#importing dataset

df.raw <- read.csv(file ='R Studio Task/dataset\_FreshCo.csv', fileEncoding="UTF-8-BOM", na.strings = '..')

#For viewing the datatype and information on console

str(df.raw)

install.packages("ggplot2")

library(ggplot2)

#Ranking by location

ggplot(data=df.raw,aes(x=reorder(location,age),y=age)) + geom\_bar(stat ='identity',aes(fill=age))+coord\_flip() + theme\_grey() + scale\_fill\_gradient(name="Age Level")+ labs(title = 'Age Group w.r.t to location', y='age',x='location')+

geom\_hline(yintercept = df.raw$age,size = 0.1, color = 'blue')

#age counts

ggplot(data = df.raw) + labs(title = 'Age Counts', y='count',x='age')+

geom\_bar(mapping = aes(x = age))

#location count

ggplot(data = df.raw) + labs(title = 'Location Counts', y='count',x='location')+

geom\_bar(mapping = aes(x = location))

#EMploy count

ggplot(data = df.raw) + labs(title = 'The employment status of the customer', y='count',x='Employ')+

geom\_bar(mapping = aes(x = employ))

# Creating facets

ggplot(data = df.raw) +

geom\_histogram(mapping = aes(x = employ), binwidth = 0.1)+ labs(title = 'The employee belongs to the respected locations') +

facet\_wrap(~location)

# Creating facets

ggplot(data = df.raw) +

geom\_histogram(mapping = aes(x = age), binwidth = 0.1)+ labs(title = 'The employee belongs to the respected age group') +

facet\_wrap(~employ)

# Creating frequency polygon

ggplot(data = df.raw, mapping = aes(x = homeshop, colour = size)) + labs(title = 'The homeshop w.r.t to size group') +

geom\_freqpoly(binwidth = 0.1)

# Creating frequency polygon

ggplot(data = df.raw, mapping = aes(x = storeshop, colour = size)) + labs(title = 'The storeshop w.r.t to size group') +

geom\_freqpoly(binwidth = 0.1)

# plotting a scatter plot

ggplot(data = df.raw) + labs(title = 'The storeshop w.r.t to age group') +

geom\_point(mapping = aes(x = storeshop, y = age))

# plotting a scatter plot

ggplot(data = df.raw) + labs(title = 'The homeshop w.r.t to age group') +

geom\_point(mapping = aes(x = homeshop, y = age))

# creating boxplot

ggplot(data = df.raw, mapping = aes(x = location, y = age)) +

geom\_boxplot(mapping = aes(group = cut\_width(location, 0.1)))

library(dplyr)

library(ggplot2)

library(DataExplorer)

head(df.raw)

summary(df.raw)

colSums (is.na(df.raw)) # Checking if there is any missing value or not column wise

install.packages("factoextra")

library(factoextra)

library(cluster)

convert(chr(df.raw['income']))

#make this example reproducible

set.seed(1)

#perform k-means clustering with k = 4 clusters

km <- kmeans(df.raw[5:7], centers = 4, nstart = 25)

#view results

km

#plot results of final k-means model

fviz\_cluster(km, data = df.raw[5:7])

#perform k-means clustering with k = 4 clusters

km <- kmeans(df.raw[1:3], centers = 4, nstart = 25)

#view results

km

#plot results of final k-means model

fviz\_cluster(km, data = df.raw[1:3])