Bank Marketing Campaign

Marcos Ikino

Introdução

Para este estudo a fonte deste conjunto de dados foi obtida do repositório do UCI: http://archive.ics.uci.edu/ml/datasets/Bank+Marketing.

O conjunto de dados está relacionado em campanhas de marketing de uma instituição bancária, que realizou ligações telefônicas aos clientes com o intuito de que eles efetuassem depósitos em dinheiro para a instituição bancária.

O objetivo deste estudo é realizar uma análise e responder às questões de negócio sobre este conjunto de dados, buscando identificar padrões sobre as características de comportamento dos clientes e das estratégias adotadas pelas campanhas de marketing, que possam auxiliar os processos a se tornarem mais eficientes a partir da identificação destes padrões.

As questões da área de negocio abordadas na próxima buscarão ser respondidas através do tratamento e agrupamento dos dados, e também pelas formulações de modelos de aprendizado de máquina. Embora o conjunto de dados esteja bastante relacionado com a previsão se um cliente irá realizar ou não um depósito bancário em função das variáveis envolvidas, o enfoque no emprego dos modelos de aprendizado de máquina será dado em atender a área de negócios e não na previsão da variável de resposta.

1. Análise de negócios

Questões a serem desenvolvidas neste estudo:

1. Qual profissão tem mais tendência a fazer um empréstimo? De qual tipo?
2. Fazendo uma relação entre número de contatos e sucesso da campanha quais são os pontos relevantes a serem observados?
3. Baseando-se nos resultados de adesão desta campanha qual o número médio e o máximo de ligações que você indica para otimizar a adesão?
4. O resultado da campanha anterior tem relevância na campanha atual?
5. Qual o fator determinante para que o banco exija um seguro de crédito?
6. Quais são as características mais proeminentes de um cliente que possua empréstimo imobiliário?

2. Variáveis de entrada e saída

Input variables:

bank client data:

1 - age (numeric)

2 - job : type of job (categorical:“admin.”,“unknown”,“unemployed”,“management”,“housemaid”,“entrepreneur”,“student”, “blue-collar”,“self-employed”,“retired”,“technician”,“services”)

3 - marital : marital status (categorical: “married”,“divorced”,“single”; note: “divorced” means divorced or widowed)

4 - education (categorical: “unknown”,“secondary”,“primary”,“tertiary”)

5 - default: has credit in default? (binary: “yes”,“no”)

6 - balance: average yearly balance, in euros (numeric)

7 - housing: has housing loan? (binary: “yes”,“no”)

8 - loan: has personal loan? (binary: “yes”,“no”)

related with the last contact of the current campaign:

9 - contact: contact communication type (categorical: “unknown”,“telephone”,“cellular”)

10 - day: last contact day of the month (numeric)

11 - month: last contact month of year (categorical: “jan”, “feb”, “mar”, …, “nov”, “dec”)

12 - duration: last contact duration, in seconds (numeric)

other attributes:

13 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

14 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)

15 - previous: number of contacts performed before this campaign and for this client (numeric)

16 - poutcome: outcome of the previous marketing campaign (categorical: “unknown”,“other”,“failure”,“success”)

Output variable (desired target):

17 - y - has the client subscribed a term deposit? (binary: “yes”,“no”)

3. Análise exploratória dos dados

3.1 Importação dos dados

```
library(readr)

dados <- read.csv('bank-full.csv', sep = ';')
```

3.2 Transformação e visualização dos dados

Transformação da variável day em formato categórico.

```
dados$day <- as.factor(dados$day)
str(dados)

## 'data.frame':    45211 obs. of  17 variables:
##  $ age      : int  58 44 33 47 33 35 28 42 58 43 ...
##  $ job      : Factor w/ 12 levels "admin.","blue-collar",..: 5 10 3 2 12 5 5 3
6 10 ...
##  $ marital  : Factor w/ 3 levels "divorced","married",..: 2 3 2 2 3 2 3 1 2 3 .
..
##  $ education: Factor w/ 4 levels "primary","secondary",..: 3 2 2 4 4 3 3 3 1 2
...
##  $ default  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 2 1 1 ...
##  $ balance  : int  2143 29 2 1506 1 231 447 2 121 593 ...
##  $ housing  : Factor w/ 2 levels "no","yes": 2 2 2 2 1 2 2 2 2 2 ...
##  $ loan     : Factor w/ 2 levels "no","yes": 1 1 2 1 1 1 2 1 1 1 ...
##  $ contact  : Factor w/ 3 levels "cellular","telephone",..: 3 3 3 3 3 3 3 3 3 3
...
##  $ day      : Factor w/ 31 levels "1","2","3","4",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ month    : Factor w/ 12 levels "apr","aug","dec",..: 9 9 9 9 9 9 9 9 9 9 ...
##  $ duration : int  261 151 76 92 198 139 217 380 50 55 ...
##  $ campaign : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ pdays    : int  -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...
##  $ previous : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ poutcome : Factor w/ 4 levels "failure","other",..: 4 4 4 4 4 4 4 4 4 4 ...
##  $ y        : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...

summary(dados)

##       age                   job           marital          education
##  Min.   :18.00   blue-collar:9732   divorced: 5207   primary  : 6851
##  1st Qu.:33.00   management :9458   married :27214   secondary:23202
##  Median :39.00   technician :7597   single  :12790   tertiary :13301
##  Mean   :40.94   admin.     :5171                    unknown  : 1857
##  3rd Qu.:48.00   services   :4154
##  Max.   :95.00   retired    :2264
##                  (Other)    :6835
##  default        balance       housing      loan           contact
##  no :44396   Min.   : -8019   no :20081   no :37967   cellular :29285
```

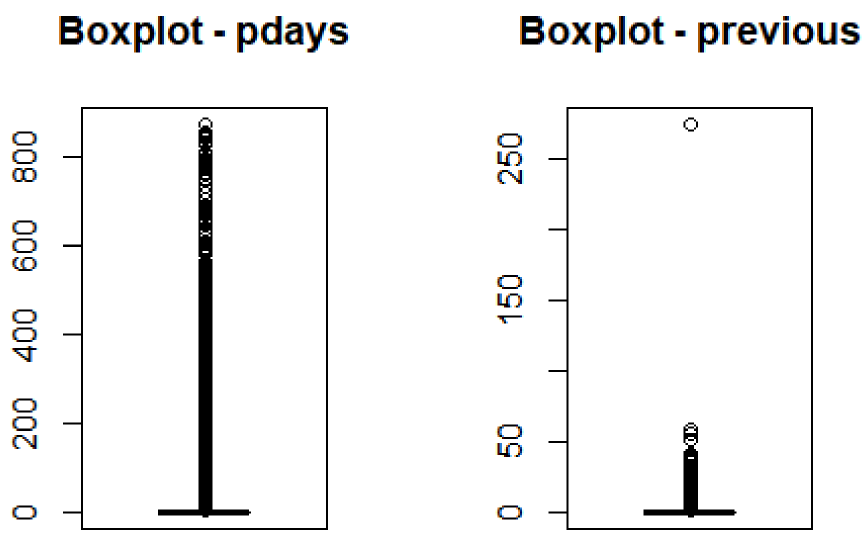
```
##  yes:  815   1st Qu.:    72   yes:25130   yes: 7244   telephone: 2906
##              Median :   448                           unknown  :13020
##              Mean   :  1362
##              3rd Qu.:  1428
##              Max.   :102127
##
##       day            month          duration         campaign
##  20     : 2752   may    :13766   Min.   :   0.0   Min.   : 1.000
##  18     : 2308   jul    : 6895   1st Qu.: 103.0   1st Qu.: 1.000
##  21     : 2026   aug    : 6247   Median : 180.0   Median : 2.000
##  17     : 1939   jun    : 5341   Mean   : 258.2   Mean   : 2.764
##  6      : 1932   nov    : 3970   3rd Qu.: 319.0   3rd Qu.: 3.000
##  5      : 1910   apr    : 2932   Max.   :4918.0   Max.   :63.000
##  (Other):32344   (Other): 6060
##      pdays          previous          poutcome       y
##  Min.   : -1.0   Min.   :  0.0000   failure: 4901   no :39922
##  1st Qu.: -1.0   1st Qu.:  0.0000   other  : 1840   yes: 5289
##  Median : -1.0   Median :  0.0000   success: 1511
##  Mean   : 40.2   Mean   :  0.5803   unknown:36959
##  3rd Qu.: -1.0   3rd Qu.:  0.0000
##  Max.   :871.0   Max.   :275.0000
##
```

3.3 Divisão e nomeação dos dados

Houve a separação dos dados entre clientes que participaram e não da campanha de marketing anterior, pois os dados originais possuem grande quantidade de valores iguais a zero e -1 respectivamente às variáveis previous e pdays, o que acarretaria, caso os dados fossem mantidos na forma original, eles influenciariam e encobertariam as análises estatísticas dos clientes que participaram ou não da campanha anterior (ver gráfico logo abaixo), gerando grande quantidade de valores classificados como outliers.

Para os clientes que não participaram da campanha de marketing anterior, denominou-se esta subdivisão dos dados originais como: dados.notprev. Já para os clientes que participaram da campanha anterior, esta outra subdivisão foi denominada como: dados.prev.

```r
par(mfrow=c(1,2))
boxplot(dados$pdays, main = 'Boxplot - pdays', col = 'darkgreen')
boxplot(dados$previous, main = 'Boxplot - previous', col = 'darkred')
```



