

Análise Nextel

Rui Romanini

29 de julho de 2017

Contents

Objetivo de negócios	1
O ambiente	1
Entendimento dos dados	1
Análise do mercado de Telefonia Móvel	1
Limpeza dos dados	2
Plotagem	2
Funcionários na NEXTEL	3
Text Mining	5
Conclusão	7

Objetivo de negócios

O objetivo desta análise é obter um entendimento do posicionamento da Nextel no mercado de Telefonia Móvel, número de empregados da Nextel ao longo dos últimos 2 anos e efetuar uma demonstração de text mining sobre um pequeno dataset, composto de mensagens do LinkedIn.

O ambiente

Para realização deste estudo, foi utilizado o R version 3.4.0 and R Studio version 1.0.143

O sistema operacional utilizado foi o Windows 10 Home.

Entendimento dos dados

Este estudo possui 3 vertentes: - Análise do mercado de telefonia móvel, baseado em dados abertos fornecidos pela ANATEL, para posicionamento da Nextel - Análise do tamanho da empresa Nextel, com base em dados fornecidos pelo LinkedIn - Exemplo de Text Mining, com base em conversas trocadas durante uma negociação de emprego, em estágio inicial, para uma vaga de cientista de dados em uma grande empresa do setor de telecom.

Análise do mercado de Telefonia Móvel

Os dados utilizados neste estudo foram fornecidos pela ANATEL, que possui um grande número de datasets. Mais precisamente, utilizamos um dataset com a quantidade de acessos a telefonia movel, onde temos uma série temporal com alguns anos de extensão, detalhada por Empresa.

O link para o dataset pode ser consultado no código R a seguir.

```
#Acessando os dados
```

```
dadosEmpresas <- read.csv(url("https://cloud.anatel.gov.br/index.php/s/TpaFAwSw7RPfBa8/download?path=%2F"))
```

O dataset contém colunas para os meses futuros, até dezembro/2017. Como nos interessa apenas as variáveis preenchidas, eliminaremos as colunas de julho/2017 em diante.

Mantivemos os últimos 8 meses, até o mês de junho/2017.

Outra informação que trataremos são os registros nulos... algumas empresas não possuem dados e estes registros serão excluídos do dataset.

Limpeza dos dados

```
#Eliminar colunas desnecessarias
keeps <- c("Empresa", "X2016.11", "X2016.12", "X2017.01", "X2017.02", "X2017.03", "X2017.04", "X2017.05", "X2017.06")
dadosEmpresas <- dadosEmpresas[keeps]
```

```
#Eliminar registros que contenham valores nulos
dadosEmpresas <- dadosEmpresas[rowSums(is.na(dadosEmpresas)) == 0,]
```

Plotagem

Para podermos comparar visualmente as empresas e a base de clientes com acesso a telefonia móvel, escolhemos a biblioteca ggplot.

```
install.packages("ggplot2", repos = c(CRAN="https://cran.r-project.org/"))
```

```
## Installing package into 'C:/Users/Rui/Documents/R/win-library/3.4'
## (as 'lib' is unspecified)
```

```
## package 'ggplot2' successfully unpacked and MD5 sums checked
##
```

```
## The downloaded binary packages are in
## C:\Users\Rui\AppData\Local\Temp\RtmpmeJ0mu\downloaded_packages
```

```
install.packages("reshape", repos = c(CRAN="https://cran.r-project.org/"))
```

```
## Installing package into 'C:/Users/Rui/Documents/R/win-library/3.4'
## (as 'lib' is unspecified)
```

```
## package 'reshape' successfully unpacked and MD5 sums checked
##
```

```
## The downloaded binary packages are in
## C:\Users\Rui\AppData\Local\Temp\RtmpmeJ0mu\downloaded_packages
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.1
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 3.4.1
```

