**PRACTICAL NO.4**

**Perform the data clustering using clustering algorithm using R/Python**

**a]Data clustering using the R:**

Code:

newiris <- iris

newiris$Species <- NULL

(kc <- kmeans(newiris, 3))

table(iris$Species, kc$cluster)

plot(newiris[c("Sepal.Length", "Sepal.Width")], col = kc$cluster)

points(kc$centers[, c("Sepal.Length", "Sepal.Width")], col = 1:3, pch = 8, cex = 2)

Output:

> newiris <- iris

> newiris$Species <- NULL

> (kc <- kmeans(newiris, 3))

K-means clustering with 3 clusters of sizes 50, 62, 38

Cluster means:

Sepal.Length Sepal.Width Petal.Length Petal.Width

1 5.006000 3.428000 1.462000 0.246000

2 5.901613 2.748387 4.393548 1.433871

3 6.850000 3.073684 5.742105 2.071053

Clustering vector:

[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

[38] 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2

[75] 2 2 2 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 3 3 3 3 2 3 3 3 3

[112] 3 3 2 2 3 3 3 3 2 3 2 3 2 3 3 2 2 3 3 3 3 3 2 3 3 3 3 2 3 3 3 2 3 3 3 2 3

[149] 3 2

Within cluster sum of squares by cluster:

[1] 15.15100 39.82097 23.87947

(between\_SS / total\_SS = 88.4 %)

Available components:

[1] "cluster" "centers" "totss" "withinss" "tot.withinss"

[6] "betweenss" "size" "iter" "ifault"

>

> table(iris$Species, kc$cluster)

1 2 3

setosa 50 0 0

versicolor 0 48 2

virginica 0 14 36

>

> plot(newiris[c("Sepal.Length", "Sepal.Width")], col = kc$cluster)

> points(kc$centers[, c("Sepal.Length", "Sepal.Width")], col = 1:3, pch = 8, cex = 2)

**Data plotting:**