3.4 dados.notprev

Selecionando os dados dos clientes que não participaram da campanha anterior.

```
library(dplyr)

dados.notprev <- dados %>%
  filter(pdays == -1)

dados.notprev$previous <- NULL
dados.notprev$pdays <- NULL
dados.notprev$poutcome <- NULL

str(dados.notprev)

## 'data.frame':    36954 obs. of  14 variables:
##  $ age      : int  58 44 33 47 33 35 28 42 58 43 ...
##  $ job      : Factor w/ 12 levels "admin.","blue-collar",..: 5 10 3 2 12 5 5 3
6 10 ...
##  $ marital  : Factor w/ 3 levels "divorced","married",..: 2 3 2 2 3 2 3 1 2 3 .
..
##  $ education: Factor w/ 4 levels "primary","secondary",..: 3 2 2 4 4 3 3 3 1 2
...
##  $ default  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 2 1 1 ...
##  $ balance  : int  2143 29 2 1506 1 231 447 2 121 593 ...
```

```
##  $ housing  : Factor w/ 2 levels "no","yes": 2 2 2 2 1 2 2 2 2 2 ...
##  $ loan     : Factor w/ 2 levels "no","yes": 1 1 2 1 1 1 2 1 1 1 ...
##  $ contact  : Factor w/ 3 levels "cellular","telephone",..: 3 3 3 3 3 3 3 3 3 3
...
##  $ day      : Factor w/ 31 levels "1","2","3","4",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ month    : Factor w/ 12 levels "apr","aug","dec",..: 9 9 9 9 9 9 9 9 9 9 ...
##  $ duration : int  261 151 76 92 198 139 217 380 50 55 ...
##  $ campaign : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ y        : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

3.5 dados.prev

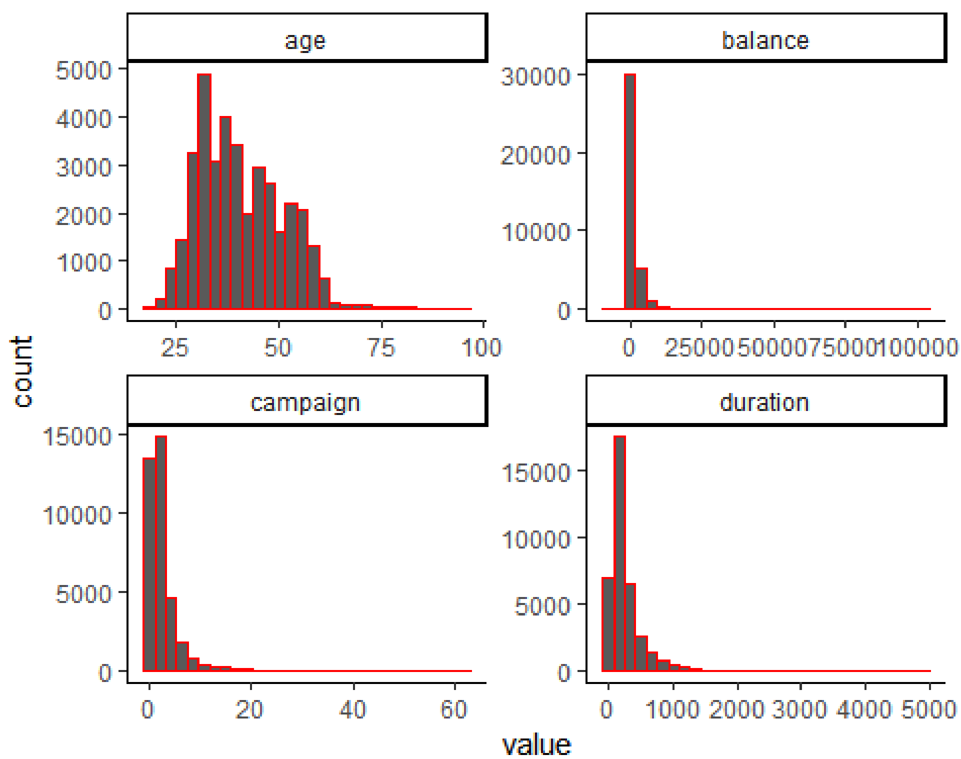
Selecionando os dados dos clientes que participaram da campanha anterior.

```
dados.prev <- anti_join(dados,dados.notprev)

## Joining, by = c("age", "job", "marital", "education", "default", "balance", "ho
using", "loan", "contact", "day", "month", "duration", "campaign", "y")

str(dados.prev)

## 'data.frame':    8257 obs. of  17 variables:
##  $ age      : int  33 42 33 36 36 56 44 26 51 34 ...
##  $ job      : Factor w/ 12 levels "admin.","blue-collar",..: 1 1 8 5 5 10 2 10
1 5 ...
##  $ marital  : Factor w/ 3 levels "divorced","married",..: 2 3 2 2 2 2 2 3 3 2 .
..
##  $ education: Factor w/ 4 levels "primary","secondary",..: 3 2 2 3 3 2 2 3 2 3
...
##  $ default  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ balance  : int  882 -247 3444 2415 0 589 1324 172 3132 1770 ...
##  $ housing  : Factor w/ 2 levels "no","yes": 1 2 2 2 2 2 2 1 1 2 ...
##  $ loan     : Factor w/ 2 levels "no","yes": 1 2 1 1 1 1 1 2 1 1 ...
##  $ contact  : Factor w/ 3 levels "cellular","telephone",..: 2 2 2 2 2 3 2 2 2 3
...
##  $ day      : Factor w/ 31 levels "1","2","3","4",..: 21 21 21 22 23 23 25 4 5
6 ...
##  $ month    : Factor w/ 12 levels "apr","aug","dec",..: 11 11 11 11 11 11 11 10
10 10 ...
##  $ duration : int  39 519 144 73 140 518 119 21 449 26 ...
##  $ campaign : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ pdays    : int  151 166 91 86 143 147 89 140 176 101 ...
##  $ previous : int  3 1 4 4 3 2 2 4 1 11 ...
##  $ poutcome : Factor w/ 4 levels "failure","other",..: 1 2 1 2 1 3 2 2 1 2 ...
##  $ y        : Factor w/ 2 levels "no","yes": 1 2 2 1 2 2 1 1 1 1 ...
```

3.6 Distribuição dos dados - dados.notprev

Visualizações das distribuições dos dados dos clientes que não participaram da campanha anterior. Observa-se um forte comportamento assimétrico dos dados com características right skewed (positive skewness).

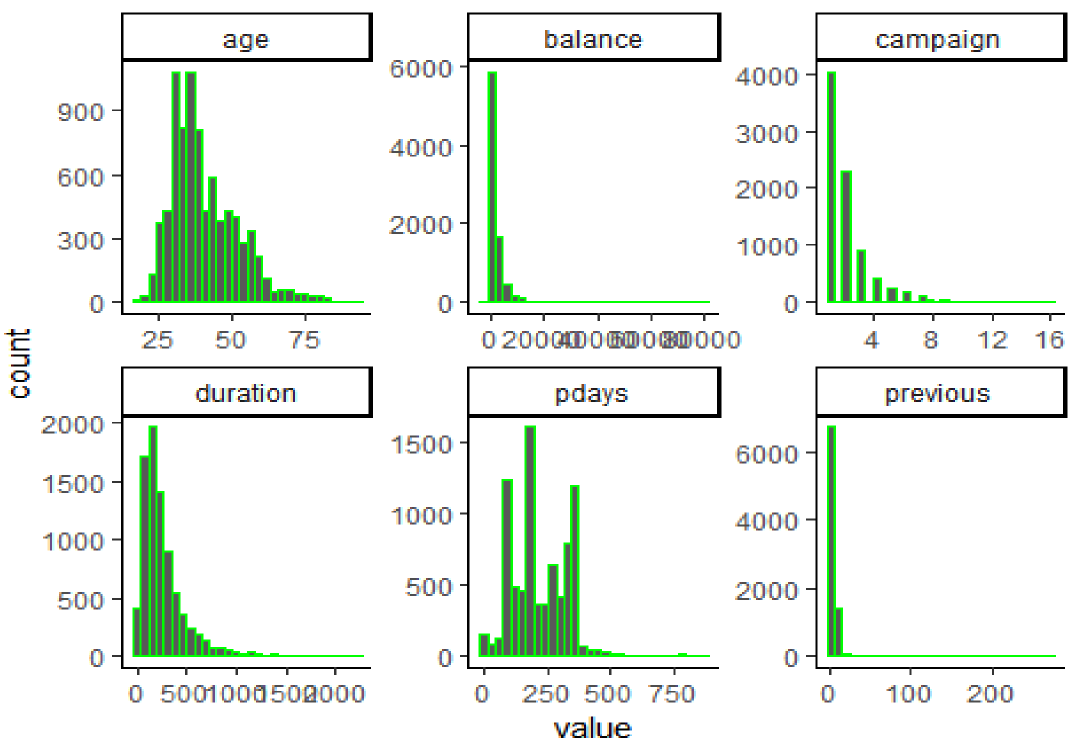
```
library(tidyr)

## Warning: package 'tidyr' was built under R version 3.4.4

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.4

dados.notprev %>%
   select("age", "balance", "duration", "campaign") %>%
   gather() %>%
   ggplot(aes(value)) +
   facet_wrap(~ key, scales = "free") +
   geom_histogram(color = 'red')  +
   theme_classic()
```



3.7 Distribuição dos dados - dados.prev

Visualizações das distribuições dos dados dos clientes que participaram da campanha anterior. Observa-se um forte comportamento assimétrico dos dados com características right skewed (positive skewness).

```
library(purrr)

library(tidyr)
library(ggplot2)

dados.prev %>%
  keep(is.numeric) %>%
  gather() %>%
  ggplot(aes(value)) +
   facet_wrap(~ key, scales = "free") +
   geom_histogram(color = 'green')  +
   theme_classic()
```



3.8 Identificação e remoção dos outliers - dados.notprev

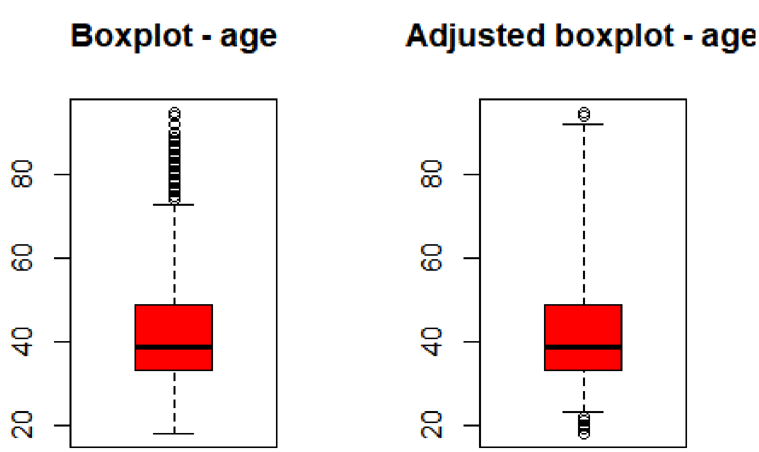
Conforme observado nas duas séries de gráficos anteriores, tanto para esta seção quanto para a próxima (3.8), utilizara-se da técnica Adjusted boxplot for skewed distributions* desenvolvida para criar boxplots que se adaptam à premissa de que os dados das variáveis contínuas apresentam comportamentos assimétricos. Com o uso desta técnica, a partir da identificação do intervalo entre os valores mínimo e máximo, haverá a remoção de todos os dados que estejam fora deste intervalo, sendo estes definidos como outlier. A aplicação ocorrerá tanto para os dados.notprev quanto para os dados.notprev.

* Hubert, M., Vandervieren, E. An Adjusted Boxplot for Skewed Distributions,

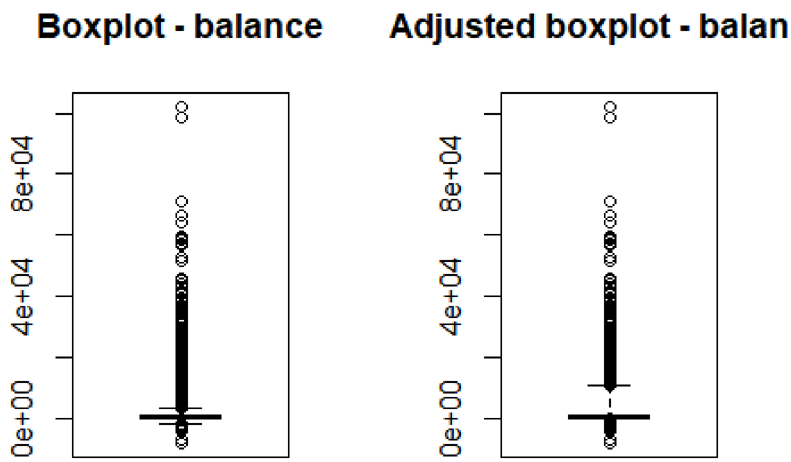
3.8.1 Visualização dos dados

Boxplots sem e com aplicação da técnica adjusted boxplot.

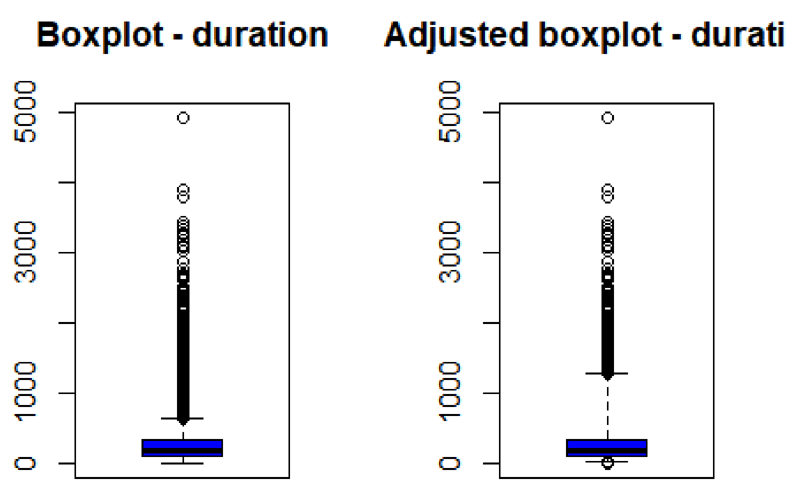
```
library(robustbase)

