# Assignment (11.1) 13- Jan 2018

1. Use the given link and locate the bank marketing dataset. Data Set Link

Perform the below operations:

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

Sol :- unzip("C:/Users/acer/Desktop/bank-additional/bank (1).zip")

> bank<-read.csv("C:/Users/acer/Desktop/bank-additional/bank.csv")

There are no missing values in dataset. This is being tested using summary() command and also by complete.cases () method.

length(complete.cases(bank))

[1] 4521

> dim(bank)

[1] 4521 1

complete.cases() returns a logical vector indicating absence or presence of missing values. “True” if no missing value. “False” if there is a missing value.

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

Sol :- library(ggplot2)

ggplot(data=f1,aes(x=job)) + geom\_bar(color="black",fill="white")

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

Sol: - Both variables Job & Marital status are categorical. Whenever we want to test the relation or association between two such variables we use CHI-SQUARE Test.

> chisq.test(bank$job, bank$marital, correct = FALSE)

**Pearson’s Chi – squared test**

data: bank$job and bank$marital

x-squared = 373.18, df = 22, p-value < 2.2e-16

Since p- value is smaller than 0.05, we reject the null hypothesis, it means that they are not related. Thus the two variables have association.

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

Sol :- Both variables Job & Marital status are categorical. Whenever we want to test the relation or association between two such variables we use CHI-SQUARE Test.

>chisq.test(bank$job, bank$education, correct = FALSE)

**Pearson’s Chi – squared test**

data= bank$job, and bank$education

x- squared = 2840, df = 33, p-value < 202e -16

Since p value is smaller than 0.05, we reject the null hypothesis, it means that they are not related. Thus the two variables have association.