**# Import BankMArketing Data**

bank <- read.csv("E:/Data Analytics with RET/Assignment/Bank/bank-additional.csv", sep=";")

View(bank)

dim(bank) # 4119 observations with 21 attributes

str(bank) # All have the correct class

**# a. Create a visual for representing missing values in the dataset.**

psych::describe(bank)

library(VIM)

missing <- bank

missing[missing == "unknown"] <- NA

aggr(missing, col=c('blue', 'red'),

numbers=TRUE, sortvars= TRUE,

labels=names(missing), cex.axis=0.5,

gap=3, ylab=c("missing data","pattern"))

sapply(missing, function(x) sum(is.na(x)))

**# b. Show a distribution of clients based on a Job.**

t <- table(bank$job)

# distribution in tabular form

t

**# distribution in graphical form**

title <- barplot(t, xlab = "Job", ylab = "Numbers", main = "Clients based on Job",

col = heat.colors(12), las=3)

text(title, 0, t, pos = 3, srt = 90)

**# c. Check whether is there any relation between Job and Marital Status?**

**# Ho : There is NO association between Job and Marital Status**

chisq.test(missing$job, missing$marital)

**# Since P Value is less than 0.05 ,**

**# there is association between Job and Marital status at 95% confidence level**

**# Since NA values are very less, are omitted**

**# d. Check whether is there any association between Job and Education?**

**# Ho : There is NO association between Job and Education.**

chisq.test(missing$job, missing$education)

**# Since the P value is less than 0.05,**

**# there is association between Job and Education at 95% confidence level**

**# Since NA values are very less, are omitted**