var_cont_notprev <- c("age", "balance", "duration", "campaign")


par(mfrow=c(1,2))
boxplot(dados.notprev$age, main = 'Boxplot - age', col = 'red')

adjbox(dados.notprev[, var_cont_notprev[c(1)]], col = 'red', main = 'Adjusted boxp
lot - age')
```



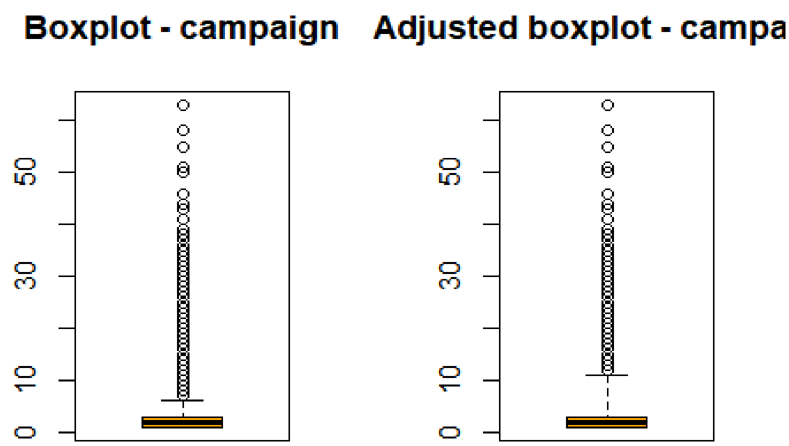
```
par(mfrow=c(1,2))
boxplot(dados.notprev$balance, main = 'Boxplot - balance', col = 'green')
adjbox(dados.notprev[, var_cont_notprev[c(2)]], col = 'green', main = 'Adjusted bo
xplot - balance')
```




```
par(mfrow=c(1,2))
boxplot(dados.notprev$duration, main = 'Boxplot - duration', col = 'blue')
adjbox(dados.notprev[, var_cont_notprev[c(3)]], col = 'blue', main = 'Adjusted box
plot - duration')
```



```r
par(mfrow=c(1,2))
boxplot(dados.notprev$campaign, main = 'Boxplot - campaign', col = 'orange')
adjbox(dados.notprev[, var_cont_notprev[c(4)]], col = 'orange', main = 'Adjusted b
oxplot - campaign')
```



3.8.2 Remoção dos outliers

```r
filtering1 <- adjboxStats(dados.notprev$age)
filtering2 <- adjboxStats(dados.notprev$balance)
filtering3 <- adjboxStats(dados.notprev$campaign)
filtering4 <- adjboxStats(dados.notprev$duration)

dados.notprev <- dados.notprev %>%
   filter((age > 23 & age < 92) & (balance > -182 & balance < 10846) &
         (campaign > 0.21 & campaign < 11.15) & (duration > 24 & duration < 1269))

str(dados.notprev)

## 'data.frame':    31894 obs. of  14 variables:
##  $ age      : int  58 44 33 47 33 35 28 42 58 43 ...
##  $ job      : Factor w/ 12 levels "admin.","blue-collar",..: 5 10 3 2 12 5 5 3
6 10 ...
##  $ marital  : Factor w/ 3 levels "divorced","married",..: 2 3 2 2 3 2 3 1 2 3 .
..
##  $ education: Factor w/ 4 levels "primary","secondary",..: 3 2 2 4 4 3 3 3 1 2
...
##  $ default  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 2 1 1 ...
##  $ balance  : int  2143 29 2 1506 1 231 447 2 121 593 ...
##  $ housing  : Factor w/ 2 levels "no","yes": 2 2 2 2 1 2 2 2 2 2 ...
##  $ loan     : Factor w/ 2 levels "no","yes": 1 1 2 1 1 1 2 1 1 1 ...
##  $ contact  : Factor w/ 3 levels "cellular","telephone",..: 3 3 3 3 3 3 3 3 3 3
```

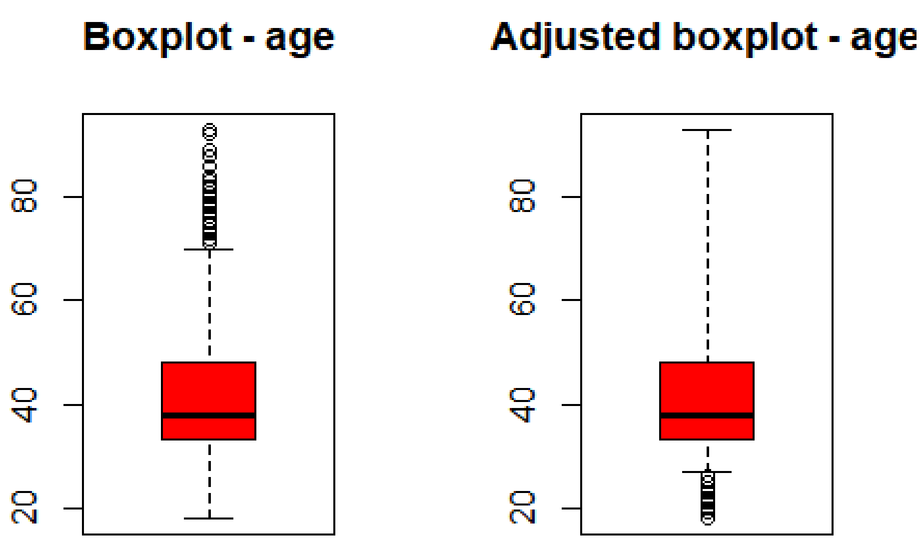
```
...
##  $ day      : Factor w/ 31 levels "1","2","3","4",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ month    : Factor w/ 12 levels "apr","aug","dec",..: 9 9 9 9 9 9 9 9 9 9 ...
##  $ duration : int  261 151 76 92 198 139 217 380 50 55 ...
##  $ campaign : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ y        : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

3.9 Identificação e remoção dos outliers - dados.prev

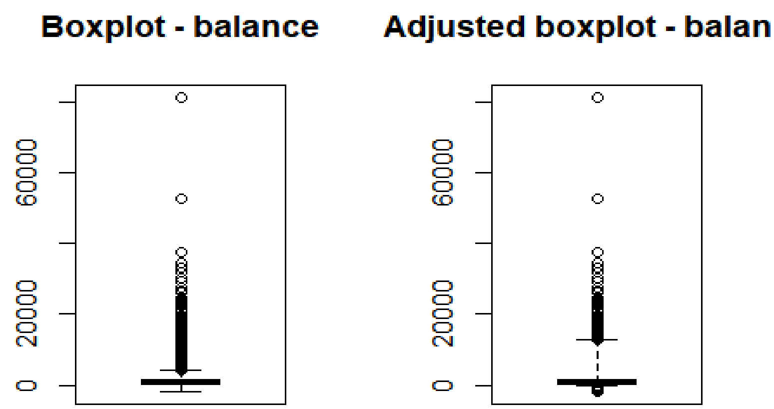
3.9.1 Visualização dos dados

Boxplots sem e com aplicação da técnica adjusted boxplot.

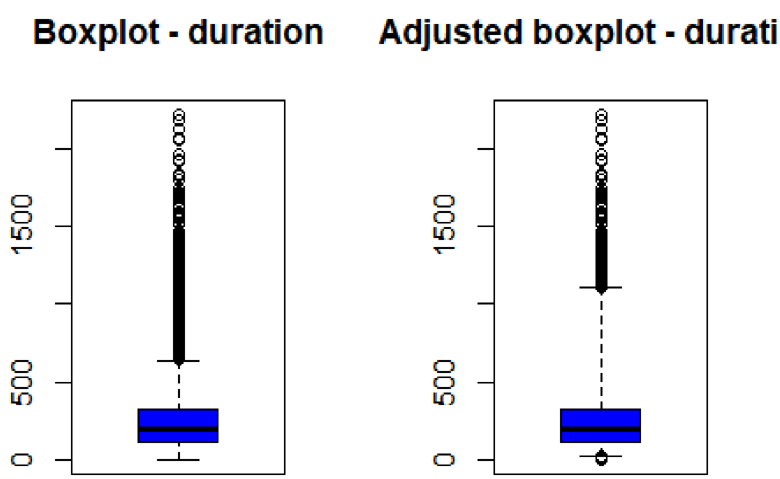
```
library(robustbase)

var_cont_prev <- c("age", "balance", "duration", "campaign", "pdays", "previous")

par(mfrow=c(1,2))
boxplot(dados.prev$age, main = 'Boxplot - age', col = 'red')
adjbox(dados.prev[, var_cont_prev[c(1)]], col = 'red', main = 'Adjusted boxplot -
age')
```



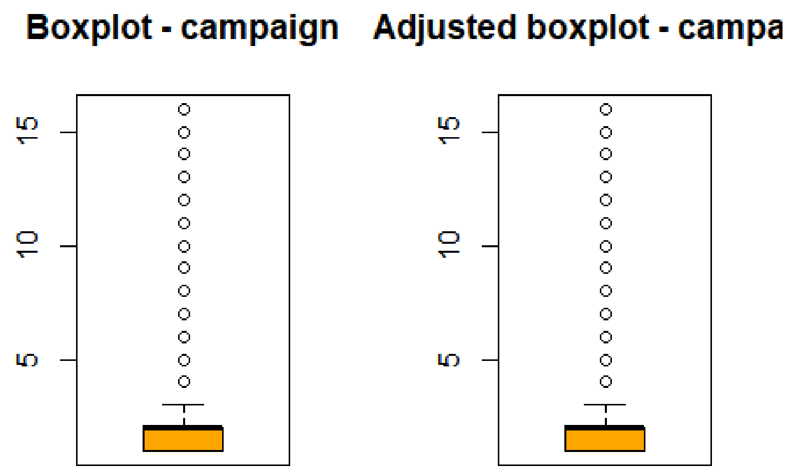
```r
par(mfrow=c(1,2))
boxplot(dados.prev$balance, main = 'Boxplot - balance', col = 'green')
adjbox(dados.prev[, var_cont_prev[c(2)]], col = 'green', main = 'Adjusted boxplot
- balance')
```



