Portuguese Bank

Andrew Truong

2/1/2021

```
# created a dataframe for the data set


bank_final.df<-read_excel("Bank_Data_R.xlsx")
head(bank_final.df)
```

```
## # A tibble: 6 x 26
##     age age_range job   jobs  marital education educations default housing loan
##   <dbl> <chr>     <chr> <chr> <chr>   <chr>     <chr>      <chr>   <chr>   <chr>
## 1    56 old       hous~ pink~ married basic.4y  primary    no      no      no
## 2    57 old       serv~ pink~ married high.sch~ secondary  unknown no      no
## 3    37 middle    serv~ pink~ married high.sch~ secondary  no      yes     no
## 4    40 middle    admi~ whit~ married basic.6y  primary    no      no      no
## 5    56 old       serv~ pink~ married high.sch~ secondary  no      no      yes
## 6    45 middle    serv~ pink~ married basic.9y  primary    unknown no      no
## # ... with 16 more variables: contact <chr>, month <chr>, Season <chr>,
## #   day_of_week <chr>, Year <dbl>, duration <dbl>, campaign <dbl>, pdays <dbl>,
## #   previous <dbl>, poutcome <chr>, emp.var.rate <dbl>, cons.price.idx <dbl>,
## #   cons.conf.idx <dbl>, euribor3m <dbl>, nr.employed <dbl>, y <chr>
```

```
summary(bank_final.df)
```

```
##       age         age_range             job                jobs          
##  Min.   :17.00   Length:41188       Length:41188       Length:41188      
##  1st Qu.:32.00   Class :character   Class :character   Class :character  
##  Median :38.00   Mode  :character   Mode  :character   Mode  :character  
##  Mean   :40.02                                                           
##  3rd Qu.:47.00                                                           
##  Max.   :98.00                                                           
##    marital           education          educations          default         
##  Length:41188       Length:41188       Length:41188       Length:41188      
##  Class :character   Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character  
##                                                                             
##                                                                             
##                                                                             
##    housing              loan             contact             month          
##  Length:41188       Length:41188       Length:41188       Length:41188      
##  Class :character   Class :character   Class :character   Class :character  
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character  
##                                                                             
##                                                                             
##                                                                             
##     Season          day_of_week             Year         duration     
##  Length:41188       Length:41188       Min.   :2008   Min.   :   0.0  
##  Class :character   Class :character   1st Qu.:2008   1st Qu.: 102.0  
##  Mode  :character   Mode  :character   Median :2008   Median : 180.0  
##                                        Mean   :2008   Mean   : 258.3  
##                                        3rd Qu.:2009   3rd Qu.: 319.0  
##                                        Max.   :2010   Max.   :4918.0  
##     campaign          pdays          previous       poutcome        
##  Min.   : 1.000   Min.   :  0.0   Min.   :0.000   Length:41188      
##  1st Qu.: 1.000   1st Qu.:999.0   1st Qu.:0.000   Class :character  
##  Median : 2.000   Median :999.0   Median :0.000   Mode  :character  
##  Mean   : 2.568   Mean   :962.5   Mean   :0.173                     
##  3rd Qu.: 3.000   3rd Qu.:999.0   3rd Qu.:0.000                     
##  Max.   :56.000   Max.   :999.0   Max.   :7.000                     
##   emp.var.rate      cons.price.idx  cons.conf.idx     euribor3m    
##  Min.   :-3.40000   Min.   :92.20   Min.   :-50.8   Min.   :0.634  
##  1st Qu.:-1.80000   1st Qu.:93.08   1st Qu.:-42.7   1st Qu.:1.344  
##  Median : 1.10000   Median :93.75   Median :-41.8   Median :4.857  
##  Mean   : 0.08189   Mean   :93.58   Mean   :-40.5   Mean   :3.621  
##  3rd Qu.: 1.40000   3rd Qu.:93.99   3rd Qu.:-36.4   3rd Qu.:4.961  
##  Max.   : 1.40000   Max.   :94.77   Max.   :-26.9   Max.   :5.045  
##   nr.employed        y            
##  Min.   :4964   Length:41188      
##  1st Qu.:5099   Class :character  
##  Median :5191   Mode  :character  
##  Mean   :5167                     
##  3rd Qu.:5228                     
##  Max.   :5228                     
```

```
a<-colSums(is.na(bank_final.df))
a
```

```
##           age      age_range            job           jobs        marital
##             0              0              0              0              0
##     education     educations        default        housing           loan
##             0              0              0              0              0
##       contact          month         Season    day_of_week           Year
##             0              0              0              0              0
##      duration       campaign          pdays       previous       poutcome
##             0              0              0              0              0
##   emp.var.rate cons.price.idx  cons.conf.idx      euribor3m    nr.employed
##             0              0              0              0              0
##             y
##             0
```

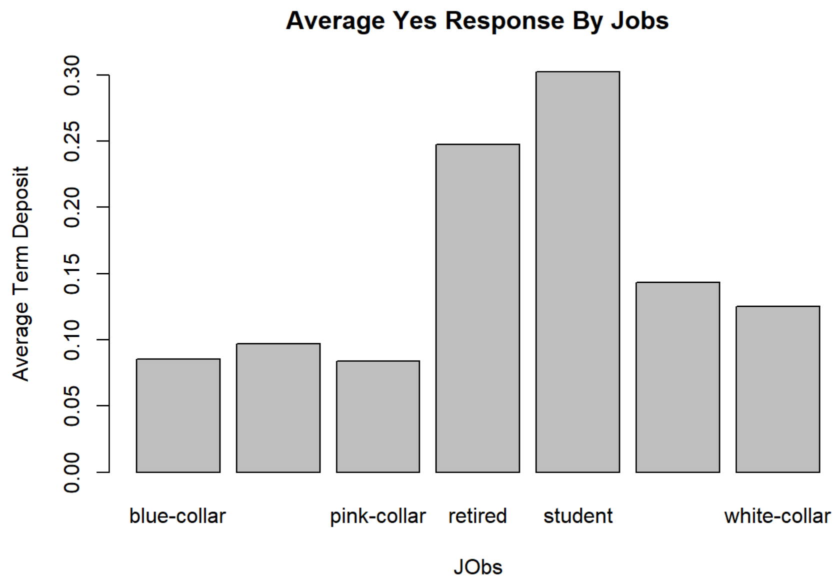
```
#removing default
bank_final.df<-bank_final.df[,-8]
head(bank_final.df)
```

```
## # A tibble: 6 x 25
##     age age_range job   jobs  marital education educations housing loan  contact
##   <dbl> <chr>     <chr> <chr> <chr>   <chr>     <chr>      <chr>   <chr> <chr>
## 1    56 old       hous~ pink~ married basic.4y  primary    no      no    teleph~
## 2    57 old       serv~ pink~ married high.sch~ secondary  no      no    teleph~
## 3    37 middle    serv~ pink~ married high.sch~ secondary  yes     no    teleph~
## 4    40 middle    admi~ whit~ married basic.6y  primary    no      no    teleph~
## 5    56 old       serv~ pink~ married high.sch~ secondary  no      yes   teleph~
## 6    45 middle    serv~ pink~ married basic.9y  primary    no      no    teleph~
## # ... with 15 more variables: month <chr>, Season <chr>, day_of_week <chr>,
## #   Year <dbl>, duration <dbl>, campaign <dbl>, pdays <dbl>, previous <dbl>,
## #   poutcome <chr>, emp.var.rate <dbl>, cons.price.idx <dbl>,
## #   cons.conf.idx <dbl>, euribor3m <dbl>, nr.employed <dbl>, y <chr>
```

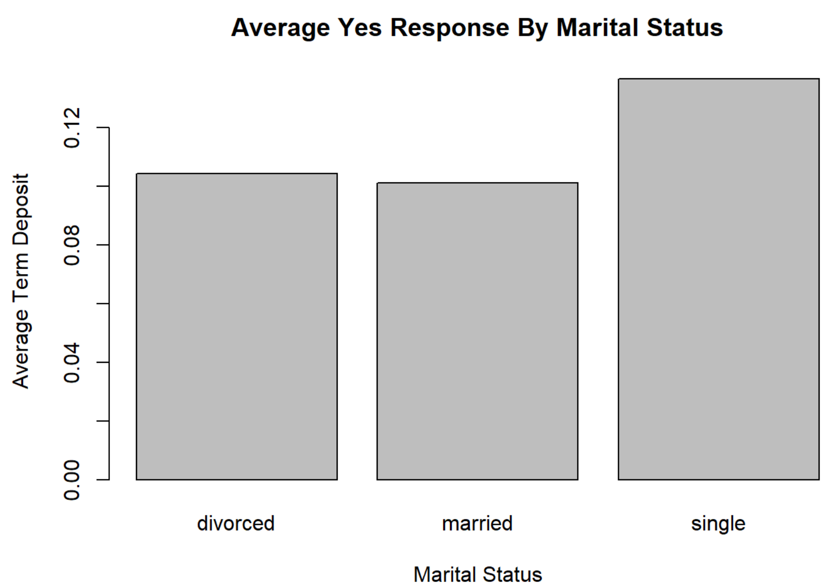
```
#removing rows with unknowns
#removing unknown jobs
bank_final.df<-bank_final.df[bank_final.df$jobs !="unknown",]

# removing unknown marital
bank_final.df<-bank_final.df[bank_final.df$marital !="unknown",]

#removing unknown educations
bank_final.df<-bank_final.df[ bank_final.df$educations !="unknown",]
bank_final.df<-bank_final.df[ bank_final.df$educations !="unkown",]

# removing unknown housing
bank_final.df<-bank_final.df[ bank_final.df$housing !="unknown", ]

#removing unknown loan
bank_final.df<-bank_final.df[bank_final.df$loan != "unknown",]
```

```
#exported data for PowerBI for additional Descriptive Statistics
write_xlsx(bank_final.df,"C:\\Users\\lilvi\\OneDrive\\Documents\\Marketing & Social Media Analytics\\Project\\Portuguese Ban
king Project\\Bank_Data_BI.xlsx")
```

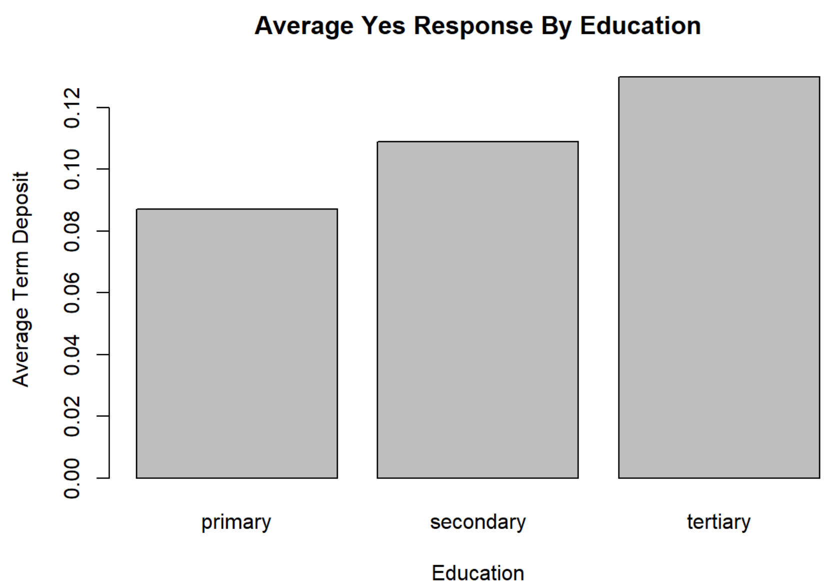
```
#bar plot Avg term deposit on job
jobs<-count(bank_final.df, jobs)

barplot(aggregate(bank_final.df$y=="yes", by=list(bank_final.df$jobs), mean, rm.na=T)[,2], xlab = "JObs", ylab="Average Term
Deposit", names.arg = jobs$jobs, main="Average Yes Response By Jobs")
```



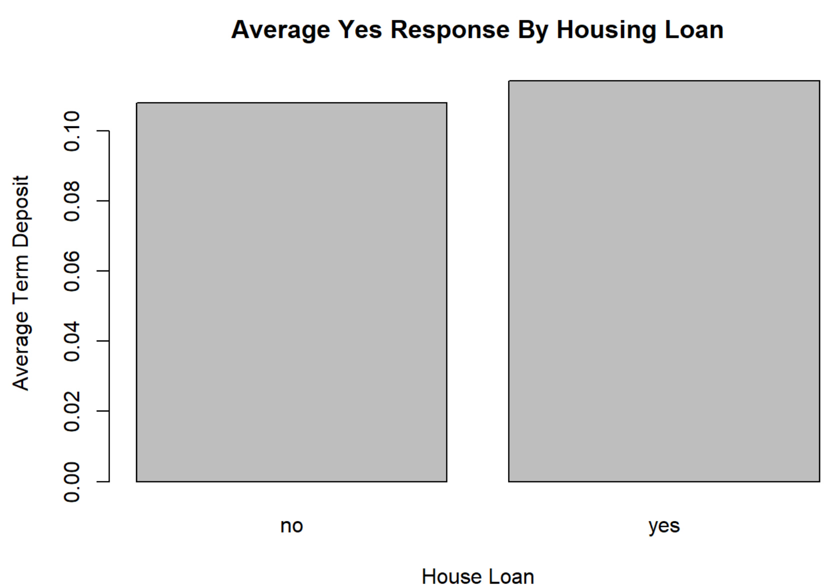

```
#bar plot Avg term deposit on marital
marital<-count(bank_final.df, marital)

barplot(aggregate(bank_final.df$y=="yes", by=list(bank_final.df$marital), mean, rm.na=T)[,2], xlab = "Marital Status", ylab=
"Average Term Deposit", names.arg = marital$marital, main="Average Yes Response By Marital Status")
```



