Gramener\_Assignment

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

**CampaignData.csv** inludes the data of a Direct Marketing campaign of a European Banking institution. The marketing campaigns were conducted through phone calls. Often, more than one contact to the same client was required, in order to access if the product subscribed (bank term deposit) would be ('yes') or ('no')

A brief analysis has been performed on the data procured and documented as an R Markdown document.

**Loading and processing Dataset**

myy <- read.table("New.txt", header = TRUE, sep = ";")  
  
m1 <- data.frame(myy)  
  
m1$ID<-seq.int(nrow(m1)) ## Adding a new column ID  
  
m2 <- m1[c(22,1:21)] ## Re-positioning the column ID to the beginning  
  
summary(m2)

## ID age job marital   
## Min. : 1 Min. :17.00 admin. :10422 divorced: 4612   
## 1st Qu.:10298 1st Qu.:32.00 blue-collar: 9254 married :24928   
## Median :20595 Median :38.00 technician : 6743 single :11568   
## Mean :20595 Mean :40.02 services : 3969 unknown : 80   
## 3rd Qu.:30891 3rd Qu.:47.00 management : 2924   
## Max. :41188 Max. :98.00 retired : 1720   
## (Other) : 6156   
## education default housing   
## university.degree :12168 no :32588 no :18622   
## high.school : 9515 unknown: 8597 unknown: 990   
## basic.9y : 6045 yes : 3 yes :21576   
## professional.course: 5243   
## basic.4y : 4176   
## basic.6y : 2292   
## (Other) : 1749   
## loan contact month day\_of\_week  
## no :33950 cellular :26144 may :13769 fri:7827   
## unknown: 990 telephone:15044 jul : 7174 mon:8514   
## yes : 6248 aug : 6178 thu:8623   
## jun : 5318 tue:8090   
## nov : 4101 wed:8134   
## apr : 2632   
## (Other): 2016   
## duration campaign pdays previous   
## Min. : 0.0 Min. : 1.000 Min. : 0.0 Min. :0.000   
## 1st Qu.: 102.0 1st Qu.: 1.000 1st Qu.:999.0 1st Qu.:0.000   
## Median : 180.0 Median : 2.000 Median :999.0 Median :0.000   
## Mean : 258.3 Mean : 2.568 Mean :962.5 Mean :0.173   
## 3rd Qu.: 319.0 3rd Qu.: 3.000 3rd Qu.:999.0 3rd Qu.:0.000   
## Max. :4918.0 Max. :56.000 Max. :999.0 Max. :7.000   
##   
## poutcome emp.var.rate cons.price.idx cons.conf.idx   
## failure : 4252 Min. :-3.40000 Min. :92.20 Min. :-50.8   
## nonexistent:35563 1st Qu.:-1.80000 1st Qu.:93.08 1st Qu.:-42.7   
## success : 1373 Median : 1.10000 Median :93.75 Median :-41.8   
## Mean : 0.08189 Mean :93.58 Mean :-40.5   
## 3rd Qu.: 1.40000 3rd Qu.:93.99 3rd Qu.:-36.4   
## Max. : 1.40000 Max. :94.77 Max. :-26.9   
##   
## euribor3m nr.employed response   
## Min. :0.634 Min. :4964 no :36548   
## 1st Qu.:1.344 1st Qu.:5099 yes: 4640   
## Median :4.857 Median :5191   
## Mean :3.621 Mean :5167   
## 3rd Qu.:4.961 3rd Qu.:5228   
## Max. :5.045 Max. :5228   
##

**Distribution of Age w.r.t the positive response**

The column age.group includes the age of all the clients contacted for the term deposit. Now, let's map the distribution of age of the clients over the positive response.

library(data.table); library(ggplot2); library(plyr);

## Warning: package 'data.table' was built under R version 3.3.1

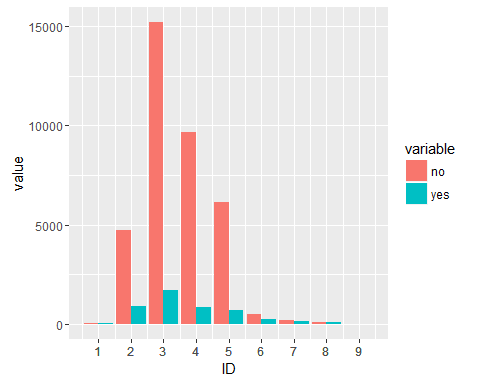
## Warning: package 'ggplot2' was built under R version 3.3.1