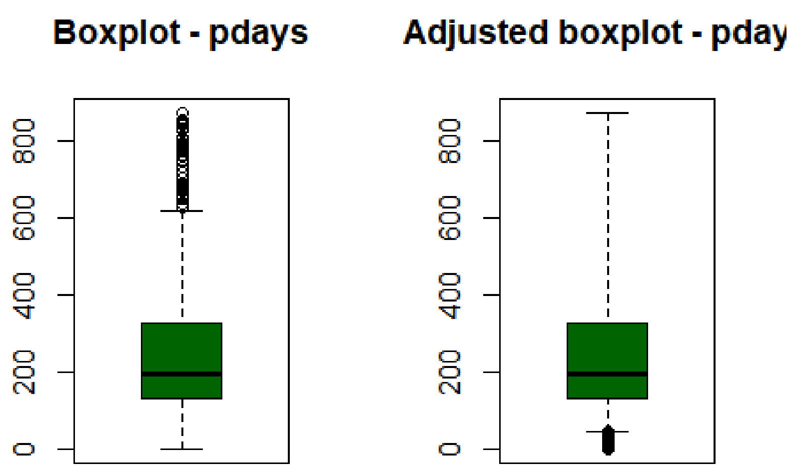

```r
par(mfrow=c(1,2))
boxplot(dados.prev$duration, main = 'Boxplot - duration', col = 'blue')
adjbox(dados.prev[, var_cont_prev[c(3)]], col = 'blue', main = 'Adjusted boxplot -
duration')
```



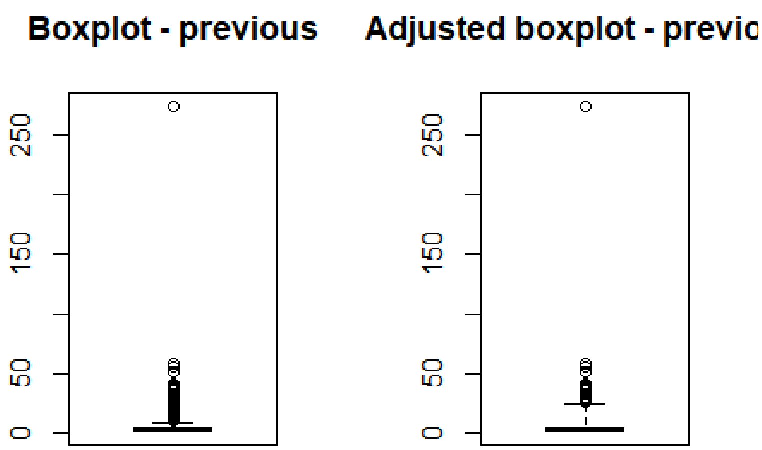
```
par(mfrow=c(1,2))
boxplot(dados.prev$campaign, main = 'Boxplot - campaign', col = 'orange')
adjbox(dados.prev[, var_cont_prev[c(4)]], col = 'orange', main = 'Adjusted boxplot
- campaign')
```




```
par(mfrow=c(1,2))
boxplot(dados.prev$pdays, main = 'Boxplot - pdays', col = 'darkgreen')
adjbox(dados.prev[, var_cont_prev[c(5)]], col = 'darkgreen', main = 'Adjusted boxp
lot - pdays')
```



```
par(mfrow=c(1,2))
boxplot(dados.prev$previous, main = 'Boxplot - previous', col = 'darkred')
adjbox(dados.prev[, var_cont_prev[c(6)]], col = 'darkred', main = 'Adjusted boxplo
t - previous')
```



3.9.2 Remoção dos outliers

```
filtering5 <- adjboxStats(dados.prev$age)
filtering6 <- adjboxStats(dados.prev$balance)
filtering7 <- adjboxStats(dados.prev$campaign)
filtering8 <- adjboxStats(dados.prev$duration)
filtering9 <- adjboxStats(dados.prev$pdays)
filtering10 <- adjboxStats(dados.prev$previous)

dados.prev <- dados.prev %>%
   filter((age > 26 & age < 102) & (balance > -135 & balance < 12748) &
         (campaign > 0.01 & campaign < 3.5) & (duration > 19 & duration < 1106) &
         (pdays > 48 & pdays < 1054) & (previous > 0.4 & previous < 24))

str(dados.prev)

## 'data.frame':    6163 obs. of  17 variables:
##  $ age      : int  33 33 36 36 56 44 51 34 33 34 ...
##  $ job      : Factor w/ 12 levels "admin.","blue-collar",..: 1 8 5 5 10 2 1 5 1
1 1 ...
##  $ marital  : Factor w/ 3 levels "divorced","married",..: 2 2 2 2 2 2 3 2 1 2 .
..
##  $ education: Factor w/ 4 levels "primary","secondary",..: 3 2 3 3 2 2 2 3 2 3
...
##  $ default  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ balance  : int  882 3444 2415 0 589 1324 3132 1770 1005 899 ...
```

```
##  $ housing  : Factor w/ 2 levels "no","yes": 1 2 2 2 2 2 1 2 2 2 ...
##  $ loan     : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact  : Factor w/ 3 levels "cellular","telephone",..: 2 2 2 2 3 2 2 3 2 3
...
##  $ day      : Factor w/ 31 levels "1","2","3","4",..: 21 21 22 23 23 25 5 6 10
12 ...
##  $ month    : Factor w/ 12 levels "apr","aug","dec",..: 11 11 11 11 11 11 10 10
10 10 ...
##  $ duration : int  39 144 73 140 518 119 449 26 175 114 ...
##  $ campaign : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ pdays    : int  151 91 86 143 147 89 176 101 174 170 ...
##  $ previous : int  3 4 4 3 2 2 1 11 2 3 ...
##  $ poutcome : Factor w/ 4 levels "failure","other",..: 1 1 2 1 3 2 1 2 1 1 ...
##  $ y        : Factor w/ 2 levels "no","yes": 1 2 1 2 2 1 1 1 1 2 ...
```

4. Questões para análise

As questões formuladas inicialmente na análise de negócios serão tratadas sequencialmente, com as elaborações das devidas técnicas para a resolução de cada questão.

4.1 Qual profissão tem mais tendência a fazer um empréstimo? De qual tipo?

4.1.1 Lista de profissões que mais efetuam empréstimo (loan) - dados.notprev

Clientes que não participaram da campanha de marketing anterior

```
job_loan_notprev <- dados.notprev %>%
   select(job, loan) %>%
   group_by(job) %>%
   filter(loan == 'yes') %>%
   count() %>%
   arrange(desc(n))

job_loan_notprev

## # A tibble: 12 x 2
## # Groups:   job [12]
##    job               n
##    <fct>         <int>
##  1 blue-collar    1145
##  2 management      891
##  3 technician      875
##  4 admin.          663
##  5 services        591
##  6 entrepreneur    252
```

```
##  7 retired         242
##  8 self-employed   162
##  9 housemaid       118
## 10 unemployed       81
## 11 student           7
## 12 unknown           2
```

4.1.2 Lista de profissões que mais efetuam empréstimo (loan) - dados.prev

Clientes que participaram da campanha de marketing anterior

```
job_loan_prev <- dados.prev %>%
   select(job, loan) %>%
   group_by(job) %>%
   filter(loan == 'yes') %>%
   count() %>%
   arrange(desc(n))

job_loan_prev
```

```
## # A tibble: 11 x 2
## # Groups:   job [11]
##    job               n
##    <fct>         <int>
##  1 blue-collar     176
##  2 technician      162
##  3 management      149
##  4 admin.          135
##  5 services         79
##  6 entrepreneur     38
##  7 retired          25
##  8 self-employed    22
##  9 unemployed       14
## 10 housemaid        10
## 11 student           1
```

4.1.3 Lista de profissões que mais efetuam empréstimo habitacional (housing) - dados.notprev

Clientes que não participaram da campanha de marketing anterior

```
job_housing_notprev <- dados.notprev %>%
   select(job, housing) %>%
   group_by(job) %>%
   filter(housing == 'yes') %>%
   count() %>%
   arrange(desc(n))
```

```
job_housing_notprev
```

```
## # A tibble: 12 x 2
## # Groups:   job [12]
##    job               n
##    <fct>         <int>
##  1 blue-collar    4809
##  2 management     3130
##  3 technician     2713
##  4 admin.         2102
##  5 services       1858
##  6 entrepreneur    622
##  7 self-employed   514
##  8 unemployed      388
##  9 retired         366
## 10 housemaid       294
## 11 student         135
## 12 unknown          16
```

4.1.4 Lista de profissões que mais efetuam empréstimo habitacional (housing) - dados.prev

Clientes que participaram da campanha de marketing anterior

```
job_housing_prev <- dados.prev %>%
   select(job, housing) %>%
   group_by(job) %>%
   filter(housing == 'yes') %>%
   count() %>%
   arrange(desc(n))

job_housing_prev
```

```
## # A tibble: 12 x 2
## # Groups:   job [12]
##    job               n
##    <fct>         <int>
##  1 blue-collar     976
##  2 management      765
##  3 technician      664
##  4 admin.          554
##  5 services        390
##  6 self-employed   126
##  7 entrepreneur    125
##  8 unemployed       73
##  9 retired          61
## 10 housemaid        48
## 11 student          25
## 12 unknown           4
```

4.2 Fazendo uma relação entre número de contatos e sucesso da campanha quais são os pontos relevantes a serem observados?

4.2.1 Cálculo para a taxa de sucesso - dados.prev

Taxa de sucesso da razão entre os clientes que efetuaram o depósito (y = yes) e o número total de contatos da campanha.

Clientes que participaram da campanha de marketing anterior.

```
success_prev <- dados.prev %>%
   filter(y == 'yes') %>%
   summarise(n = n())
success_prev

##      n
## 1 1505

campaign_prev <- dados.prev %>%
   select(campaign) %>%
   summarise_all(funs(sum))

## Warning: funs() is soft deprecated as of dplyr 0.8.0
## please use list() instead
##
## # Before:
## funs(name = f(.)
##
## # After:
## list(name = ~f(.))
## This warning is displayed once per session.

campaign_prev

##   campaign
## 1     9617

rate_success_prev <- round(success_prev/campaign_prev, 2)
rate_success_prev

##      n
## 1 0.16
```

Taxa de sucesso = 0.16

Construção do modelo de aprendizado de máquina para verificação do nível de significância da variável campaign.