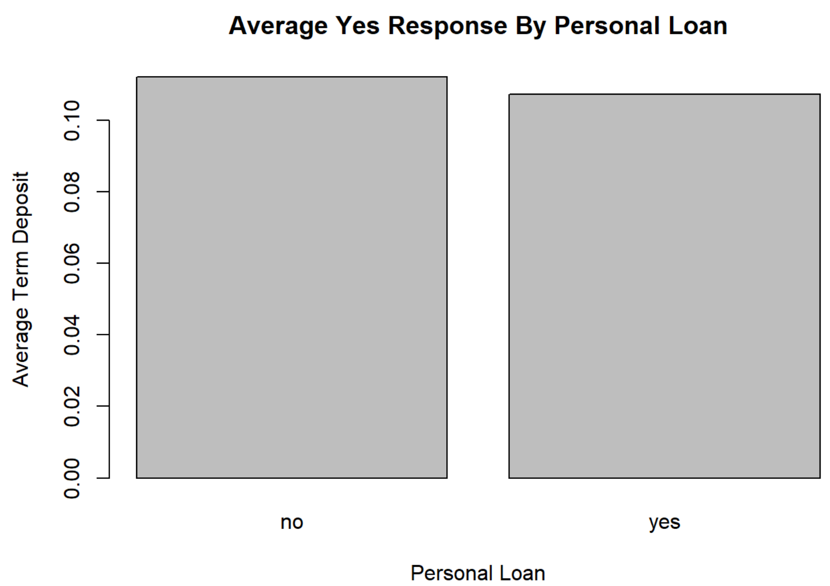

```
#bar plot Avg term deposit on educations
educations<-count(bank_final.df, educations)

barplot(aggregate(bank_final.df$y=="yes", by=list(bank_final.df$educations), mean, rm.na=T)[,2], xlab = "Education", ylab="A
verage Term Deposit", names.arg = educations$educations, main="Average Yes Response By Education" )
```



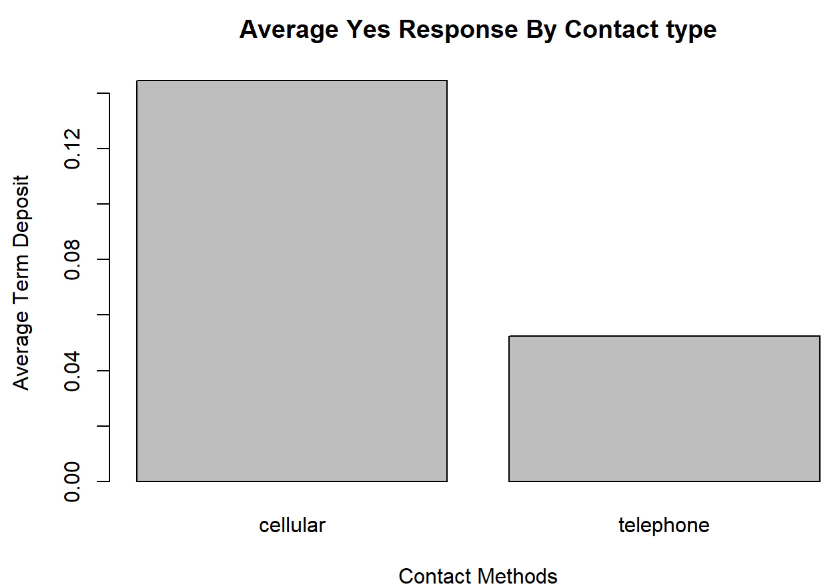

```
#bar plot Avg term deposit on house loan
house<-count(bank_final.df, housing)

barplot(aggregate(bank_final.df$y=="yes", by=list(bank_final.df$housing), mean, rm.na=T)[,2], xlab = "House Loan", ylab="Ave
rage Term Deposit", names.arg = house$housing, main="Average Yes Response By Housing Loan")
```



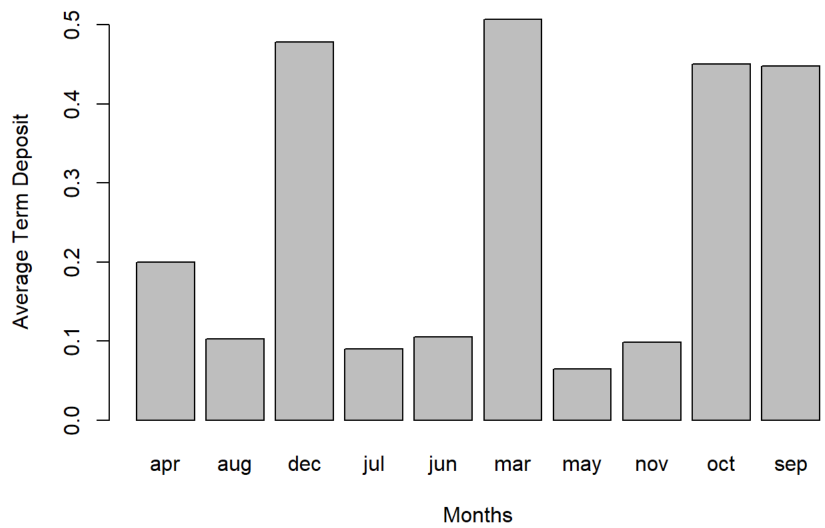

```
#bar plot Avg term deposit on personal loan
loan<-count(bank_final.df, loan)

barplot(aggregate(bank_final.df$y=="yes", by=list(bank_final.df$loan), mean, rm.na=T)[,2], xlab = "Personal Loan", ylab="Ave
rage Term Deposit", names.arg = loan$loan, main="Average Yes Response By Personal Loan")
```



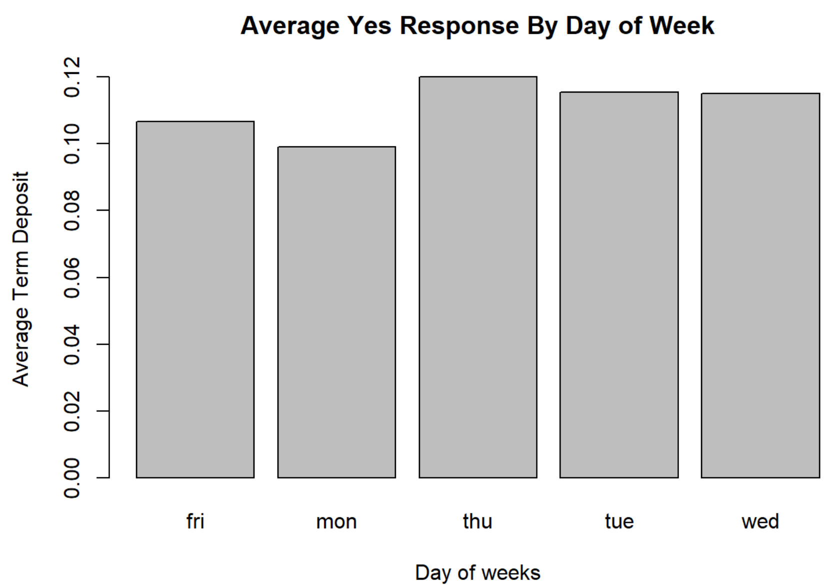

```
#bar plot Avg term deposit on contact method
contact<-count(bank_final.df, contact)

barplot(aggregate(bank_final.df$y=="yes", by=list(bank_final.df$contact), mean, rm.na=T)[,2], xlab = "Contact Methods", ylab
="Average Term Deposit", names.arg = contact$contact, main="Average Yes Response By Contact type")
```



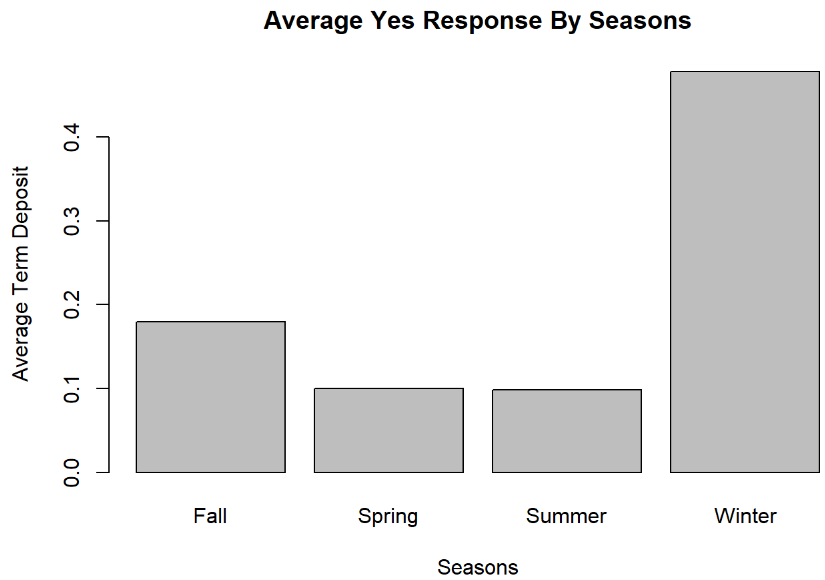

```
#bar plot Avg term deposit on months
month<-count(bank_final.df, month)

barplot(aggregate(bank_final.df$y=="yes", by=list(bank_final.df$month), mean, rm.na=T)[,2], xlab = "Months", ylab="Average T
erm Deposit", names.arg = month$month)
```



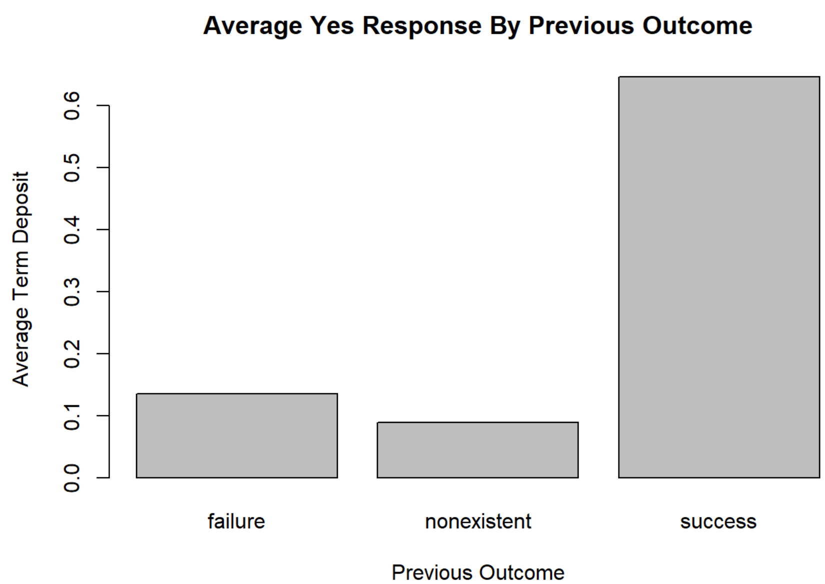

```
#bar plot Avg term deposit on day of week
dow<-count(bank_final.df, day_of_week)

barplot(aggregate(bank_final.df$y=="yes", by=list(bank_final.df$day_of_week), mean, rm.na=T)[,2], xlab = "Day of weeks", yla
b="Average Term Deposit", names.arg = dow$day_of_week, main="Average Yes Response By Day of Week")
```



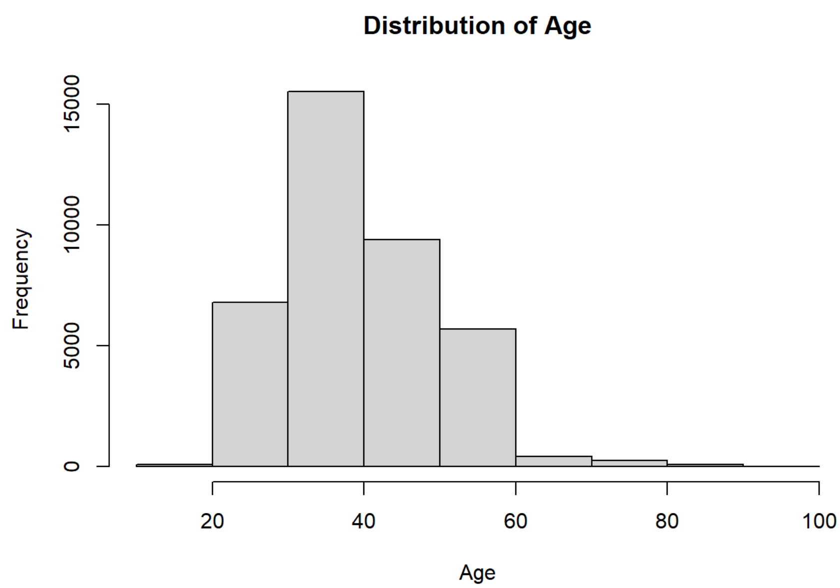

```
#bar plot Avg term deposit on Season
season<-count(bank_final.df, Season)

barplot(aggregate(bank_final.df$y=="yes", by=list(bank_final.df$Season), mean, rm.na=T)[,2], xlab = "Seasons", ylab="Average
Term Deposit", names.arg = season$Season, main="Average Yes Response By Seasons")
```



