Course Name: Business Intelligence Lab **Course Code:** 20CSP-421

Experiment - 3.3

Aim:

Write program to perform k-means clustering using clustering algorithm

Software Required:

- Python

Description:

The experiment involves writing a program to implement the k-means clustering algorithm for business intelligence purposes. Participants will learn how to preprocess data, initialize clusters, assign data points to clusters, update cluster centroids, and iterate until convergence.

Pseudo code/Algorithms/Flowchart/Steps:

1.Import and Preprocess Data:

a. Select a dataset suitable for clustering analysis.

b. Preprocess the data by handling missing values, scaling features, and addressing any outliers.

2.Initialize Cluster Centroids:

a. Determine the desired number of clusters, k.

b. Randomly select k data points from the dataset as initial cluster centroids.

3.Assign Data Points to Clusters:

a. Calculate the distance between each data point and the cluster centroids (e.g., using Euclidean distance).

b. Assign each data point to the cluster with the closest centroid.

4.Update Cluster Centroids:

a. Calculate the mean of each cluster's data points to obtain updated centroids.

b. Update the centroids based on the mean values.

Course Name: Business Intelligence Lab **Course Code:** 20CSP-421

5.Iterate until Convergence:

a. Repeat steps 3 and 4 until convergence is achieved (when the centroids stabilize or the maximum number of iterations is reached).

b. Monitor the change in centroids and data point assignments to determine convergence.

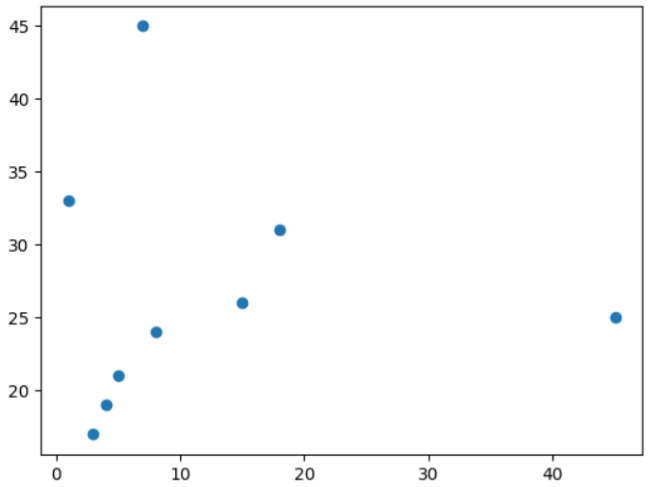
6.Analyze Clustering Results:

a. Evaluate the quality of the clustering by assessing metrics such as within-cluster sum of squares (WCSS) or silhouette score.

b. Interpret and analyze the resulting clusters to gain insights and identify patterns within the data.

Implementation

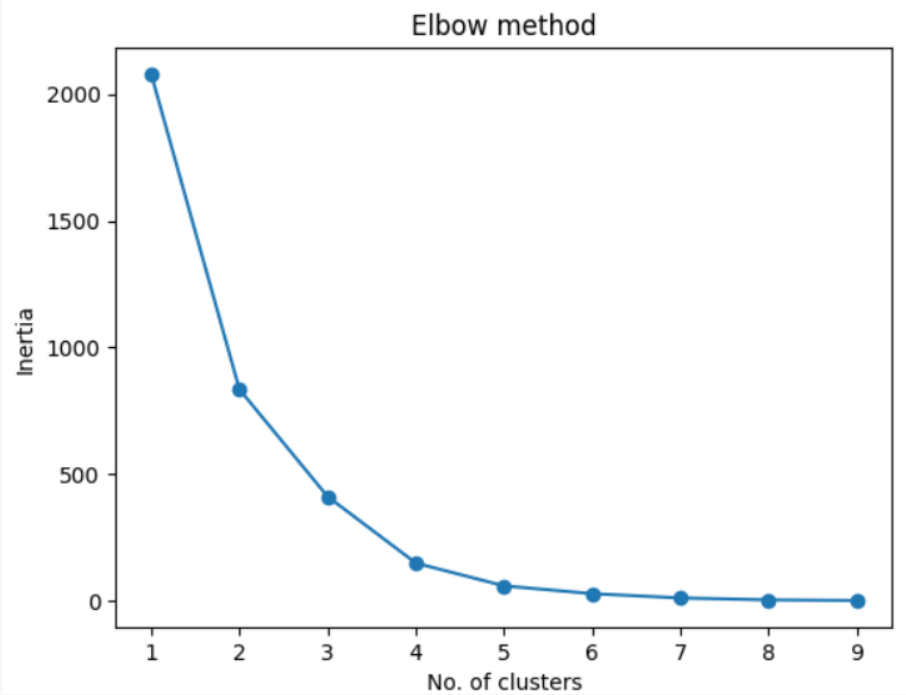
```
import matplotlib.pyplot as plt
x=[5,8,7,4,3,45,18,15,1]
y=[21,24,45,19,17,25,31,26,33]
plt.scatter(x, y)
plt.show()
```




```
from sklearn.cluster import KMeans
```

