Implement clustering techniques – Hierarchical and K-Means

AIM:

To Implement clustering techniques – Hierarchical and K-Means using R.

PROCEDURE:

- Collect and load the dataset from sources like CSV files or databases.
- Clean and preprocess the data, including handling missing values and scalingfeatures.
- Determine the number of clusters (K) for K-Means, or decide on the stoppingcriterion for Hierarchical Clustering.
- Choose the appropriate clustering algorithm: K-Means for partitioning,Hierarchical for nested clustering.
- Applythe K-Means algorithm usingfit_predict toassign data pointsto clusters.
- Apply the Hierarchical Clustering algorithm using Agglomerative Clustering for hierarchical clusters.
- Visualize the clusters with scatter plots for K-Means, and dendrograms forHierarchical Clustering.
- Evaluate clustering performance using metrics like silhouette score or inertia(for KMeans).
- Fine-tune the clustering by adjusting the number of clusters or linkage criteria.
- Interpret the results to understand the structure and relationships within thedata.

CODE:

Hierarchical Clustering.R:

```
# Load the iris dataset
data(iris)
# Use only the numeric columns for clustering (exclude the Species column)
iris_data <- iris[, -5]
# Standardize the data
iris_scaled <- scale(iris_data)
# Compute the distance matrix
distance_matrix <- dist(iris_scaled, method = "euclidean")
# Perform hierarchical clustering using the "complete" linkage method
hc_complete <- hclust(distance_matrix, method = "complete")
# Plot the dendrogram
plot(hc_complete, main = "Hierarchical Clustering Dendrogram", xlab = "", sub = "", cex = 0.6)
# Cut the tree to form 3 clusters
clusters <- cutree(hc_complete, k = 3)
# Print the cluster memberships
print(clusters)
```

```
# Add the clusters to the original dataset
iris$Cluster <- as.factor(clusters)
# Display the first few rows of the updated dataset
head(iris)
```
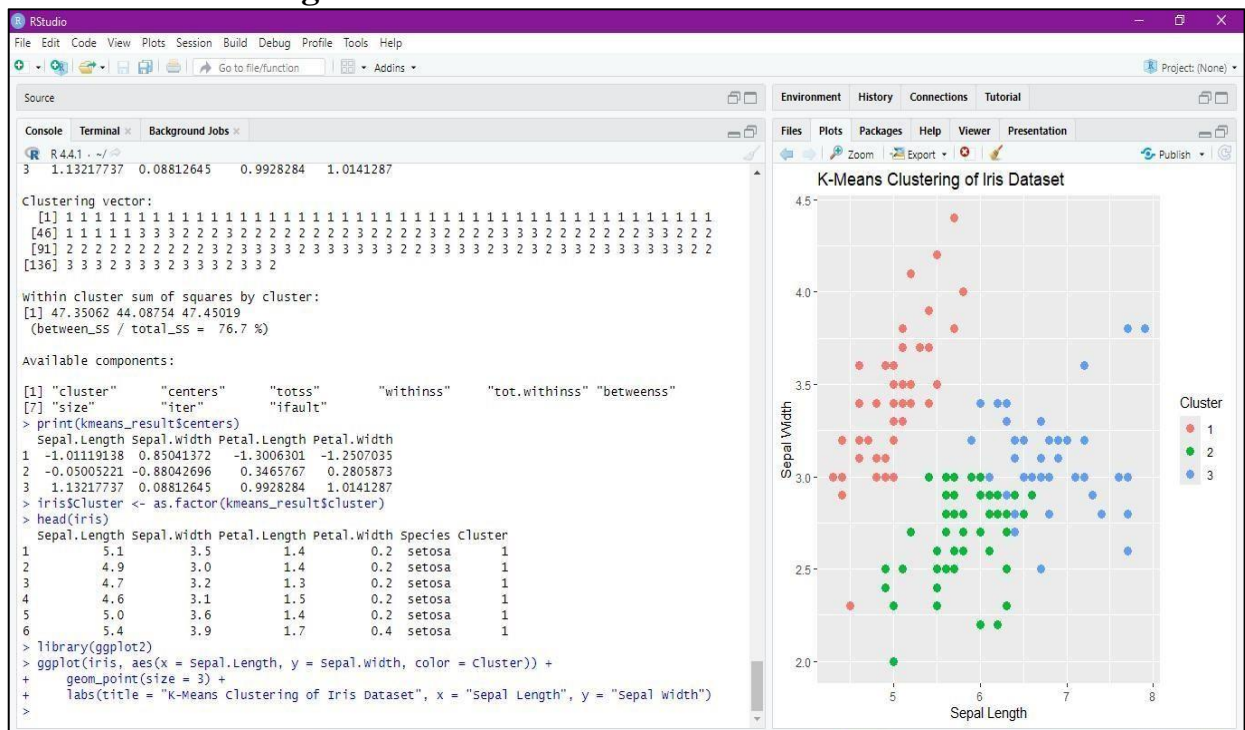**K-Means Clustering.R:**

```
# Load the iris dataset
data(iris)
# Use only the numeric columns for clustering (exclude the Species column)
iris_data <- iris[, -5]
# Standardize the data iris_scaled
<- scale(iris_data) # Set the
number of clusters set.seed(123) #
For reproducibility k <- 3 #
Number of clusters # Perform K-
Means clustering
kmeans_result <- kmeans(iris_scaled, centers = k, nstart = 25)
# Print the K-Means result
print(kmeans_result) # Print
the cluster centers
print(kmeans_result$centers)
# Add the cluster assignments to the original dataset
iris$Cluster <- as.factor(kmeans_result$cluster) #
Display the first few rows of the updated dataset
head(iris)
# Plot the clusters library(ggplot2) ggplot(iris, aes(x = Sepal.Length, y
= Sepal.Width, color = Cluster)) +
  geom_point(size = 3) + labs(title = "K-Means Clustering of Iris Dataset", x = "Sepal
  Length", y = "Sepal Width")
```

