**K Means Clustering**

In **Unsupervised Learning** we have different type of algorithms such as:

1. Clustering
2. Association Rules
3. Recommendation Engine
4. PCA
5. Text mining
6. NLP

In **Clustering** we have :

1. Hierarchial Clustering
2. K-Means Clustering
3. DBSCAN Clustering

In this repository we will discuss mainly about *K-Means Clustering*

Step 1: K= 3, we can take different values for K, here K=3 so entire data is randomly divided in to 3 parts need not be equal.  
Step 2: Centroid computation - calculate centroid for each part.  
Step 3: find distance from centroid to all data points in each part.  
Step 4: Move data point in to nearest centroid.  
Step 5: Recompute centroids.  
Step 6: Repeat steps from 3 to 5 until there is no need to move data points from one cluster to other cluster.

By using Elbow graph or screw plot we will decide proper K-value.

**Advantages:**

1. Partition of data accurately fast
2. Suitable for larger datasets

**Diasadvantages:**

1. If we have outliers, it will give false clusters

**Data used:**

Universities: Clustering of similar universities

Crime\_data: Clustering of Murderers based on their similarity

EastWestAirlines: Clustering of Airlines based on their similarity

**Programming:**

Python

The Codes regarding this K-Means Clustering with three different business problems *Clustering of universities* ,*Clustering of murderers*, *Clustering of Airlines* with their datasets are present in this Repository in detail