

## **Experiment No:- 5**

### **Title:**

Implementation of Clustering algorithm (K-means)

**Tool used:** Python

### **Theory:**

K-means clustering is an unsupervised learning algorithm that partitions data into K clusters by minimizing the variance within each cluster. The algorithm begins with the random initialization of K centroids. Each data point is then assigned to the nearest centroid based on a distance metric, typically the Euclidean distance. After assignment, the centroids are recalculated as the mean of all data points in each cluster. This iterative process of assignment and centroid recalculation continues until convergence, which occurs when the centroids no longer change significantly or a maximum number of iterations is reached. Key concepts include:

- **Centroids:** The central point of a cluster, representing the mean of all data points in the cluster.
- **Distance Metrics:** Euclidean distance is commonly used to measure the proximity of data points to centroids.
- **Iterative Process:** The algorithm iteratively refines the centroids and cluster assignments to optimize the clustering outcome.

### **Conclusion:**

Hence, we implemented the overview of K-means clustering using python.