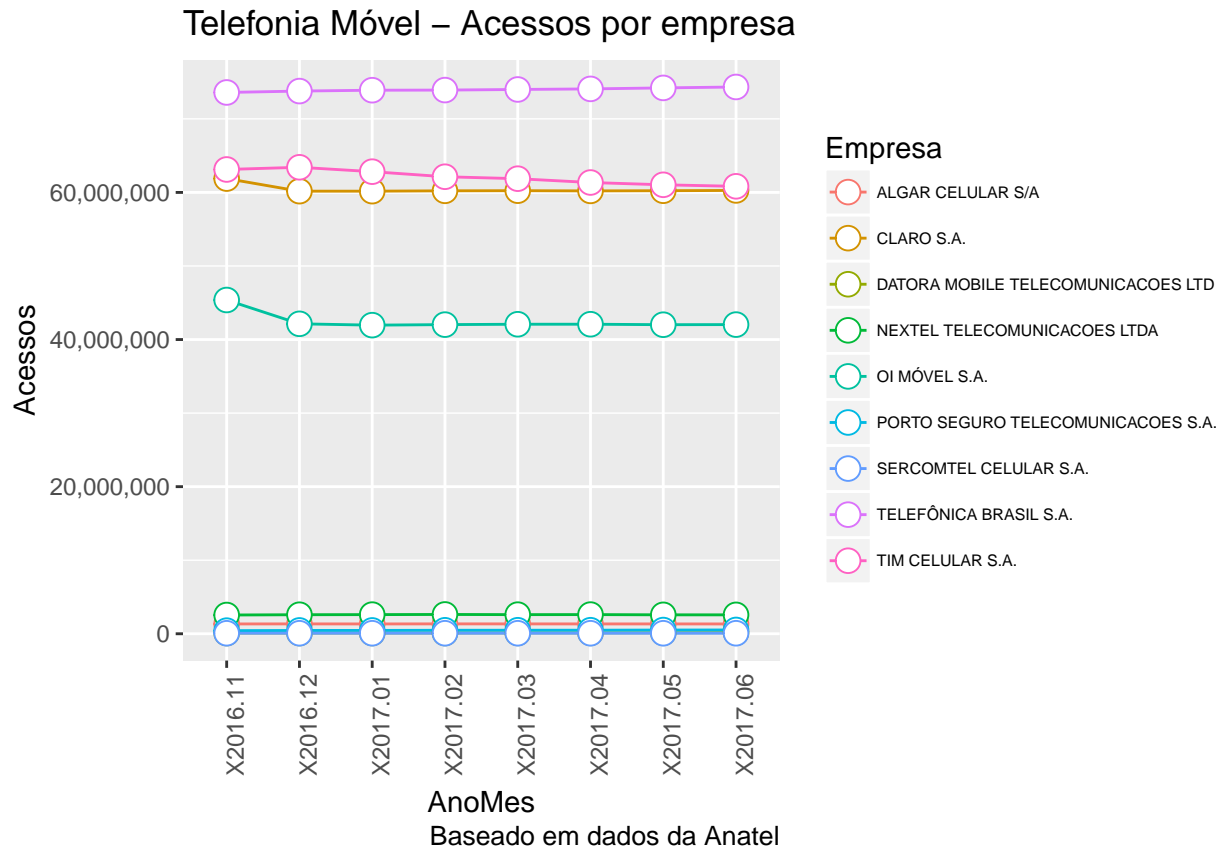
```
#Transformar colunas de cada anomes em uma unica coluna
#para uso no ggplot
```

```
dadosEmpresas <- melt(dadosEmpresas, id.vars="Empresa", value.name="value", variable.name="AnoMes")
```

```
library(scales)
```

```
ggplot(data=dadosEmpresas, aes(x=AnoMes, y=value, group = Empresa, colour = Empresa)) + geom_line() +
```

```
geom_point( size=4, shape=21, fill="white") + theme(legend.text=element_text(size=6)) + theme(ax
labs(title = "Telefonia Móvel - Acessos por empresa",caption = "Baseado em dados da Anatel") +
ylab("Acessos") +
scale_y_continuous(labels = comma)
```



Como podemos acompanhar no gráfico, a Nextel ocupa um pequeno espaço do mercado, mas se mantém lider entre as telecoms de menor porte.

Funcionários na NEXTEL

Com base em dados extraídos do site LinkedIn, é possível ter uma idéia da evolução do número de funcionários na NEXTEL.

O gráfico mostra uma leve tendência de queda, provavelmente por conta da grave crise que aflige o Brasil, tanto em ambito politico como econômico.

A confiança do mercado está voltando aos poucos. Recentemente houve um investimento de US\$ 50 milhões por parte da AINMT e existe ainda a expectativa de mais investimentos no mês de novembro na Nextel Brasil.