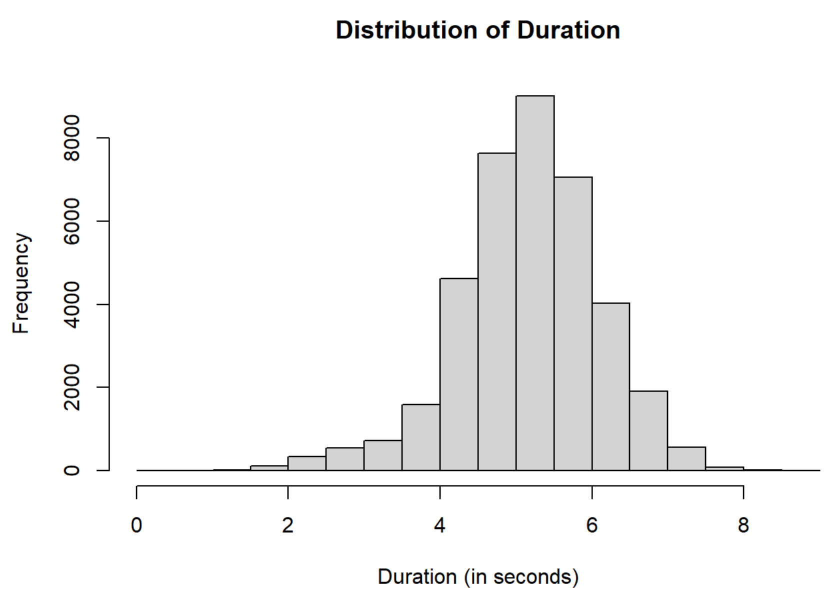

```
#bar plot Avg term deposit on previous outcome
poutcome<-count(bank_final.df, poutcome)

barplot(aggregate(bank_final.df$y=="yes", by=list(bank_final.df$poutcome), mean, rm.na=T)[,2], xlab = "Previous Outcome", yl
ab="Average Term Deposit", names.arg = poutcome$poutcome, main="Average Yes Response By Previous Outcome")
```



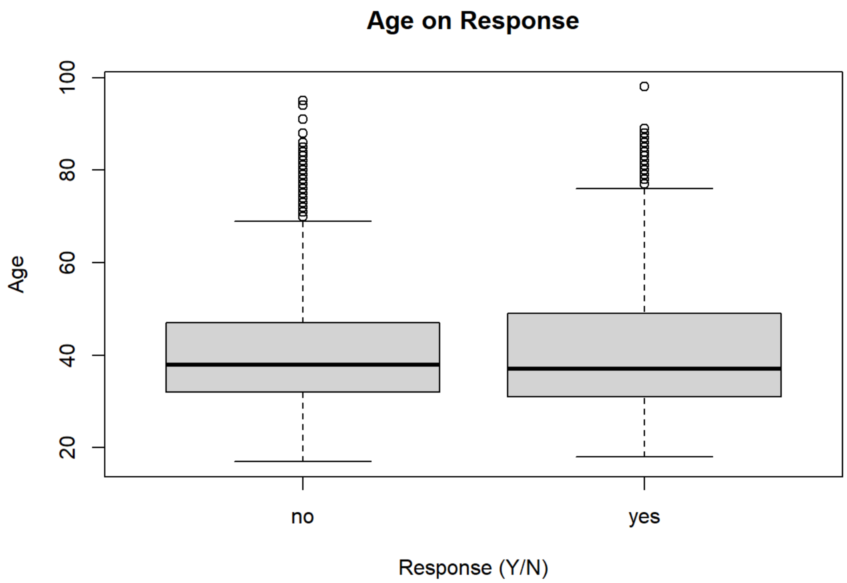

```
#distribution of age

hist(bank_final.df$age, xlab = "Age", breaks = 10, main = "Distribution of Age")
```




```
#distribution of duration

hist(log(bank_final.df$duration), xlab = "Duration (in seconds)", breaks = 20, main = "Distribution of Duration")
```

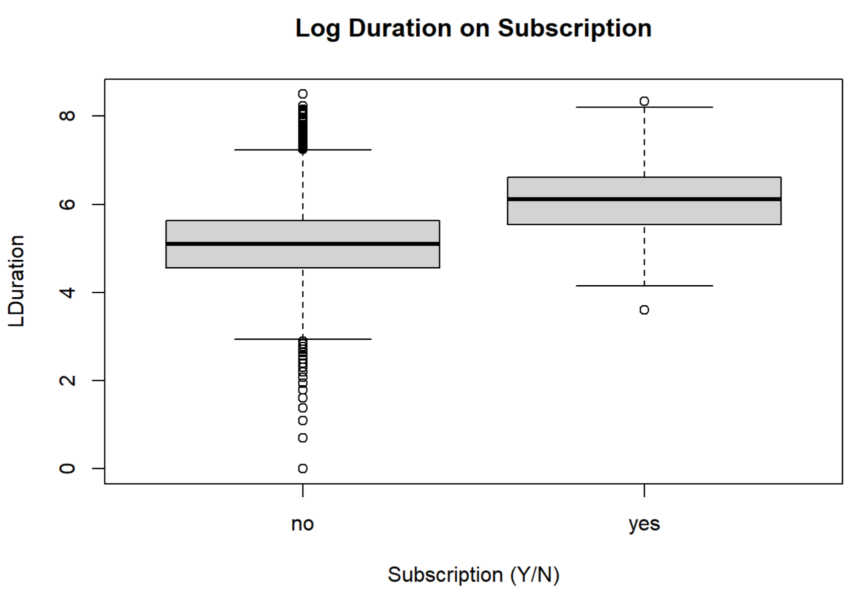



```
# box plots

# box plots of age on subscription
boxplot(bank_final.df$age~bank_final.df$y, ylab="Age", xlab="Response (Y/N)", main="Age on Response")
```

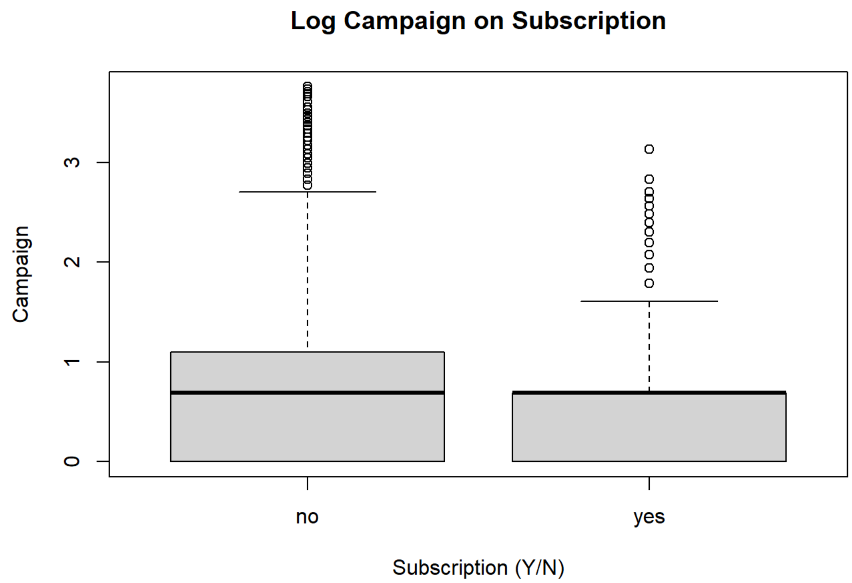



```
# box plots of call duration on subscription
boxplot(log(bank_final.df$duration)~bank_final.df$y, ylab="LDuration", xlab = "Subscription (Y/N)", main="Log Duration on Su
bscription")
```

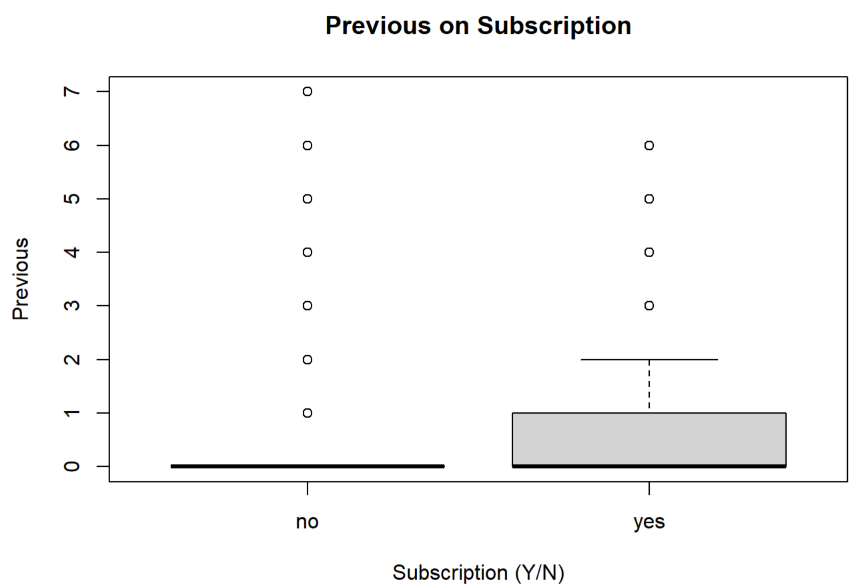
```
## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out = z$out[z$group
## == : Outlier (-Inf) in boxplot 1 is not drawn
```




```
# box plots of campaign on subscription
boxplot(log(bank_final.df$campaign)~bank_final.df$y, ylab="Campaign", xlab="Subscription (Y/N)", main="Log Campaign on Subsc
ription")
```




```
# box plots of previous on subscription
boxplot(bank_final.df$previous~bank_final.df$y, ylab="Previous", xlab="Subscription (Y/N)", main="Previous on Subscription")
```




```
# create dummy variables
df_dummy<-dummy_cols(bank_final.df, select_columns = c("jobs", "marital", "educations", "housing", "loan","contact","Season"
, "day_of_week","poutcome","y"))

df_dummy<-df_dummy[,-c(2:14, 19:25, 39, 41, 44, 57)]
```

```
# correlation matrix

b<-cor(df_dummy)
kable(b, digits=3)
```

	age	duration	campaign	pdays	previous	jobs_blue-collar	jobs_entrepreneur/self	jobs_pink-collar	jobs_retired	jobs_student	jobs_unemployed	jobs
age	1.000	0.002	0.006	-0.038	0.028	-0.069	0.023	-0.017	0.442	-0.186	-0.002	
duration	0.002	1.000	-0.071	-0.044	0.016	0.001	0.006	-0.002	0.014	0.011	-0.004	
campaign	0.006	-0.071	1.000	0.052	-0.078	0.001	0.002	0.001	-0.005	-0.024	-0.001	
pdays	-0.038	-0.044	0.052	1.000	-0.581	0.059	0.021	0.024	-0.074	-0.082	-0.027	
previous	0.028	0.016	-0.078	-0.581	1.000	-0.057	-0.016	-0.014	0.064	0.087	0.010	
jobs_blue-collar	-0.069	0.001	0.001	0.059	-0.057	1.000	-0.221	-0.300	-0.166	-0.108	-0.129	
jobs_entrepreneur/self	0.023	0.006	0.002	0.021	-0.016	-0.221	1.000	-0.103	-0.057	-0.037	-0.044	

	age	duration	campaign	pdays	previous	jobs_blue-collar	jobs_entrepreneur/self	jobs_pink-collar	jobs_retired	jobs_student	jobs_unemployed	jobs
jobs_pink-collar	-0.017	-0.002	0.001	0.024	-0.014	-0.300	-0.103	1.000	-0.078	-0.051	-0.060	
jobs_retired	0.442	0.014	-0.005	-0.074	0.064	-0.166	-0.057	-0.078	1.000	-0.028	-0.033	
jobs_student	-0.186	0.011	-0.024	-0.082	0.087	-0.108	-0.037	-0.051	-0.028	1.000	-0.022	
jobs_unemployed	-0.002	-0.004	-0.001	-0.027	0.010	-0.129	-0.044	-0.060	-0.033	-0.022	1.000	
jobs_white-collar	-0.062	-0.010	0.006	-0.026	0.023	-0.563	-0.194	-0.263	-0.146	-0.095	-0.113	
marital_divorced	0.171	-0.003	0.003	0.009	0.001	-0.052	-0.005	0.031	0.066	-0.044	0.005	
marital_married	0.266	-0.001	0.005	0.029	-0.041	0.071	0.046	0.005	0.054	-0.152	0.006	
marital_single	-0.409	0.003	-0.007	-0.037	0.044	-0.041	-0.046	-0.027	-0.105	0.197	-0.010	
educations_primary	0.151	0.014	-0.002	0.041	-0.040	0.342	-0.026	-0.024	0.085	-0.036	0.006	
educations_secondary	-0.102	0.006	0.000	0.003	0.018	-0.238	-0.074	0.307	-0.035	0.084	0.007	
educations_tertiary	-0.053	-0.018	0.002	-0.041	0.022	-0.115	0.088	-0.242	-0.049	-0.039	-0.012	
housing_yes	0.000	-0.009	-0.011	-0.009	0.021	-0.004	0.004	-0.008	-0.001	0.000	0.009	
loan_yes	-0.006	0.000	0.004	0.002	-0.003	-0.009	-0.010	-0.001	-0.009	0.004	-0.001	
contact_cellular	-0.005	0.024	-0.079	-0.115	0.209	-0.043	-0.017	-0.037	0.033	0.032	-0.009	
Season_Fall	0.065	-0.003	-0.101	-0.116	0.171	-0.087	0.047	-0.033	0.043	0.013	0.036	
Season_Spring	-0.055	0.025	-0.062	0.046	0.049	0.055	-0.010	0.039	-0.037	0.008	-0.026	
Season_Summer	0.004	-0.025	0.132	0.044	-0.173	0.008	-0.022	-0.015	0.001	-0.022	-0.001	
Season_Winter	0.043	0.019	-0.011	-0.079	0.060	-0.025	-0.009	-0.010	0.047	0.043	0.015	
day_of_week_fri	0.006	-0.008	0.029	0.013	0.006	-0.004	0.002	-0.006	0.001	-0.002	0.003	
day_of_week_mon	0.020	-0.022	0.014	0.002	-0.003	-0.012	0.009	0.008	0.000	-0.002	-0.004	
day_of_week_thu	-0.023	0.013	0.007	-0.011	0.003	0.005	0.014	-0.002	-0.013	0.006	0.003	
day_of_week_tue	0.020	0.004	-0.027	-0.005	-0.001	-0.004	-0.012	0.005	0.012	-0.002	0.002	
day_of_week_wed	-0.024	0.014	-0.023	0.002	-0.004	0.015	-0.013	-0.005	0.000	0.001	-0.005	
poutcome_failure	0.000	-0.016	-0.069	0.008	0.690	-0.021	0.000	0.001	0.020	0.031	-0.006	
poutcome_nonexistent	-0.020	-0.008	0.087	0.486	-0.882	0.048	0.009	0.011	-0.054	-0.065	-0.009	
poutcome_success	0.038	0.043	-0.050	-0.953	0.520	-0.056	-0.019	-0.022	0.070	0.072	0.028	
y_yes	0.030	0.406	-0.065	-0.319	0.221	-0.066	-0.013	-0.033	0.090	0.082	0.016	

```
# data reduction after correlation matrix previous and pdays

bank_final1.df<-df_dummy[, -c(4:5)]
```

```
# correlation matrix of reduced dataframe
c<-cor(bank_final1.df)
kable(c, digits=3)
```