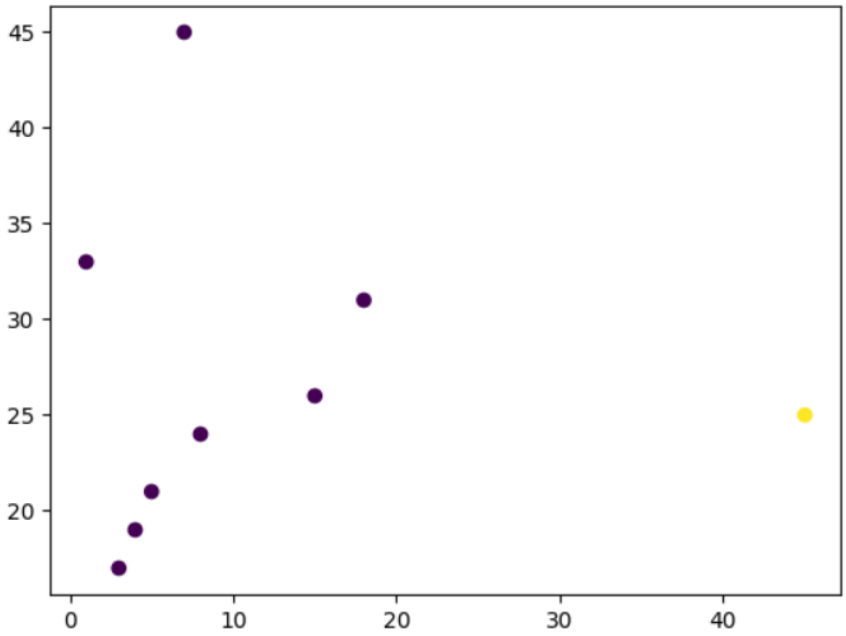
Course Name: Business Intelligence Lab **Course Code:** 20CSP-421

```
data=list(zip(x,y))
inertias=[]
for i in range (1,10):
  kmeans= KMeans(n_clusters=i)
  kmeans.fit(data)
  inertias.append(kmeans.inertia_)
plt.plot(range(1,10),inertias,marker="o")
plt.title("Elbow method")
plt.xlabel("No. of clusters")
plt.ylabel("Inertia")
plt.show()
```




```
kmeans=KMeans(n_clusters=2)
kmeans.fit(data)
plt.scatter(x, y, c=kmeans.labels_)
plt.show()
```

Course Name: Business Intelligence Lab **Course Code:** 20CSP-421



Learning Outcomes:

1. Develop programming skills to implement the k-means clustering algorithm.
2. Understand the importance of data preprocessing in preparing data for clustering analysis.
3. Learn the steps involved in initializing cluster centroids and assigning data points to clusters.
4. Gain proficiency in updating cluster centroids and iterating until convergence is achieved.