install.packages("psych")

install.packages("cluster")

install.packages("GGally")

install.packages("olsrr")

install.packages("ggplot2")

library(tidyverse)

library(factoextra)

library(psych)

library(cluster)

library(GGally)

library(olsrr)

library(ggplot2)

set.seed(851)

# Read the sample dataset

data\_dab <- read\_csv("~/diabetes1.csv")

View(data\_dab)

# TASK 1: Multivariate Statistical Analysis (MLO2)

## QUESTION 1 (a)

# Select the first 8 input variables

X <- data[,1:8]

# Perform PCA

pca <- prcomp(X, center = TRUE, scale. = TRUE)

# Scree plot

fviz\_eig(pca, addlabels = TRUE)

# Loadings plot

fviz\_pca\_var(pca, col.var = "contrib",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE)

# Biplot using PC1 and PC2

fviz\_pca\_biplot(pca, col.var = "contrib",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, axes = c(1,2))

## QUESTION 1 (b)

# Select the first 8 input variables

X <- data[,1:8]

# Perform PCA

pca <- prcomp(X, center = TRUE, scale. = TRUE)

# Biplot using PC2 and PC3

fviz\_pca\_biplot(pca, col.var = "contrib",

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, axes = c(2,3))

## QUESTION 2 (a)

# Select the first 8 input variables

X <- data[,1:8]

# Perform Factor Analysis

fa <- fa(X, nfactors = 3, rotate = "varimax")

# Obtain Loadings

loadings <- fa$loadings

# Obtain Communalities

communalities <- fa$communalities

# Obtain Specific Variances

specific\_variances <- fa$uniqueness

print(loadings)

print(communalities)

print(specific\_variances)

## QUESTION 2 (b)

# Obtain and interpret the rotated version of the loadings

rotated\_loadings <- fa$rotated.loadings

print(rotated\_loadings)

## QUESTION 3 (a)

# Select the 8 input variables

X <- data[,1:8]

dist\_matrix <- dist(X, method = "euclidean")

hc <- hclust(dist\_matrix, method = "ward.D2")

plot(hc)

## QUESTION 3 (b)

corr\_matrix <- cor(X)

dist\_matrix <- as.dist(1 - corr\_matrix)

hc <- hclust(dist\_matrix, method = "complete")

plot(hc)