## Warning: package 'plyr' was built under R version 3.3.1

m2$age.group <- findInterval(m1$age, c(10, 20, 30, 40, 50, 60, 70, 80, 90, 100))  
  
new.m<- table(m2$age.group, m2$response)  
  
new.m1 = as.data.frame.matrix(new.m)  
  
new.m1$ID<-seq.int(nrow(new.m1))  
  
new.m1 <- new.m1[c(3,1:2)]  
  
mm <- melt(new.m1[,c('ID','no','yes')],id.vars = 1)  
  
new.m1$age.group <- c("10-19","20-29","30-39","40-49", "50-59", "60-69", "70-79", "80-89", "90-99")  
  
new.m1 <- new.m1[c(1,4,2,3)]  
  
print(new.m1[, 1:2])

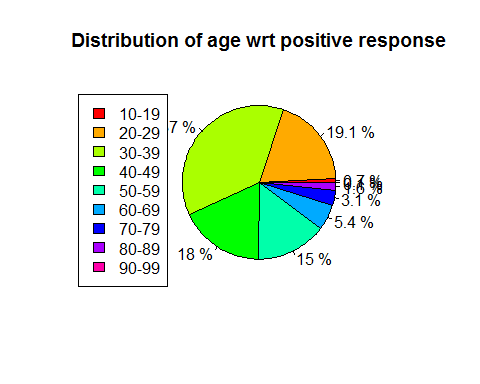
## ID age.group  
## 1 1 10-19  
## 2 2 20-29  
## 3 3 30-39  
## 4 4 40-49  
## 5 5 50-59  
## 6 6 60-69  
## 7 7 70-79  
## 8 8 80-89  
## 9 9 90-99

ggplot(mm,aes(x = ID,y = value, fill = variable)) +   
 geom\_bar(stat="identity", position = "dodge") + scale\_x\_continuous(breaks= mm$ID)



The above plot shows the distribution of Age w.r.t the responses attained. Let's plot a pie chart for more clearer view at the age of the subscribers.

piepercent<- round(100\*new.m1$yes/sum(new.m1$yes), 1)  
  
piepercent <- paste(piepercent, "%")  
  
pie(new.m1$yes, labels = piepercent, main = "Distribution of age wrt positive response", col = rainbow(length(new.m1$yes)))  
legend("bottomleft", c("10-19","20-29","30-39","40-49", "50-59", "60-69", "70-79", "80-89", "90-99"), cex = 1.0,  
 fill = rainbow(length(new.m1$yes)))



From the above plot we could clearly see which Age-Group has been subscribed for the term deposit which is *30-39 age group* which is nearly 37 % of the whole subscriptions.

**New clients who subscribed to the term deposit**

The dataset includes the responses of more than 40000 clients. The Columns *default*, *housing* and *loan* includes the responses whether the clients is already a part of the credit plan or not.

Let's analyse how many new clients have subscribed to the *term deposit*.

for(i in m2[,1])  
{  
 if(m2$default[i] == "no")  
 {  
 if(m2$housing[i] == "no")  
 {  
 if(m2$loan[i] == "no")  
 {  
 m2$new.user[i] = "yes"  
 }  
 }  
 else  
 m2$new.user[i] = "no"  
 }  
 else  
 m2$new.user[i] = "no"  
}  
  
table(m2$new.user)

##   
## no yes   
## 26574 14614

The results shows that out of the total subscribers, only 14614 clients are the fresh subscribers who doesn't have housing loan, credit default or anyother person loan.

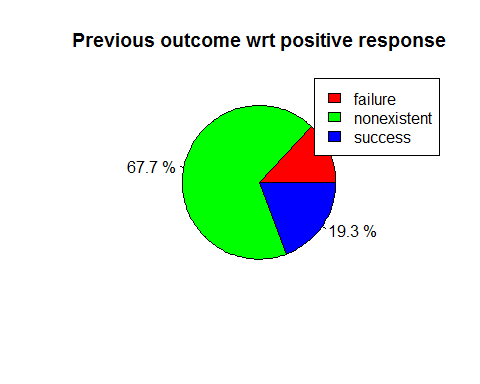
**Previous outcome w.r.t the positive response**

The clients have been previously contacted or not, If yes what is outcome has been recorded as the *poutcome* column of the dataset. Here we are going to plot the result of the previous outcome of the clients w.r.t the outcome of the response of the same user for the term deposit scheme.

p2 <- table(m2$poutcome, m2$response)  
p2 = as.data.frame.matrix(p2)  
  
setDT(p2, keep.rownames = TRUE)[]