	age	duration	campaign	jobs_blue-collar	jobs_entrepreneur/self	jobs_pink-collar	jobs_retired	jobs_student	jobs_unemployed	jobs_white-collar	marital_
age	1.000	0.002	0.006	-0.069	0.023	-0.017	0.442	-0.186	-0.002	-0.062	
duration	0.002	1.000	-0.071	0.001	0.006	-0.002	0.014	0.011	-0.004	-0.010	
campaign	0.006	-0.071	1.000	0.001	0.002	0.001	-0.005	-0.024	-0.001	0.006	
jobs_blue-collar	-0.069	0.001	0.001	1.000	-0.221	-0.300	-0.166	-0.108	-0.129	-0.563	
jobs_entrepreneur/self	0.023	0.006	0.002	-0.221	1.000	-0.103	-0.057	-0.037	-0.044	-0.194	
jobs_pink-collar	-0.017	-0.002	0.001	-0.300	-0.103	1.000	-0.078	-0.051	-0.060	-0.263	
jobs_retired	0.442	0.014	-0.005	-0.166	-0.057	-0.078	1.000	-0.028	-0.033	-0.146	
jobs_student	-0.186	0.011	-0.024	-0.108	-0.037	-0.051	-0.028	1.000	-0.022	-0.095	
jobs_unemployed	-0.002	-0.004	-0.001	-0.129	-0.044	-0.060	-0.033	-0.022	1.000	-0.113	
jobs_white-collar	-0.062	-0.010	0.006	-0.563	-0.194	-0.263	-0.146	-0.095	-0.113	1.000	
marital_divorced	0.171	-0.003	0.003	-0.052	-0.005	0.031	0.066	-0.044	0.005	0.018	
marital_married	0.266	-0.001	0.005	0.071	0.046	0.005	0.054	-0.152	0.006	-0.084	
marital_single	-0.409	0.003	-0.007	-0.041	-0.046	-0.027	-0.105	0.197	-0.010	0.078	
educations_primary	0.151	0.014	-0.002	0.342	-0.026	-0.024	0.085	-0.036	0.006	-0.352	
educations_secondary	-0.102	0.006	0.000	-0.238	-0.074	0.307	-0.035	0.084	0.007	0.062	
educations_tertiary	-0.053	-0.018	0.002	-0.115	0.088	-0.242	-0.049	-0.039	-0.012	0.276	
housing_yes	0.000	-0.009	-0.011	-0.004	0.004	-0.008	-0.001	0.000	0.009	0.005	
loan_yes	-0.006	0.000	0.004	-0.009	-0.010	-0.001	-0.009	0.004	-0.001	0.019	
contact_cellular	-0.005	0.024	-0.079	-0.043	-0.017	-0.037	0.033	0.032	-0.009	0.060	
Season_Fall	0.065	-0.003	-0.101	-0.087	0.047	-0.033	0.043	0.013	0.036	0.054	
Season_Spring	-0.055	0.025	-0.062	0.055	-0.010	0.039	-0.037	0.008	-0.026	-0.057	
Season_Summer	0.004	-0.025	0.132	0.008	-0.022	-0.015	0.001	-0.022	-0.001	0.020	
Season_Winter	0.043	0.019	-0.011	-0.025	-0.009	-0.010	0.047	0.043	0.015	0.000	
day_of_week_fri	0.006	-0.008	0.029	-0.004	0.002	-0.006	0.001	-0.002	0.003	0.006	
day_of_week_mon	0.020	-0.022	0.014	-0.012	0.009	0.008	0.000	-0.002	-0.004	0.004	
day_of_week_thu	-0.023	0.013	0.007	0.005	0.014	-0.002	-0.013	0.006	0.003	-0.009	
day_of_week_tue	0.020	0.004	-0.027	-0.004	-0.012	0.005	0.012	-0.002	0.002	0.002	
day_of_week_wed	-0.024	0.014	-0.023	0.015	-0.013	-0.005	0.000	0.001	-0.005	-0.003	
poutcome_failure	0.000	-0.016	-0.069	-0.021	0.000	0.001	0.020	0.031	-0.006	0.006	
poutcome_nonexistent	-0.020	-0.008	0.087	0.048	0.009	0.011	-0.054	-0.065	-0.009	-0.018	
poutcome_success	0.038	0.043	-0.050	-0.056	-0.019	-0.022	0.070	0.072	0.028	0.025	
y_yes	0.030	0.406	-0.065	-0.066	-0.013	-0.033	0.090	0.082	0.016	0.031	

```
# partition

set.seed(1)

train.rows<-sample(rownames(bank_final1.df), dim(bank_final1.df)*.7)

train.data<-bank_final1.df[train.rows,]

valid.rows<-setdiff(rownames(bank_final1.df), train.rows)

valid.data<-bank_final1.df[valid.rows,]
```

#Logistic Regression

```
#model to test if duration need higher order
mod0<-glm(y_yes~age + duration  + campaign + + `jobs_blue-collar` + `jobs_white-collar` + jobs_retired + jobs_student + jobs
_unemployed + `jobs_entrepreneur/self` + marital_single + marital_divorced + educations_secondary + educations_tertiary + ho
using_yes + loan_yes + contact_cellular + Season_Fall + Season_Spring + Season_Winter + day_of_week_tue + day_of_week_wed +
 day_of_week_thu + day_of_week_fri + poutcome_nonexistent + poutcome_success, data = train.data, family="binomial")

#testing if higher order of variable is needed for duration
resettest(mod0, power = 2, type = "regressor")
```

```
##
##  RESET test
##
## data:  mod0
## RESET = 21.67, df1 = 25, df2 = 26720, p-value < 2.2e-16
```

```
# building logistic regression model

mod1<- glm(y_yes~age+ duration + I(duration^2) + campaign + `jobs_blue-collar` + `jobs_white-collar` + jobs_retired + jobs_s
tudent + jobs_unemployed + `jobs_entrepreneur/self` + marital_single + marital_divorced + educations_secondary + educations_
tertiary + housing_yes + loan_yes + contact_cellular + Season_Fall + Season_Spring + Season_Winter + day_of_week_tue + day_o
f_week_wed + day_of_week_thu + day_of_week_fri + poutcome_nonexistent + poutcome_success, data = train.data, family="binomia
l")

mod11<-glm(y_yes~duration + I(duration^2) + campaign + I(campaign^2), data= train.data, family="binomial")
resettest(mod11, power = 2, type = "regressor")
```

```
##
##  RESET test
##
## data:  mod11
## RESET = 8.9696, df1 = 4, df2 = 26762, p-value = 3.099e-07
```

```
summary(mod1)
```

```
##
## Call:
## glm(formula = y_yes ~ age + duration + I(duration^2) + campaign +
##     `jobs_blue-collar` + `jobs_white-collar` + jobs_retired +
##     jobs_student + jobs_unemployed + `jobs_entrepreneur/self` +
##     marital_single + marital_divorced + educations_secondary +
##     educations_tertiary + housing_yes + loan_yes + contact_cellular +
##     Season_Fall + Season_Spring + Season_Winter + day_of_week_tue +
##     day_of_week_wed + day_of_week_thu + day_of_week_fri + poutcome_nonexistent +
##     poutcome_success, family = "binomial", data = train.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3473  -0.3595  -0.2375  -0.1494   4.1998
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -5.769e+00  2.039e-01 -28.288  < 2e-16 ***
## age                      7.596e-03  2.849e-03   2.666 0.007680 **
## duration                 7.343e-03  1.842e-04  39.873  < 2e-16 ***
## I(duration^2)           -2.352e-06  1.152e-07 -20.420  < 2e-16 ***
## campaign                -8.932e-02  1.422e-02  -6.283 3.32e-10 ***
## `jobs_blue-collar`       1.027e-01  9.143e-02   1.123 0.261355
## `jobs_white-collar`      3.799e-01  9.026e-02   4.209 2.56e-05 ***
## jobs_retired             1.282e+00  1.344e-01   9.541  < 2e-16 ***
## jobs_student             1.219e+00  1.571e-01   7.759 8.56e-15 ***
## jobs_unemployed          6.752e-01  1.579e-01   4.277 1.89e-05 ***
## `jobs_entrepreneur/self`  9.624e-02  1.250e-01   0.770 0.441320
## marital_single           2.452e-01  5.971e-02   4.106 4.03e-05 ***
## marital_divorced        -6.630e-02  7.939e-02  -0.835 0.403613
## educations_secondary     2.090e-01  7.350e-02   2.844 0.004453 **
## educations_tertiary      4.013e-01  6.330e-02   6.340 2.30e-10 ***
## housing_yes             -6.702e-03  4.803e-02  -0.140 0.889037
## loan_yes                -9.946e-02  6.709e-02  -1.483 0.138196
## contact_cellular         9.177e-01  6.261e-02  14.658  < 2e-16 ***
## Season_Fall              3.782e-01  6.841e-02   5.529 3.23e-08 ***
## Season_Spring            5.343e-02  5.630e-02   0.949 0.342621
## Season_Winter            1.390e+00  2.438e-01   5.702 1.18e-08 ***
## day_of_week_tue          1.517e-01  7.620e-02   1.991 0.046466 *
## day_of_week_wed          1.632e-01  7.559e-02   2.159 0.030883 *
## day_of_week_thu          1.102e-01  7.468e-02   1.476 0.139962
## day_of_week_fri          9.035e-02  7.862e-02   1.149 0.250528
## poutcome_nonexistent    -2.636e-01  7.435e-02  -3.545 0.000392 ***
## poutcome_success         2.618e+00  1.039e-01  25.196  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18579  on 26770  degrees of freedom
## Residual deviance: 12380  on 26744  degrees of freedom
## AIC: 12434
##
## Number of Fisher Scoring iterations: 6
```

```
mod1.pred <- predict(mod1,valid.data,type="response")
summary(mod1.pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.01786 0.03775 0.11140 0.09840 0.99644
```

```
pred <- as.factor(ifelse(mod1.pred >=0.2,1,0))
confusionMatrix(pred,factor(valid.data$y_yes))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    0    1
##          0 9234  480
##          1  933  827
##                                           
##                Accuracy : 0.8769          
##                  95% CI : (0.8707, 0.8828)
##     No Information Rate : 0.8861          
##     P-Value [Acc > NIR] : 0.999           
##                                           
##                   Kappa : 0.47            
##                                           
##  Mcnemar's Test P-Value : <2e-16          
##                                           
##             Sensitivity : 0.9082          
##             Specificity : 0.6327          
##          Pos Pred Value : 0.9506          
##          Neg Pred Value : 0.4699          
##              Prevalence : 0.8861          
##          Detection Rate : 0.8048          
##    Detection Prevalence : 0.8466          
##       Balanced Accuracy : 0.7705          
##                                           
##        'Positive' Class : 0               
## 
```

```
#without day of week
mod2<-glm(y_yes~age + duration + I(duration^2) + campaign  + `jobs_blue-collar` + `jobs_white-collar` + jobs_retired + jobs_student + jobs_unemployed + `jobs_entrepreneur/self` + marital_single + marital_divorced + educations_secondary + educations_tertiary +  housing_yes + loan_yes + contact_cellular + Season_Fall + Season_Spring+ Season_Winter + poutcome_nonexistent + poutcome_success , data = train.data, family="binomial")

summary(mod2)
```

```
##
## Call:
## glm(formula = y_yes ~ age + duration + I(duration^2) + campaign +
##     `jobs_blue-collar` + `jobs_white-collar` + jobs_retired +
##     jobs_student + jobs_unemployed + `jobs_entrepreneur/self` +
##     marital_single + marital_divorced + educations_secondary +
##     educations_tertiary + housing_yes + loan_yes + contact_cellular +
##     Season_Fall + Season_Spring + Season_Winter + poutcome_nonexistent +
##     poutcome_success, family = "binomial", data = train.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3784  -0.3602  -0.2373  -0.1501   4.1997
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -5.647e+00  1.963e-01 -28.770  < 2e-16 ***
## age                      7.357e-03  2.848e-03   2.584  0.00978 **
## duration                 7.346e-03  1.840e-04  39.934  < 2e-16 ***
## I(duration^2)           -2.353e-06  1.150e-07 -20.450  < 2e-16 ***
## campaign                -9.067e-02  1.423e-02  -6.374 1.84e-10 ***
## `jobs_blue-collar`       1.035e-01  9.138e-02   1.133  0.25731
## `jobs_white-collar`      3.814e-01  9.019e-02   4.229 2.35e-05 ***
## jobs_retired             1.290e+00  1.342e-01   9.607  < 2e-16 ***
## jobs_student             1.216e+00  1.572e-01   7.739 9.98e-15 ***
## jobs_unemployed          6.763e-01  1.578e-01   4.285 1.83e-05 ***
## `jobs_entrepreneur/self`  9.170e-02  1.249e-01   0.734  0.46301
## marital_single           2.452e-01  5.970e-02   4.107 4.01e-05 ***
## marital_divorced        -6.415e-02  7.936e-02  -0.808  0.41889
## educations_secondary     2.054e-01  7.346e-02   2.796  0.00518 **
## educations_tertiary      3.977e-01  6.325e-02   6.288 3.22e-10 ***
## housing_yes             -7.796e-03  4.802e-02  -0.162  0.87103
## loan_yes                -9.853e-02  6.705e-02  -1.469  0.14172
## contact_cellular         9.139e-01  6.251e-02  14.620  < 2e-16 ***
## Season_Fall              3.812e-01  6.832e-02   5.580 2.41e-08 ***
## Season_Spring            4.909e-02  5.617e-02   0.874  0.38217
## Season_Winter            1.376e+00  2.434e-01   5.653 1.58e-08 ***
## poutcome_nonexistent    -2.640e-01  7.430e-02  -3.553  0.00038 ***
## poutcome_success         2.618e+00  1.038e-01  25.216  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18579  on 26770  degrees of freedom
## Residual deviance: 12386  on 26748  degrees of freedom
## AIC: 12432
##
## Number of Fisher Scoring iterations: 6
```

```
kable(xtable(mod2))
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-5.6466670	0.1962679	-28.7702077	0.0000000
age	0.0073570	0.0028476	2.5835794	0.0097781
duration	0.0073462	0.0001840	39.9340211	0.0000000
I(duration^2)	-0.0000024	0.0000001	-20.4503458	0.0000000
campaign	-0.0906721	0.0142250	-6.3741234	0.0000000
`jobs_blue-collar`	0.1035166	0.0913828	1.1327801	0.2573066
`jobs_white-collar`	0.3814317	0.0901943	4.2290008	0.0000235
jobs_retired	1.2896669	0.1342356	9.6074874	0.0000000
jobs_student	1.2163108	0.1571572	7.7394525	0.0000000
jobs_unemployed	0.6763038	0.1578211	4.2852564	0.0000183
`jobs_entrepreneur/self`	0.0916980	0.1249466	0.7338974	0.4630113
marital_single	0.2451765	0.0596983	4.1069283	0.0000401
marital_divorced	-0.0641497	0.0793587	-0.8083512	0.4188884
educations_secondary	0.2053935	0.0734637	2.7958498	0.0051763
educations_tertiary	0.3976876	0.0632474	6.2878133	0.0000000
housing_yes	-0.0077959	0.0480184	-0.1623523	0.8710284
loan_yes	-0.0985275	0.0670519	-1.4694213	0.1417186
contact_cellular	0.9138675	0.0625096	14.6196370	0.0000000
Season_Fall	0.3812318	0.0683238	5.5797838	0.0000000
Season_Spring	0.0490904	0.0561735	0.8739061	0.3821694
Season_Winter	1.3756802	0.2433642	5.6527624	0.0000000
poutcome_nonexistent	-0.2640352	0.0743047	-3.5534113	0.0003803
poutcome_success	2.6180929	0.1038280	25.2156654	0.0000000

