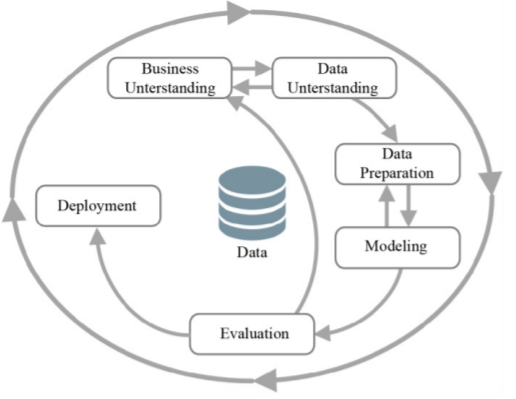
# Introduction

A common technique for growing an organization is to run marketing selling campaigns. Direct marketing is used by businesses to reach certain categories of customers to achieve a specific goal. Customer distant interactions may be integrated into a support department, making campaign administration easier. Customers can communicate with these centers via a variety of means, the most common of which is the cellphone. (Moro, et al., 2014) The data analyzed in this paper is connected to a banking institution's direct marketing initiatives. Customers were encouraged to sign up for a term deposit through a telephone marketing campaign. To determine the prediction old dataset was used if the product (bank term deposit) would be subscribed ('yes') or not ('no'). There were 41188 observations with 22 variables out of which 10 are numerical variables. 4640 customers had subscribed for bank term deposits which are around 8.9% customers of the whole data set. We must first import and pre-process data before we can start building the model.

Before being analyzed for outliers and missing values, the dataset is first to read into a variable. The data is cleaned and separated into two files, test and train, before being used to build a regression model. With the use of a summary table and assumptions, insights are provided. Caret, corrplot, lmtest, psych, ggplot2, dplyr, and car are some of the R programs utilized. Model 1, Model 2, and Model 3 are three logistic regression models using 3, 6, and 9 variables, respectively. The dataset is explored using a logistic regression model with ten variables in total. The dependent variable is 'y,' which represents whether a term deposit is subscribed or not. The variables that are utilized as independent variables include job, default, contact, month, duration, poutcome, emp.var.rate, euribor3m and nr.employed.

# Methodology

The CRISP-DM (Cross-Industry Standard Process for Data Mining) is a widely used approach for improving the success of data mining operations. The technique specifies a non-linear series of six steps that enable the creation and deployment of a DM model in a real-world setting, assisting with business choices. (Chapman, et al., 2000) CRISP-DM would be considered to understand the step by step process which was utilized in this paper to analyse the banking data set.

  
*Fig. 1 Shows the CRISP-DM process*

## Business Understanding: It is important to note that banks are under enormous pressure to raise their financial assets related to internal competitiveness. To address this problem, one technique used is to promote appealing long-term deposit applications with high-interest rates, particularly through targeted marketing initiatives. The same factors are also pushing for cost and time savings. As a result, efficiency must be improved: fewer contacts should be made, but an approximate number of successes (customers subscribing to the deposit) should be retained. (Moro & Laureano, 2011)

## Data Understanding: Once test and train banking data is downloaded into the system. It is then loaded into R-Studio. Attributes, properties, size, and structure are analysed to get better insights. The 2 excel file – test and train banking data sets were merged

## For better insights, data quality issues and observations were revealed by a more thorough investigation. colSums(is.na()) and summary() function was used to analyze whether there are any NA values and basic summary of the data set. Looking for new frequency patterns in data would be a relevant approach in this case of a predictive deposit scenario with the goal of detecting the number of subscribers and analyzing the important attributes.

## Data Preparation: The engineer gathers necessary data and prepares it for the real task during the "Data Preparation" phase. This comprises data reduction and filtering, as well as preparation. (Hubera, et al., 2019) Summarizing the age of customers it was observed that the lowest age was 3 and the highest age was 170, it seemed to be uncertain values. So, the age below 17 and above 98 were replaced by the mean values. ‘Education’ attribute ‘basic.4y’, ‘basic.6y’ and ‘basic.9y’ were replaced with ‘basic’ to get proper understanding. For month attribute re-leveling of months was done to get proper insights in ggplot visualization. The y variable with ‘yes’ or ‘no’ was replaced with ‘Subscribed’ and ‘Not Subscribed’. A new table was created using the filter() function which consisted of the values where term deposit (y) was only ‘yes’. It was also made sure that all the character variables were converted into factors. The chisq.test() function (*Pearson's Chi-squared test)*was used to test hypotheses about the relationship between categorical variables. Pearson’s correlation was also used to understand the relationship between 2 numerical variables. Once, the data quality issue and sorting were done the data was split into 80-20 partition which was named as test and train data.

## Modelling The "modeling" step is designed to determine the required parameter values for the chosen algorithms and to run the data analytics task on the preprocessed data. Using the test and train dataset 3 models are generated with assumptions and checks MODEL 1 consisting of ‘default’, ‘contact’ and ‘poutcome’. Model 2 consists of ‘default’, ‘contact’, ‘poutcome’, ‘month’, ‘duration’ and ‘emp.var.rate’ and Model 3 consist of ‘job’, ‘euribor3m’ and ‘nr.employed’ extra variable to model 2. NOTE: ‘duration’ variable is taken for benchmark purpose, to increase the accuracy. If this model was about to be implemented in the industry then it won’t be a fair practice to take duration attribute. When the target (dependent) variable is categorical and has two categories, logistic regression should be applied. To do a logistic regression in R, should use glm() function. The postResample() method is used to determine accuracy and Kappa values. The models that will be used will be examined to assess their correctness, and the confusion matrix will be used to determine this accuracy. The residuals are calculated using the resid() function. The VIF() function is used to determine whether there is a problem with multicollinearity.

## Evaluation The trained model is tested against real data sets in a production situation during the "Evaluation" step, and the outcomes are evaluated against the underlying business objectives. Test data sets are created for this purpose by following the processes outlined in the "Data Preparation" and "Modeling" stages. (Hubera, et al., 2019) The model with high accuracy would be taken into consideration. *The coefficients are interpreted in the same way as linear regression coefficients are. The coefficient indicates the change in the logit of the outcome variable caused by a one-unit change in the predictor variable. A pseudo R square can be used to evaluate a logistic regression model (it has a similar interpretation to the R squared in R).*The residuals are obtained using the resid() function. The residuals are useful for determining how well the model matches the data. Residuals above 1.96 were calculated. As a rule of thumb, only 5% should lie outside of ± 1.96.

## Deployment

## Creation of the Logistic regression model is not the endpoint of the project. Typically, the knowledge gathered must be arranged and presented in such a way that the consumer can make use of it. The deployment step might be as easy as creating a report or as sophisticated as establishing a repeatable data mining process, depending on the needs. In many circumstances, the user, not the data analyst, will do the deployment processes. In any instance, it is critical to understand what steps must be taken ahead of time to use the models that have been built.

