DATA ANALYTICS WITH R, EXCEL AND TABLAEU

ASSIGNMENT R integration with Hadoop 22.1

ANSWERS

By ASHISH S SHANBHAG

ashishshanbhag108@gmail.com

- 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

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.