#Acessando os dados

```
mydata = read.csv("c:/Rui/funcionariosNextel.csv",header = TRUE,sep = ";")
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote
## = quote, : incomplete final line found by readTableHeader on 'c:/Rui/
## funcionariosNextel.csv'
```

```
head(mydata)
```

```
## periodo funcionarios
## 1 2016/01          5261
## 2 2016/07          5113
## 3 2017/01          5086
## 4 2017/07          5027
```

```
#Estatística Descritiva
```

```
#Principais estatísticas envolvendo o número de funcionários
```

```
summary(mydata$funcionarios)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      5027   5071   5100   5122   5150   5261
```

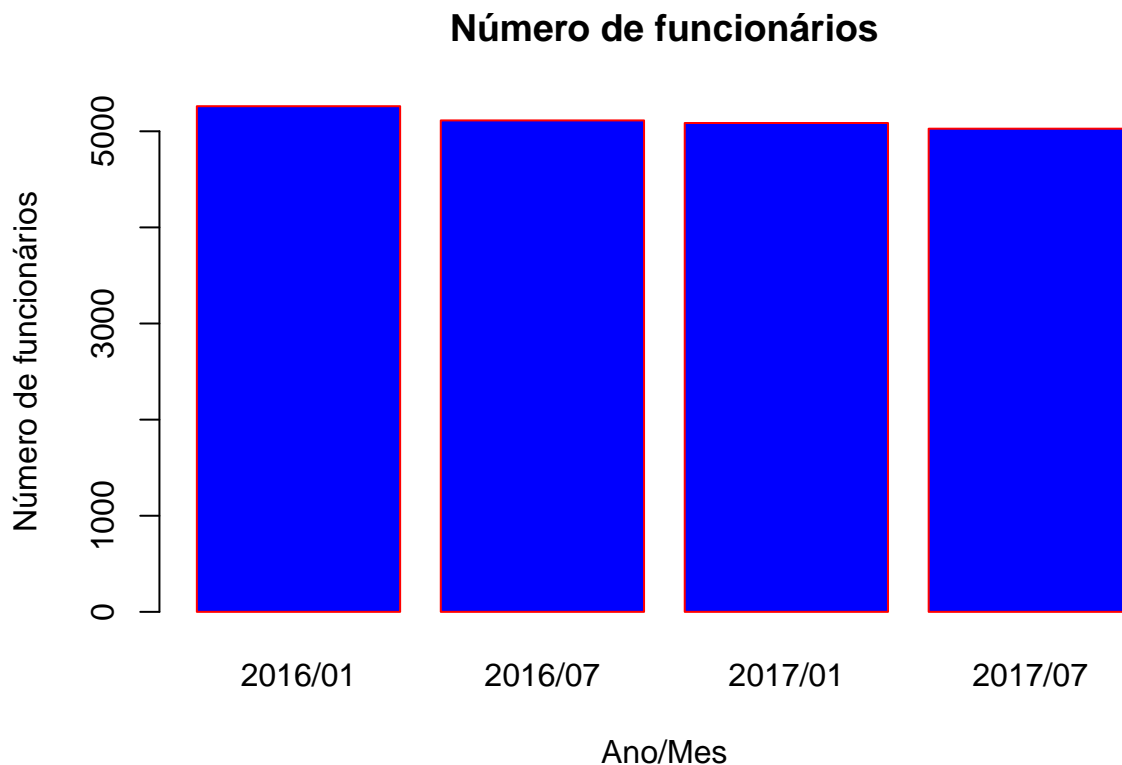
```
# Create the data for the chart.
```

```
H <- mydata$funcionarios
```

```
M <- mydata$periodo
```

```
# Plot the bar chart.
```

```
barplot(H,names.arg = M,xlab = "Ano/Mes",ylab = "Número de funcionários",col = "blue",main = "Número de
```



Text Mining

Para este processo de text mining, criamos um arquivo csv com as mensagens trocadas via LinkedIn. Infelizmente, o LinkedIn não abre a possibilidade de consultar os mails trocados via API, mas isso não impede de consultar o email pessoal via API ou mesmo manualmente.

```
## Installing package into 'C:/Users/Rui/Documents/R/win-library/3.4'
## (as 'lib' is unspecified)

## package 'tm' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\Rui\AppData\Local\Temp\RtmpmeJ0mu\downloaded_packages

## Installing package into 'C:/Users/Rui/Documents/R/win-library/3.4'
## (as 'lib' is unspecified)

## package 'SnowballC' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\Rui\AppData\Local\Temp\RtmpmeJ0mu\downloaded_packages

## Installing package into 'C:/Users/Rui/Documents/R/win-library/3.4'
## (as 'lib' is unspecified)

## package 'wordcloud' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\Rui\AppData\Local\Temp\RtmpmeJ0mu\downloaded_packages

## Warning: package 'tm' was built under R version 3.4.1

## Loading required package: NLP

##
## Attaching package: 'NLP'

## The following object is masked from 'package:ggplot2':
##
##   annotate

## Warning: package 'wordcloud' was built under R version 3.4.1

## Loading required package: RColorBrewer
```

