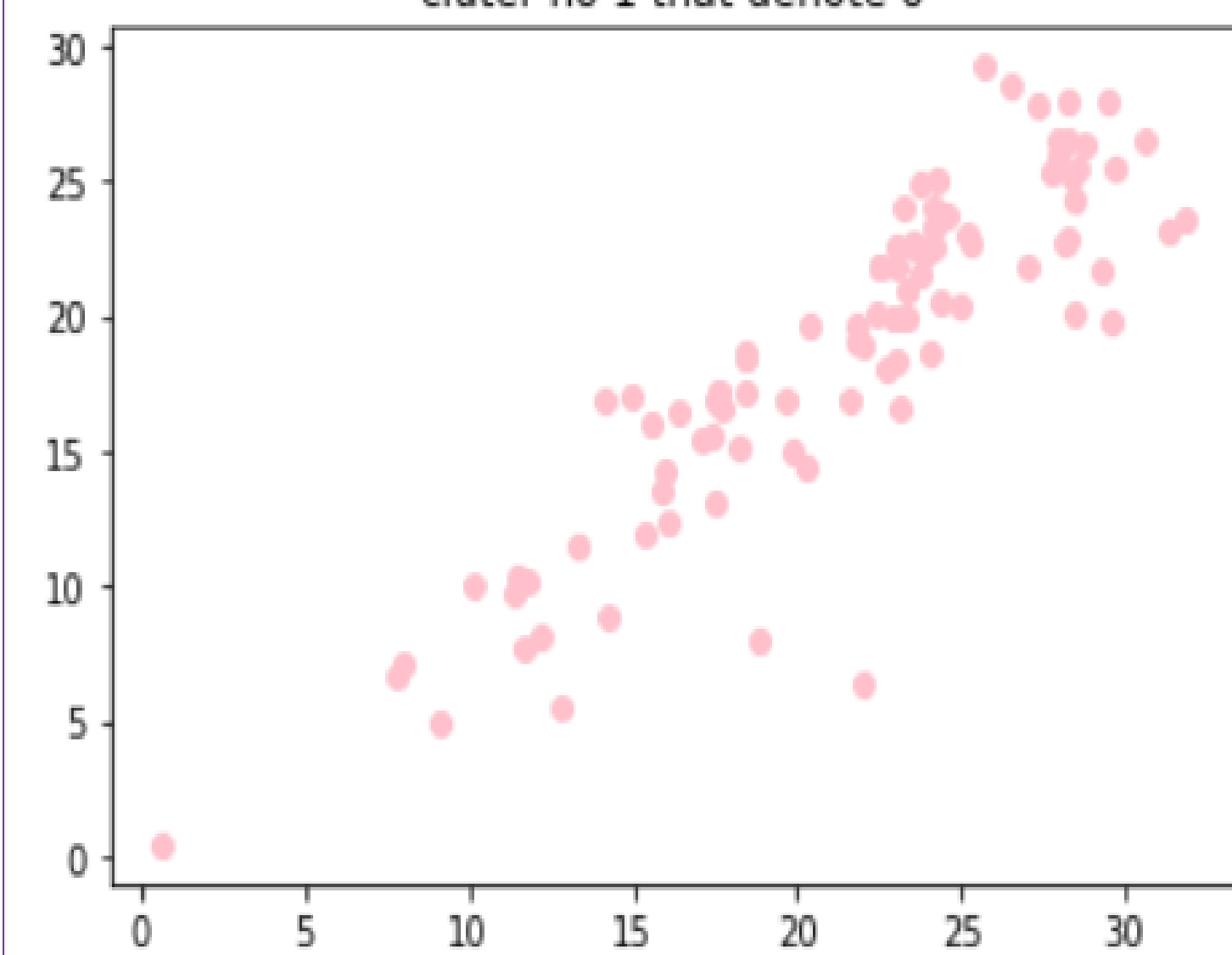
## K Mean Clustering



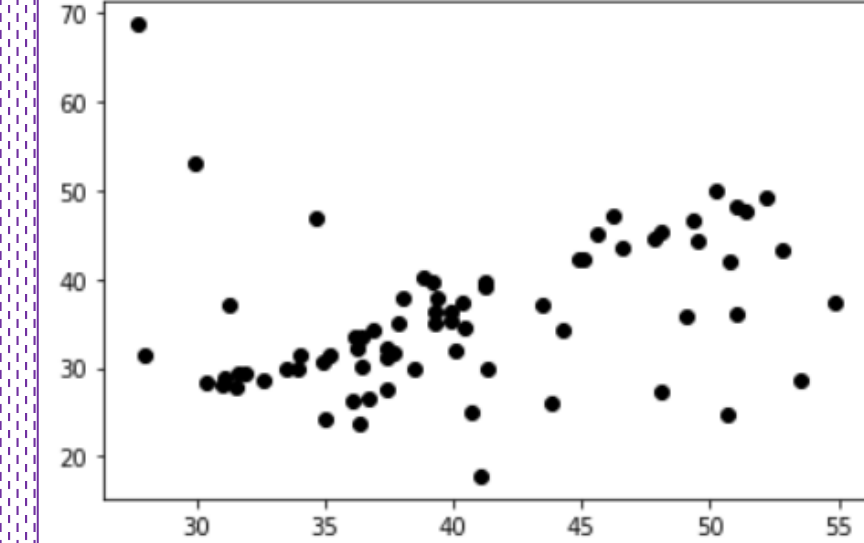
## K mean cluster

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 4 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