# Results (Descriptive statistics, visualisation, and measures of association)

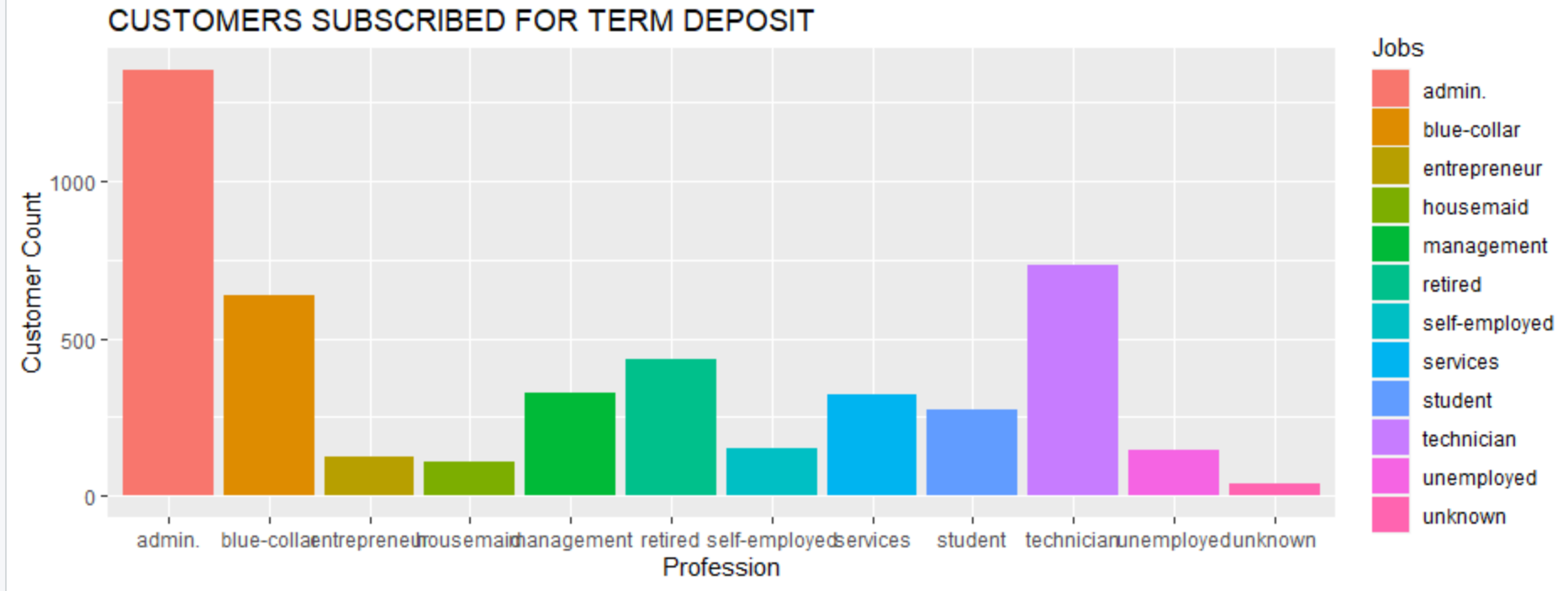
## **Descriptive Statistics**

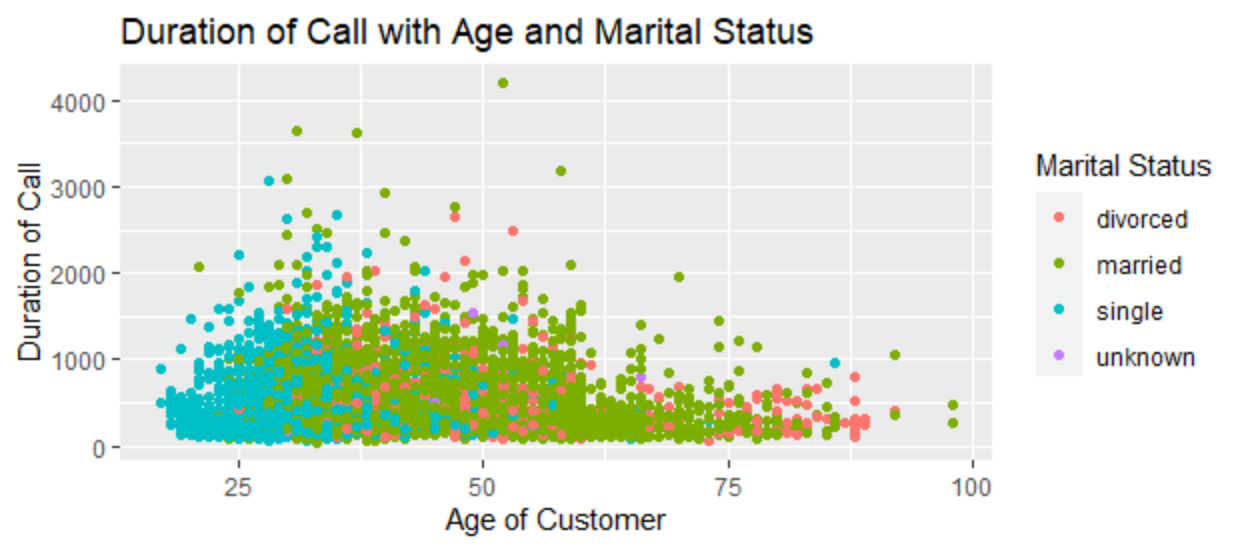
  
*Table 1. Shows descriptive statistics of jobs of customers with respect to age.*

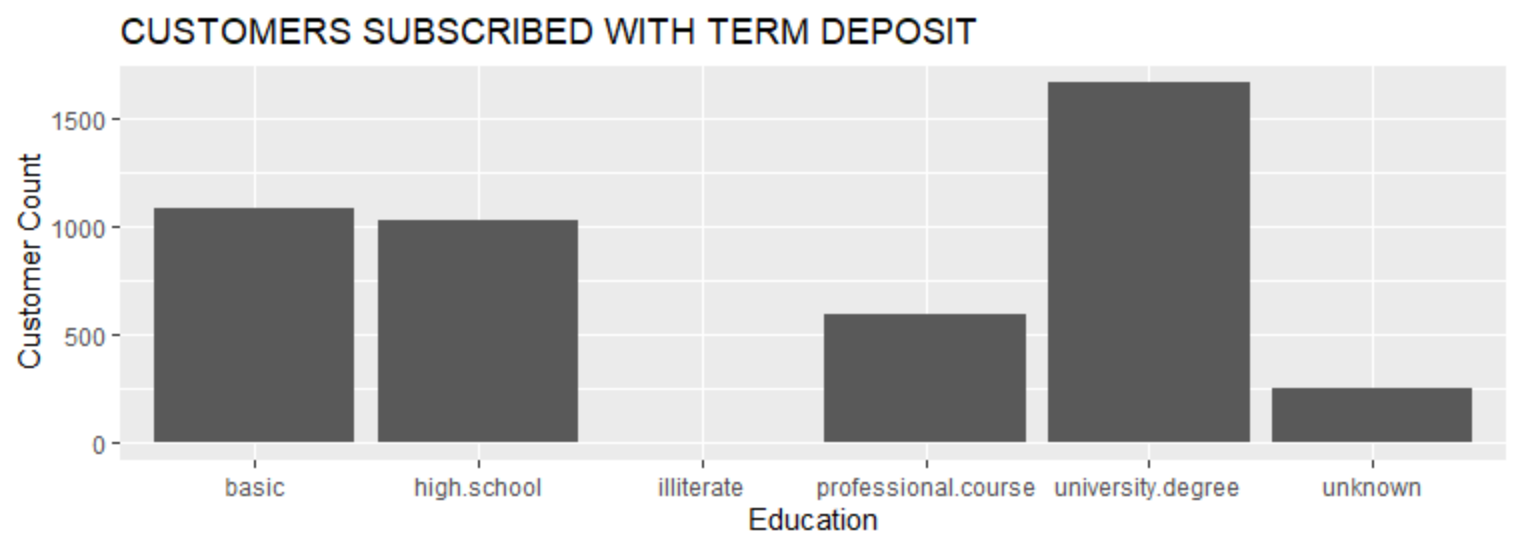
 *Table 2. Shows Customer count with loan*