```
mod2.pred <- predict(mod2,valid.data,type="response")
summary(mod2.pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.01782 0.03784 0.11138 0.09908 0.99622
```

```
pred <- as.factor(ifelse(mod2.pred >=0.15,1,0))
confusionMatrix(pred,factor(valid.data$y_yes))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 8940  371
##          1 1227  936
##
##                Accuracy : 0.8607
##                  95% CI : (0.8543, 0.867)
##     No Information Rate : 0.8861
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.4633
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.8793
##             Specificity : 0.7161
##          Pos Pred Value : 0.9602
##          Neg Pred Value : 0.4327
##              Prevalence : 0.8861
##          Detection Rate : 0.7792
##    Detection Prevalence : 0.8115
##       Balanced Accuracy : 0.7977
##
##        'Positive' Class : 0
##
```

```
#switch blue collar to pink
mod3<-glm(y_yes~duration + I(duration^2) + campaign + `jobs_pink-collar` + `jobs_white-collar` + jobs_retired + jobs_student
+ jobs_unemployed + `jobs_entrepreneur/self` + marital_single + marital_divorced + educations_secondary + educations_tertiar
y + housing_yes + loan_yes + contact_cellular + Season_Fall + Season_Winter +  poutcome_nonexistent + poutcome_success, data
= train.data, family="binomial")
summary(mod3)
```

```
##
## Call:
## glm(formula = y_yes ~ duration + I(duration^2) + campaign + `jobs_pink-collar` +
##     `jobs_white-collar` + jobs_retired + jobs_student + jobs_unemployed +
##     `jobs_entrepreneur/self` + marital_single + marital_divorced +
##     educations_secondary + educations_tertiary + housing_yes +
##     loan_yes + contact_cellular + Season_Fall + Season_Winter +
##     poutcome_nonexistent + poutcome_success, family = "binomial",
##     data = train.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3761  -0.3598  -0.2379  -0.1500   4.1887
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -5.185e+00  1.200e-01 -43.221  < 2e-16 ***
## duration                 7.348e-03  1.839e-04  39.948  < 2e-16 ***
## I(duration^2)           -2.356e-06  1.151e-07 -20.474  < 2e-16 ***
## campaign                -9.182e-02  1.418e-02  -6.477 9.36e-11 ***
## `jobs_pink-collar`      -9.519e-02  9.133e-02  -1.042 0.297296
## `jobs_white-collar`      2.862e-01  6.147e-02   4.656 3.23e-06 ***
## jobs_retired             1.341e+00  9.979e-02  13.437  < 2e-16 ***
## jobs_student             1.057e+00  1.431e-01   7.385 1.53e-13 ***
## jobs_unemployed          5.783e-01  1.429e-01   4.048 5.16e-05 ***
## `jobs_entrepreneur/self`  1.835e-03  1.039e-01   0.018 0.985905
## marital_single           1.891e-01  5.544e-02   3.410 0.000649 ***
## marital_divorced        -4.605e-02  7.899e-02  -0.583 0.559943
## educations_secondary     1.837e-01  7.303e-02   2.516 0.011884 *
## educations_tertiary      3.813e-01  6.286e-02   6.066 1.31e-09 ***
## housing_yes             -5.585e-03  4.798e-02  -0.116 0.907337
## loan_yes                -9.869e-02  6.704e-02  -1.472 0.140983
## contact_cellular         9.016e-01  6.112e-02  14.750  < 2e-16 ***
## Season_Fall              3.685e-01  6.300e-02   5.850 4.93e-09 ***
## Season_Winter            1.382e+00  2.421e-01   5.708 1.14e-08 ***
## poutcome_nonexistent    -2.818e-01  7.222e-02  -3.902 9.55e-05 ***
## poutcome_success         2.611e+00  1.028e-01  25.411  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18579  on 26770  degrees of freedom
## Residual deviance: 12393  on 26750  degrees of freedom
## AIC: 12435
##
## Number of Fisher Scoring iterations: 6
```

```
mod3.pred <- predict(mod3,valid.data,type="response")
summary(mod3.pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.01783 0.03804 0.11137 0.09902 0.99647
```

```
pred <- as.factor(ifelse(mod3.pred >=0.15,1,0))
confusionMatrix(pred,factor(valid.data$y_yes))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 8937  381
##          1 1230  926
##
##                Accuracy : 0.8596
##                  95% CI : (0.8531, 0.8659)
##     No Information Rate : 0.8861
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.4579
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.8790
##             Specificity : 0.7085
##          Pos Pred Value : 0.9591
##          Neg Pred Value : 0.4295
##              Prevalence : 0.8861
##          Detection Rate : 0.7789
##    Detection Prevalence : 0.8121
##       Balanced Accuracy : 0.7938
##
##        'Positive' Class : 0
##
```

```
# w/o contact, season and poutcome
mod4<-glm(y_yes~duration + I(duration^2) + campaign + `jobs_blue-collar` + `jobs_white-collar` + jobs_retired + jobs_student
+ jobs_unemployed + `jobs_entrepreneur/self` + marital_single + marital_divorced + educations_secondary + educations_tertiar
y + housing_yes + loan_yes , data = train.data, family="binomial")

summary(mod4)
```

```
##
## Call:
## glm(formula = y_yes ~ duration + I(duration^2) + campaign + `jobs_blue-collar` +
##     `jobs_white-collar` + jobs_retired + jobs_student + jobs_unemployed +
##     `jobs_entrepreneur/self` + marital_single + marital_divorced +
##     educations_secondary + educations_tertiary + housing_yes +
##     loan_yes, family = "binomial", data = train.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3154  -0.4161  -0.2861  -0.1998   4.3015
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -4.500e+00  1.069e-01 -42.098  < 2e-16 ***
## duration                 6.926e-03  1.702e-04  40.688  < 2e-16 ***
## I(duration^2)           -2.294e-06  1.085e-07 -21.138  < 2e-16 ***
## campaign                -1.400e-01  1.363e-02 -10.270  < 2e-16 ***
## `jobs_blue-collar`       7.025e-02  8.603e-02   0.817   0.4141
## `jobs_white-collar`      4.441e-01  8.444e-02   5.260 1.44e-07 ***
## jobs_retired             1.736e+00  1.111e-01  15.635  < 2e-16 ***
## jobs_student             1.558e+00  1.408e-01  11.064  < 2e-16 ***
## jobs_unemployed          8.448e-01  1.439e-01   5.871 4.34e-09 ***
## `jobs_entrepreneur/self`  5.938e-02  1.177e-01   0.505   0.6138
## marital_single           2.593e-01  5.168e-02   5.018 5.23e-07 ***
## marital_divorced        -6.419e-02  7.386e-02  -0.869   0.3848
## educations_secondary     2.175e-01  6.849e-02   3.175   0.0015 **
## educations_tertiary      5.261e-01  5.858e-02   8.981  < 2e-16 ***
## housing_yes              8.550e-02  4.465e-02   1.915   0.0555 .
## loan_yes                -8.863e-02  6.249e-02  -1.418   0.1561
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18579  on 26770  degrees of freedom
## Residual deviance: 14166  on 26755  degrees of freedom
## AIC: 14198
##
## Number of Fisher Scoring iterations: 6
```

```
mod4.pred <- predict(mod4,valid.data,type="response")
summary(mod4.pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.02801 0.05201 0.11273 0.11844 0.93217
```

```
pred <- as.factor(ifelse(mod4.pred >=0.15,1,0))
confusionMatrix(pred,factor(valid.data$y_yes))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 8696  458
##          1 1471  849
##
##                Accuracy : 0.8319
##                  95% CI : (0.8249, 0.8387)
##     No Information Rate : 0.8861
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.3774
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.8553
##             Specificity : 0.6496
##          Pos Pred Value : 0.9500
##          Neg Pred Value : 0.3659
##              Prevalence : 0.8861
##          Detection Rate : 0.7579
##    Detection Prevalence : 0.7978
##       Balanced Accuracy : 0.7524
##
##        'Positive' Class : 0
##
```

```
#w/o age, poutxome
mod5<-glm(y_yes~+ duration + I(duration^2) + campaign + `jobs_blue-collar` + `jobs_white-collar` + jobs_retired + jobs_stude
nt + jobs_unemployed + `jobs_entrepreneur/self` + marital_single + marital_divorced + educations_secondary + educations_tert
iary + housing_yes + loan_yes + contact_cellular + Season_Fall + Season_Winter, data = train.data, family="binomial")


summary(mod5)
```

```
##
## Call:
## glm(formula = y_yes ~ +duration + I(duration^2) + campaign +
##     `jobs_blue-collar` + `jobs_white-collar` + jobs_retired +
##     jobs_student + jobs_unemployed + `jobs_entrepreneur/self` +
##     marital_single + marital_divorced + educations_secondary +
##     educations_tertiary + housing_yes + loan_yes + contact_cellular +
##     Season_Fall + Season_Winter, family = "binomial", data = train.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6176  -0.4063  -0.2679  -0.1620   4.1311
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -5.341e+00  1.186e-01 -45.015  < 2e-16 ***
## duration                  7.039e-03  1.756e-04  40.098  < 2e-16 ***
## I(duration^2)            -2.281e-06  1.124e-07 -20.289  < 2e-16 ***
## campaign                 -1.097e-01  1.387e-02  -7.912 2.54e-15 ***
## `jobs_blue-collar`        4.903e-02  8.738e-02   0.561   0.5747
## `jobs_white-collar`       3.760e-01  8.614e-02   4.365 1.27e-05 ***
## jobs_retired              1.549e+00  1.134e-01  13.653  < 2e-16 ***
## jobs_student              1.393e+00  1.437e-01   9.695  < 2e-16 ***
## jobs_unemployed           7.812e-01  1.470e-01   5.315 1.07e-07 ***
## `jobs_entrepreneur/self`  2.070e-02  1.201e-01   0.172   0.8632
## marital_single            2.209e-01  5.274e-02   4.189 2.81e-05 ***
## marital_divorced         -6.442e-02  7.529e-02  -0.856   0.3922
## educations_secondary      1.702e-01  6.968e-02   2.442   0.0146 *
## educations_tertiary       4.167e-01  5.993e-02   6.952 3.59e-12 ***
## housing_yes               3.678e-03  4.565e-02   0.081   0.9358
## loan_yes                 -9.958e-02  6.378e-02  -1.561   0.1184
## contact_cellular          1.076e+00  5.866e-02  18.340  < 2e-16 ***
## Season_Fall               5.444e-01  5.838e-02   9.326  < 2e-16 ***
## Season_Winter             1.805e+00  2.234e-01   8.077 6.65e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18579  on 26770  degrees of freedom
## Residual deviance: 13565  on 26752  degrees of freedom
## AIC: 13603
##
## Number of Fisher Scoring iterations: 6
```

```
mod5.pred <- predict(mod5,valid.data,type="response")
summary(mod5.pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.02159 0.04816 0.11251 0.11800 0.98166
```