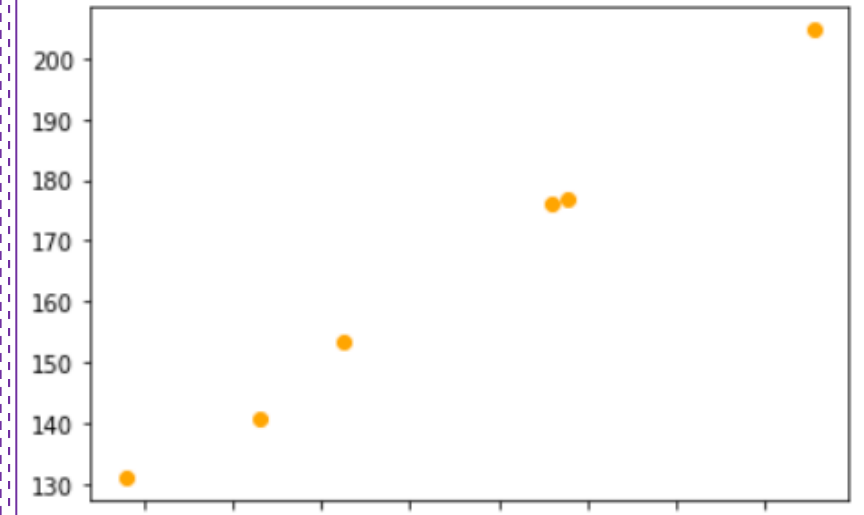
cluter no 1 that denote 0



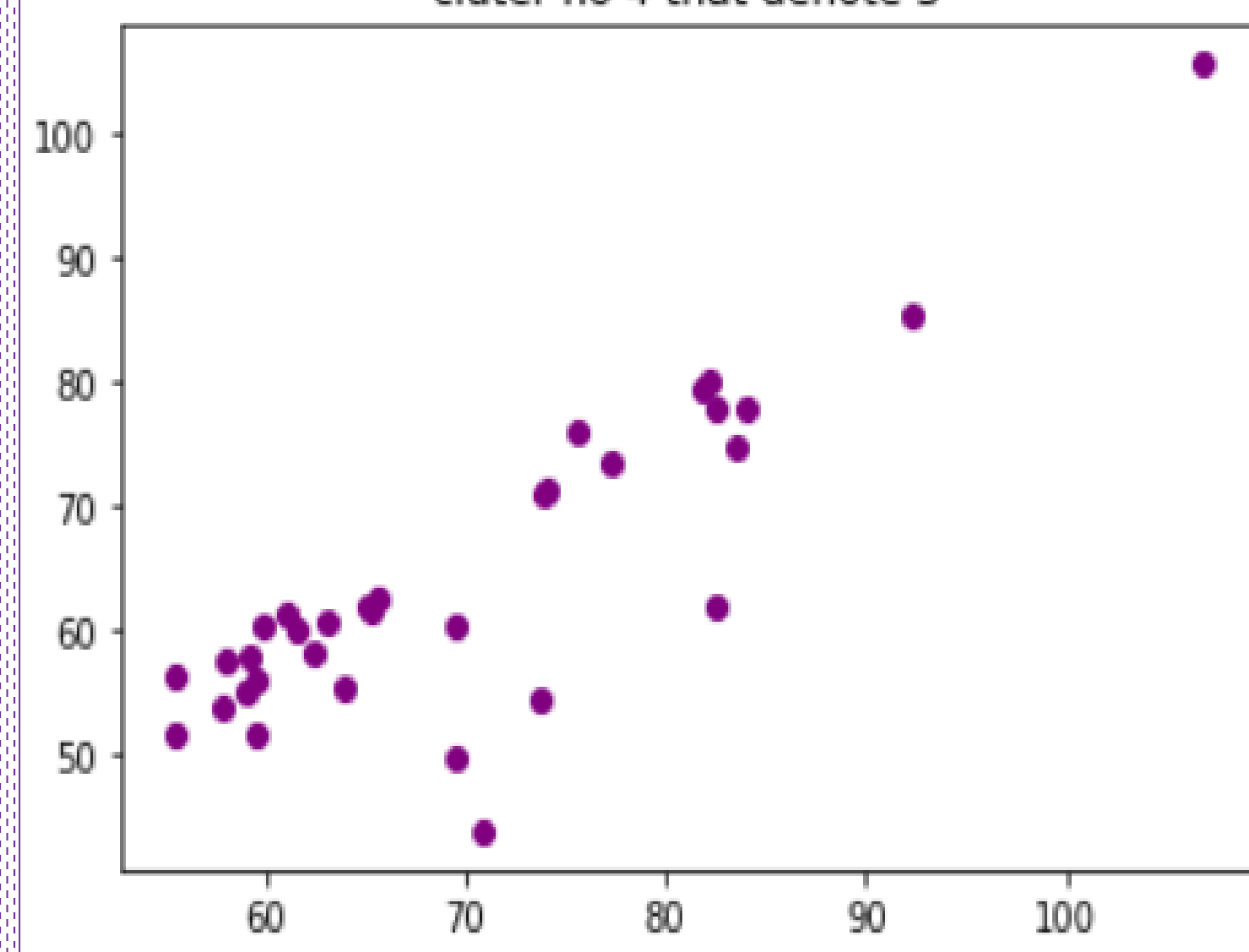
cluter no 3 that denote 2



cluter no 2 that denote 1



cluter no 4 that denote 3



## Interpretation of Results.

according to prediction of k mean most of values are 0 then 2 then 3 and 1 is at last  
first cluster is of 0  
second cluster is of 2  
third cluster is of 3  
fourth cluster is of 1

<https://github.com/ZayyamFida/assignment3>