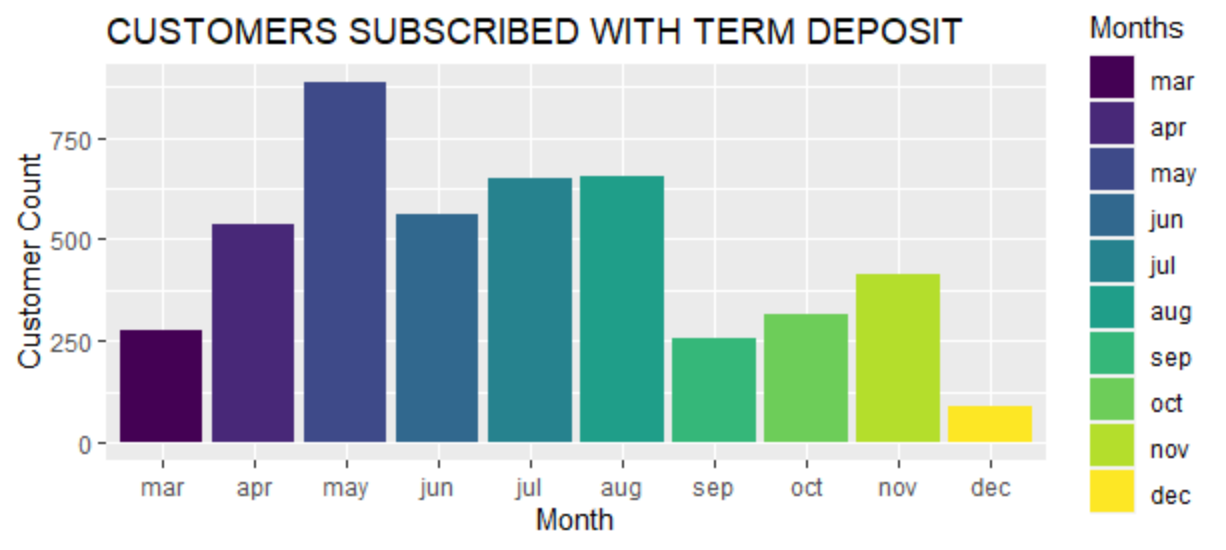
 *Table 3. Shows Marital Status*

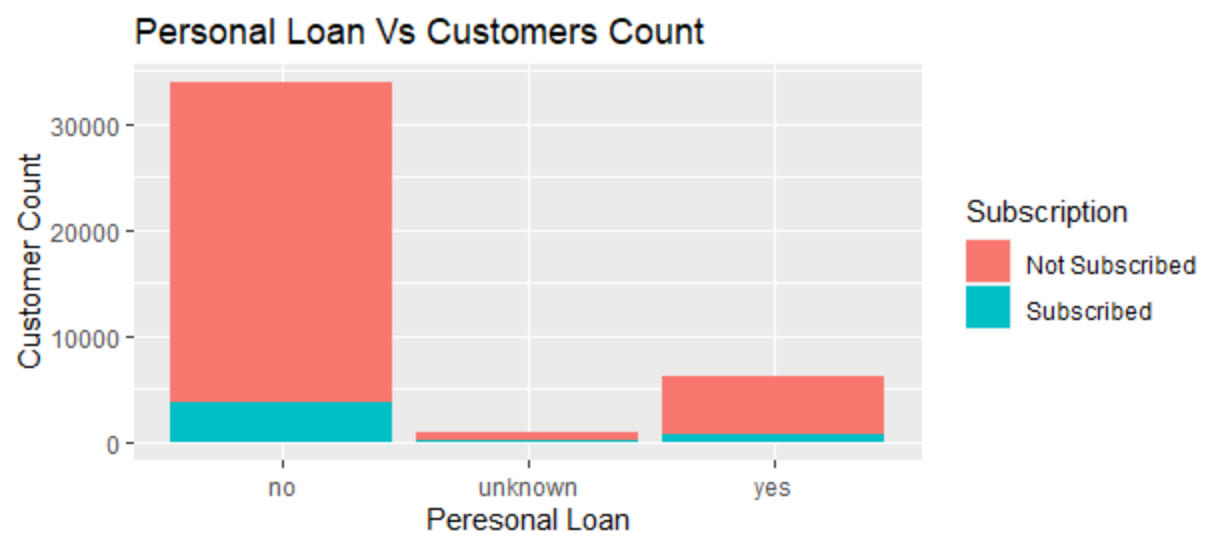
## **Data Visualisations**

*  
Fig 2. Shows Customers subscribed for term deposit with respect to profession (BAR)*

 *Fig 3. Shows subscribed customer with term deposit with marital status and age (POINT)*

 *Fig 4. Shows subscribed customer with term deposit and education background (BAR)*

 *Fig 5. Shows subscribed customer with term deposit and Months (BAR)*

 *Fig 6. Shows Customers who have taken personal loan with respect to subscription*

## **Measures of Association**

  
*Table 4. Shows p-value and tabular comparison of Month and subscription*

 *Table 5. Shows p-value and tabular comparison of Outcome of the previous marketing and subscription*

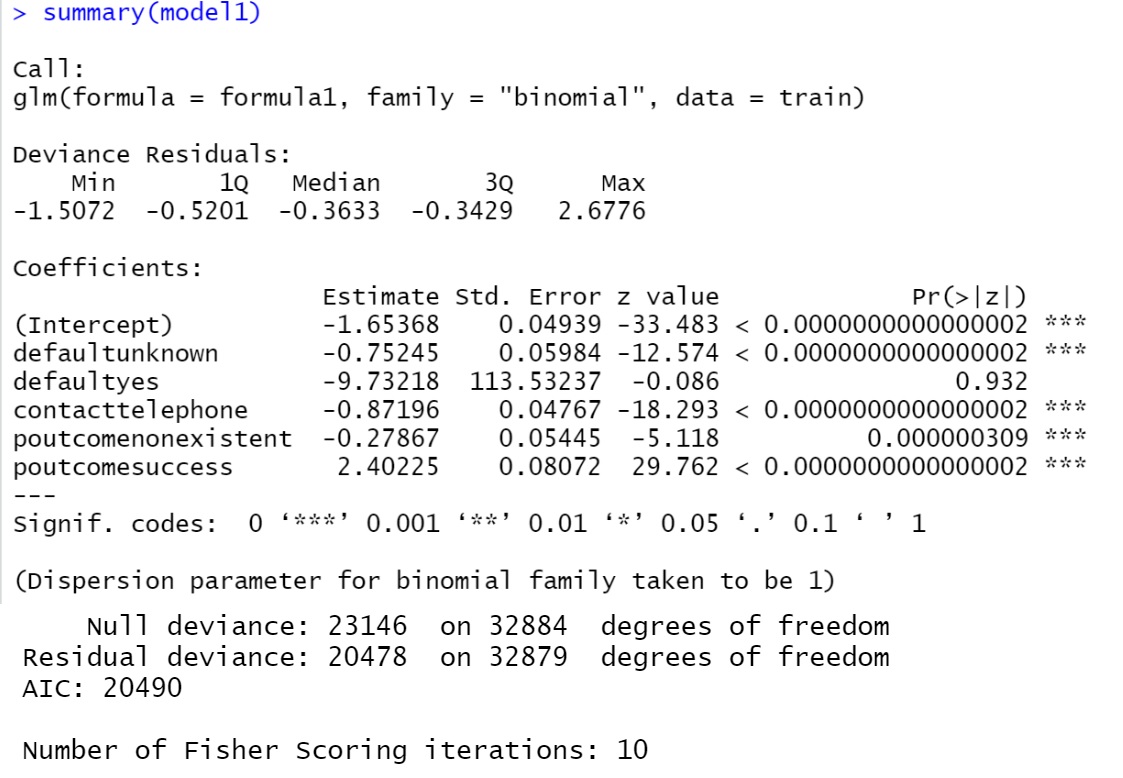
 *Table 6. Shows p-value and tabular comparison of Marital Status and subscription*

 *Table 7. Shows p-value and tabular comparison of contact and subscription*

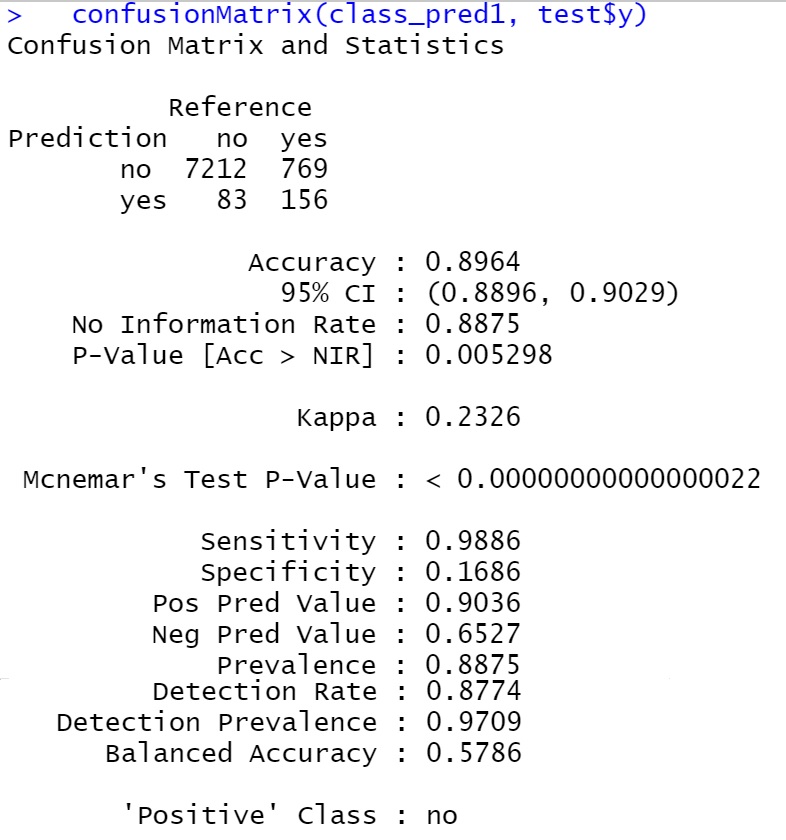
 *Table 8. Shows correlation between Consumer price index and Consumer confidence index*