## rn no yes  
## 1: failure 3647 605  
## 2: nonexistent 32422 3141  
## 3: success 479 894

colnames(p2)[1] <- "outcome"  
  
piepercent11<- round(100\*p2$yes/sum(p2$yes), 1)  
  
piepercent11 <- paste(piepercent11, "%")  
  
pie(p2$yes, labels = piepercent11, main = "Previous outcome wrt positive response", col = rainbow(length(p2$yes)))  
legend("topright", c(p2$outcome), cex = 1.0, fill = rainbow(length(p2$yes)))



*The plot shows the Success rate of the Term deposit campaign w.r.t. the outcome of the previous campaign.*

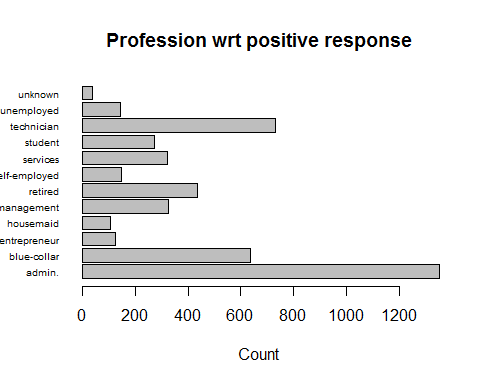
**Profession of the Subscibers**

Variation of positive response w.r.t the count of the profession of the term subscribers.

j2 <- table(m1$job, m1$response)  
  
j2 = as.data.frame.matrix(j2)  
  
setDT(j2, keep.rownames = TRUE)[]

## rn no yes  
## 1: admin. 9070 1352  
## 2: blue-collar 8616 638  
## 3: entrepreneur 1332 124  
## 4: housemaid 954 106  
## 5: management 2596 328  
## 6: retired 1286 434  
## 7: self-employed 1272 149  
## 8: services 3646 323  
## 9: student 600 275  
## 10: technician 6013 730  
## 11: unemployed 870 144  
## 12: unknown 293 37

colnames(j2)[1] <- "job"  
  
barplot(j2$yes, names.arg=j2$job, horiz=TRUE, las=1, cex.names=0.6, border=TRUE, xlab = "Count", main = "Profession wrt positive response", axes = "TRUE")

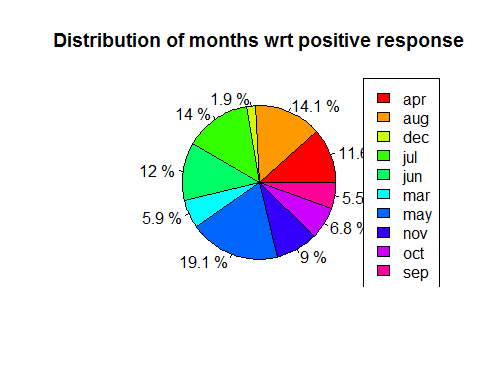


**Subcription rate Month**

n2 <- table(m1$month, m1$response)  
  
n2 = as.data.frame.matrix(n2)  
  
setDT(n2, keep.rownames = TRUE)[]

## rn no yes  
## 1: apr 2093 539  
## 2: aug 5523 655  
## 3: dec 93 89  
## 4: jul 6525 649  
## 5: jun 4759 559  
## 6: mar 270 276  
## 7: may 12883 886  
## 8: nov 3685 416  
## 9: oct 403 315  
## 10: sep 314 256

colnames(n2)[1] <- "month"  
  
nn2 <- melt(n2[,c('month','no','yes')],id.vars = 1)  
  
piepercent1<- round(100\*n2$yes/sum(n2$yes), 1)  
  
piepercent1 <- paste(piepercent1, "%")  
  
pie(n2$yes, labels = piepercent1, main = "Distribution of months wrt positive response", col = rainbow(length(n2$yes)))  
legend("topright", c(n2$month), cex = 1.0, fill = rainbow(length(n2$yes)))



*Variation of monthly subscriptions with positive response on the campaign.*

**Other Relevant Obsevations**

table(m2$housing, m2$response) ## Shows how many clients have housing loan and has subscribed for the term deposit

##   
## no yes  
## no 16596 2026  
## unknown 883 107  
## yes 19069 2507

table(m2$loan, m2$response) ## Shows how many clients have loan and has subscribed for the term deposit

##   
## no yes  
## no 30100 3850  
## unknown 883 107  
## yes 5565 683

table(m2$default, m2$response) ## Shows how many clients have credit in default and has subscribed for the term deposit

##   
## no yes  
## no 28391 4197  
## unknown 8154 443  
## yes 3 0

Complete Analysis has been performed using R studio.