Case_9CaBer

Rafael Bicudo Rosa

12 de agosto de 2018

Case 9CaBer

Esse relatório tem por finalidade identificar os principais influenciadores para um cliente realizar todos os passos do funil de conversão (install > car_added > first_ride). Optei por utilizar o R pela sua flexibilidade no tratamento dos dados e facilidade de geração de reports.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
campanha_df <- readxl::read_excel("Case_Estag_Estat.xlsx", sheet = 2)

campanha_fact <- data.frame(sapply(campanha_df[2:11], factor))
# variaveis descartadas por possuirem pouco informacao relevante
campanha_tbl <- cbind(campanha_df[1],campanha_fact)

glimpse(campanha_tbl)
```

```
## Observations: 27,199
## Variables: 11
## $ `Postback Timestamp` <dttm> 2018-05-31 23:56:15, 2018-05-31 23:55:27...
## $ Click.ID             <fct> w9AJMV0C671MQ33EHLHVRH92, w2D5EHI6SH04BQA...
## $ Transaction.ID       <fct> install, install, first_ride, install, in...
## $ Country              <fct> Brazil, Brazil, Brazil, Brazil, Brazil, B...
## $ Traffic.Source.ID    <fct> 9a2a1675-584d-491d-9fc9-aeaa674c6e2d, 5cf...
## $ OS                   <fct> Android, Android, Android, Android, Andro...
## $ Isp                  <fct> Telefonica Brasil s.a., Brasil Telecom Sa...
## $ Mobile.Carrier       <fct> Vivo, NA, Claro, NA, NA, Claro, NA, NA, N...
## $ Connection.Type      <fct> Mobile, Xdsl, Mobile, Broadband, Broadban...
## $ IP                   <fct> 177118105122, 201.89.13.176, 187.26.74.12...
## $ Site.ID              <fct> 24138, 7c0ajw76178_6761_652_60601652, 120...
```

```
summary(campanha_tbl)
```

```
##  Postback Timestamp                              Click.ID
##  Min.   :2018-05-01 01:08:52   w02NRHPDHE4F396E1THTQG4S:    3
##  1st Qu.:2018-05-15 03:06:03   w04NDKJ3DU44FQSD16HLBSB4:    3
##  Median :2018-05-20 01:03:00   w07VVKO4VQSE20AE1I9VUL9O:    3
##  Mean   :2018-05-19 23:24:31   w09NCC0129SO4AOEHAGG9MBE:    3
##  3rd Qu.:2018-05-24 03:40:27   w0AQNHRU7IANKTOEHRNJCDKS:    3
##  Max.   :2018-05-31 23:56:15   w0E2IHKBH2KO9P2EH5HI90B6:    3
```