# Regression Analysis Results

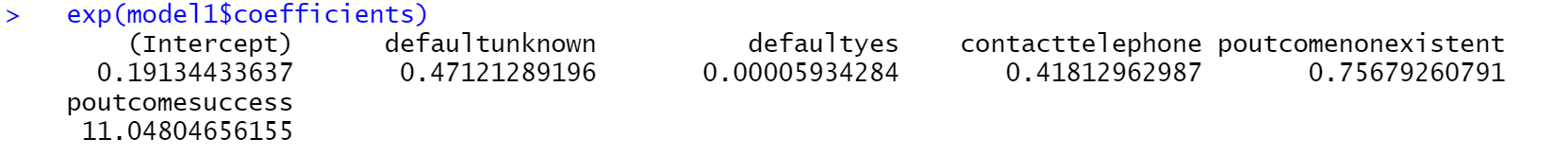
1. **Logistics Regression Model 1**

*  
Fig.7 Shows Summary of Model 1*

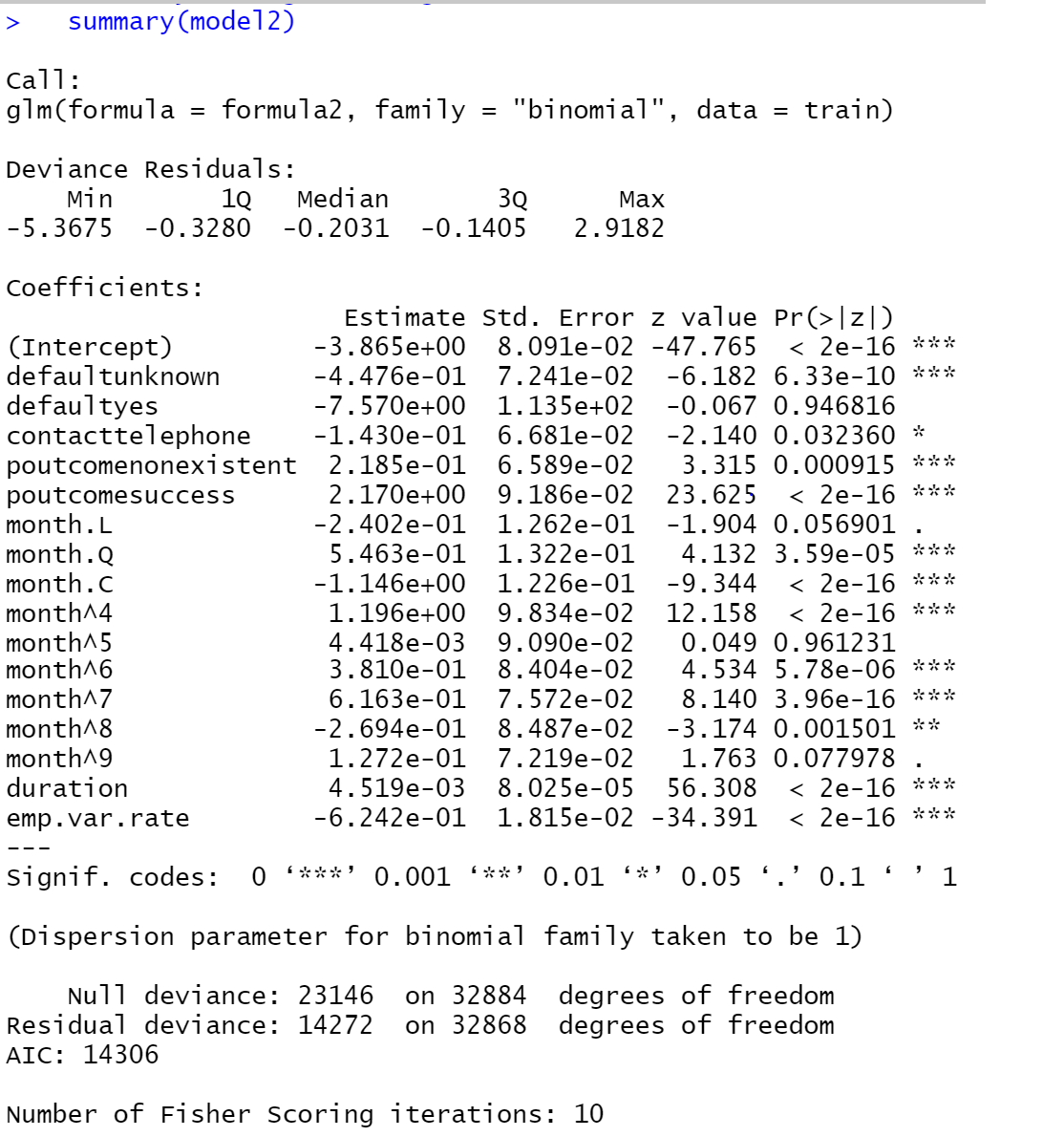
 *Table 9. Shows Accuracy and Kappa value for Model 1*

**

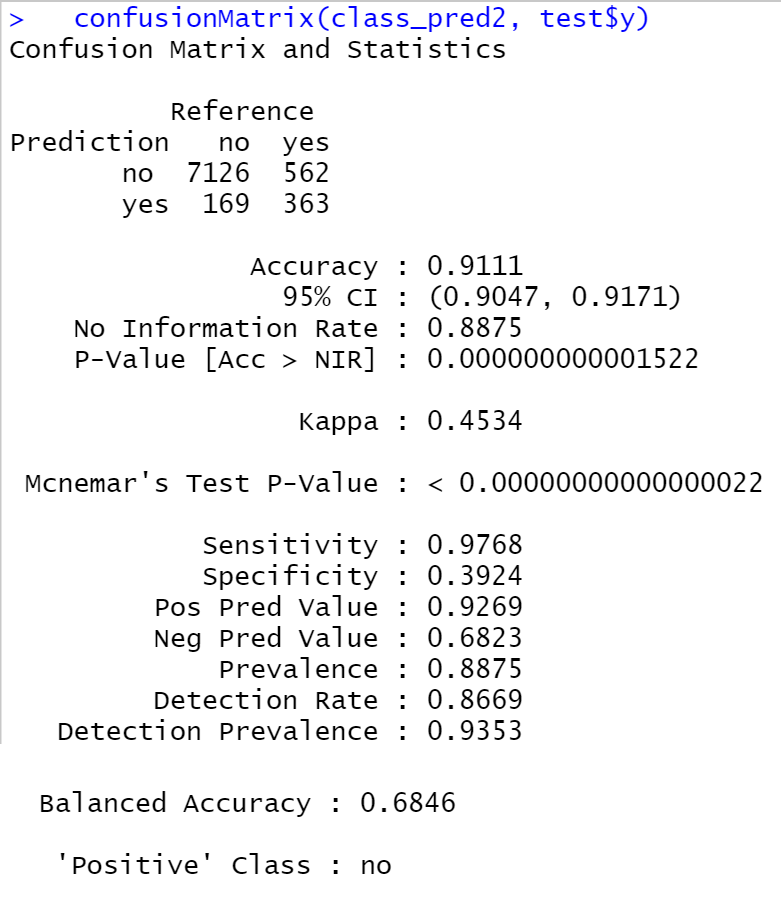
*Fig. 8 Shows Confusion Matrix for Model 1*

