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

DataSet

Ans:

library(readr)

setwd("E:\Acadgild\R\Assignment test")

zipF<- "E:\\Acadgild\\R\\Assignment test\\epi\_r.csv.zip"

outDir<-"epi\_r.csv"

h<-unzip(zipF,exdir=outDir)

myData <- read.csv(h)

data<-data.frame(myData )

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

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

Ans:

library(Hmisc)

data$calories<- with(data, impute(calories, 'random'))

data$protein<- with(data, impute(protein, 'random'))

data$fat<- with(data, impute(fat, 'random'))

data$sodium<- with(data, impute(sodium, 'random'))

names(data)

#scatter plot

plot(rating~calories,data)

#normilization

h<-data[,-1]

head(h)

m<-apply(h,2,mean)

s<-apply(h,2,sd)

h<-scale(h,m,s)

#calculate equlidien distance

distance<-dist(h)

wss<-(nrow(h)-1)\*sum(apply(h,2,var))

for(i in 2:20)wss[i]<-sum(kmeans(h,centers = i)$withinss)

plot(1:20,wss,type="b",xlab = "number of clusters",ylab = "with in group ss")

km = kmeans(h,5)

km$withinss

km$tot.withinss

km

clus<-data.frame(data$title,km$cluster)

head(clus)

table(clus$km.cluster)

table(km$cluster)

#simalar recipies

library(tidyverse)

simlar\_recipes<-clus %>% filter( km$cluster == "3")

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

Ans:

a<-data[,-1]

m<-apply(a,1,mean)

s<-apply(a,1,sd)

a<-scale(t(a),m,s)

km1 = kmeans(a,5)

km1$withinss

km1$tot.withinss

km1

similar\_attributes<-table(km1$cluster)

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

Ans:

table(data$title)

library(dplyr)

unique<-data %>% distinct(title, .keep\_all = TRUE)

unique$title

**d. what are their typical profiles**

Ans:

typical\_profiles<-data %>% distinct(title, .keep\_all = TRUE)