```
##                                (Other)                 :27181
##     Transaction.ID           Country
##  card_added: 1005   Brazil       :27080
##  first_ride:  693   United States:   78
##  install   :25501   Colombia     :    8
##                     Peru         :    8
##                     Ecuador      :    7
##                     (Other)      :   17
##                     NA's         :    1
##                              Traffic.Source.ID                OS
##  9a2a1675-584d-491d-9fc9-aeaa674c6e2d:20345   Android         :26671
##  06305dce-21d8-4412-bc81-6b1f5ecf63c9: 1531   IOS             :  402
##  7826a798-e3ce-4462-9248-900af565efff: 1385   Other desktop OS:   75
##  0bab9113-87a4-4da8-8a46-c2d1f18ce4c8: 1195   RIM OS          :    1
##  db6c7a1a-450f-4cdf-bf3c-d7efd71f35cf:  735   Windows         :   50
##  6b3d5dad-15d8-47a2-8109-2a85d80df0fc:  656
##  (Other)                                : 1352
##                                          Isp       Mobile.Carrier
##  Claro Brazil                              :6997   Vivo   : 2554
##  Telefonica Brasil s.a.                    :6739   TIM    : 2232
##  Telemar Norte Leste s.a.                  :2500   Claro  : 1631
##  Tim Celular S.A.                          :2232   Oi     :  568
##  Brasil Telecom Sa - Filial Distrito Federal:1064   Nextel :  334
##  (Other)                                   :7663   (Other):   41
##  NA's                                      :   4   NA's   :19839
##   Connection.Type               IP           Site.ID
##  Broadband:19307   189105176108  :   72   12088  : 4575
##  Cable    :   13   177102238191  :   46   24138  : 3924
##  Dialup   :   78   191142134143  :   36   26920  : 2806
##  Mobile   : 7362   201.8.168.166 :   29   23103  : 1472
##  Satellite:   32   189.105.178.89:   28   26896  : 1196
##  Wireless :   17   191180120106  :   27   (Other):13216
##  Xdsl     :  390   (Other)       :26961   NA's   :   10
```

O primeiro passo foi a preparação do dataframe para execução de análise. Assim como visto acima, nem todas as variáveis contribuirão da mesma maneira. Desconsiderei as variáveis "Click ID", "IP" e "Site ID" por serem de identificação, portanto não possuindo muita informação de interesse à análise, e com potencial de gerar viés negativo em um modelo preditivo. Outra desconsiderada foi "Postback Timestamp" pelo intervalo de tempo curto. Em relação à variável 'Isp', por possuir uma concentração muito grande em algumas empresas e muitas outras dispersas, optei por agrupar todas essas com o label "other". Por fim, como a conversão passa obrigatoriamente por todas as fases, criei uma classe binária de "converted" e "n_converted".

```
campanha_fact2 <- data.frame(sapply(campanha_df[c(4,5,6,8,9)], factor))
# variaveis descartadas por possuirem pouco informacao relevante
Isp_enxuta = ifelse(campanha_df$Isp %in% c('Claro Brazil', 'Telefonica Brasil s.a.', 'Telemar Norte Les
                                           'Tim Celular S.A.', 'Brasil Telecom Sa - Filial Distrito Federal'
                    campanha_df$Isp, 'other')
Isp_f = factor(Isp_enxuta)

Transac_enxuta = ifelse(campanha_df[[3]] %in% c('install',"card_added"),
                        "n_converted", 'converted')
Transac_f = factor(Transac_enxuta)

campanha_tbl <- cbind(Transac_f, campanha_fact2, Isp_f)
names(campanha_tbl) = c('Transaction', 'Country', 'Source', 'OS', 'Mobile', 'Connection', 'Isp')
```

```
campanha_tbl$Mobile <- addNA(campanha_tbl$Mobile)
campanha_tbl$Country <- addNA(campanha_tbl$Country)