Neste estudo conforme mencionado na introdução, o foco será dado na simplificação e nas questões da análise de negócio. Os modelos de aprendizado de máquina que forem criados, eles se concentrarão na determinação e interpretação do nível de significância das variáveis dos modelos. Não serão adotadas as tradicionais e importantes técnicas de subdivisão dos dados em treino e teste, bem como na construção e maximização no nível de acurácia para cada modelo construído. Esta estratégia será aplicada neste modelo e nos demais que vierem a ser elaborados.

```
model.glm <- glm(y ~., data = dados.prev, family = 'binomial')
summary(model.glm)

##
## Call:
## glm(formula = y ~ ., family = "binomial", data = dados.prev)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.9140  -0.4809  -0.2711  -0.1241   2.7003
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -4.974e+00  4.853e-01 -10.249  < 2e-16 ***
## age                 6.249e-03  4.907e-03   1.273 0.202843
## jobblue-collar     -2.073e-01  1.599e-01  -1.297 0.194743
## jobentrepreneur    -6.801e-01  3.104e-01  -2.191 0.028442 *
## jobhousemaid       -3.740e-01  3.003e-01  -1.245 0.212955
## jobmanagement       2.526e-02  1.510e-01   0.167 0.867136
## jobretired         -1.384e-01  2.114e-01  -0.654 0.512831
## jobself-employed   -3.001e-01  2.422e-01  -1.239 0.215228
## jobservices        -6.920e-02  1.863e-01  -0.371 0.710366
## jobstudent          2.500e-01  2.748e-01   0.910 0.363023
## jobtechnician      -1.646e-01  1.440e-01  -1.143 0.252989
## jobunemployed       1.451e-01  2.428e-01   0.598 0.550154
## jobunknown          1.059e-02  4.895e-01   0.022 0.982741
## maritalmarried      1.635e-01  1.340e-01   1.221 0.222170
## maritalsingle       1.831e-01  1.534e-01   1.194 0.232610
## educationsecondary  3.918e-01  1.545e-01   2.535 0.011244 *
## educationtertiary   5.625e-01  1.764e-01   3.188 0.001431 **
## educationunknown    3.489e-01  2.425e-01   1.439 0.150127
## defaultyes          2.974e-01  6.384e-01   0.466 0.641291
## balance             7.863e-06  1.943e-05   0.405 0.685740
## housingyes         -6.829e-01  9.273e-02  -7.364 1.78e-13 ***
## loanyes            -5.927e-01  1.473e-01  -4.025 5.70e-05 ***
## contacttelephone   -3.229e-01  1.643e-01  -1.965 0.049399 *
## contactunknown      7.345e-01  4.129e-01   1.779 0.075233 .
## day2                4.520e-01  3.533e-01   1.279 0.200775
```

```
## day3                 1.228e+00  3.647e-01   3.368 0.000758 ***
## day4                 1.216e+00  3.488e-01   3.487 0.000488 ***
## day5                 7.723e-01  3.668e-01   2.105 0.035261 *
## day6                 8.704e-01  3.679e-01   2.366 0.017982 *
## day7                 2.865e-01  3.789e-01   0.756 0.449533
## day8                 6.166e-01  3.629e-01   1.699 0.089296 .
## day9                 1.442e+00  3.748e-01   3.847 0.000119 ***
## day10                1.777e+00  4.114e-01   4.320 1.56e-05 ***
## day11                1.008e+00  3.593e-01   2.804 0.005049 **
## day12                1.184e+00  3.463e-01   3.418 0.000631 ***
## day13                1.175e+00  3.487e-01   3.368 0.000756 ***
## day14                1.081e+00  3.618e-01   2.987 0.002817 **
## day15                1.228e+00  3.503e-01   3.506 0.000455 ***
## day16                9.389e-01  3.651e-01   2.571 0.010132 *
## day17                5.490e-02  3.725e-01   0.147 0.882819
## day18                4.456e-01  3.705e-01   1.203 0.229072
## day19                4.121e-01  4.036e-01   1.021 0.307262
## day20                2.695e-01  3.859e-01   0.698 0.484959
## day21                9.029e-01  3.878e-01   2.328 0.019909 *
## day22                1.780e+00  3.924e-01   4.537 5.71e-06 ***
## day23                1.910e+00  4.450e-01   4.292 1.77e-05 ***
## day24                9.741e-01  4.788e-01   2.035 0.041898 *
## day25                1.720e+00  3.959e-01   4.344 1.40e-05 ***
## day26                1.015e+00  4.004e-01   2.536 0.011209 *
## day27                2.123e+00  3.975e-01   5.342 9.18e-08 ***
## day28                1.345e+00  4.087e-01   3.291 0.000998 ***
## day29                9.846e-01  3.911e-01   2.518 0.011804 *
## day30                1.411e+00  3.723e-01   3.790 0.000151 ***
## day31                1.474e+00  5.583e-01   2.640 0.008283 **
## monthaug             1.031e+00  1.912e-01   5.392 6.95e-08 ***
## monthdec             1.337e+00  3.141e-01   4.257 2.07e-05 ***
## monthfeb             4.143e-01  1.941e-01   2.135 0.032743 *
## monthjan            -5.082e-01  2.479e-01  -2.050 0.040330 *
## monthjul             1.370e+00  2.279e-01   6.012 1.83e-09 ***
## monthjun             1.246e+00  2.182e-01   5.711 1.12e-08 ***
## monthmar             1.365e+00  2.468e-01   5.531 3.18e-08 ***
## monthmay            -2.252e-01  1.677e-01  -1.343 0.179120
## monthnov             2.173e-01  1.867e-01   1.164 0.244612
## monthoct             1.123e+00  2.034e-01   5.520 3.39e-08 ***
## monthsep             1.472e+00  2.151e-01   6.842 7.82e-12 ***
## duration             4.954e-03  2.130e-04  23.258  < 2e-16 ***
## campaign            -7.165e-02  5.798e-02  -1.236 0.216548
## pdays                5.305e-04  3.711e-04   1.430 0.152845
## previous             3.746e-02  1.431e-02   2.618 0.008837 **
## poutcomeother        2.467e-01  1.069e-01   2.307 0.021037 *
## poutcomesuccess      2.018e+00  9.495e-02  21.254  < 2e-16 ***
## poutcomeunknown     -1.158e+01  1.783e+02  -0.065 0.948233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6851.7  on 6162  degrees of freedom
## Residual deviance: 4196.0  on 6091  degrees of freedom
## AIC: 4340
##
## Number of Fisher Scoring iterations: 11
```

Para os clientes que participaram da campanha anterior, a taxa de sucesso da campanha atual é de 16%, e o valor do p-value é de 0.21, o que denota que a campanha atual não possui uma significância tão elevada.

4.2.2 Cálculo para a taxa de sucesso - dados.notprev

Clientes que não participaram da campanha de marketing anterior

```
success_notprev <- dados.notprev %>%
   filter(y == 'yes') %>%
   summarise(n = n())
success_notprev

##      n
## 1 2867

campaign_notprev <- dados.notprev %>%
   select(campaign) %>%
   summarise_all(funs(sum))
campaign_notprev

##   campaign
## 1    78549

rate_success_notprev <- round(success_notprev/campaign_notprev, 2)
rate_success_notprev

##      n
## 1 0.04
```

Taxa de sucesso = 0.04

Construção do modelo de aprendizado de máquina para verificação do nível de significância da variável campaign.

```
model.glm1 <- glm(y ~., data = dados.notprev, family = 'binomial')
summary(model.glm1)
```

```
##
## Call:
## glm(formula = y ~ ., family = "binomial", data = dados.notprev)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3249  -0.3214  -0.2029  -0.1267   3.6346
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -2.049e+00  2.908e-01  -7.046 1.84e-12 ***
## age                 1.578e-03  2.911e-03   0.542 0.587644
## jobblue-collar     -2.758e-01  9.485e-02  -2.908 0.003634 **
## jobentrepreneur    -4.257e-01  1.637e-01  -2.601 0.009289 **
## jobhousemaid       -5.979e-01  1.793e-01  -3.335 0.000852 ***
## jobmanagement      -2.359e-01  9.769e-02  -2.415 0.015740 *
## jobretired          2.107e-01  1.271e-01   1.658 0.097378 .
## jobself-employed   -3.545e-01  1.475e-01  -2.403 0.016261 *
## jobservices        -2.947e-01  1.117e-01  -2.638 0.008348 **
## jobstudent          1.482e-01  1.652e-01   0.897 0.369479
## jobtechnician      -1.558e-01  9.129e-02  -1.707 0.087901 .
## jobunemployed      -3.657e-01  1.457e-01  -2.510 0.012060 *
## jobunknown         -6.348e-01  3.203e-01  -1.982 0.047452 *
## maritalmarried     -3.067e-01  7.581e-02  -4.046 5.21e-05 ***
## maritalsingle      -1.464e-02  8.668e-02  -0.169 0.865917
## educationsecondary  2.216e-01  8.362e-02   2.650 0.008049 **
## educationtertiary   4.607e-01  9.762e-02   4.720 2.36e-06 ***
## educationunknown    2.110e-01  1.383e-01   1.526 0.127093
## defaultyes          1.194e-01  2.223e-01   0.537 0.591169
## balance             5.139e-05  1.262e-05   4.071 4.68e-05 ***
## housingyes         -7.325e-01  5.887e-02 -12.443  < 2e-16 ***
## loanyes            -3.853e-01  7.865e-02  -4.899 9.66e-07 ***
## contacttelephone   -2.019e-02  9.750e-02  -0.207 0.835957
## contactunknown     -1.796e+00  9.071e-02 -19.805  < 2e-16 ***
## day2               -5.059e-01  2.442e-01  -2.072 0.038260 *
## day3               -4.117e-01  2.469e-01  -1.668 0.095400 .
## day4               -7.015e-01  2.450e-01  -2.863 0.004198 **
## day5               -6.905e-01  2.373e-01  -2.909 0.003623 **
## day6               -6.962e-01  2.448e-01  -2.844 0.004456 **
## day7               -6.296e-01  2.440e-01  -2.581 0.009860 **
## day8               -4.446e-01  2.397e-01  -1.855 0.063657 .
## day9               -6.729e-01  2.512e-01  -2.679 0.007390 **
## day10               8.798e-03  2.788e-01   0.032 0.974824
## day11              -5.768e-01  2.464e-01  -2.341 0.019248 *
## day12              -3.115e-01  2.435e-01  -1.279 0.200859
## day13              -5.660e-02  2.435e-01  -0.232 0.816205
## day14              -3.959e-01  2.426e-01  -1.632 0.102736
## day15              -4.021e-01  2.444e-01  -1.645 0.099952 .
## day16              -4.406e-01  2.444e-01  -1.803 0.071368 .
## day17              -1.150e+00  2.432e-01  -4.729 2.25e-06 ***
```

```
## day18              -4.358e-01  2.363e-01  -1.844 0.065202 .
## day19              -1.193e+00  2.592e-01  -4.602 4.18e-06 ***
## day20              -8.528e-01  2.399e-01  -3.555 0.000378 ***
## day21              -4.772e-01  2.438e-01  -1.957 0.050336 .
## day22              -3.398e-01  2.544e-01  -1.336 0.181630
## day23              -1.141e-01  2.638e-01  -0.432 0.665401
## day24              -5.820e-01  3.062e-01  -1.901 0.057324 .
## day25              -2.239e-01  2.640e-01  -0.848 0.396477
## day26              -1.029e-01  2.637e-01  -0.390 0.696279
## day27               6.871e-03  2.586e-01   0.027 0.978805
## day28              -6.146e-01  2.608e-01  -2.357 0.018432 *
## day29              -8.820e-01  2.641e-01  -3.340 0.000839 ***
## day30              -5.451e-02  2.378e-01  -0.229 0.818675
## day31              -5.724e-01  3.386e-01  -1.690 0.090967 .
## monthaug           -1.411e+00  1.115e-01 -12.662  < 2e-16 ***
## monthdec            9.746e-01  2.579e-01   3.779 0.000158 ***
## monthfeb           -4.506e-01  1.302e-01  -3.461 0.000538 ***
## monthjan           -1.615e+00  1.840e-01  -8.777  < 2e-16 ***
## monthjul           -1.534e+00  1.061e-01 -14.460  < 2e-16 ***
## monthjun            1.791e-01  1.273e-01   1.406 0.159596
## monthmar            1.633e+00  1.597e-01  10.228  < 2e-16 ***
## monthmay           -6.962e-01  1.070e-01  -6.509 7.59e-11 ***
## monthnov           -1.096e+00  1.264e-01  -8.672  < 2e-16 ***
## monthoct            9.942e-01  1.538e-01   6.465 1.01e-10 ***
## monthsep            7.974e-01  1.832e-01   4.353 1.34e-05 ***
## duration            5.365e-03  9.291e-05  57.739  < 2e-16 ***
## campaign           -8.063e-02  1.579e-02  -5.107 3.27e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 19282  on 31893  degrees of freedom
## Residual deviance: 12674  on 31827  degrees of freedom
## AIC: 12808
##
## Number of Fisher Scoring iterations: 6
```