```
pred <- as.factor(ifelse(mod5.pred >=0.15,1,0))
confusionMatrix(pred,factor(valid.data$y_yes))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    0    1
##          0 8747  422
##          1 1420  885
##                                           
##                Accuracy : 0.8395          
##                  95% CI : (0.8326, 0.8461)
##     No Information Rate : 0.8861          
##     P-Value [Acc > NIR] : 1               
##                                           
##                   Kappa : 0.4033          
##                                           
##  Mcnemar's Test P-Value : <2e-16          
##                                           
##             Sensitivity : 0.8603          
##             Specificity : 0.6771          
##          Pos Pred Value : 0.9540          
##          Neg Pred Value : 0.3839          
##              Prevalence : 0.8861          
##          Detection Rate : 0.7623          
##    Detection Prevalence : 0.7991          
##       Balanced Accuracy : 0.7687          
##                                           
##        'Positive' Class : 0               
## 
```

```
#w/o age and season
mod6<-glm(y_yes~ duration + I(duration^2) + campaign + `jobs_blue-collar` + `jobs_white-collar` + jobs_retired + jobs_studen
t + jobs_unemployed  + marital_married + marital_divorced + educations_secondary + educations_tertiary + housing_yes + loan_
yes + poutcome_nonexistent + poutcome_success, data = train.data, family="binomial")


summary(mod6)
```

```
## 
## Call:
## glm(formula = y_yes ~ duration + I(duration^2) + campaign + `jobs_blue-collar` + 
##     `jobs_white-collar` + jobs_retired + jobs_student + jobs_unemployed + 
##     marital_married + marital_divorced + educations_secondary + 
##     educations_tertiary + housing_yes + loan_yes + poutcome_nonexistent + 
##     poutcome_success, family = "binomial", data = train.data)
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -3.2726  -0.3662  -0.2462  -0.1721   4.3403  
## 
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)    
## (Intercept)          -4.110e+00  1.225e-01 -33.561  < 2e-16 ***
## duration              7.306e-03  1.807e-04  40.426  < 2e-16 ***
## I(duration^2)        -2.378e-06  1.125e-07 -21.131  < 2e-16 ***
## campaign             -1.111e-01  1.399e-02  -7.945 1.94e-15 ***
## `jobs_blue-collar`    5.899e-02  7.375e-02   0.800  0.42374    
## `jobs_white-collar`   3.791e-01  7.205e-02   5.262 1.43e-07 ***
## jobs_retired          1.504e+00  1.075e-01  13.990  < 2e-16 ***
## jobs_student          1.185e+00  1.468e-01   8.069 7.09e-16 ***
## jobs_unemployed       6.626e-01  1.465e-01   4.523 6.10e-06 ***
## marital_married      -2.177e-01  5.473e-02  -3.979 6.93e-05 ***
## marital_divorced     -2.656e-01  8.458e-02  -3.140  0.00169 ** 
## educations_secondary  2.063e-01  7.161e-02   2.881  0.00397 ** 
## educations_tertiary   4.659e-01  6.165e-02   7.557 4.12e-14 ***
## housing_yes           5.960e-02  4.728e-02   1.261  0.20740    
## loan_yes             -8.833e-02  6.617e-02  -1.335  0.18188    
## poutcome_nonexistent -5.628e-01  7.045e-02  -7.989 1.37e-15 ***
## poutcome_success      2.598e+00  1.021e-01  25.457  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 18579  on 26770  degrees of freedom
## Residual deviance: 12723  on 26754  degrees of freedom
## AIC: 12757
## 
## Number of Fisher Scoring iterations: 6
```

```
mod6.pred <- predict(mod6,valid.data,type="response")
summary(mod6.pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.02078 0.03944 0.11121 0.09977 0.98641
```

```
pred <- as.factor(ifelse(mod6.pred >=0.15,1,0))
confusionMatrix(pred,factor(valid.data$y_yes))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 8913  414
##          1 1254  893
##
##                Accuracy : 0.8546
##                  95% CI : (0.848, 0.861)
##     No Information Rate : 0.8861
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.4374
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.8767
##             Specificity : 0.6832
##          Pos Pred Value : 0.9556
##          Neg Pred Value : 0.4159
##              Prevalence : 0.8861
##          Detection Rate : 0.7768
##    Detection Prevalence : 0.8129
##       Balanced Accuracy : 0.7800
##
##        'Positive' Class : 0
##
```

```
#w/o contact cellular and season spring
mod7<-glm(y_yes~age+ duration + I(duration^2) + campaign + `jobs_blue-collar` + `jobs_white-collar`  + jobs_student + jobs_u
nemployed  + marital_married + marital_divorced + educations_secondary + educations_tertiary + housing_yes + loan_yes + Seas
on_Winter + Season_Fall + poutcome_nonexistent + poutcome_success, data = train.data, family="binomial")


summary(mod7)
```

```
##
## Call:
## glm(formula = y_yes ~ age + duration + I(duration^2) + campaign +
##     `jobs_blue-collar` + `jobs_white-collar` + jobs_student +
##     jobs_unemployed + marital_married + marital_divorced + educations_secondary +
##     educations_tertiary + housing_yes + loan_yes + Season_Winter +
##     Season_Fall + poutcome_nonexistent + poutcome_success, family = "binomial",
##     data = train.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3914  -0.3666  -0.2475  -0.1738   4.4055
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -4.692e+00  1.582e-01 -29.661  < 2e-16 ***
## age                   2.025e-02  2.516e-03   8.048 8.45e-16 ***
## duration              7.298e-03  1.805e-04  40.435  < 2e-16 ***
## I(duration^2)        -2.377e-06  1.125e-07 -21.127  < 2e-16 ***
## campaign             -1.025e-01  1.392e-02  -7.368 1.73e-13 ***
## `jobs_blue-collar`   -1.794e-01  6.703e-02  -2.676  0.00744 **
## `jobs_white-collar`   1.067e-01  6.525e-02   1.635  0.10209
## jobs_student          1.090e+00  1.479e-01   7.371 1.69e-13 ***
## jobs_unemployed       3.451e-01  1.438e-01   2.400  0.01639 *
## marital_married      -3.461e-01  5.862e-02  -5.905 3.53e-09 ***
## marital_divorced     -4.193e-01  8.913e-02  -4.705 2.54e-06 ***
## educations_secondary  1.678e-01  7.154e-02   2.345  0.01900 *
## educations_tertiary   4.298e-01  6.171e-02   6.965 3.27e-12 ***
## housing_yes           4.714e-02  4.729e-02   0.997  0.31887
## loan_yes             -8.323e-02  6.607e-02  -1.260  0.20776
## Season_Winter         1.491e+00  2.376e-01   6.273 3.53e-10 ***
## Season_Fall           4.990e-01  6.248e-02   7.987 1.39e-15 ***
## poutcome_nonexistent -4.833e-01  7.121e-02  -6.786 1.15e-11 ***
## poutcome_success      2.590e+00  1.021e-01  25.362  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18579  on 26770  degrees of freedom
## Residual deviance: 12736  on 26752  degrees of freedom
## AIC: 12774
##
## Number of Fisher Scoring iterations: 6
```

```
mod7.pred <- predict(mod7,valid.data,type="response")
summary(mod7.pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.02098 0.03948 0.11073 0.09874 0.99578
```

```
pred <- as.factor(ifelse(mod7.pred >=0.15,1,0))
confusionMatrix(pred,factor(valid.data$y_yes))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    0    1
##          0 8934  400
##          1 1233  907
##                                           
##                Accuracy : 0.8577          
##                  95% CI : (0.8512, 0.864) 
##     No Information Rate : 0.8861          
##     P-Value [Acc > NIR] : 1               
##                                           
##                   Kappa : 0.4482          
##                                           
##  Mcnemar's Test P-Value : <2e-16          
##                                           
##             Sensitivity : 0.8787          
##             Specificity : 0.6940          
##          Pos Pred Value : 0.9571          
##          Neg Pred Value : 0.4238          
##              Prevalence : 0.8861          
##          Detection Rate : 0.7786          
##    Detection Prevalence : 0.8135          
##       Balanced Accuracy : 0.7863          
##                                           
##        'Positive' Class : 0               
## 
```

```
#without season spring, day of week and age
mod8<- glm(y_yes~ duration + I(duration^2) + campaign + `jobs_blue-collar` + `jobs_white-collar` + jobs_retired + jobs_stude
nt + jobs_unemployed + `jobs_entrepreneur/self` + marital_single + marital_divorced + educations_secondary + educations_tert
iary + housing_yes + loan_yes + contact_cellular + Season_Fall + Season_Winter+ poutcome_nonexistent + poutcome_success, dat
a = train.data, family="binomial")

summary(mod8)
```

```
##
## Call:
## glm(formula = y_yes ~ duration + I(duration^2) + campaign + `jobs_blue-collar` +
##     `jobs_white-collar` + jobs_retired + jobs_student + jobs_unemployed +
##     `jobs_entrepreneur/self` + marital_single + marital_divorced +
##     educations_secondary + educations_tertiary + housing_yes +
##     loan_yes + contact_cellular + Season_Fall + Season_Winter +
##     poutcome_nonexistent + poutcome_success, family = "binomial",
##     data = train.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3761  -0.3598  -0.2379  -0.1500   4.1887
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -5.280e+00  1.404e-01 -37.614  < 2e-16 ***
## duration                 7.348e-03  1.839e-04  39.948  < 2e-16 ***
## I(duration^2)           -2.356e-06  1.151e-07 -20.474  < 2e-16 ***
## campaign                -9.182e-02  1.418e-02  -6.477 9.36e-11 ***
## `jobs_blue-collar`       9.519e-02  9.133e-02   1.042 0.297296
## `jobs_white-collar`      3.814e-01  9.017e-02   4.230 2.34e-05 ***
## jobs_retired             1.436e+00  1.208e-01  11.888  < 2e-16 ***
## jobs_student             1.152e+00  1.552e-01   7.421 1.16e-13 ***
## jobs_unemployed          6.735e-01  1.577e-01   4.272 1.94e-05 ***
## `jobs_entrepreneur/self`  9.702e-02  1.248e-01   0.777 0.437064
## marital_single           1.891e-01  5.544e-02   3.410 0.000649 ***
## marital_divorced        -4.605e-02  7.899e-02  -0.583 0.559943
## educations_secondary     1.837e-01  7.303e-02   2.516 0.011884 *
## educations_tertiary      3.813e-01  6.286e-02   6.066 1.31e-09 ***
## housing_yes             -5.585e-03  4.798e-02  -0.116 0.907337
## loan_yes                -9.869e-02  6.704e-02  -1.472 0.140983
## contact_cellular         9.016e-01  6.112e-02  14.750  < 2e-16 ***
## Season_Fall              3.685e-01  6.300e-02   5.850 4.93e-09 ***
## Season_Winter            1.382e+00  2.421e-01   5.708 1.14e-08 ***
## poutcome_nonexistent    -2.818e-01  7.222e-02  -3.902 9.55e-05 ***
## poutcome_success         2.611e+00  1.028e-01  25.411  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18579  on 26770  degrees of freedom
## Residual deviance: 12393  on 26750  degrees of freedom
## AIC: 12435
##
## Number of Fisher Scoring iterations: 6
```

```
#w/o spring, day of week, age
mod8.pred <- predict(mod8,valid.data,type="response")
summary(mod8.pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.01783 0.03804 0.11137 0.09902 0.99647
```

```
pred <- as.factor(ifelse(mod8.pred >=0.15,1,0))
confusionMatrix(pred,factor(valid.data$y_yes))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    0    1
##          0 8937  381
##          1 1230  926
##                                           
##                Accuracy : 0.8596          
##                  95% CI : (0.8531, 0.8659)
##     No Information Rate : 0.8861          
##     P-Value [Acc > NIR] : 1               
##                                           
##                   Kappa : 0.4579          
##                                           
##  Mcnemar's Test P-Value : <2e-16          
##                                           
##             Sensitivity : 0.8790          
##             Specificity : 0.7085          
##          Pos Pred Value : 0.9591          
##          Neg Pred Value : 0.4295          
##              Prevalence : 0.8861          
##          Detection Rate : 0.7789          
##    Detection Prevalence : 0.8121          
##       Balanced Accuracy : 0.7938          
##                                           
##        'Positive' Class : 0               
## 
```