glimpse(campanha_tbl)
```

```
## Observations: 27,199
## Variables: 7
## $ Transaction <fct> n_converted, n_converted, converted, n_converted, ...
## $ Country     <fct> Brazil, Brazil, Brazil, Brazil, Brazil, Brazil, Br...
## $ Source      <fct> 9a2a1675-584d-491d-9fc9-aeaa674c6e2d, 5cff49a9-50d...
## $ OS          <fct> Android, Android, Android, Android, Android, Andro...
## $ Mobile      <fct> Vivo, NA, Claro, NA, NA, Claro, NA, NA, NA, NA, NA...
## $ Connection  <fct> Mobile, Xdsl, Mobile, Broadband, Broadband, Mobile...
## $ Isp         <fct> Telefonica Brasil s.a., Brasil Telecom Sa - Filial...
```

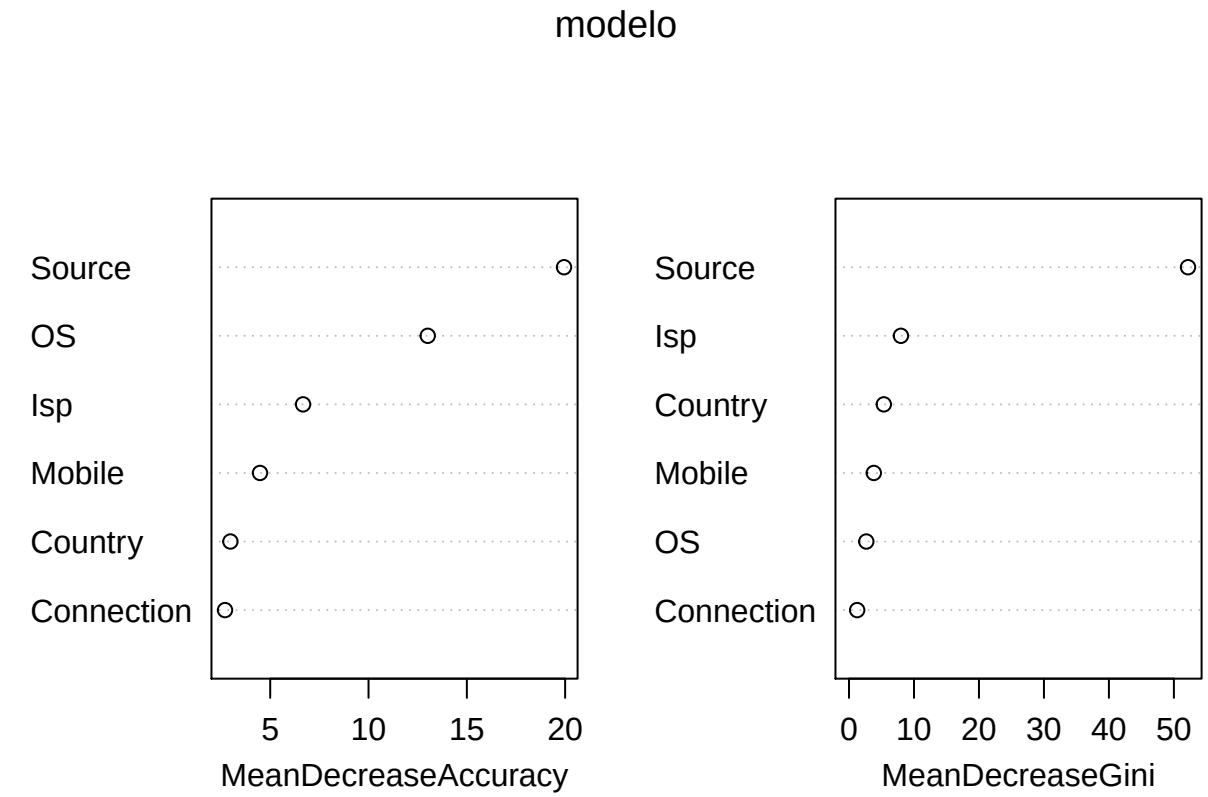
```
summary(campanha_tbl)
```

```
##       Transaction             Country
##  converted  :  693   Brazil       :27080
##  n_converted:26506   United States:   78
##                      Colombia     :    8
##                      Peru         :    8
##                      Ecuador      :    7
##                      Argentina    :    4
##                      (Other)      :   14
##                                    Source                     OS
##  9a2a1675-584d-491d-9fc9-aeaa674c6e2d:20345   Android         :26671
##  06305dce-21d8-4412-bc81-6b1f5ecf63c9: 1531   IOS             :  402
##  7826a798-e3ce-4462-9248-900af565efff: 1385   Other desktop OS:   75
##  0bab9113-87a4-4da8-8a46-c2d1f18ce4c8: 1195   RIM OS          :    1
##  db6c7a1a-450f-4cdf-bf3c-d7efd71f35cf:  735   Windows         :   50
##  6b3d5dad-15d8-47a2-8109-2a85d80df0fc:  656
##  (Other)                             : 1352
##      Mobile          Connection
##  NA     :19839   Broadband:19307
##  Vivo   : 2554   Cable    :   13
##  TIM    : 2232   Dialup   :   78
##  Claro  : 1631   Mobile   : 7362
##  Oi     :  568   Satellite:   32
##  Nextel :  334   Wireless :   17
##  (Other):   41   Xdsl     :  390
##                                          Isp
##  Brasil Telecom Sa - Filial Distrito Federal:1064
##  Claro Brazil                               :6997
##  other                                      :7667
##  Telefonica Brasil s.a.                     :6739
##  Telemar Norte Leste s.a.                   :2500
##  Tim Celular S.A.                           :2232
##
```

Variáveis mais relevantes

Por ser um problema de classificação, usarei um modelo RandomForest para descobrir quais são as características mais infleuntes na determinação da de uma conversão

