NAME - VIKRAM SURYA R
Reg No - 19BCE1490

Tasks for Week-7: Partitioning Based Clustering
Understand the following operations/functions on 'iris' data and perform similar operations on 'USArrests' dataset based on given instructions.

Aim: To develop an Kmeans clustering model for the given data using R programming and perform operations on the data "USArrests".

Algorithm:
• Set working directory and read data.
• Create a scaled data from the original data.
• Train the Kmeans model and review all the terms of the model.
• Find WCSS values and store them in a list.
• Plot the graph to find the elbow point.
• Train the K-Medoids model and review all the terms of the model.

Inference/Results:
• The elbow point is 4 so we need to select 4 clusters.
• The plots are available at the end of the submission.

CODE -

```
rm(list = ls())
mydata<-read.csv(file.choose(),header=T)
View(mydata)
df <- scale(mydata[2:5])
fit<- kmeans(df,centers = 2)
fit$size
fit$withinss
fit$tot.withinss
Kmax <- 15
WCSS <- rep(NA,Kmax)
nClust <- list()
for (i in 1:Kmax){
    fit<- kmeans(df,i)
    WCSS[i] <- fit$tot.withinss
    nClust[[i]] <- fit$size
}
plot(1:Kmax,WCSS,type="b",pch=19)
fit <- kmeans(df, centers = 4)
fit$cluster
fit$size
# install.packages("factoextra")
library(factoextra)
library(ggplot2)
fviz_nbclust(df, kmeans, method = "wss")
fviz_cluster(fit, mydata[2:5])
# install.packages("cluster")
library(cluster)
fitm <- pam(df, 3, metric = "manhattan") # K-
Medoids fitm
fitm$medoids
fitm$clustering
```

SCREENSHOTS -

```
R  R 4.1.0 · ~/
> rm(list = ls())
> mydata<-read.csv(file.choose(),header=T)
> view(mydata)
> df <- scale(mydata[2:5])
> fit<- kmeans(df,centers = 2)
> fit$size
[1] 20 30
> fit$withinss
[1] 46.74796 56.11445
> fit$tot.withinss
[1] 102.8624
> Kmax <- 15
> WCSS <- rep(NA,Kmax)
> nClust <- list()
> for (i in 1:Kmax){
+    fit<- kmeans(df,i)
+    WCSS[i] <- fit$tot.withinss
+    nClust[[i]] <- fit$size
+ }
> plot(1:Kmax,WCSS,type="b",pch=19)
> fit <- kmeans(df, centers = 4)
> fit$cluster
 [1] 3 4 4 3 4 4 1 1 4 3 1 2 4 1 2 1 2 3 2 4 1 4 2 3 4 2 2 4 2 1 4 4 3 2 1 1 1 1 1 3 2 3 4 1 2 1 1 2 2 1
> fit$size
[1] 16 13  8 13
> # install.packages("factoextra")
> library(factoextra)
> library(ggplot2)
> fviz_nbclust(df, kmeans, method = "wss")
> fviz_cluster(fit, mydata[2:5])
> # install.packages("cluster")
> library(cluster)
> fitm <- pam(df, 3, metric = "manhattan") # K-Medoids
> fitm
Medoids:
     ID   Murder    Assault   UrbanPop      Rape
[1,] 31  0.8292944  1.3708088  0.3081225  1.1603196
[2,] 36 -0.2727580 -0.2371077  0.1699510 -0.1315342
[3,] 15 -1.2829727 -1.3770485 -0.5899924 -1.0603878
Clustering vector:
 [1] 1 1 1 2 1 1 3 2 1 1 2 3 1 2 3 2 2 1 3 1 2 1 3 1 2 1 1 1 3 2 2 2 2 2 1 3 1 1 2 3 2 2 3 3 2
Objective function:
   build    swap
2.006126 2.006126

Available components:
 [1] "medoids"    "id.med"    "clustering" "objective" "isolation" "clusinfo"  "silinfo"   "diss"      "call"      "data"
> fitm$medoids
       Murder    Assault   UrbanPop      Rape
[1,]  0.8292944  1.3708088  0.3081225  1.1603196
[2,] -0.2727580 -0.2371077  0.1699510 -0.1315342
[3,] -1.2829727 -1.3770485 -0.5899924 -1.0603878
> fitm$clustering
 [1] 1 1 1 2 1 1 3 2 1 1 2 3 1 2 3 2 2 1 3 1 2 1 3 1 2 1 1 1 3 2 2 2 2 2 1 3 1 1 2 3 2 2 3 3 2
>
```

Optimal number of clusters

Cluster plot