OUTPUT:

Hierarchical Clustering:
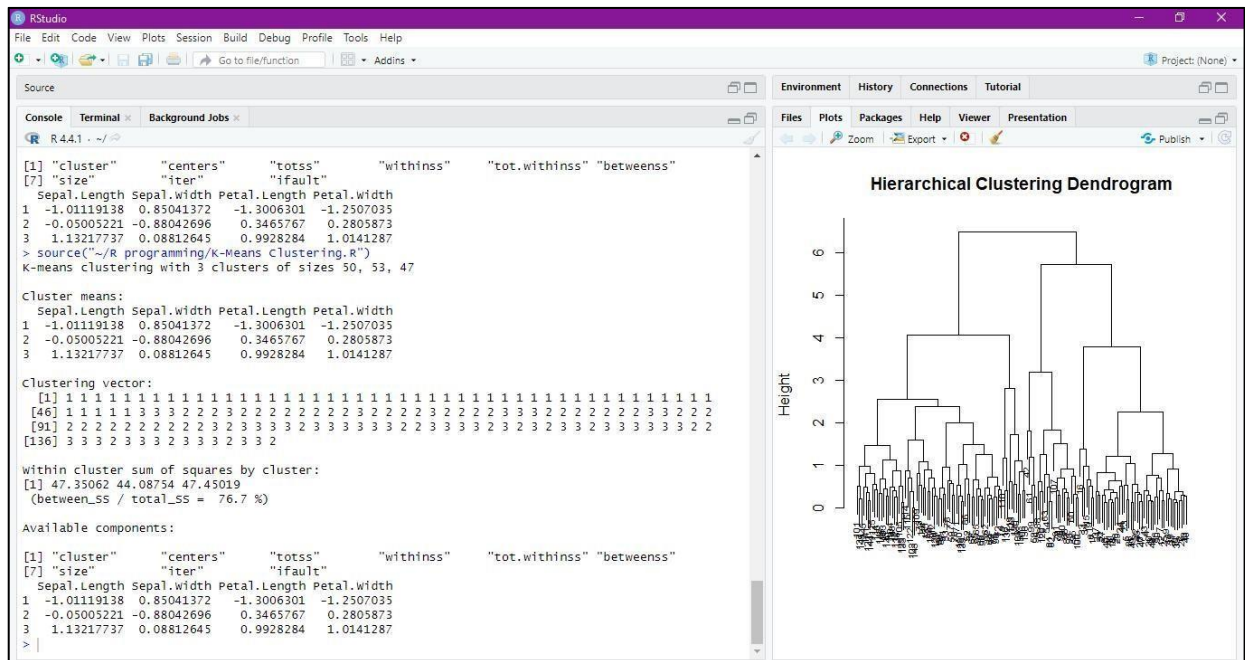
K-Means Clustering:

RESULT:

Thus, to implement clustering techniques – Hierarchical and K-Means using R has been successfully executed.