visual inspection helps understand the relationships between the datapoints within and across the cluster.

* Interpreting Result :

The result of k-means clustering are interpreted in the content of the dataset, identifying cluster characteristics & any meaningful patterns or insights.

For example, clusters with similar combinations of VAR1 & VAR2 may represent distinct groups or categories within the data.

* Conclusion :

The concept of k-means clustering is studied and the program is executed successively.

Applying K-means to VAR1 & VAR2 :
In this experiment, nine combinations of VAR1 & VAR2 variables will be clustered using K-means based on their similarities. Each combination represents a datapoint in a two-dimensional feature space.

Initial centroid Initialization:
Cluster centroids are initially assigned randomly using techniques like K-means to improve converging to suboptimal solutions. Proper initialization helps to prevent the algorithm from converging to cluster quality.

K-means Iterative Process:
During each iteration, data points are assigned to the nearest centroid & centroids are updated based on the mean of data points in each cluster. The process continues until centroid no longer change significantly or maximum number of of iterations is reached.

Evaluation metrics:
Evaluation metrics like within-cluster sum of squares (WCSS) or Silhouette score quantify the quality of clusters produced by K-means.
Lower wcss indicates tighter clusters, while higher silhouette score suggests better cluster separation.

Vizualizing clusters:
Vizualization techniques like satter plots or heatmaps can be used to vizualize the clusters and gain insights into the underlying structure of the data.

Experiment - 4

RAISONI GROUP
— a vision beyond —

* **Aim :** Classify nine combinations of VAR1 & VAR2 using K-means clustering.

* **Theory :**

Introduction to K-means clustering :

- K-means is a popular unsupervised learning algorithm used for clustering. It is used to partition a dataset into K clusters based on similarity.
- It is widely used for many applications including customer segmentation, image compression & anamoly detection.
- K means iteratively assigns a datapoint to nearest cluster centroid & updates centroid based on the mean of data points in each other.
- The process continues until convergence or until a max. no. of iterations is reached.

Data Preprocessing :

Data preprocessing involves steps like normalization, standardization & handling missing values or outliers to improve clustering performance. For example, scaling the data to have zero mean & unit variance can prevent features with larger scales from dominating the clustering.

Choosing the number of clusters (K) :

Determining the optimal no. of clusters (K) is crucial for meaningful clustering result. Techniques like elbow method, silhouette score or domain knowledge can help in selecting an appropriate K value.