 *Fig. 9 Shows Odds Ratios for Model 1*

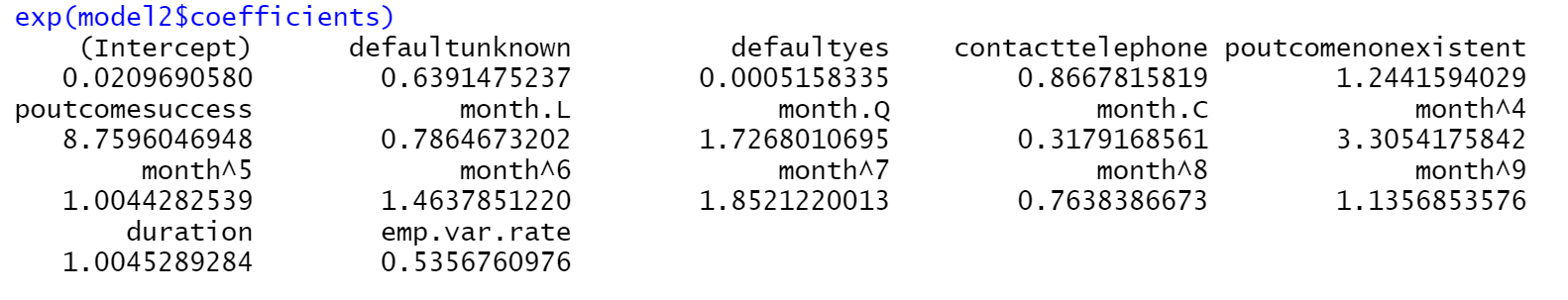
1. ***Logistics Regression for Model 2***

*  
Fig. 10 Shows Summary of Model 2*

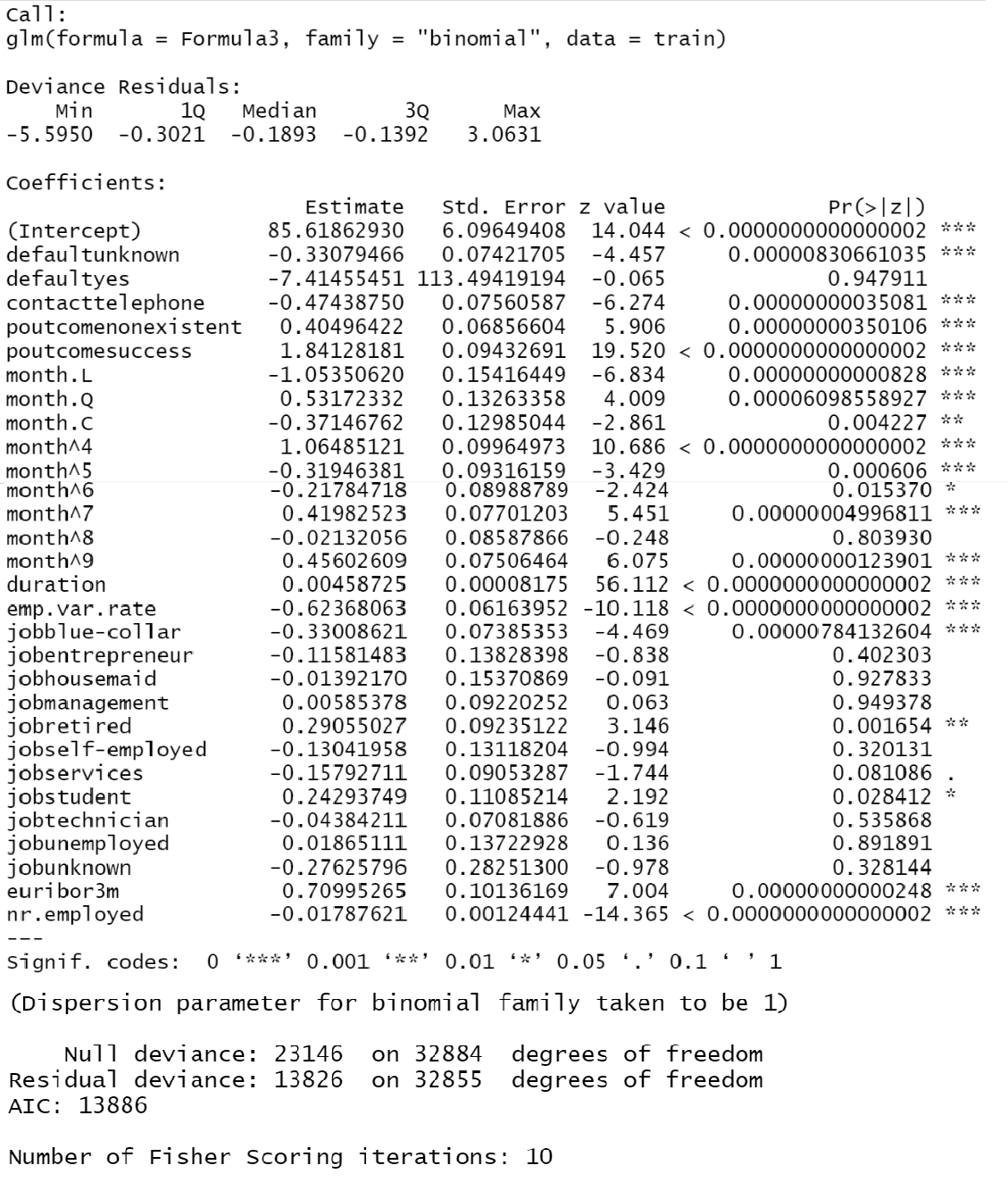
 *Table 10. Shows Accuracy and Kappa value for Model 2*

**

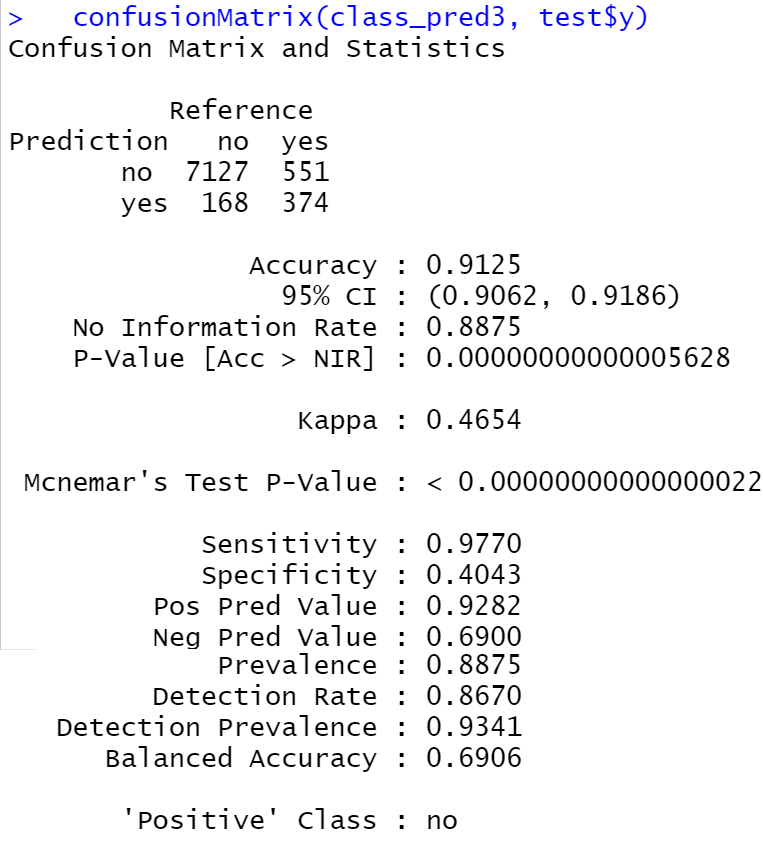
*Fig. 11 Shows Confusion Matrix for Model 1*

 *Fig. 12 Shows Odds Ratios for Model 2*

1. ***Logistics Regression for Model 3***

*  
Fig. 13 Shows Summary of Model 3*

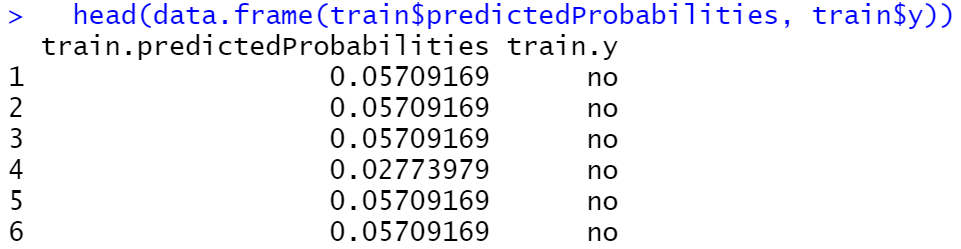
 *Table 11. Shows Accuracy and Kappa value for Model 3*

