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

**Data Set**

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

**a. Apply PCA to the dataset and show proportion of variance**

Ans:

library(readr)

setwd("C:/Users/hemakumar/Downloads")

zipF<- "C:\\Users\\hemakumar\\Downloads\\epi\_r.csv.zip"

outDir<-"epi\_r.csv"

h<-unzip(zipF,exdir=outDir)

myData <- read.csv(h)

data<-data.frame(myData )

head(data)

summary(data)

colSums(is.na(data))

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'))

str(data)

#finding the corelation os attributes

cor(data$rating,data[,-1])

cor(data$calories,data[,-1])

cor(data$protein,data[,-1])

cor(data$fat,data[,-1])

cor(data$sodium,data[,-1])

cor(data$X.cakeweek,data[,-1])

#by seeing the corelations between all the attributes we observe there is a strong corelation between only

#calories,protein,fat,sodium. so we take only these four for our model

data1<-data[,c(1,2,3,4,5,6)]

head(data1)

data$rating<-as.factor(data$rating)

pc<-prcomp(data1[ ,-c(1,2)],

center = TRUE,

scale. = TRUE)

pc

summary(pc)

cor(pc$x)

#for predctions with principle components

prd<-predict(pc,data1)

prd<-data.frame(prd,data1[c(2,1)])

table(prd$rating)

#multinomial logistic regression

library(nnet)

prd$rating<-relevel(prd$rating,ref = "0")

mymodel<-multinom(rating~PC1+PC2+PC3,data=prd)

summary(mymodel)

#confusion matrix and accuracy

p<-predict(mymodel,prd)

tab<-table(p,prd$rating)

accuracy=sum(diag(tab))/sum(tab)

**b. Perform PCA using SVD approach**

Ans:

SVD<-svd(data1[,-c(1,2)])

**c. Show the graphs of PCA components**

Ans:

library(devtools)

library(ggbiplot)

g<-ggbiplot(pc,

obs.scale = 0.1,

var.scale = 0.1,

groups = data1$rating,

ellipse = TRUE,

circle = TRUE,

ellipse.prob = 0.90)

g<-g+scale\_color\_discrete(name="")

g<-g+theme(legend.direction = 'horizontal',

legend.position = 'top')

g