**DATA ANALYTICS WITH R, EXCEL AND TABLAEU**

**ASSIGNMENTR integration with Hadoop 22.1**

**ANSWERS**

**5. Problem Statement**

**1. Use the below given data set**

**DataSet**

**2. Perform the below given activities:**

**a. apply K-means clustering to identify similar recipies**

**Ans**

set.seed(20)

epiCluster<- kmeans(epi[, 3:4], 3, nstart = 20)

epiCluster

K-means clustering with 3 clusters of sizes 46, 54, 50

**b. apply K-means clustering to identify similar attributes**

**Ans** *K-means clustering with 3 clusters of sizes 46, 54, 50*

*Cluster means:*

*Petal.LengthPetal.Width*

*1 5.626087 2.047826*

*2 4.292593 1.359259*

*3 1.462000 0.246000*

*Clustering vector:*

*[1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3*

*[35] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2*

*[69] 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1*

*[103] 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 2 2 1 1 1 1 1 1 1 1*

*[137] 1 1 2 1 1 1 1 1 1 1 1 1 1 1*

*Within cluster sum of squares by cluster:*

*[1] 15.16348 14.22741 2.02200*

*(between\_SS / total\_SS= 94.3 %)*

**c. how many unique recipies that people order often**

**Ans 5.**