**

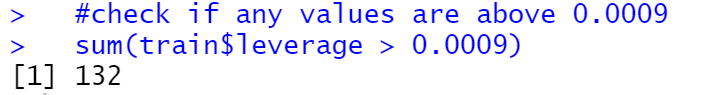
*Fig. 14 Shows Confusion Matrix for Model 1*

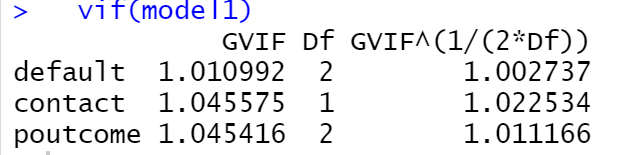
## **Prediction and Accuracy**

1. **Prediction and Accuracy for Model 1**

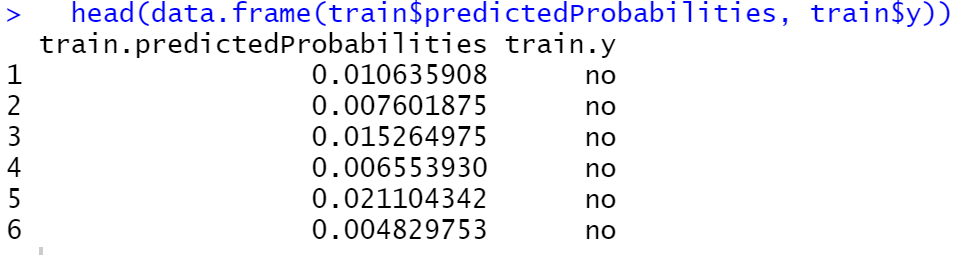
 *Fig. 15 shows probability of y, and the actual outcome*

 *Fig. 15 Shows Residual above 1.96*

  
Fig. 16 Shows Leverage for Model 1

 *Fig. 17 Shows VIF for model 1*

1. **Prediction and Accuracy for Model 2**

 *Fig. 18 shows probability of y, and the actual outcome for model 2*

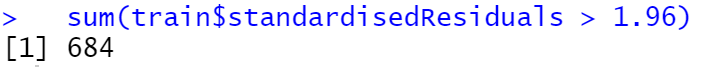
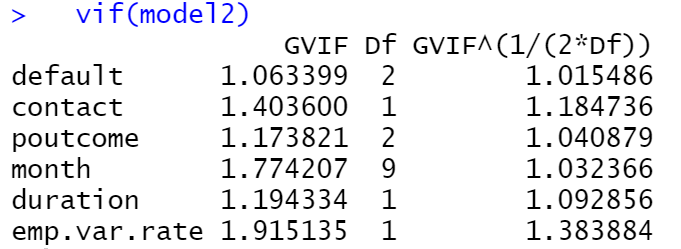
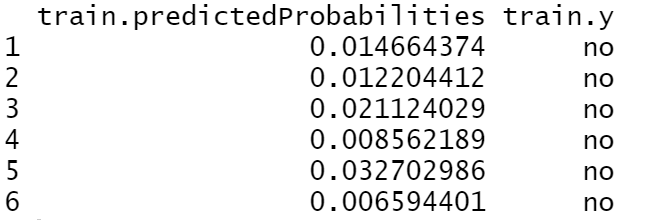
 *Fig. 19 Shows Residual above 1.96 for model 2*

  
Fig. 20 Shows Leverage for Model 2

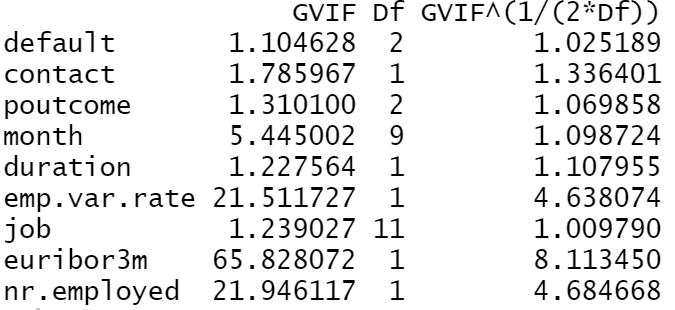
 *Fig. 21 Shows VIF for model 2*

1. **Prediction and Accuracyfor Model 3**

 *Fig. 22 Shows probability of y, and the actual outcome for model 3*

 *Fig. 23 Shows Residual above 1.96 for model 3*

  
Fig. 24 Shows Leverage for Model 3

 *Fig. 25 Shows VIF for model 2*

# Discussion

It can be observed from Fig 1. that customer who has a job as Admin are the most subscribed customers for a term deposit. But only 1352 admins have subscribed out of 10422 which is around 13%. This is because a maximum of the customers have a job profile as admin. Observing Fig 2. it can be noted that maximum customers having a duration of call had marital status as married. This is due to maximum customers are married. The highest subscription rate is 14% with singles whereas for divorcees and married it's around 10%. Fig 4. depicts that the education background of customers having illiterate did not subscribe to the term deposit at all Customer with University degree as their education background were the maximum subscriber for the term deposit.

Observing Fig. 5, it can be analyzed that the maximum customers who subscribed to the term deposits are in May. It is the least subscription rate with 6.43%. The least customers subscribed month is December. We can also observe that there is around a 50% subscription rate in March and December. The bank office operations might not be operative in December due to the Christmas holidays with a 100% workforce, this might be one of the reason customer might not get support from the bank to subscribe to term deposits. Fig. 6, depicts that a customer who has not taken any personal loan has more subscribers than the one who has taken a personal loan. It can also be noted that customers with the personal loan are in minority.

Observing Table 4,5,6 and 7. It can be observed that month, previous outcome, marital status, and contact are statistically significant concerning variable y. As the p-value for all of them is 2.2e-16 which is lesser than 0.05. The lower the p-value, the greater the statistical significance of the observed difference.

While considering all the logistic regression models referring to figures 7, 10, and 13 it can be observed that the null deviance is greater than residual deviance. The null deviance is the deviance of the model with no predictors, while the residual deviance is the deviance of the model with the predictor. As a result, the null deviance should be greater than the residual deviation. It can also be observed from Table 16. 17. and 18. that the value of R-square is increasing as there is an increase in the attributes of a logistics regression model. From fig 12. it can be observed that the confidence interval is above 1 for 'poutcome', 'month', and 'duration'. If the confidence interval exceeds one, we cannot be certain of the relationship's direction (and the b will probably not be statistically significant).

  
*Table 12. Shows comparison of the model*

Analysing Table 12. it can be observed for all the 3 models that accuracy and kappa value increases as the number of attributes increases. For model 1 the kappa value was 0.23 from model 2 onwards the kappa value is above 0.3. The larger the kappa value the better is the model. Residual above 1.96 was analyzed and it is noticed that residual count was decreasing as the model accuracy was increasing and the leverage value above 0.0009 increased. GVIF variable was very high for 3 variables.

# Conclusion

For the next marketing campaign, it would be better for the bank to focus on March and December instead of May. As the subscription rate is seen to be 50% around these 2 months. More Singles and youngsters in their 20s must be targeted as the probability of them subscribing to the term deposit is higher. Customers with no prior history of credit default have a greater probability of subscribing to the bank's term deposit. Customer with no default history must only be targeted for a better probability rate. To know the bank customers it needs to perform analysis timely basis so that the banking products/services offered to meet their demands and the sale is assured. As a result, the bank should be as involved in their clients' operations as possible, providing financial and logistical support, specialized consultation, and help. The bank must aim for a long-term competitive advantage through promoting good interest rates for term deposits, and the growth of customer loyalty. (Catalina, 2010)

# Appendix 1: R Code Used

#Install the required packages

#Read the Packages

library(readxl)

library(psych)

library(ggplot2)

library(tidyverse)

library(dplyr)

library(caret) #to split the data

library(Hmisc) #For rcorr() function

library(corrplot)

library(lmtest)

library(car)

#Set Workind Directory

setwd('D:/Business Analytics/Statistics For Business/Assignment 2')

#Read the excel sheet into variable test and train

test <- read\_excel('bank\_test.xlsx')

train <- read\_excel('bank\_train.xlsx')

#Combine the train and test data into BANK

bank <- rbind(train,test)

#To remove 10E values

options(scipen = 10000)

# :: Summarize the Data ::

#Analyze the Columns with NA

colSums(is.na(bank))

#Summarise the data

summary(bank)

#::::: Data Quality Issues and Action ::::::

#AGE: Maximum Age is 170 and minimmum is 3. Change the age at appropriate value.