```
## randomForest 4.6-14
## Type rfNews() to see new features/changes/bug fixes.
##
## Attaching package: 'randomForest'
## The following object is masked from 'package:dplyr':
##
##     combine
```



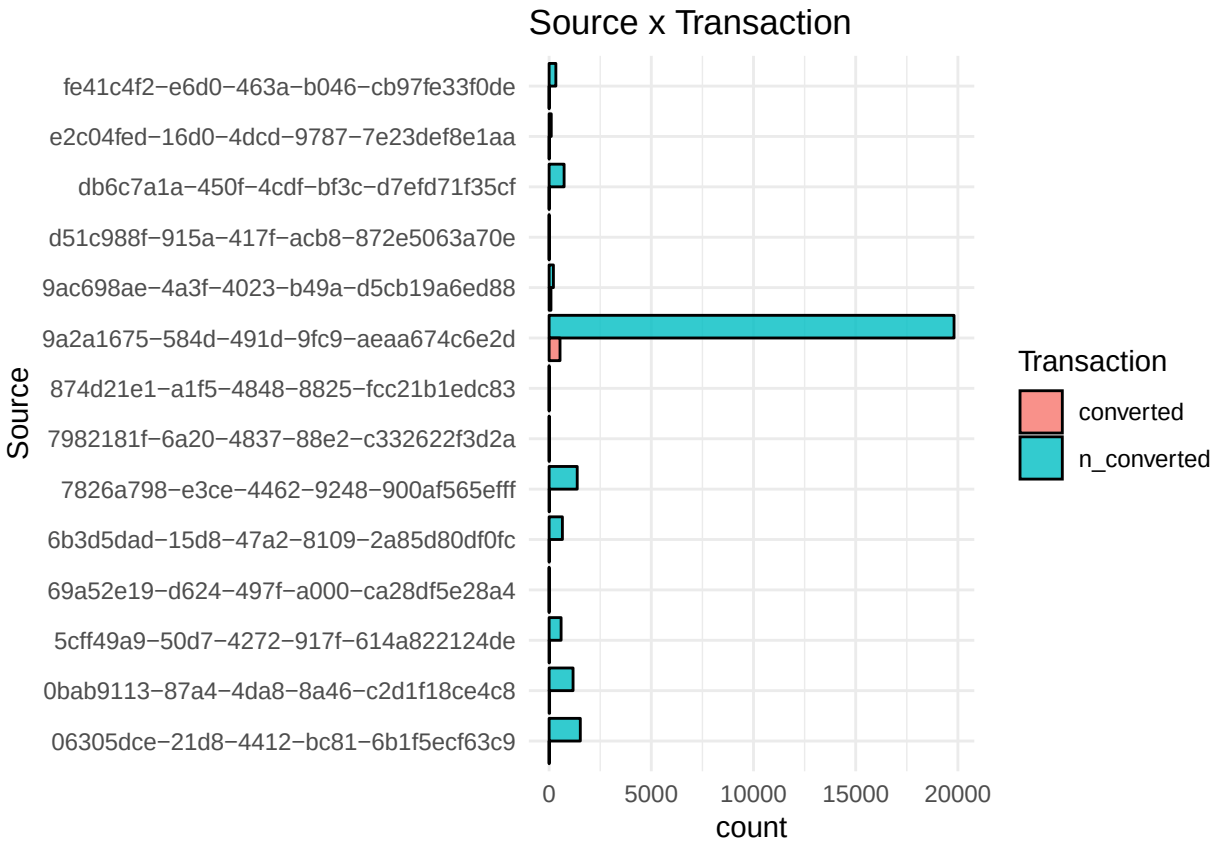
Assim como visto acima, "Source" parece, de forma hegemônica, ser a variável de maior importância nos dados fornecidos, seguido por OS, Isp e Mobile com relevância