```
#w/o age, day of week, job self
mod9<- glm(y_yes~ duration + I(duration^2) + campaign + `jobs_blue-collar` + `jobs_white-collar` + jobs_retired + jobs_stude
nt + jobs_unemployed  + marital_single + marital_divorced + educations_secondary + educations_tertiary + housing_yes + loan_
yes + contact_cellular + Season_Fall + Season_Spring + Season_Winter + poutcome_nonexistent + poutcome_success, data = trai
n.data, family="binomial")

summary(mod9)
```

```
##
## Call:
## glm(formula = y_yes ~ duration + I(duration^2) + campaign + `jobs_blue-collar` +
##     `jobs_white-collar` + jobs_retired + jobs_student + jobs_unemployed +
##     marital_single + marital_divorced + educations_secondary +
##     educations_tertiary + housing_yes + loan_yes + contact_cellular +
##     Season_Fall + Season_Spring + Season_Winter + poutcome_nonexistent +
##     poutcome_success, family = "binomial", data = train.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3784  -0.3593  -0.2382  -0.1502   4.1868
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -5.289e+00  1.428e-01 -37.044  < 2e-16 ***
## duration              7.346e-03  1.840e-04  39.928  < 2e-16 ***
## I(duration^2)        -2.354e-06  1.151e-07 -20.452  < 2e-16 ***
## campaign             -9.064e-02  1.424e-02  -6.367 1.92e-10 ***
## `jobs_blue-collar`    5.541e-02  7.470e-02   0.742 0.458289
## `jobs_white-collar`   3.423e-01  7.316e-02   4.678 2.89e-06 ***
## jobs_retired          1.400e+00  1.091e-01  12.834  < 2e-16 ***
## jobs_student          1.116e+00  1.480e-01   7.542 4.64e-14 ***
## jobs_unemployed       6.359e-01  1.490e-01   4.268 1.97e-05 ***
## marital_single        1.869e-01  5.540e-02   3.374 0.000740 ***
## marital_divorced     -4.790e-02  7.897e-02  -0.607 0.544131
## educations_secondary  1.777e-01  7.243e-02   2.454 0.014141 *
## educations_tertiary   3.900e-01  6.277e-02   6.213 5.20e-10 ***
## housing_yes          -6.374e-03  4.800e-02  -0.133 0.894354
## loan_yes             -9.903e-02  6.703e-02  -1.477 0.139556
## contact_cellular      9.123e-01  6.248e-02  14.600  < 2e-16 ***
## Season_Fall           3.924e-01  6.819e-02   5.755 8.67e-09 ***
## Season_Spring         4.700e-02  5.614e-02   0.837 0.402516
## Season_Winter         1.406e+00  2.437e-01   5.769 7.98e-09 ***
## poutcome_nonexistent -2.667e-01  7.425e-02  -3.591 0.000329 ***
## poutcome_success      2.623e+00  1.038e-01  25.274  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18579  on 26770  degrees of freedom
## Residual deviance: 12393  on 26750  degrees of freedom
## AIC: 12435
##
## Number of Fisher Scoring iterations: 6
```

```
mod9.pred <- predict(mod9,valid.data,type="response")
summary(mod9.pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.01778 0.03789 0.11139 0.09855 0.99650
```

```
pred <- as.factor(ifelse(mod9.pred >=0.15,1,0))
confusionMatrix(pred,factor(valid.data$y_yes))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 8931  378
##          1 1236  929
##
##                Accuracy : 0.8593
##                  95% CI : (0.8528, 0.8656)
##     No Information Rate : 0.8861
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.4582
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.8784
##             Specificity : 0.7108
##          Pos Pred Value : 0.9594
##          Neg Pred Value : 0.4291
##              Prevalence : 0.8861
##          Detection Rate : 0.7784
##    Detection Prevalence : 0.8113
##       Balanced Accuracy : 0.7946
##
##        'Positive' Class : 0
##
```

```
#without day of week, house loan, and loan
mod10<-glm(y_yes~age + duration + I(duration^2) + campaign + `jobs_blue-collar` + `jobs_white-collar` + jobs_retired + jobs_student + jobs_unemployed + `jobs_entrepreneur/self` + marital_single + marital_divorced + educations_secondary + educations_tertiary  + contact_cellular + Season_Fall + Season_Spring+ Season_Winter + poutcome_nonexistent + poutcome_success , data = train.data, family="binomial")

summary(mod10)
```

```
##
## Call:
## glm(formula = y_yes ~ age + duration + I(duration^2) + campaign +
##     `jobs_blue-collar` + `jobs_white-collar` + jobs_retired +
##     jobs_student + jobs_unemployed + `jobs_entrepreneur/self` +
##     marital_single + marital_divorced + educations_secondary +
##     educations_tertiary + contact_cellular + Season_Fall + Season_Spring +
##     Season_Winter + poutcome_nonexistent + poutcome_success,
##     family = "binomial", data = train.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3724  -0.3605  -0.2379  -0.1500   4.2192
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -5.662e+00  1.953e-01 -28.997  < 2e-16 ***
## age                      7.351e-03  2.847e-03   2.582 0.009835 **
## duration                 7.352e-03  1.839e-04  39.974  < 2e-16 ***
## I(duration^2)           -2.359e-06  1.150e-07 -20.520  < 2e-16 ***
## campaign                -9.113e-02  1.423e-02  -6.404 1.51e-10 ***
## `jobs_blue-collar`       1.035e-01  9.137e-02   1.133 0.257251
## `jobs_white-collar`      3.806e-01  9.018e-02   4.220 2.44e-05 ***
## jobs_retired             1.290e+00  1.342e-01   9.613  < 2e-16 ***
## jobs_student             1.215e+00  1.571e-01   7.733 1.05e-14 ***
## jobs_unemployed          6.754e-01  1.578e-01   4.281 1.86e-05 ***
## `jobs_entrepreneur/self`  9.344e-02  1.249e-01   0.748 0.454306
## marital_single           2.454e-01  5.969e-02   4.111 3.93e-05 ***
## marital_divorced        -6.405e-02  7.936e-02  -0.807 0.419611
## educations_secondary     2.052e-01  7.346e-02   2.794 0.005213 **
## educations_tertiary      3.966e-01  6.323e-02   6.273 3.55e-10 ***
## contact_cellular         9.116e-01  6.229e-02  14.633  < 2e-16 ***
## Season_Fall              3.825e-01  6.829e-02   5.601 2.13e-08 ***
## Season_Spring            4.903e-02  5.614e-02   0.873 0.382481
## Season_Winter            1.372e+00  2.434e-01   5.639 1.71e-08 ***
## poutcome_nonexistent    -2.645e-01  7.427e-02  -3.562 0.000369 ***
## poutcome_success         2.618e+00  1.038e-01  25.219  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18579  on 26770  degrees of freedom
## Residual deviance: 12388  on 26750  degrees of freedom
## AIC: 12430
##
## Number of Fisher Scoring iterations: 6
```

```
mod10.pred <- predict(mod10,valid.data,type="response")
summary(mod10.pred)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.01785 0.03783 0.11137 0.09899 0.99615
```

```
pred <- as.factor(ifelse(mod10.pred >=0.15,1,0))
confusionMatrix(pred,factor(valid.data$y_yes))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    0    1
##          0 8935  374
##          1 1232  933
##                                           
##                Accuracy : 0.86            
##                  95% CI : (0.8535, 0.8663)
##     No Information Rate : 0.8861          
##     P-Value [Acc > NIR] : 1               
##                                           
##                   Kappa : 0.4609          
##                                           
##  Mcnemar's Test P-Value : <2e-16          
##                                           
##             Sensitivity : 0.8788          
##             Specificity : 0.7138          
##          Pos Pred Value : 0.9598          
##          Neg Pred Value : 0.4309          
##              Prevalence : 0.8861          
##          Detection Rate : 0.7787          
##    Detection Prevalence : 0.8113          
##       Balanced Accuracy : 0.7963          
##                                           
##        'Positive' Class : 0               
## 
```

#Naive Bayes

```
# naive bayes function


# change numerical variables to categorical
#bank_final.df$jobs<-factor(bank_final.df$jobs)
#bank_final.df$marital<-factor(bank_final.df$marital)
#bank_final.df$educations<-factor(bank_final.df$educations)
#bank_final.df$housing<-factor(bank_final.df$housing)
#bank_final.df$loan<-factor(bank_final.df$loan)
#bank_final.df$Season<-factor(bank_final.df$Season)
#bank_final.df$contact<factor(bank_final.df$contact)


#create training and validation sets
#selected.var<-c(4, 5, 7:10, 12, 19, 25)
#train.rows1<-sample(rownames(bank_final.df), dim(bank_final.df)*.7)

#train.data1<-bank_final.df[train.rows1, selected.var]

#valid.rows1<-setdiff(rownames(bank_final.df), train.rows1)

#valid.data1<-bank_final.df[valid.rows1, selected.var]




#run naive bayes
#bank_final.nb<-naiveBayes(y~., data = train.data1)
#bank_final.nb
```

```
#First naive bayes model with no day of week
selected.var1<-c(4:23,29:32)
train.rows11<-sample(rownames(bank_final1.df), dim(bank_final1.df)*.7)

train.data11<-bank_final1.df[train.rows11, selected.var1]

valid.rows11<-setdiff(rownames(bank_final1.df), train.rows11)

valid.data11<-bank_final1.df[valid.rows11, selected.var1]

#run naive bayes
bank_final.nb1<-naiveBayes(y_yes~., data = train.data11)
bank_final.nb1
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
## Y
##         0         1
## 0.8873408 0.1126592
##
## Conditional probabilities:
##    jobs_blue-collar
## Y        [,1]      [,2]
##   0 0.4009682 0.4901049
##   1 0.3037135 0.4599367
##
##    jobs_entrepreneur/self
## Y         [,1]      [,2]
##   0 0.07249000 0.2593030
##   1 0.06034483 0.2381641
##
##    jobs_pink-collar
## Y         [,1]      [,2]
##   0 0.12734161 0.3333623
##   1 0.09250663 0.2897879
##
##    jobs_retired
## Y         [,1]      [,2]
##   0 0.03544517 0.1849061
##   1 0.09084881 0.2874416
##
##    jobs_student
## Y         [,1]      [,2]
##   0 0.01435487 0.1189513
##   1 0.04675066 0.2111393
##
##    jobs_unemployed
## Y         [,1]      [,2]
##   0 0.02412124 0.1534288
##   1 0.03249337 0.1773358
##
##    jobs_white-collar
## Y        [,1]      [,2]
##   0 0.3252789 0.4684888
##   1 0.3733422 0.4837720
##
##    marital_divorced
## Y        [,1]      [,2]
##   0 0.1124816 0.3159647
##   1 0.1064324 0.3084413
##
##    marital_married
## Y        [,1]      [,2]
##   0 0.6126710 0.4871501
##   1 0.5464191 0.4979232
##
##    marital_single
## Y        [,1]      [,2]
##   0 0.2748474 0.4464467
##   1 0.3471485 0.4761424
##
##    educations_primary
## Y        [,1]      [,2]
##   0 0.3225426 0.4674592
##   1 0.2456897 0.4305667
##
##    educations_secondary
## Y        [,1]      [,2]
##   0 0.2446643 0.4298970
##   1 0.2390584 0.4265792
##
##    educations_tertiary
## Y        [,1]      [,2]
##   0 0.4327931 0.4954731
##   1 0.5152520 0.4998502
##
##    housing_yes
## Y        [,1]      [,2]
##   0 0.5358872 0.4987209
##   1 0.5474138 0.4978294
```

```
##
##    loan_yes
## Y        [,1]      [,2]
##   0 0.1534414 0.3604200
##   1 0.1505305 0.3576499
##
##    contact_cellular
## Y        [,1]      [,2]
##   0 0.6139760 0.4868464
##   1 0.8232759 0.3814983
##
##    Season_Fall
## Y        [,1]      [,2]
##   0 0.1224163 0.3277729
##   1 0.2148541 0.4107892
##
##    Season_Spring
## Y        [,1]      [,2]
##   0 0.4162913 0.4929534
##   1 0.3620690 0.4806783
##
##    Season_Summer
## Y        [,1]      [,2]
##   0 0.4591875 0.4983420
##   1 0.4051724 0.4910068
##
##    Season_Winter
## Y         [,1]       [,2]
##   0 0.00210482 0.04583097
##   1 0.01790451 0.13262643
##
##    poutcome_failure
## Y         [,1]      [,2]
##   0 0.09968428 0.2995849
##   1 0.12168435 0.3269751
##
##    poutcome_nonexistent
## Y        [,1]      [,2]
##   0 0.8874763 0.3160163
##   1 0.6876658 0.4635222
##
##    poutcome_success
## Y        [,1]      [,2]
##   0 0.0128394 0.1125837
##   1 0.1906499 0.3928787
```

```
#library(caret)



#mod1.pred <- predict(mod1,valid.data,type="response")
#summary(mod1.pred)
#pred <- as.factor(ifelse(mod1.pred >=0.2,1,0))
#confusionMatrix(pred,factor(valid.data$y_yes))

# training
#pred.class1 <- predict(bank_final.nb, newdata = train.data1, type="raw")
#pred1<-as.factor(ifelse(pred.class1[,2]>=.15,1,0))

#y<-factor(train.data1$y)
#confusionMatrix(pred.class1, y)

# validation
#pred.class2 <- predict(bank_final.nb, newdata = valid.data1, type="response")
#confusionMatrix(pred.class2, valid.data1$y)
```

```
# training
pred.class11 <- predict(bank_final.nb1, newdata = train.data11, type="raw")
pred11<-as.factor(ifelse(pred.class11[,2]>=.15,1,0))

x<-factor(train.data11$y_yes)

confusionMatrix(pred11, factor(train.data11$y_yes))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction     0     1
##          0 21005  1884
##          1  2750  1132
##                                           
##                Accuracy : 0.8269          
##                  95% CI : (0.8223, 0.8314)
##     No Information Rate : 0.8873          
##     P-Value [Acc > NIR] : 1               
##                                           
##                   Kappa : 0.2307          
##                                           
##  Mcnemar's Test P-Value : <2e-16          
##                                           
##             Sensitivity : 0.8842          
##             Specificity : 0.3753          
##          Pos Pred Value : 0.9177          
##          Neg Pred Value : 0.2916          
##              Prevalence : 0.8873          
##          Detection Rate : 0.7846          
##    Detection Prevalence : 0.8550          
##       Balanced Accuracy : 0.6298          
##                                           
##        'Positive' Class : 0               
## 
```

```
# validation
pred.class22 <- predict(bank_final.nb1, newdata = valid.data11, type="raw")
pred22<-as.factor(ifelse(pred.class22[,2]>=.15,1,0))


confusionMatrix(pred22, factor(valid.data11$y_yes))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    0    1
##          0 9040  783
##          1 1192  459
##                                           
##                Accuracy : 0.8279          
##                  95% CI : (0.8208, 0.8347)
##     No Information Rate : 0.8918          
##     P-Value [Acc > NIR] : 1               
##                                           
##                   Kappa : 0.2211          
##                                           
##  Mcnemar's Test P-Value : <2e-16          
##                                           
##             Sensitivity : 0.8835          
##             Specificity : 0.3696          
##          Pos Pred Value : 0.9203          
##          Neg Pred Value : 0.2780          
##              Prevalence : 0.8918          
##          Detection Rate : 0.7879          
##    Detection Prevalence : 0.8561          
##       Balanced Accuracy : 0.6265          
##                                           
##        'Positive' Class : 0               
## 
```

