Aim:

To build a k-Means clustering algorithm for the given test datasets using Python Scikit-learn package.

Datasets

Clustering Datasets:

1. **Sample data *(D1):*** *Perform k-means clustering manually with k=2 on the given example and plot the observations, label the clusers.*

Obs.	X_1	X_2
1	1	4
2	1	3
3	0	4
4	5	1
5	6	2
6	4	0

2. **Air Traffic Passenger Data *(D2):*** San Francisco International Airport Report on Monthly Passenger Traffic Statistics by Airline. Airport data is seasonal in nature, therefore any comparative analyses should be done on a period-over-period basis (i.e. January 2010 vs. January 2009) as opposed to period-to-period (i.e. January 2010 vs. February 2010). It is also important to note that fact and attribute field relationships are not always 1-to-1. For example, Passenger Counts belonging to United Airlines will appear in multiple attribute fields and are additive, which provides flexibility for the user to derive categorical **Passenger Counts** as desired.

Implement the k - Means Clustering

1. The script must load the Air Traffic Passenger dataset.
2. Find out how these airlines can be assigned to clusters using the K-Means algorithm.
3. The range of clusters is defined as range (2, 6). For each number of clusters, the clustering algorithm is run and the WCSS and Silhouette Scores are saved into a list.
4. The optimal number of clusters is evaluated using the **Silhouette Score.**
5. The results to be shown on a plot (WCSS, Silhouette Score) for each k (number of clusters)