considerável. Abaixo, um plot ilustrando graficamente o que o modelo concluiu.

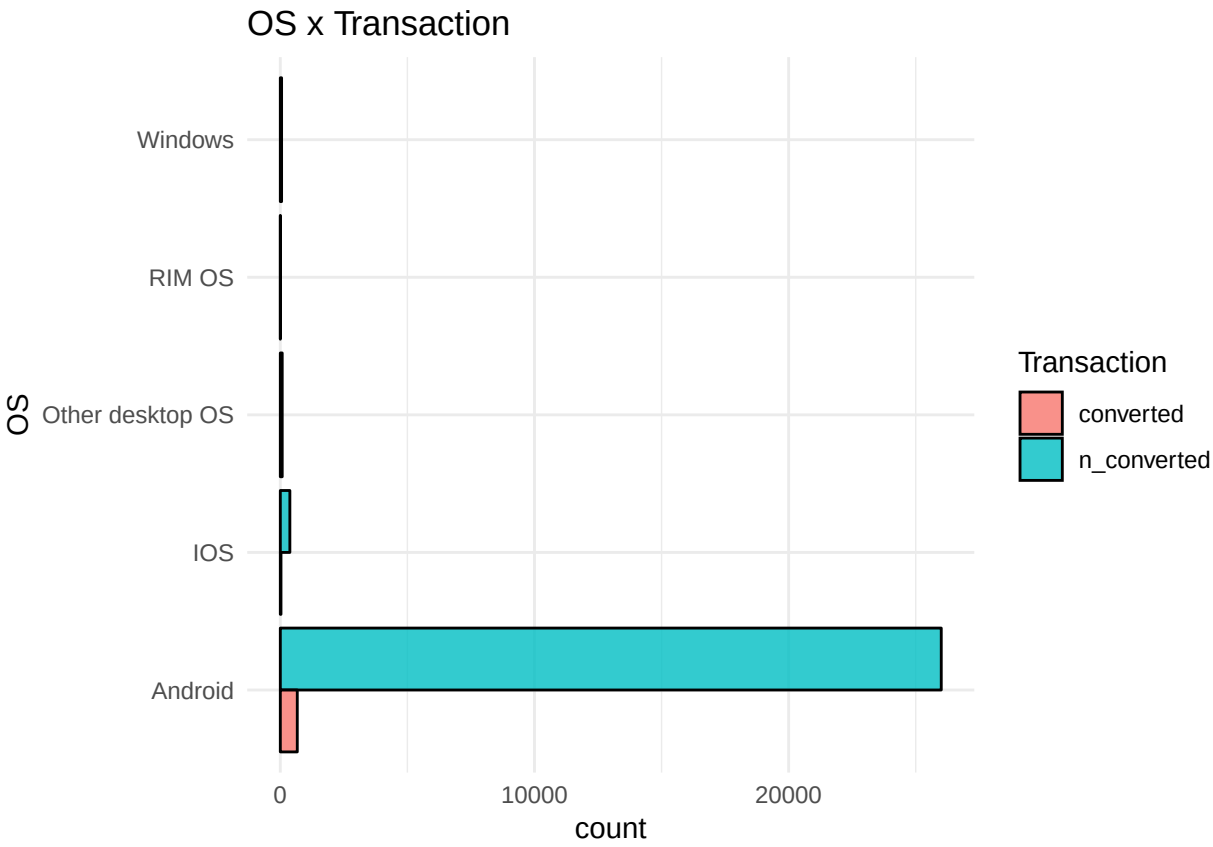
```
library(ggplot2)
```

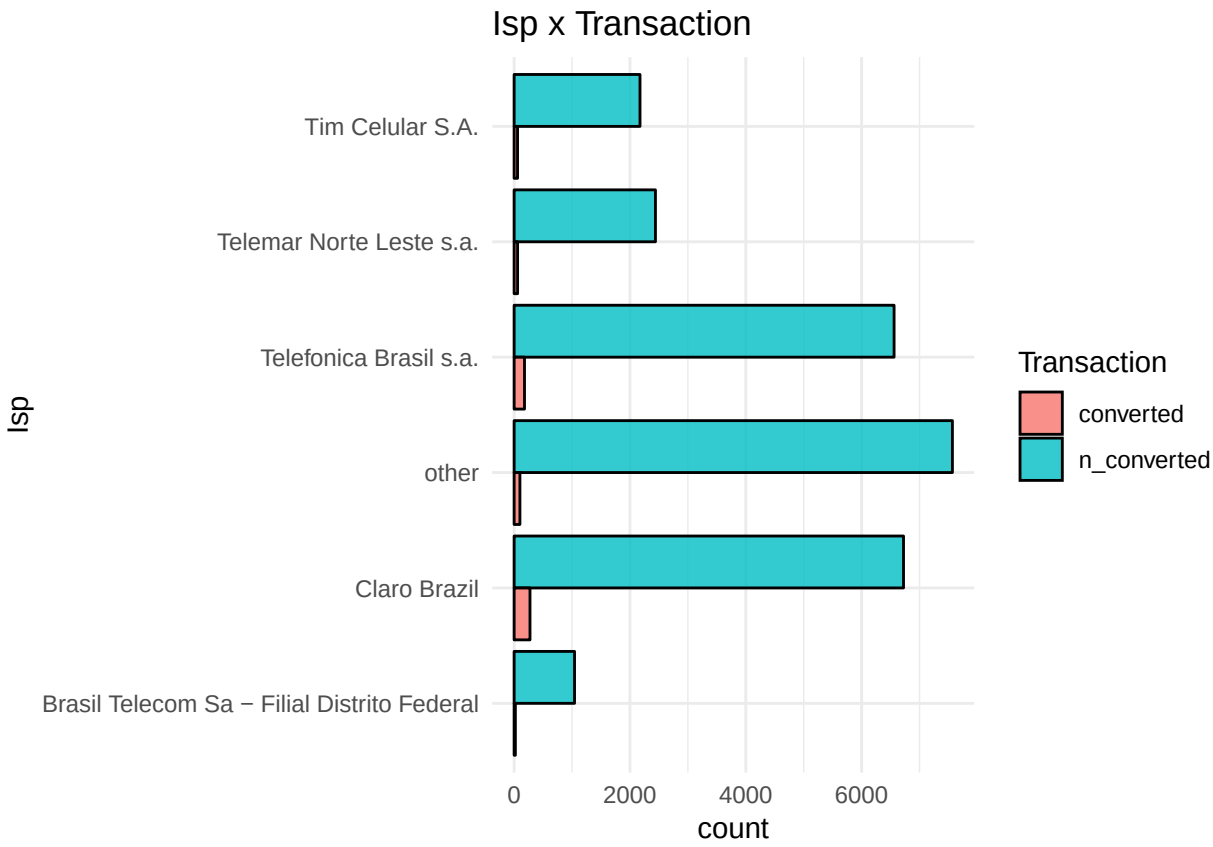
```
##
## Attaching package: 'ggplot2'
## The following object is masked from 'package:randomForest':
##
##     margin
```

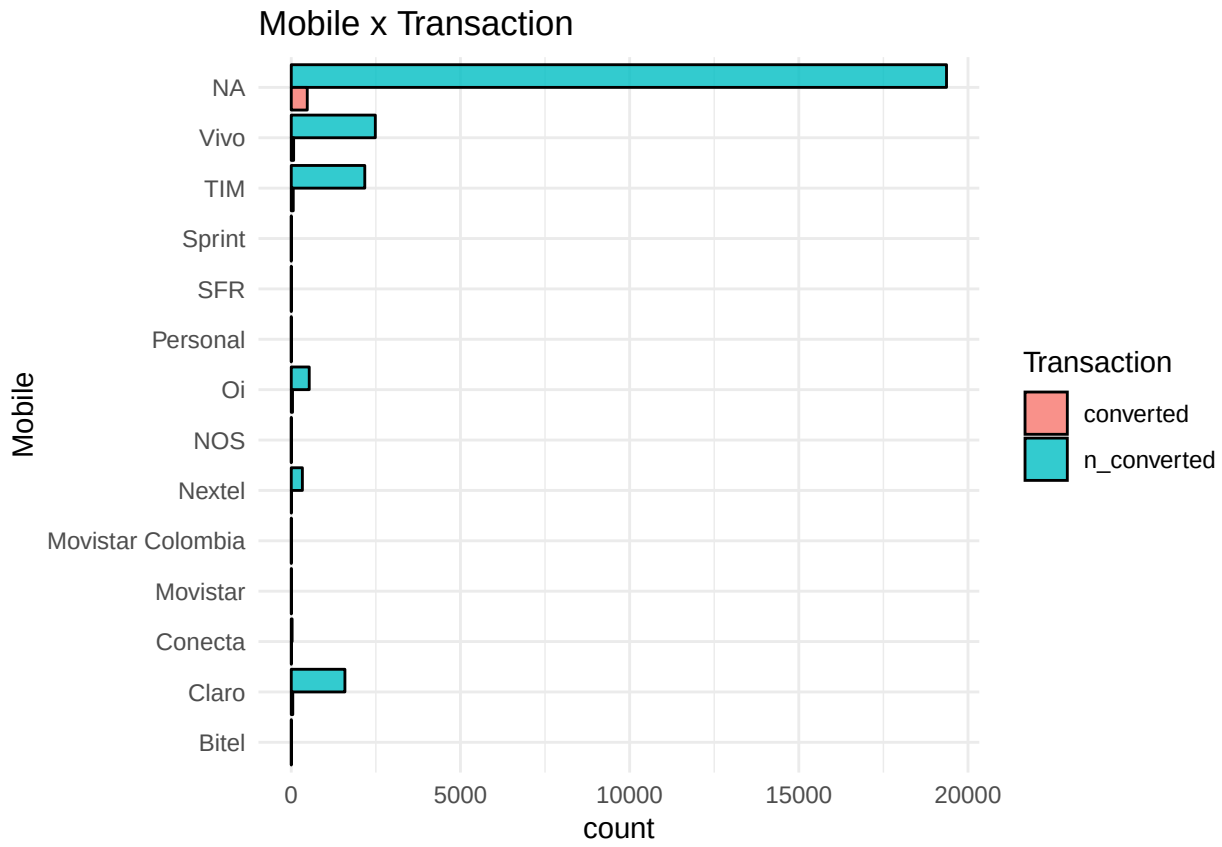
```
plots<- list()
for (i in c("Source", "OS", "Isp", "Mobile")) {
  plots[[i]] <- ggplot(campanha_tbl, aes_string(x = i, fill = 'Transaction')) +
    geom_bar(alpha=0.8, colour='black', position = 'dodge') + ggtitle(paste(i, 'x Transaction')) +
    theme_minimal() +
    coord_flip()
```

```
  print(plots[[i]])
}
```




OS x Transaction
Windows
RIM OS
Other desktop OS
IOS
Android
OS
0
10000
20000
count
Transaction
converted
n_converted


Isp x Transaction
Tim Celular S.A.
Telemar Norte Leste s.a.
Telefonica Brasil s.a.
other
Claro Brazil
Brasil Telecom Sa − Filial Distrito Federal
Isp
0
2000
4000
6000
count
Transaction
converted
n_converted



Em uma nova rodada de investimento, seria interessante aumentar o direcionamento para a fonte de tráfego "9a2a1675-584d-491d-9fc9-aeaa674c6e2d". Em relação ao perfil dos usuários, há uma maior adesão entre usuários de Andoid, clientes de internet da Claro Brazil.