Para os clientes que não participaram da campanha anterior, a taxa de sucesso da campanha atual é de apenas 0.04, e o p-value foi de 3.27e-07, possuindo grande relevância e influência a campanha atual, porém com resultado inverso ao desejado.

4.3 Baseando-se nos resultados de adesão desta campanha qual o número médio e o máximo de ligações que você indica para otimizar a adesão?

4.3.1 Cálculo do número médio e máximo de ligações indicado - dados.notprev

Clientes que não participaram da campanha de marketing anterior

```
success_mean_notprev <- dados.notprev %>%
   filter(y == 'yes') %>%
   summarise(total = n(), mean_notprev = mean(campaign)) %>%
   round(2)
success_mean_notprev

##   total mean_notprev
## 1  2867         2.16

success_max_notprev <- dados.notprev %>%
   filter(y == 'yes') %>%
   group_by(campaign) %>%
   summarise(n = n()) %>%
   mutate(perc_success = n/sum(n)) %>%
   round(2)
success_max_notprev

## # A tibble: 11 x 3
##    campaign     n perc_success
##       <dbl> <dbl>        <dbl>
##  1        1  1332         0.46
##  2        2   755         0.26
##  3        3   339         0.12
##  4        4   208         0.07
##  5        5    91         0.03
##  6        6    58         0.02
##  7        7    28         0.01
##  8        8    22         0.01
##  9        9    11         0
## 10       10    11         0
## 11       11    12         0
```

Número médio de ligações: 2.16.

Por apresentar maior taxa de sucesso (0.46), recomenda-se **realizar apenas uma ligação** aos clientes.

4.3.2 Cálculo do número médio e máximo de ligações indicado - dados.prev

Clientes que participaram da campanha de marketing anterior

```
success_mean_prev <- dados.prev %>%
   filter(y == 'yes') %>%
   summarise(total = n(), mean_prev = mean(campaign)) %>%
   round(2)
success_mean_prev

##   total mean_prev
## 1  1505      1.54

success_max_prev <- dados.prev %>%
   filter(y == 'yes') %>%
   group_by(campaign) %>%
   summarise(n = n()) %>%
   mutate(perc_success = n/sum(n)) %>%
   round(2)
success_max_prev

## # A tibble: 3 x 3
##   campaign     n perc_success
##      <dbl> <dbl>        <dbl>
## 1        1   881        0.59
## 2        2   442        0.290
## 3        3   182        0.12
```

Número médio de ligações: 1.54

Por apresentar maior taxa de sucesso, recomenda-se **realizar apenas uma ligação** aos clientes.

4.4 O resultado da campanha anterior tem relevância na campanha atual?

4.4.1 Estatísticas e cálculos - dados.prev

Estatísticas e cálculos apenas para os clientes que participaram da campanha de marketing anterior. Por não terem participado da campanha anterior, não existe cálculo para os clientes dos dados.notprev.

```
model.glm2 <- glm(y ~., data = dados.prev, family = 'binomial')
summary(model.glm2)

##
## Call:
## glm(formula = y ~ ., family = "binomial", data = dados.prev)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -2.9140  -0.4809  -0.2711  -0.1241   2.7003
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -4.974e+00  4.853e-01 -10.249  < 2e-16 ***
## age                 6.249e-03  4.907e-03   1.273 0.202843
## jobblue-collar     -2.073e-01  1.599e-01  -1.297 0.194743
## jobentrepreneur    -6.801e-01  3.104e-01  -2.191 0.028442 *
## jobhousemaid       -3.740e-01  3.003e-01  -1.245 0.212955
## jobmanagement       2.526e-02  1.510e-01   0.167 0.867136
## jobretired         -1.384e-01  2.114e-01  -0.654 0.512831
## jobself-employed   -3.001e-01  2.422e-01  -1.239 0.215228
## jobservices        -6.920e-02  1.863e-01  -0.371 0.710366
## jobstudent          2.500e-01  2.748e-01   0.910 0.363023
## jobtechnician      -1.646e-01  1.440e-01  -1.143 0.252989
## jobunemployed       1.451e-01  2.428e-01   0.598 0.550154
## jobunknown          1.059e-02  4.895e-01   0.022 0.982741
## maritalmarried      1.635e-01  1.340e-01   1.221 0.222170
## maritalsingle       1.831e-01  1.534e-01   1.194 0.232610
## educationsecondary  3.918e-01  1.545e-01   2.535 0.011244 *
## educationtertiary   5.625e-01  1.764e-01   3.188 0.001431 **
## educationunknown    3.489e-01  2.425e-01   1.439 0.150127
## defaultyes          2.974e-01  6.384e-01   0.466 0.641291
## balance             7.863e-06  1.943e-05   0.405 0.685740
## housingyes         -6.829e-01  9.273e-02  -7.364 1.78e-13 ***
## loanyes            -5.927e-01  1.473e-01  -4.025 5.70e-05 ***
## contacttelephone   -3.229e-01  1.643e-01  -1.965 0.049399 *
## contactunknown      7.345e-01  4.129e-01   1.779 0.075233 .
## day2                4.520e-01  3.533e-01   1.279 0.200775
## day3                1.228e+00  3.647e-01   3.368 0.000758 ***
## day4                1.216e+00  3.488e-01   3.487 0.000488 ***
## day5                7.723e-01  3.668e-01   2.105 0.035261 *
## day6                8.704e-01  3.679e-01   2.366 0.017982 *
## day7                2.865e-01  3.789e-01   0.756 0.449533
## day8                6.166e-01  3.629e-01   1.699 0.089296 .
## day9                1.442e+00  3.748e-01   3.847 0.000119 ***
## day10               1.777e+00  4.114e-01   4.320 1.56e-05 ***
## day11               1.008e+00  3.593e-01   2.804 0.005049 **
## day12               1.184e+00  3.463e-01   3.418 0.000631 ***
## day13               1.175e+00  3.487e-01   3.368 0.000756 ***
## day14               1.081e+00  3.618e-01   2.987 0.002817 **
## day15               1.228e+00  3.503e-01   3.506 0.000455 ***
## day16               9.389e-01  3.651e-01   2.571 0.010132 *
## day17               5.490e-02  3.725e-01   0.147 0.882819
## day18               4.456e-01  3.705e-01   1.203 0.229072
## day19               4.121e-01  4.036e-01   1.021 0.307262
## day20               2.695e-01  3.859e-01   0.698 0.484959
## day21               9.029e-01  3.878e-01   2.328 0.019909 *
## day22               1.780e+00  3.924e-01   4.537 5.71e-06 ***
```

```
## day23                  1.910e+00  4.450e-01   4.292 1.77e-05 ***
## day24                  9.741e-01  4.788e-01   2.035 0.041898 *
## day25                  1.720e+00  3.959e-01   4.344 1.40e-05 ***
## day26                  1.015e+00  4.004e-01   2.536 0.011209 *
## day27                  2.123e+00  3.975e-01   5.342 9.18e-08 ***
## day28                  1.345e+00  4.087e-01   3.291 0.000998 ***
## day29                  9.846e-01  3.911e-01   2.518 0.011804 *
## day30                  1.411e+00  3.723e-01   3.790 0.000151 ***
## day31                  1.474e+00  5.583e-01   2.640 0.008283 **
## monthaug               1.031e+00  1.912e-01   5.392 6.95e-08 ***
## monthdec               1.337e+00  3.141e-01   4.257 2.07e-05 ***
## monthfeb               4.143e-01  1.941e-01   2.135 0.032743 *
## monthjan              -5.082e-01  2.479e-01  -2.050 0.040330 *
## monthjul               1.370e+00  2.279e-01   6.012 1.83e-09 ***
## monthjun               1.246e+00  2.182e-01   5.711 1.12e-08 ***
## monthmar               1.365e+00  2.468e-01   5.531 3.18e-08 ***
## monthmay              -2.252e-01  1.677e-01  -1.343 0.179120
## monthnov               2.173e-01  1.867e-01   1.164 0.244612
## monthoct               1.123e+00  2.034e-01   5.520 3.39e-08 ***
## monthsep               1.472e+00  2.151e-01   6.842 7.82e-12 ***
## duration               4.954e-03  2.130e-04  23.258  < 2e-16 ***
## campaign              -7.165e-02  5.798e-02  -1.236 0.216548
## pdays                  5.305e-04  3.711e-04   1.430 0.152845
## previous               3.746e-02  1.431e-02   2.618 0.008837 **
## poutcomeother          2.467e-01  1.069e-01   2.307 0.021037 *
## poutcomesuccess        2.018e+00  9.495e-02  21.254  < 2e-16 ***
## poutcomeunknown       -1.158e+01  1.783e+02  -0.065 0.948233
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6851.7  on 6162  degrees of freedom
## Residual deviance: 4196.0  on 6091  degrees of freedom
## AIC: 4340
##
## Number of Fisher Scoring iterations: 11
```

Cálculos das taxas de sucesso para cada subcategoria de poutcome (failure, other e success)

```
# failure
df1 <- dados.prev %>%
   filter(poutcome == 'failure') %>%
   summarise(n = n())

result_df1 <- dados.prev %>%
   filter(y == 'yes' & poutcome == 'failure') %>%
   summarise(n = n())
```

```
rate1 <- result_df1/df1
rate1

##           n
## 1 0.1279958



# other
df2 <- dados.prev %>%
   filter(poutcome == 'other') %>%
   summarise(n = n())

result_df2 <- dados.prev %>%
   filter(y == 'yes' & poutcome == 'other') %>%
   summarise(n = n())

rate2 <- result_df2/df2
rate2

##           n
## 1 0.1874463



# success
df3 <- dados.prev %>%
   filter(poutcome == 'success') %>%
   summarise(n = n())

result_df3 <- dados.prev %>%
   filter(y == 'yes' & poutcome == 'success') %>%
   summarise(n = n())

rate3 <- result_df3/df3
rate3

##        n
## 1 0.6675
```

Através da análise estatística do modelo preditivo criado, a variável poutcome demonstra possuir um nível de significância relevante, vistos nos p-values, e pelos cálculos de taxas de sucesso, ou seja, quando o resultado da campanha anterior foi falho (failure), a probabilidade de sucesso da campanha atual é baixo (0.12). Por outro lado, quando o resultado da campanha anterior obteve sucesso (success), a probabilidade de haver sucesso na campanha atual aumenta consideravelmente (0.66).