#Re summarize Age above 98 and below 17 as mean value

bank$age[(bank$age > 98)] <- mean(bank$age)

bank$age[(bank$age < 17)] <- mean(bank$age)

#Convert job into factor

bank$job <- as.factor(bank$job)

#Convert marital into factor

bank$marital <- as.factor(bank$marital)

#Convert education into factor and Combine basic education

bank$education[bank$education == 'basic.4y'] <- 'basic'

bank$education[bank$education == 'basic.6y'] <- 'basic'

bank$education[bank$education == 'basic.9y'] <- 'basic'

bank$education <- as.factor(bank$education)

#In Contact change the name of 1 Mobile to cellular and contact as factor

bank$contact[bank$contact == 'mobile'] <- 'cellular'

bank$contact <- as.factor(bank$contact)

#Convert housing into factor

bank$housing <- as.factor(bank$housing)

#Convert loan into factor and change the name pf NA's to unknown

bank$loan <- as.factor(bank$loan)

bank$loan[is.na(bank$loan)] <- 'unknown'

#Convert Month in factor and level-up in sequence to show proper interpretation

bank$month <- as.factor(bank$month)

bank$month <- factor(bank$month, levels = c("jan", "feb", "mar", "apr", "may", "jun", "jul", "aug", "sep", "oct", "nov", "dec"), ordered = TRUE)

#Convert day\_of\_week in factor and level-up in sequence to show proper interpretation (it consists 83 NA's)

bank$day\_of\_week <- as.factor(bank$day\_of\_week)

bank$day\_of\_week <- factor(bank$day\_of\_week, levels = c("mon", "tue", "wed", "thu", "fri"), ordered = TRUE)

#Convert y variable into factor as 'yes' or 'no'

#bank$y[bank$y == 'yes'] <- 'Subscribed'

#bank$y[bank$y == 'no'] <- 'Not Subscribed'

bank$y <- as.factor(bank$y)

#bank with y column as yes

bank %>% filter(y == 'Subscribed') -> yes

#::: Descriptive Statistics :::

#1. descriptive statistics of jobs of customers with respect to age.

describeBy(x = bank$age, group = bank$job)

#2. Credit Default with respect to Marketing Campaign duration

describeBy(x = bank$duration, group = bank$default, na.rm = TRUE)

#3. Subscribed Customers with respect to Marketing Campaign duration

describeBy(x = bank$duration, group = bank$y, na.rm = TRUE)

#4. Customers who have taken Housing and Personal Loan

summary(bank$loan)

summary(bank$housing)

#5. Customer Count of Marital status

summary(bank$marital)

#6. Eduaction status count of customer

summary(bank$education)

#:: GGPLOT VISUALISATIONS::

#1. Customers subscribed for term deposit with respect to profession (BAR)

yes %>% ggplot(aes(x=job,,fill = job))+

geom\_bar()+

labs(title = "CUSTOMERS SUBSCRIBED FOR TERM DEPOSIT", x="Profession", ,

y= "Customer Count", fill = 'Jobs')+

scale\_y\_continuous(labels = function(x) format(x, scientific = FALSE))

#2. subscribed customer with term deposit with marital status and age

yes %>% ggplot(aes(x = age, y=duration, color = marital),stat = "Summary", fun.y = "mean")+

geom\_point()+

labs(title = "Duration of Call with Age and Marital Status", x="Age of Customer", y= "Duration of Call", color = "Marital Status")+

scale\_y\_continuous(labels = function(x) format(x, scientific = FALSE))

scale\_fill\_brewer(palette = "Dark2")

#3. subscribed customer with term deposit and education background

yes %>% ggplot(aes(x=education))+

geom\_bar()+

labs(title = "CUSTOMERS SUBSCRIBED WITH TERM DEPOSIT", x="Education",

y= "Customer Count")+

scale\_y\_continuous(labels = function(x) format(x, scientific = FALSE))

#4.subscribed customer with term deposit and Months

yes %>% ggplot(aes(x=month, fill = month))+

geom\_bar()+

labs(title = "CUSTOMERS SUBSCRIBED WITH TERM DEPOSIT", x="Month",

y= "Customer Count", fill = 'Months')+

scale\_y\_continuous(labels = function(x) format(x, scientific = FALSE))

#5. Customers who have taken personal loan with respect to subscription

bank %>% ggplot(aes(x = loan, fill = y))+

geom\_bar()+

labs(title = "Personal Loan Vs Customers Count", x="Peresonal Loan", y= "Customer Count", fill = "Subscription")+

scale\_y\_continuous(labels = function(x) format(x, scientific = FALSE))

scale\_fil\_brewer(palette = "Dark2")

#::: MEASURES OF ASSOSCIATION::::

#Chi Square Tests in R

#1.cross tabs for the variables MONTH

table(bank$y,bank$month)

#chisq.test() function to perform the test

chisq.test(bank$y, bank$month, correct = FALSE)

#2.cross tabs for the variables Outcome of previous Marketing

table(bank$y,bank$poutcome)

#chisq.test() function to perform the test

chisq.test(bank$y, bank$poutcome, correct = FALSE)

#3.cross tabs for the variables MONTH

table(bank$y,bank$marital)

#chisq.test() function to perform the test

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

#4. cross tabs for the variables MONTH

table(bank$y,bank$contact)

#chisq.test() function to perform the test

chisq.test(bank$y, bank$contact, correct = FALSE)

#5. Relationship between Consumer price index and Consumer confidence index

cor.test(x=bank$cons.conf.idx, y=bank$cons.price.idx)

#::::: SPLIT THE BANK DATA INTO TRAINING AND TEST::::

#Delete TEST and TRAIN data first, it would set it as empty

test <- NULL

train <- NULL

#to create a partition with 80%

bank <- bank %>% filter(!is.na(day\_of\_week))

bank<- bank %>% mutate\_if(is.character, as.factor)

set.seed(123) #generate a sequence of random numbers

index <- createDataPartition(bank$y, p = 0.8, list = FALSE,)

train <- bank[index, ] #first 80% for training

test <- bank[-index, ] #bottom 20% for testing

# ::: BUILD THE MODEL :::

#1. :::: Logistic Regression MODEL 1 ::::

formula1 <- y ~ default + contact + poutcome

model1 <- glm(formula1, data = train, family = "binomial")

#Summary of Logistic Regression MODEL 1

summary(model1)

#prediction using the model

predictions1 <- predict(model1,test,type ="response")

#Convert probabilities to yes or no

class\_pred1 <-as.factor(ifelse(predictions1 > 0.5,"yes","no"))

#evaluate the accuracy of the predictions

postResample(class\_pred1,test$y)

#Confusion Matrix

confusionMatrix(class\_pred1, test$y)

#2. :::: Logistic Regression MODEL 2 ::::

formula2 <- y ~ default + contact + poutcome + month + duration + emp.var.rate

model2 <- glm(formula2, data = train, family = "binomial")

#Summary of Logistic Regression MODEL 2

summary(model2)

#prediction using the model

predictions2 <- predict(model2,test,type ="response")

#Convert probabilities to yes or no

class\_pred2<-as.factor(ifelse(predictions2 > 0.5,"yes","no"))

#evaluate the accuracy of the predictions

postResample(class\_pred2,test$y)

#Confusion Matrix

confusionMatrix(class\_pred2, test$y)

#3. :::: Logistic Regression MODEL 3 ::::

