library(readr)

library(stats)

library(psych)

library(regclass)

library(fastDummies)

library(caTools)

library(glm2)

library(tidyverse)

library(car)

african\_crises <- read\_csv("~/Downloads/african\_crises.csv")

#looking for missing values

sapply(african\_crises, function(x) sum(is.na(x)))

describe(african\_crises)

#removing unnecessary variables

african\_crises[,1:3]<-list(NULL)

african\_crises$independence<-NULL

#removed the inflation\_annual\_cpi variable because it was causing a perfect prediction warning

african\_crises$inflation\_annual\_cpi<-NULL

#altering the banking\_crisis variable to a binary variable with values 0/1, 1 for crisis, 0 for no crisis

african\_crises$banking\_crisis <- as.factor(ifelse(african\_crises$banking\_crisis=='crisis',1,0))

#The currency crises variable is a binary variable with values 0/1, there were some observations with 2 as values, I removed these

african\_crises<-african\_crises[!(african\_crises$currency\_crises==2),]

#checking for outliers in the data using boxplots

boxplot(african\_crises)

#removing the extreme values found in the exch\_usd variable

african\_crises<-african\_crises[!(african\_crises$exch\_usd > 700),]

#univariate plots

ggplot(african\_crises, aes(y=inflation\_crises)) + geom\_bar()

ggplot(african\_crises, aes(x= exch\_usd)) + geom\_boxplot()

ggplot(african\_crises, aes(x= sovereign\_external\_debt\_default)) + geom\_bar()

ggplot(african\_crises, aes(x= currency\_crises)) + geom\_bar()

ggplot(african\_crises, aes(x= banking\_crisis)) + geom\_bar()

#bivariate plots

ggplot(data = african\_crises, aes(x = exch\_usd, y = inflation\_crises)) +

geom\_boxplot (stat = "boxplot") + geom\_jitter(width = 0.2)

ggplot(african\_crises) +

geom\_bar(aes(x = sovereign\_external\_debt\_default, fill = inflation\_crises), position = "dodge")

ggplot(african\_crises) +

geom\_bar(aes(x = currency\_crises, fill = inflation\_crises), position = "dodge")

ggplot(african\_crises) +

geom\_bar(aes(x = banking\_crisis, fill = inflation\_crises), position = "dodge")

#summary statistics

describe(african\_crises)

#reviewing data types

sapply(african\_crises, class)

#transforming categorical variables to factors

african\_crises$inflation\_crises <- as.factor(african\_crises$inflation\_crises)

african\_crises$systemic\_crisis <- as.factor(african\_crises$systemic\_crisis)

african\_crises$domestic\_debt\_in\_default <- as.factor(african\_crises$domestic\_debt\_in\_default)

african\_crises$sovereign\_external\_debt\_default <- as.factor(african\_crises$sovereign\_external\_debt\_default)

african\_crises$currency\_crises <- as.factor(african\_crises$currency\_crises)

#splitting the data into a training/testing set

split = sample.split(african\_crises$inflation\_crises, SplitRatio = 0.8)

train\_set = subset(african\_crises, split == TRUE)

test\_set = subset(african\_crises, split == FALSE)

view(train\_set)

Model <- glm(inflation\_crises ~.,data = train\_set, family = binomial(link = "logit"))

print(summary(Model))

Model2 <- glm(inflation\_crises ~ exch\_usd + sovereign\_external\_debt\_default + currency\_crises

+ banking\_crisis,data = train\_set, family = binomial())

print(summary(Model2))

#testing the multi-collinearity assumption

VIF(Model2)

#testing the fourth assumption of the logistic regression

#the numeric variable is exch\_usd

probs <- predict(Model2, type = "response")

logit = log(probs/(1-probs))

ggplot(train\_set, aes(logit, exch\_usd))+

geom\_point(size = 0.5, alpha = 0.5) +

geom\_smooth(method = "loess") +

theme\_bw()

#accuracy scores for both models

fitted.results <- predict(Model, newdata = test\_set, type = 'response')

fitted.results <- ifelse(fitted.results > 0.5,1,0)

error <- mean(fitted.results != test\_set$inflation\_crises)

print(paste('Logistic Regression Accuracy score for Model', 1-error))

fitted.resultslg2 <- predict(Model2, newdata = test\_set, type = 'response')

fitted.resultslg2 <- ifelse(fitted.resultslg2 > 0.5,1,0)

error <- mean(fitted.resultslg2 != test\_set$inflation\_crises)

print(paste('Logistic Regression Accuracy score for Model2', 1-error))

#Confusing matrix

print("Confusion matrix\_Model"); table(Predicted = fitted.results, Actual = test\_set$inflation\_crises)

print("Confusion matrix\_Model2"); table(Predicted = fitted.resultslg2, Actual = test\_set$inflation\_crises)

#provides the coefficients of the variables in the final model

coef(Model2)

#final data export, this resulted in an extra column counting the rows, I then did some excel formatting

#to submit the final data used

write.table(african\_crises, file= "african\_crises\_final.csv", sep = ",")