A.I.S.S.M.S
COLLEGE OF ENGINEERING
NO - 1, Kennedy Road, Near R.T.O Pune - 411001

ML.

Practical - 06.

* Aim: Implement K means clustering / heirarchical clustering on sales data .csv. dataset. determine the number of clusters using elbow method.

* Theory.

1] K-means clustering Overview -

- K means is an unsupervised learning algorithm used to group data into k clusters. based on feature similarity.
- It minimizes the within cluster variance (sum of squared distances between points and their cluster centroid)
- Steps -
  - Select number of clusters k.
  - Initialize k cluster centroids randomly.
  - Assign each data point to the nearest centroid.
  - Recalculate centroids as the mean of all points in each cluster.
  - Repeat until centroids no longer change significantly.

A.I.S.S.M.S
COLLEGE OF ENGINEERING
NO - 1, Kennedy Road, Near R.T.O Pune - 411001

2] Heirarchical Clustering overview -

- Builds a heirarchy of clusters without pre-specifyin k.
- Two main types -
  - Agglomerative - Start with each point as its own cluster and merge clusters iteratively.
  - Divisive - Start with one cluster and split iteratively.
- Results are visualized using a dendrogram, the cut point on the dendrogram determines the number of clusters.

3] Algorithm Steps - for K means with elbow method.

Step 1: Load the dataset using pandas.

Step 2: Select the numerical features you want to cluster (eg. sales amount, quantity, ordered, MSRP).

Step 3: Handle missing values if any.

Step 4: Normalize/ standardize the features to ensure equal importance.

Step 5: Use the elbow method.
- Run k means for a range of k values (eg 1 to 10).
- Compute the within cluster sum of squares (WCSS) for each k.
- Plot k vs. WCSS.
- The "elbow point" on the plot indicates the optimal k.

Step 6: Run k means with the chosen k to obtain cluster labels.

Step 7: For heirarchical clustering, compute a linkage matrix (eg. Wards method), plot a dendrogram and cut it at a certain height to define clusters.

* CONCLUSION:

By applying k-means or Heirarchical clustering to the sales dataset.

- We can group sales data points into distinct clusters based on sales-related features

→

28.

- The elbow method helps determine the optimal number of clusters by balancing model complexity with explained variance.

- Heirarchical clustering techniques allow businesses to segment customers or transactions for targeted marketing, inventory optimization and strategic decision making.

* * * * * * *