Formula3 <- y ~ default + contact + poutcome + month + duration + emp.var.rate + job + euribor3m + nr.employed

model3 <- glm(Formula3, data = train, family = "binomial")

#Summary of Logistic Regression MODEL 3

summary(model3)

#prediction using the model

Predictions3 <- predict(model3,test,type ="response")

#Convert probabilities to yes or no

class\_pred3<-as.factor(ifelse(Predictions3 > 0.5,"yes","no"))

#evaluate the accuracy of the predictions

postResample(class\_pred3,test$y)

#Confusion Matrix

confusionMatrix(class\_pred3, test$y)

#Assessing Model R-Square

logisticPseudoR2s <- function(LogModel) {

dev <- LogModel$deviance

nullDev <- LogModel$null.deviance

modelN <- length(LogModel$fitted.values)

R.l <- 1 - dev / nullDev

R.cs <- 1- exp ( -(nullDev - dev) / modelN)

R.n <- R.cs / ( 1 - ( exp (-(nullDev / modelN))))

cat("Pseudo R^2 for logistic regression\n")

cat("Hosmer and Lemeshow R^2 ", round(R.l, 3), "\n")

cat("Cox and Snell R^2 ", round(R.cs, 3), "\n")

cat("Nagelkerke R^2 ", round(R.n, 3), "\n")

}

logisticPseudoR2s(model1)

#Odds Ratio (Exponential of coefficient)

exp(model1$coefficients)

exp(model2$coefficients)

exp(model3$coefficients)

#confidence interval

exp(confint(model1))

exp(confint(model2))

exp(confint(model3))

#:::::evaluate the model assumption::::

#::MODEL 1 ASSUMPTIONS::

#Add the predicted probabilities to the data frame

train$predictedProbabilities <- fitted(model1)

#This shows the probability of churn, and the actual outcome.

head(data.frame(train$predictedProbabilities, train$y))

#Add the standardised and Studentised residuals can be added to the data frame

train$standardisedResiduals <- rstandard(model1)

train$studentisedResiduals <- rstudent(model1)

#count the residuals above 1.96

sum(train$standardisedResiduals > 1.96)

#COOKs Distance

train$cook <- cooks.distance(model1)

sum(train$cook > 1)

train$leverage <- hatvalues(model1)

#check if any values are above 0.0009

sum(train$leverage > 0.0009)

#VIF to identify if there is a potential problem with multicolinearity

vif(model1)

#::MODEL 2 ASSUMPTIONS::

#Add the predicted probabilities to the data frame

train$predictedProbabilities <- fitted(model2)

#This shows the probability of churn, and the actual outcome.

head(data.frame(train$predictedProbabilities, train$y))

#Add the standardised and Studentised residuals can be added to the data frame

train$standardisedResiduals <- rstandard(model2)

train$studentisedResiduals <- rstudent(model2)

#count the residuals above 1.96

sum(train$standardisedResiduals > 1.96)

#COOKs Distance

train$cook <- cooks.distance(model2)

sum(train$cook > 1)

train$leverage <- hatvalues(model2)

#check if any values are above 0.0009

sum(train$leverage > 0.0009)

#VIF to identify if there is a potential problem with multicolinearity

vif(model2)

#::MODEL 3 ASSUMPTIONS::

#Add the predicted probabilities to the data frame

train$predictedProbabilities <- fitted(model3)

#This shows the probability of churn, and the actual outcome.

head(data.frame(train$predictedProbabilities, train$y))

#Add the standardised and Studentised residuals can be added to the data frame

train$standardisedResiduals <- rstandard(model3)

train$studentisedResiduals <- rstudent(model3)

#count the residuals above 1.96

sum(train$standardisedResiduals > 1.96)

#COOKs Distance

train$cook <- cooks.distance(model3)

sum(train$cook > 1)

train$leverage <- hatvalues(model3)

#check if any values are above 0.0009

sum(train$leverage > 0.0009)

#VIF to identify if there is a potential problem with multicolinearity

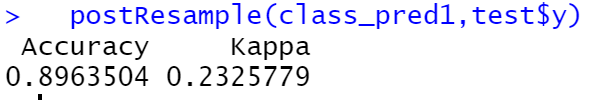
vif(model3)

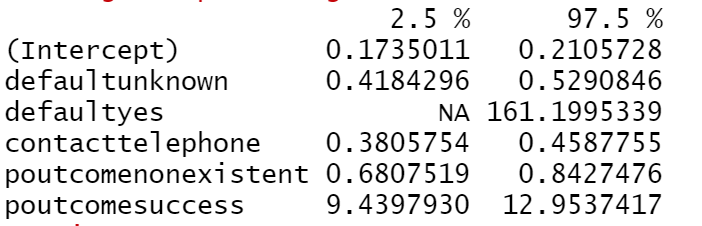
# Appendix 2: R/tables Screenshot

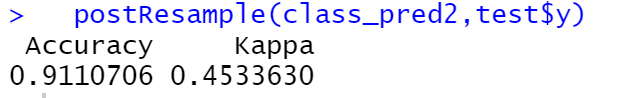
 *Table 13. Shows Credit Default with respect to Marketing Campaign duration*

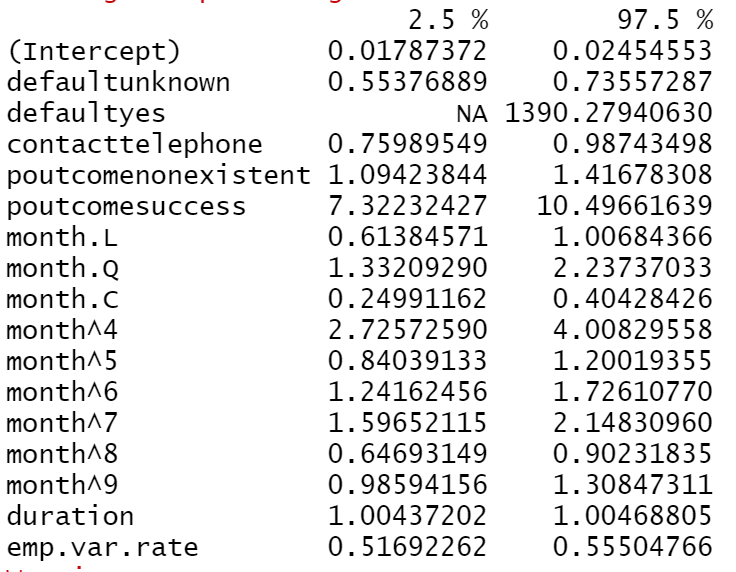
 *Table 14. Subscription of Customers with respect to Marketing Campaign duration*

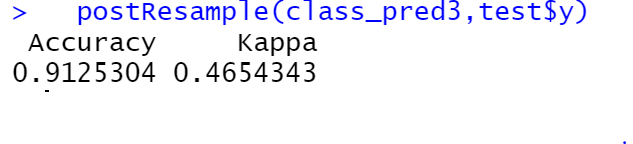
 *Table 15. Education background of customers*

*  
Fig. 27 Show Accuracy and kappa for Model 1*

 *Fig 8. Shows Confidence Interval for Model 1*

*  
Fig. 28 Show Accuracy and kappa for Model 3*

 *Fig. 12 Shows Confidence Interval for Model 2*

**

*Fig 29. Show Accuracy and kappa for Model 3*

 *Table 16. Shows R square for model 1*

 *Table 17. Shows R square for model 2*

 *Table 18. Shows R square for model 3*

# References

Catalina, T. M., 2010. CONCEPT AND EVOLUTION OF BANK MARKETING. *Research Gate.*

Chapman, P. et al., 2000. *CRISP-DM 1.0: Step-by-step data mining guide.* [Online]   
Available at: https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf  
[Accessed 5 January 2022].

Hubera, S., Wiemer, H., Schneider, D. & Ihlenfeldt, S., 2019. DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP DM Model. *Elsevier B.V,* p. 403–408.

Moro, S., Cortez, P. & Rita, P., 2014. A data-driven approach to predict the success of bank telemarketing. *ELSEVIER,* pp. 22-31.

Moro, S. & Laureano, R. M. S., 2011. USING DATA MINING FOR BANK DIRECT MARKETING: AN APPLICATION OF THE CRISP-DM METHODOLOGY. *EUROSIS-ETI.*