```
#naive bayes with all categorical variables

selected.var2<-c(4:32)
train.rows2<-sample(rownames(bank_final1.df), dim(bank_final1.df)*.7)

train.data2<-bank_final1.df[train.rows2, selected.var2]

valid.rows2<-setdiff(rownames(bank_final1.df), train.rows2)

valid.data2<-bank_final1.df[valid.rows2, selected.var2]

#run naive bayes
bank_final.nb2<-naiveBayes(y_yes~., data = train.data2)
bank_final.nb2
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
## Y
##         0         1
## 0.8891338 0.1108662
##
## Conditional probabilities:
##    jobs_blue-collar
## Y        [,1]      [,2]
##   0 0.4006218 0.4900347
##   1 0.3042453 0.4601646
##
##    jobs_entrepreneur/self
## Y         [,1]      [,2]
##   0 0.07162963 0.2578791
##   1 0.06266846 0.2424065
##
##    jobs_pink-collar
## Y         [,1]      [,2]
##   0 0.12607654 0.3319426
##   1 0.09097035 0.2876154
##
##    jobs_retired
## Y         [,1]      [,2]
##   0 0.03495358 0.1836661
##   1 0.09063342 0.2871355
##
##    jobs_student
## Y         [,1]      [,2]
##   0 0.01449397 0.1195178
##   1 0.04615903 0.2098647
##
##    jobs_unemployed
## Y         [,1]      [,2]
##   0 0.02457673 0.1548345
##   1 0.03301887 0.1787159
##
##    jobs_white-collar
## Y        [,1]      [,2]
##   0 0.3276478 0.4693655
##   1 0.3723046 0.4835004
##
##    marital_divorced
## Y        [,1]      [,2]
##   0 0.1141453 0.3179943
##   1 0.1024259 0.3032586
##
##    marital_married
## Y        [,1]      [,2]
##   0 0.6116456 0.4873861
##   1 0.5535714 0.4972056
##
##    marital_single
## Y        [,1]      [,2]
##   0 0.2742091 0.4461242
##   1 0.3440027 0.4751220
##
##    educations_primary
## Y        [,1]      [,2]
##   0 0.3247070 0.4682751
##   1 0.2459569 0.4307257
##
##    educations_secondary
## Y        [,1]      [,2]
##   0 0.2431206 0.4289764
##   1 0.2355121 0.4243900
##
##    educations_tertiary
## Y        [,1]      [,2]
##   0 0.4321724 0.4953885
##   1 0.5185310 0.4997407
##
##    housing_yes
## Y        [,1]      [,2]
##   0 0.5348065 0.4987975
##   1 0.5485175 0.4977243
```

```
##
##    loan_yes
## Y        [,1]      [,2]
##   0 0.1567449 0.3635678
##   1 0.1506065 0.3577251
##
##    contact_cellular
## Y        [,1]      [,2]
##   0 0.6123598 0.4872220
##   1 0.8308625 0.3749365
##
##    Season_Fall
## Y        [,1]      [,2]
##   0 0.1222115 0.3275368
##   1 0.2159704 0.4115632
##
##    Season_Spring
## Y        [,1]      [,2]
##   0 0.4165861 0.4930034
##   1 0.3685984 0.4825060
##
##    Season_Summer
## Y        [,1]      [,2]
##   0 0.4585977 0.4982934
##   1 0.3995957 0.4898978
##
##    Season_Winter
## Y          [,1]       [,2]
##   0 0.002604714 0.05097096
##   1 0.015835580 0.12486019
##
##    day_of_week_fri
## Y        [,1]      [,2]
##   0 0.1914465 0.3934479
##   1 0.1816038 0.3855826
##
##    day_of_week_mon
## Y        [,1]      [,2]
##   0 0.2084191 0.4061866
##   1 0.1826146 0.3864153
##
##    day_of_week_thu
## Y        [,1]      [,2]
##   0 0.2067807 0.4050053
##   1 0.2254043 0.4179187
##
##    day_of_week_tue
## Y        [,1]      [,2]
##   0 0.1940092 0.3954443
##   1 0.2048518 0.4036613
##
##    day_of_week_wed
## Y        [,1]      [,2]
##   0 0.1993446 0.3995160
##   1 0.2055256 0.4041533
##
##    poutcome_failure
## Y        [,1]      [,2]
##   0 0.1020460 0.3027151
##   1 0.1233154 0.3288543
##
##    poutcome_nonexistent
## Y        [,1]      [,2]
##   0 0.8852246 0.3187575
##   1 0.6842992 0.4648727
##
##    poutcome_success
## Y         [,1]      [,2]
##   0 0.01272949 0.1121070
##   1 0.19238544 0.3942406
```

```
# training
pred.class2 <- predict(bank_final.nb2, newdata = train.data2, type="raw")
pred2<-as.factor(ifelse(pred.class2[,2]>=.15,1,0))

x<-factor(train.data2$y_yes)

confusionMatrix(pred2, factor(train.data2$y_yes))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction     0     1
##          0 21092  1872
##          1  2711  1096
##                                           
##                Accuracy : 0.8288          
##                  95% CI : (0.8242, 0.8333)
##     No Information Rate : 0.8891          
##     P-Value [Acc > NIR] : 1               
##                                           
##                   Kappa : 0.2273          
##                                           
##  Mcnemar's Test P-Value : <2e-16          
##                                           
##             Sensitivity : 0.8861          
##             Specificity : 0.3693          
##          Pos Pred Value : 0.9185          
##          Neg Pred Value : 0.2879          
##              Prevalence : 0.8891          
##          Detection Rate : 0.7879          
##    Detection Prevalence : 0.8578          
##       Balanced Accuracy : 0.6277          
##                                           
##        'Positive' Class : 0               
## 
```

```
# validation
pred.class3 <- predict(bank_final.nb2, newdata = valid.data2, type="raw")
pred3<-as.factor(ifelse(pred.class3[,2]>=.15,1,0))


confusionMatrix(pred3, factor(valid.data2$y_yes))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    0    1
##          0 9038  815
##          1 1146  475
##                                           
##                Accuracy : 0.8291          
##                  95% CI : (0.8221, 0.8359)
##     No Information Rate : 0.8876          
##     P-Value [Acc > NIR] : 1               
##                                           
##                   Kappa : 0.2299          
##                                           
##  Mcnemar's Test P-Value : 9.191e-14       
##                                           
##             Sensitivity : 0.8875          
##             Specificity : 0.3682          
##          Pos Pred Value : 0.9173          
##          Neg Pred Value : 0.2930          
##              Prevalence : 0.8876          
##          Detection Rate : 0.7877          
##    Detection Prevalence : 0.8587          
##       Balanced Accuracy : 0.6278          
##                                           
##        'Positive' Class : 0               
## 
```

```
# no contact cellular
selected.var3<-c(4:18, 20:32)
train.rows3<-sample(rownames(bank_final1.df), dim(bank_final1.df)*.7)

train.data3<-bank_final1.df[train.rows3, selected.var3]

valid.rows3<-setdiff(rownames(bank_final1.df), train.rows3)

valid.data3<-bank_final1.df[valid.rows3, selected.var3]

#run naive bayes
bank_final.nb3<-naiveBayes(y_yes~., data = train.data3)
bank_final.nb3
```

```
##
## Naive Bayes Classifier for Discrete Predictors
##
## Call:
## naiveBayes.default(x = X, y = Y, laplace = laplace)
##
## A-priori probabilities:
## Y
##         0         1
## 0.8894699 0.1105301
##
## Conditional probabilities:
##    jobs_blue-collar
## Y        [,1]      [,2]
##   0 0.4006803 0.4900466
##   1 0.2984116 0.4576384
##
##    jobs_entrepreneur/self
## Y         [,1]      [,2]
##   0 0.07319839 0.2604673
##   1 0.06488679 0.2463676
##
##    jobs_pink-collar
## Y         [,1]      [,2]
##   0 0.12959852 0.3358683
##   1 0.09564042 0.2941472
##
##    jobs_retired
## Y         [,1]      [,2]
##   0 0.03460440 0.1827795
##   1 0.08752957 0.2826573
##
##    jobs_student
## Y         [,1]      [,2]
##   0 0.01390055 0.1170808
##   1 0.05035485 0.2187131
##
##    jobs_unemployed
## Y         [,1]      [,2]
##   0 0.02523938 0.1568547
##   1 0.03109158 0.1735946
##
##    jobs_white-collar
## Y        [,1]      [,2]
##   0 0.3227784 0.4675486
##   1 0.3720852 0.4834426
##
##    marital_divorced
## Y        [,1]      [,2]
##   0 0.1133042 0.3169710
##   1 0.1081446 0.3106155
##
##    marital_married
## Y        [,1]      [,2]
##   0 0.6143961 0.4867479
##   1 0.5484961 0.4977267
##
##    marital_single
## Y        [,1]      [,2]
##   0 0.2722997 0.4451527
##   1 0.3433592 0.4749104
##
##    educations_primary
## Y        [,1]      [,2]
##   0 0.3268100 0.4690570
##   1 0.2450152 0.4301689
##
##    educations_secondary
## Y        [,1]      [,2]
##   0 0.2429447 0.4288710
##   1 0.2399459 0.4271224
##
##    educations_tertiary
## Y        [,1]      [,2]
##   0 0.4302453 0.4951208
##   1 0.5150389 0.4998583
##
##    housing_yes
## Y        [,1]      [,2]
##   0 0.5374181 0.4986084
##   1 0.5569449 0.4968307
```

```
##
##    loan_yes
## Y        [,1]      [,2]
##   0 0.1550899 0.3619979
##   1 0.1470091 0.3541749
##
##    Season_Fall
## Y        [,1]      [,2]
##   0 0.1209894 0.3261218
##   1 0.2088543 0.4065588
##
##    Season_Spring
## Y        [,1]      [,2]
##   0 0.4187804 0.4933696
##   1 0.3710713 0.4831731
##
##    Season_Summer
## Y        [,1]      [,2]
##   0 0.4576684 0.4982153
##   1 0.4004731 0.4900771
##
##    Season_Winter
## Y          [,1]       [,2]
##   0 0.002561734 0.05054976
##   1 0.019601217 0.13864886
##
##    day_of_week_fri
## Y        [,1]      [,2]
##   0 0.1897783 0.3921338
##   1 0.1811423 0.3852011
##
##    day_of_week_mon
## Y        [,1]      [,2]
##   0 0.2108601 0.4079278
##   1 0.1801284 0.3843593
##
##    day_of_week_thu
## Y        [,1]      [,2]
##   0 0.2090123 0.4066117
##   1 0.2250760 0.4177030
##
##    day_of_week_tue
## Y        [,1]      [,2]
##   0 0.1927180 0.3944417
##   1 0.2031092 0.4023811
##
##    day_of_week_wed
## Y        [,1]      [,2]
##   0 0.1976314 0.3982209
##   1 0.2105441 0.4077640
##
##    poutcome_failure
## Y        [,1]      [,2]
##   0 0.1005375 0.3007217
##   1 0.1253802 0.3312055
##
##    poutcome_nonexistent
## Y        [,1]      [,2]
##   0 0.8862338 0.3175337
##   1 0.6955052 0.4602709
##
##    poutcome_success
## Y         [,1]      [,2]
##   0 0.01322862 0.1142549
##   1 0.17911457 0.3835130
```

```
# training
pred.class4 <- predict(bank_final.nb3, newdata = train.data3, type="raw")
pred4<-as.factor(ifelse(pred.class4[,2]>=.15,1,0))

z<-factor(train.data3$y_yes)

confusionMatrix(pred4, factor(train.data3$y_yes))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction     0     1
##          0 21470  1971
##          1  2342   988
##                                           
##                Accuracy : 0.8389          
##                  95% CI : (0.8344, 0.8433)
##     No Information Rate : 0.8895          
##     P-Value [Acc > NIR] : 1               
##                                           
##                   Kappa : 0.2233          
##                                           
##  Mcnemar's Test P-Value : 1.761e-08       
##                                           
##             Sensitivity : 0.9016          
##             Specificity : 0.3339          
##          Pos Pred Value : 0.9159          
##          Neg Pred Value : 0.2967          
##              Prevalence : 0.8895          
##          Detection Rate : 0.8020          
##    Detection Prevalence : 0.8756          
##       Balanced Accuracy : 0.6178          
##                                           
##        'Positive' Class : 0               
## 
```

```
# validation
pred.class4b <- predict(bank_final.nb3, newdata = valid.data3, type="raw")
pred4b<-as.factor(ifelse(pred.class4b[,2]>=.15,1,0))


confusionMatrix(pred4b, factor(valid.data3$y_yes))
```

```
## Confusion Matrix and Statistics
## 
##           Reference
## Prediction    0    1
##          0 9216  812
##          1  959  487
##                                           
##                Accuracy : 0.8457          
##                  95% CI : (0.8389, 0.8522)
##     No Information Rate : 0.8868          
##     P-Value [Acc > NIR] : 1.0000000       
##                                           
##                   Kappa : 0.2675          
##                                           
##  Mcnemar's Test P-Value : 0.0005218       
##                                           
##             Sensitivity : 0.9057          
##             Specificity : 0.3749          
##          Pos Pred Value : 0.9190          
##          Neg Pred Value : 0.3368          
##              Prevalence : 0.8868          
##          Detection Rate : 0.8032          
##    Detection Prevalence : 0.8740          
##       Balanced Accuracy : 0.6403          
##                                           
##        'Positive' Class : 0               
## 
```