As mensagens a serem analisadas foram carregadas para um arquivo csv e consultadas na instrução a seguir:

```
mydata = read.csv("c:/Rui/mensagensLinkedIn.csv",header = TRUE,sep = ";")
#head(mydata)

myCorpus <- Corpus(VectorSource(mydata$Mensagem))

#Transformacoes para tratamento do teste
#Capitalizacao - reduzir para caracteres minusculos
myCorpus <- tm_map(myCorpus, tolower)

# Remocao de pontuacao
myCorpus <- tm_map(myCorpus, removePunctuation)

# Remover numeros
myCorpus <- tm_map(myCorpus, removeNumbers)
```

```

# Remover stopwords usando lista padrao
stopWords <- c(stopwords('portuguese'), "caso", "aqui", "entao")
myCorpus <- tm_map(myCorpus, removeWords, stopWords)
#inspect(myCorpus)

#Stem ou reducao dos termos a seus radicais
myCorpus2 <- tm_map(myCorpus, stemDocument)
#inspect(myCorpus2)

#Frequencia dos termos
corpus_tf <- TermDocumentMatrix(myCorpus, control = list(minWordLength = 1,minDocFreq=1))

#Construcao da representacao vetorial
corpus_tf_idf = weightTfIdf(corpus_tf,normalize = TRUE)

# Para gerar o wordcloud os dados devem estar em uma matriz
m <- as.matrix(corpus_tf)

```

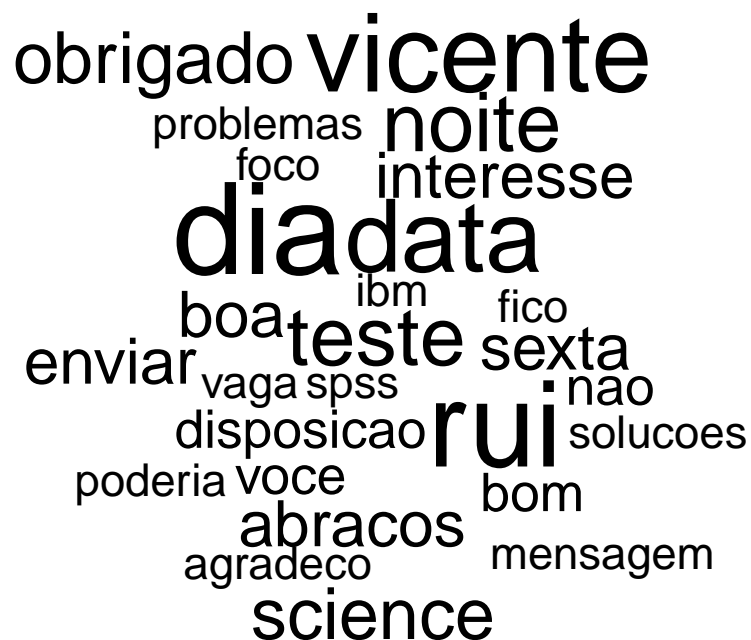
Construção da frequência das palavras para plotagem do gráfico do tipo “Word Cloud”

```

#Calculo da frequencia das palavras
freqPalavras <- sort(rowSums(m),decreasing = TRUE)

# Gera uma wordcloud dos dados
wordcloud(names(freqPalavras),freqPalavras,min.freq = 3)

```



Conclusão

Ao analisarmos o gráfico do tipo word cloud, percebemos uma frequência maior em termos como “dia”, “noite”, “data”, “science”, “interesse”, “Vicente”, “Rui”.

Provavelmente mostra uma conversa bastante educada entre as partes e um interesse particular pela área de data science.

Um segundo passo aqui seria realizar uma análise de sentimento, que em um plano real, teria muita utilidade por exemplo, na análise das comunicações dos clientes com a Nextel.

Meu objetivo aqui foi demonstrar alguns conceitos de análise, codificação R, visualização de dados, R Markdown para geração de documentação e claro, me manter competitivo para a Nextel.