***# K-Means Clustering***

install.packages("cluster")

install.packages("factoextra")

library(cluster)

library(factoextra)

# Load and preprocess the data

df <- USArrests

df <- na.omit(df) # Eliminates rows with missing data

df <- scale(df) # Standardization of the variables

dim(df)

***# Elbow Method to inspect the within-cluster sum of squares***

fviz\_nbclust(df, kmeans, method = "wss")

***# Using Gap Statistic to further evaluate the optimal number of clusters***

gapstats <- clusGap(df, FUN = kmeans, nstart = 25, K.max = 10, B = 50)

fviz\_gap\_stat(gapstats)

# Based on the analysis (k=4)

km <- kmeans(df, centers = 4, nstart = 25)

# Visualize the clusters using the 'factoextra' package

fviz\_cluster(km, data = df)

***# Examine the cluster centres by computing the mean per cluster for the original variables***

aggregate(USArrests, by = list(cluster = km$cluster), mean)

***#K-Medoids***

data(iris)

head(iris)

library(cluster)

iris\_prepared <- iris[, 1:4]

head(iris\_prepared)

*# The Partion Around Medoids (PAM) function will be used to apply K-Medoid with k=3*

kmedoids\_final <- pam(iris\_prepared, k = 3)

#print the medoids which represent the central points of each cluster

print(kmedoids\_final$medoids)

# Print the medoids

print(kmedoids\_final$medoids)

# Print the cluster assignments

print(kmedoids\_final$clustering)

# Plotting of the clusters

plot (kmedoids\_final, main = "K-Medoids Clustering")

***#Hierarchical Clustering Applications***

data(iris)

head(iris)

iris\_hierarchy <- iris[, 1:4]

# View the prepared data

head(iris\_hierarchy)

# Calculate the Distance Matrix, which will show the pairwise distances between all observations.

distance\_matrix <- dist(iris\_hierarchy) # distance matrix using Euclidean distance

head(as.matrix(distance\_matrix)) # first few values of the distance matrix

#Hierarchical clustering using the complete linkage method

hc <- hclust(distance\_matrix, method = "complete")

print(hc) # View the hierarchical clustering result

#plot the Dendrogram

plot (hc, main = "Dendrogram", xlab = "Sample Index", ylab = "Height", sub = "")

clusters <- cutree(hc, k = 3) # Here, we choose 3 clusters

pairs(iris\_hierarchy, col = clusters, pch = 20, main = "Hierarchical Clustering on IRIS Data set")