**K-means clustering**

**What is a cluster ?**

A cluster refers to a collection of data points aggregated together because of certain similarities.

**Objective :**

To find k number of clusters in the dataset

**Algorithm of K means Clustering :**

* Initialize k points randomly
* Categorize each point to its closest mean
* Update the means by finding the average of items categorized in that mean.
* Repeat the process for fixed number of iterations .

**How to choose K ?**

Elbow curve is used to choose the perfect ‘k ’.

\* Run the k means algorithm for a fixed number of iterations.

\* Calculate the Sum Of Squares of Errors (SSE) [Distortion from the mean] for each k.

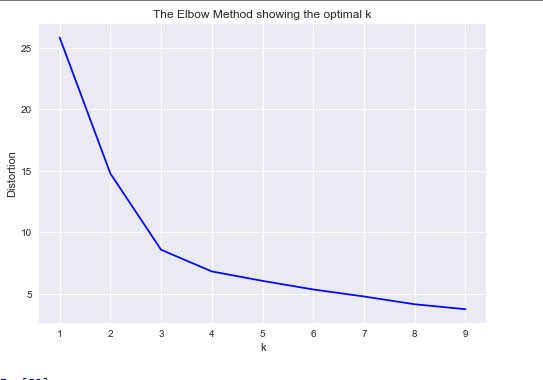
\* Plot SSE Vs K graph

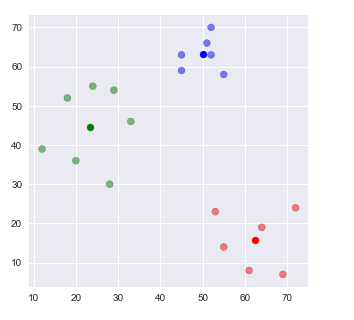
We can observe that as ‘k’ increases the SSE value decreases and tends to zero. So ,we choose the k which gives the elbow point as it will give the optimal clusters with a less error.

**IMPLEMENTATION IN PYTHON:OUTPUTS**A screenshot of a cell phone

Description automatically generated

**OUTPUT:**

****

****

**INFERENCES:**

**\*From the elbow curve , we identify the k value as 3.**

**REFERENCES:**

<https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>

<https://pythonprogramminglanguage.com/kmeans-elbow-method/>