4.5 Qual o fator determinante para que o banco exija um seguro de crédito?

4.5.1 dados.prev

Clientes que participaram da campanha de marketing anterior

```
dados.prev1 <- dados.prev %>%
   filter(default == 'yes')

dados.prev1$default <- NULL
str(dados.prev1)
```

```
## 'data.frame':    28 obs. of  16 variables:
##  $ age      : int  37 43 30 52 42 39 27 40 35 27 ...
##  $ job      : Factor w/ 12 levels "admin.","blue-collar",..: 5 10 2 5 1 5 7 10
10 2 ...
##  $ marital  : Factor w/ 3 levels "divorced","married",..: 2 3 3 2 1 2 2 3 1 3 .
..
##  $ education: Factor w/ 4 levels "primary","secondary",..: 3 2 2 1 2 3 2 2 2 2
...
##  $ balance  : int  0 685 447 57 723 47 254 45 362 81 ...
##  $ housing  : Factor w/ 2 levels "no","yes": 1 2 1 2 1 1 2 1 1 1 ...
##  $ loan     : Factor w/ 2 levels "no","yes": 1 1 1 2 1 1 2 2 1 2 ...
##  $ contact  : Factor w/ 3 levels "cellular","telephone",..: 1 1 1 1 1 1 1 1 1 1
...
##  $ day      : Factor w/ 31 levels "1","2","3","4",..: 17 18 19 20 20 21 29 29 2
9 29 ...
##  $ month    : Factor w/ 12 levels "apr","aug","dec",..: 10 10 10 10 10 10 5 5 5
5 ...
##  $ duration : int  44 78 426 45 298 28 194 261 329 123 ...
##  $ campaign : int  1 1 2 1 2 3 1 1 2 2 ...
##  $ pdays    : int  123 110 189 196 112 158 188 182 240 205 ...
##  $ previous : int  2 2 6 1 2 3 1 23 8 2 ...
##  $ poutcome : Factor w/ 4 levels "failure","other",..: 1 1 1 1 2 1 1 2 2 3 ...
##  $ y        : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

Criação do modelo preditivo

```
library(caret)

myControl <- trainControl(
          method = 'repeatedcv',
          number = 10,
          repeats = 5)

model.rf <- train(y ~.,
               data = dados.prev1,
               method = 'rf',
               preProcess = c('nzv', 'center', 'scale'),
```

```
                metric = 'Accuracy',
                tuneLength = 8,
                trControl = myControl)

model.rf

## Random Forest
##
## 28 samples
## 15 predictors
##  2 classes: 'no', 'yes'
##
## Pre-processing: centered (27), scaled (27), remove (43)
## Resampling: Cross-Validated (10 fold, repeated 5 times)
## Summary of sample sizes: 26, 26, 24, 25, 25, 26, ...
## Resampling results across tuning parameters:
##
##   mtry  Accuracy   Kappa
##    2    0.8816667   0.00000000
##   11    0.8616667  -0.02272727
##   21    0.8133333  -0.05128205
##   31    0.8066667  -0.04938272
##   40    0.8066667  -0.04938272
##   50    0.8133333  -0.05128205
##   60    0.8133333  -0.05128205
##   70    0.8133333  -0.05128205
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 2.

varImp(model.rf)

## rf variable importance
##
##   only 20 most important variables shown (out of 27)
##
##                      Overall
## duration             100.000
## balance               99.167
## pdays                 58.553
## age                   39.789
## maritalmarried        29.776
## previous              29.237
## campaign              29.162
## jobmanagement         22.761
## poutcomeother         19.746
## monthmay              19.674
## day17                 17.962
## housingyes            17.490
## day7                  14.716
## jobblue-collar        13.967
```

```
## loanyes              13.054
## educationtertiary    11.268
## educationsecondary   10.835
## maritalsingle        10.399
## monthnov              7.519
## day29                 3.475
```

Conforme observado na análise estatística, o fator determinante de exigência pelo banco é a variável balance, ou saldo bancário.

4.5.2 dados.notprev

Clientes que não participaram da campanha de marketing anterior

```
dados.notprev1 <- dados.notprev %>%
   filter(default == 'yes')

dados.notprev1$default <- NULL
```

Criação do modelo preditivo

```
library(caret)
myControl <- trainControl(
          method = 'cv',
          number = 12)

model.rf1 <- train(y ~.,
               data = dados.notprev1,
               method = 'rf',
               preProcess = c('nzv', 'center', 'scale'),
               metric = 'Accuracy',
               tuneLength = 10,
               trControl = myControl)

model.rf1

## Random Forest
##
## 453 samples
##  12 predictor
##   2 classes: 'no', 'yes'
##
## Pre-processing: centered (25), scaled (25), remove (40)
## Resampling: Cross-Validated (12 fold)
## Summary of sample sizes: 415, 414, 416, 416, 416, 415, ...
## Resampling results across tuning parameters:
##
##   mtry  Accuracy   Kappa
```

```
##    2    0.9383412  0.0000000
##    9    0.9535264  0.4203749
##   16    0.9490842  0.5483126
##   23    0.9513365  0.5577332
##   30    0.9512772  0.5577019
##   37    0.9469475  0.5383152
##   44    0.9469475  0.4931590
##   51    0.9469475  0.5312722
##   58    0.9513927  0.5571250
##   65    0.9490842  0.5483126
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 9.
```

```
varImp(model.rf1)
```

```
## rf variable importance
##
##   only 20 most important variables shown (out of 25)
##
##                     Overall
## duration           100.0000
## age                 24.3157
## balance             17.7211
## campaign             6.9340
## maritalsingle        3.4786
## educationsecondary   3.4552
## maritalmarried       3.3351
## jobblue-collar       2.6199
## housingyes           2.1859
## educationtertiary    1.9817
## contactunknown       1.8955
## loanyes              1.7936
## monthaug             1.5697
## monthmay             1.5518
## monthjul             1.3765
## jobtechnician        1.2903
## day21                1.1862
## monthjun             1.1197
## jobmanagement        1.0246
## jobservices          0.8345
```

Resultado obtido um pouco diferente em relação aos clientes que participaram da campanha anterior, tendo a variável idade com uma importância superior em relação ao saldo bancário. .

4.6 Quais são as características mais proeminentes de um cliente que possua empréstimo imobiliário?

4.6.1 dados.prev

Clientes que participaram da campanha de marketing anterior

```
dados.prev2 <- dados.prev %>%
   filter(housing == 'yes')

dados.prev2$housing <- NULL
str(dados.prev2)
```

```
## 'data.frame':    3811 obs. of  16 variables:
##  $ age      : int  33 36 36 56 44 34 33 34 30 30 ...
##  $ job      : Factor w/ 12 levels "admin.","blue-collar",..: 8 5 5 10 2 5 11 1
1 5 ...
##  $ marital  : Factor w/ 3 levels "divorced","married",..: 2 2 2 2 2 2 1 2 2 3 .
..
##  $ education: Factor w/ 4 levels "primary","secondary",..: 2 3 3 2 2 3 2 3 2 3
...
##  $ default  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ balance  : int  3444 2415 0 589 1324 1770 1005 899 873 1243 ...
##  $ loan     : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ contact  : Factor w/ 3 levels "cellular","telephone",..: 2 2 2 3 2 3 2 3 2 2
...
##  $ day      : Factor w/ 31 levels "1","2","3","4",..: 21 22 23 23 25 6 10 12 12
13 ...
##  $ month    : Factor w/ 12 levels "apr","aug","dec",..: 11 11 11 11 11 10 10 10
10 10 ...
##  $ duration : int  144 73 140 518 119 26 175 114 119 86 ...
##  $ campaign : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ pdays    : int  91 86 143 147 89 101 174 170 167 174 ...
##  $ previous : int  4 4 3 2 2 11 2 3 3 1 ...
##  $ poutcome : Factor w/ 4 levels "failure","other",..: 1 2 1 3 2 2 1 1 3 1 ...
##  $ y        : Factor w/ 2 levels "no","yes": 2 1 2 2 1 1 1 2 1 1 ...
```

Construção do modelo

```
model.glm3 <- glm(y ~., data = dados.prev2, family = 'binomial')
summary(model.glm3)
```

```
##
## Call:
## glm(formula = y ~ ., family = "binomial", data = dados.prev2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.1785  -0.3339  -0.2075  -0.1316   2.9560
```

```
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -6.890e+00  8.075e-01  -8.533  < 2e-16 ***
## age                 1.351e-02  8.592e-03   1.572 0.115938
## jobblue-collar     -1.512e-01  2.264e-01  -0.668 0.504184
## jobentrepreneur    -8.933e-01  4.963e-01  -1.800 0.071889 .
## jobhousemaid       -3.071e-01  6.202e-01  -0.495 0.620482
## jobmanagement      -6.553e-02  2.424e-01  -0.270 0.786942
## jobretired          6.897e-01  4.842e-01   1.424 0.154341
## jobself-employed   -1.664e-01  4.016e-01  -0.414 0.678609
## jobservices        -2.575e-02  2.649e-01  -0.097 0.922556
## jobstudent          7.650e-03  7.471e-01   0.010 0.991830
## jobtechnician      -3.040e-01  2.193e-01  -1.386 0.165613
## jobunemployed      -2.121e-01  4.455e-01  -0.476 0.633927
## jobunknown         -1.039e+01  4.321e+02  -0.024 0.980813
## maritalmarried      1.294e-01  2.032e-01   0.637 0.524380
## maritalsingle       2.308e-01  2.305e-01   1.001 0.316769
## educationsecondary  5.599e-01  2.462e-01   2.274 0.022949 *
## educationtertiary   6.674e-01  2.955e-01   2.259 0.023891 *
## educationunknown    5.704e-01  4.150e-01   1.374 0.169330
## defaultyes          1.670e+00  6.994e-01   2.388 0.016941 *
## balance             2.233e-05  3.499e-05   0.638 0.523342
## loanyes            -5.276e-01  2.038e-01  -2.589 0.009618 **
## contacttelephone    1.295e-01  3.180e-01   0.407 0.683960
## contactunknown      1.187e+00  5.926e-01   2.003 0.045192 *
## day2                5.943e-01  6.094e-01   0.975 0.329485
## day3                7.198e-01  6.367e-01   1.131 0.258237
## day4                1.341e+00  6.214e-01   2.158 0.030940 *
## day5                1.062e+00  6.329e-01   1.678 0.093275 .
## day6                1.253e+00  6.209e-01   2.018 0.043606 *
## day7                4.920e-01  6.490e-01   0.758 0.448381
## day8                9.678e-01  6.302e-01   1.536 0.124619
## day9                1.829e+00  6.706e-01   2.727 0.006395 **
## day10               1.206e+00  7.488e-01   1.611 0.107189
## day11               1.265e+00  6.210e-01   2.037 0.041601 *
## day12               1.290e+00  6.208e-01   2.078 0.037707 *
## day13               1.429e+00  6.118e-01   2.336 0.019476 *
## day14               1.481e+00  6.283e-01   2.357 0.018440 *
## day15               1.744e+00  6.192e-01   2.817 0.004842 **
## day16               9.644e-01  6.395e-01   1.508 0.131545
## day17              -1.737e-01  6.324e-01  -0.275 0.783551
## day18               6.588e-01  6.348e-01   1.038 0.299383
## day19               7.769e-01  6.957e-01   1.117 0.264129
## day20               3.784e-01  6.473e-01   0.585 0.558824
## day21               9.658e-01  6.858e-01   1.408 0.159059
## day22               2.802e+00  7.454e-01   3.759 0.000171 ***
## day23               2.242e+00  7.666e-01   2.925 0.003444 **
## day24               2.061e+00  9.970e-01   2.068 0.038685 *
## day25               2.120e+00  7.657e-01   2.769 0.005629 **
```

```
## day26                 1.246e+00  8.160e-01   1.527 0.126667    
## day27                 2.911e+00  7.213e-01   4.036 5.45e-05 ***
## day28                 1.304e+00  7.529e-01   1.732 0.083191 .  
## day29                 9.699e-01  7.068e-01   1.372 0.170025    
## day30                 1.941e+00  6.592e-01   2.945 0.003233 ** 
## day31                 1.151e+00  9.570e-01   1.203 0.228933    
## monthaug              2.229e+00  3.121e-01   7.144 9.08e-13 ***
## monthdec              2.516e+00  6.289e-01   4.001 6.32e-05 ***
## monthfeb              6.795e-01  3.112e-01   2.184 0.028966 *  
## monthjan             -7.836e-01  4.854e-01  -1.614 0.106466    
## monthjul              1.673e+00  4.583e-01   3.650 0.000262 ***
## monthjun              2.287e+00  3.665e-01   6.240 4.38e-10 ***
## monthmar              2.557e+00  4.741e-01   5.393 6.92e-08 ***
## monthmay             -4.072e-01  2.470e-01  -1.648 0.099257 .  
## monthnov              2.190e-01  3.005e-01   0.729 0.466080    
## monthoct              1.266e+00  3.729e-01   3.396 0.000683 ***
## monthsep              2.424e+00  3.880e-01   6.247 4.17e-10 ***
## duration              5.811e-03  3.046e-04  19.076  < 2e-16 ***
## campaign             -1.737e-01  9.413e-02  -1.846 0.064935 .  
## pdays                 9.935e-04  5.931e-04   1.675 0.093871 .  
## previous              7.562e-02  2.158e-02   3.505 0.000457 ***
## poutcomeother         1.101e-01  1.675e-01   0.657 0.510912    
## poutcomesuccess       2.152e+00  1.638e-01  13.141  < 2e-16 ***
## poutcomeunknown      -1.341e+01  8.827e+02  -0.015 0.987881    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2981.2  on 3810  degrees of freedom
## Residual deviance: 1741.9  on 3740  degrees of freedom
## AIC: 1883.9
##
## Number of Fisher Scoring iterations: 13
```

As características mais marcantes destes clientes é que eles possuem nível educacional secundário e terciário, têm empréstimos pessoais, além de possuírem perfil empreendedor.

4.6.2 dados.notprev

Clientes que não participaram da campanha de marketing anterior

```
dados.notprev2 <- dados.notprev %>%
   filter(housing == 'yes')

dados.notprev2$housing <- NULL
str(dados.notprev2)
```

```
## 'data.frame':    16947 obs. of  13 variables:
##  $ age      : int  58 44 33 47 35 28 42 58 43 41 ...
##  $ job      : Factor w/ 12 levels "admin.","blue-collar",..: 5 10 3 2 5 5 3 6 1
0 1 ...
##  $ marital  : Factor w/ 3 levels "divorced","married",..: 2 3 2 2 2 3 1 2 3 1 .
..
##  $ education: Factor w/ 4 levels "primary","secondary",..: 3 2 2 4 3 3 3 1 2 2
...
##  $ default  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 2 1 1 1 ...
##  $ balance  : int  2143 29 2 1506 231 447 2 121 593 270 ...
##  $ loan     : Factor w/ 2 levels "no","yes": 1 1 2 1 1 2 1 1 1 1 ...
##  $ contact  : Factor w/ 3 levels "cellular","telephone",..: 3 3 3 3 3 3 3 3 3 3
...
##  $ day      : Factor w/ 31 levels "1","2","3","4",..: 5 5 5 5 5 5 5 5 5 5 ...
##  $ month    : Factor w/ 12 levels "apr","aug","dec",..: 9 9 9 9 9 9 9 9 9 9 ...
##  $ duration : int  261 151 76 92 139 217 380 50 55 222 ...
##  $ campaign : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ y        : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

Construção do modelo

```
model.glm4 <- glm(y ~., data = dados.notprev2, family = 'binomial')
summary(model.glm4)

##
## Call:
## glm(formula = y ~ ., family = "binomial", data = dados.notprev2)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6481  -0.2348  -0.1498  -0.0992   4.1446
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -5.136e+00  6.565e-01  -7.824 5.12e-15 ***
## age                   5.003e-03  5.363e-03   0.933  0.35087
## jobblue-collar       -3.035e-01  1.405e-01  -2.160  0.03079 *
## jobentrepreneur      -1.200e-01  2.511e-01  -0.478  0.63277
## jobhousemaid         -1.090e+00  3.907e-01  -2.791  0.00525 **
## jobmanagement        -1.001e-01  1.650e-01  -0.607  0.54411
## jobretired           -2.063e-01  3.048e-01  -0.677  0.49860
## jobself-employed     -2.745e-01  2.525e-01  -1.087  0.27702
## jobservices          -3.900e-01  1.733e-01  -2.251  0.02438 *
## jobstudent            2.348e-01  3.830e-01   0.613  0.53989
## jobtechnician        -1.919e-01  1.467e-01  -1.308  0.19085
## jobunemployed        -5.722e-01  3.048e-01  -1.877  0.06048 .
## jobunknown            1.714e+00  1.063e+00   1.613  0.10676
## maritalmarried       -2.293e-01  1.308e-01  -1.753  0.07961 .
```

```
## maritalsingle        2.097e-01  1.448e-01   1.448  0.14759
## educationsecondary   3.616e-02  1.317e-01   0.274  0.78372
## educationtertiary   -2.602e-02  1.690e-01  -0.154  0.87766
## educationunknown    -1.463e-01  2.619e-01  -0.559  0.57640
## defaultyes           4.698e-01  3.772e-01   1.245  0.21300
## balance              2.056e-05  2.319e-05   0.886  0.37539
## loanyes             -2.941e-01  1.182e-01  -2.488  0.01285 *
## contacttelephone    -9.791e-02  2.132e-01  -0.459  0.64609
## contactunknown      -1.804e+00  1.368e-01 -13.190  < 2e-16 ***
## day2                -2.185e-01  6.054e-01  -0.361  0.71817
## day3                -4.020e-01  6.159e-01  -0.653  0.51392
## day4                 7.377e-03  6.105e-01   0.012  0.99036
## day5                -2.310e-01  6.003e-01  -0.385  0.70039
## day6                -5.167e-02  6.023e-01  -0.086  0.93164
## day7                -2.993e-01  6.033e-01  -0.496  0.61979
## day8                 5.873e-02  5.997e-01   0.098  0.92198
## day9                 2.628e-01  6.111e-01   0.430  0.66713
## day10                8.183e-01  6.499e-01   1.259  0.20793
## day11               -3.319e-01  6.106e-01  -0.544  0.58674
## day12                2.148e-02  6.068e-01   0.035  0.97177
## day13                6.325e-01  5.898e-01   1.072  0.28355
## day14                3.484e-01  5.895e-01   0.591  0.55451
## day15                2.481e-01  5.906e-01   0.420  0.67442
## day16                5.181e-01  5.917e-01   0.876  0.38116
## day17               -4.747e-01  5.988e-01  -0.793  0.42791
## day18                5.495e-01  5.835e-01   0.942  0.34630
## day19               -2.713e-01  6.211e-01  -0.437  0.66221
## day20                1.532e-01  5.918e-01   0.259  0.79575
## day21                4.939e-01  5.946e-01   0.831  0.40620
## day22                6.081e-01  6.442e-01   0.944  0.34520
## day23                6.015e-01  6.179e-01   0.973  0.33036
## day24                1.569e-01  6.628e-01   0.237  0.81285
## day25                6.381e-01  6.363e-01   1.003  0.31595
## day26                1.045e+00  6.281e-01   1.664  0.09606 .
## day27                1.092e+00  6.242e-01   1.750  0.08017 .
## day28                4.960e-01  6.267e-01   0.791  0.42872
## day29                4.356e-01  6.188e-01   0.704  0.48149
## day30                1.178e+00  5.988e-01   1.967  0.04917 *
## day31                6.445e-01  7.135e-01   0.903  0.36641
## monthaug             2.846e-01  2.497e-01   1.139  0.25453
## monthdec             3.244e+00  6.364e-01   5.097 3.44e-07 ***
## monthfeb             1.244e+00  2.592e-01   4.800 1.58e-06 ***
## monthjan            -1.491e+00  5.142e-01  -2.899  0.00375 **
## monthjul            -4.335e-01  1.935e-01  -2.241  0.02505 *
## monthjun             1.948e+00  2.485e-01   7.839 4.53e-15 ***
## monthmar             3.700e+00  3.339e-01  11.081  < 2e-16 ***
## monthmay             4.616e-01  1.870e-01   2.468  0.01359 *
## monthnov             4.971e-02  2.203e-01   0.226  0.82153
## monthoct             3.906e+00  3.259e-01  11.988  < 2e-16 ***
## monthsep             3.941e+00  4.157e-01   9.479  < 2e-16 ***
```

```
## duration            6.232e-03  1.496e-04  41.667  < 2e-16 ***
## campaign           -7.198e-02  2.630e-02  -2.737  0.00620 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 7710.3  on 16946  degrees of freedom
## Residual deviance: 4694.8  on 16881  degrees of freedom
## AIC: 4826.8
##
## Number of Fisher Scoring iterations: 7
```

As características mais marcantes destes clientes é que eles são da área de serviços e possuem empréstimos pessoais.

Conclusão

Este estudo teve por objetivo buscar respostas às questões da área de negócios através da análise dos dados, ao agrupá-los ou aplicando modelos matemáticos que pudessem identificar padrões nos dados, auxiliando às tomadas de decisões e focando em tornar os processos mais claros e eficientes.