

Data Science Academy - Projeto com Feedback 02

Henrique Jordão Figueiredo Alves

22 Maio, 2021

Projeto com Feedback 02 - Prevendo Demanda de Estoque com Base em Vendas

O objetivo do projeto é desenvolver um modelo para prever com precisão a demanda de estoque com base nos dados históricos de vendas. Isso fará com que os consumidores dos mais de 100 produtos de panificação não fiquem olhando para as prateleiras vazias, além de reduzir o valor gasto com reembolsos para os proprietários de lojas com produtos excedentes impróprios para venda.

Fonte dos dados: <https://www.kaggle.com/c/grupo-bimbo-inventory-demand>

Ao testar o projeto, favor modificar o diretório da pasta abaixo para onde os arquivos do projeto estão localizados em sua máquina

Etapa 1: Carregando os dados

```
# Carregando os pacotes necessários para o projeto
library(tidyr)
library(ggplot2)
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
```

```
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##   between, first, last
```

```
library(scales)

# Carregando os dados das tabelas

cliente <- fread("cliente_tabla.csv")
produto <- fread("producto_tabla.csv")
local <- fread("town_state.csv")

dados <- fread("train.csv")
```

Etapa 2: Manipulando os Dados

```
# Coletando uma amostra menor do dataset dados

sample1 <- sample(1:nrow(dados), 16333333)
dados <- dados[sample1,]

# Eliminando dados repetidos em cliente

cliente <- cliente[!duplicated(cliente$Cliente_ID),]

# Modificando o encoding para UTF-8 e capitalizando a coluna State

Encoding(local$State) <- "UTF-8"
local$State <- toupper(local$State)

# Realizando JOINS entre as tabelas

dados <- merge(dados, cliente, by = "Cliente_ID")
dados <- merge(dados, produto, by = "Producto_ID")
dados <- merge(dados, local, by = "Agencia_ID")

# Eliminando as variáveis IDs

ids <- c("Cliente_ID", "Agencia_ID", "Producto_ID")
dados[,ids] <- NULL

# Convertendo variáveis categóricas

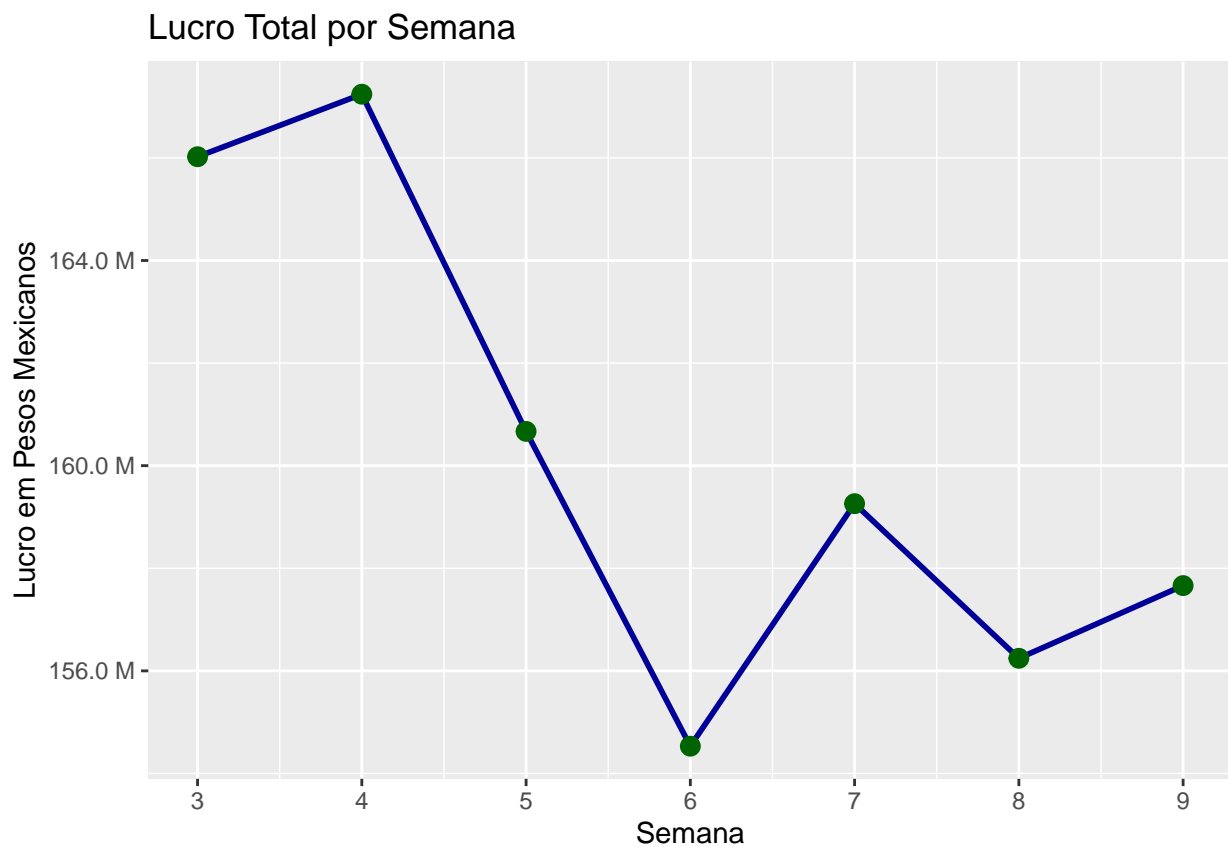
source("toFactor.R")
categorical <- c("Semana", "Canal_ID", "Ruta_SAK", "NombreCliente",
               "NombreProducto", "Town", "State")
dados <- to.factors(df = dados, variables = categorical)
str(dados)
```

```
## Classes 'data.table' and 'data.frame': 16333333 obs. of 12 variables:
## $ Semana : Factor w/ 7 levels "3","4","5","6",...: 1 2 5 4 7 7 1 1 7 6 ...
## $ Canal_ID : Factor w/ 9 levels "1","2","4","5",...: 6 6 6 6 6 6 6 9 9 9 ...
## $ Ruta_SAK : Factor w/ 3077 levels "1","2","3","4",...: 1344 1344 1344 1344 1353 1355 1349 1...
## $ Venta_uni_hoy : int 12 24 30 6 6 25 40 24 24 16 ...
## $ Venta_hoy : num 256 512 640 128 128 ...
```

```
## $ Dev_uni_proxima : int 0 0 0 0 0 25 0 0 0 0 ...
## $ Dev_proxima : num 0 0 0 0 0 ...
## $ Demanda_uni_equil: int 12 24 30 6 6 0 40 24 24 16 ...
## $ NombreCliente : Factor w/ 287760 levels "007","056 THE AIRPORT MARKET",...: 125869 125869 125869 ...
## $ NombreProducto : Factor w/ 1658 levels "100pct Whole Wheat 680g MTA ORO 43111",...: 972 972 972 ...
## $ Town : Factor w/ 257 levels "2001 AG. ATIZAPAN",...: 6 6 6 6 6 6 6 6 6 ...
## $ State : Factor w/ 33 levels "AGUASCALIENTES",...: 15 15 15 15 15 15 15 15 15 ...
## - attr(*, ".internal.selfref")=<externalptr>
```

Etapa 3: Análise Exploratória dos Dados

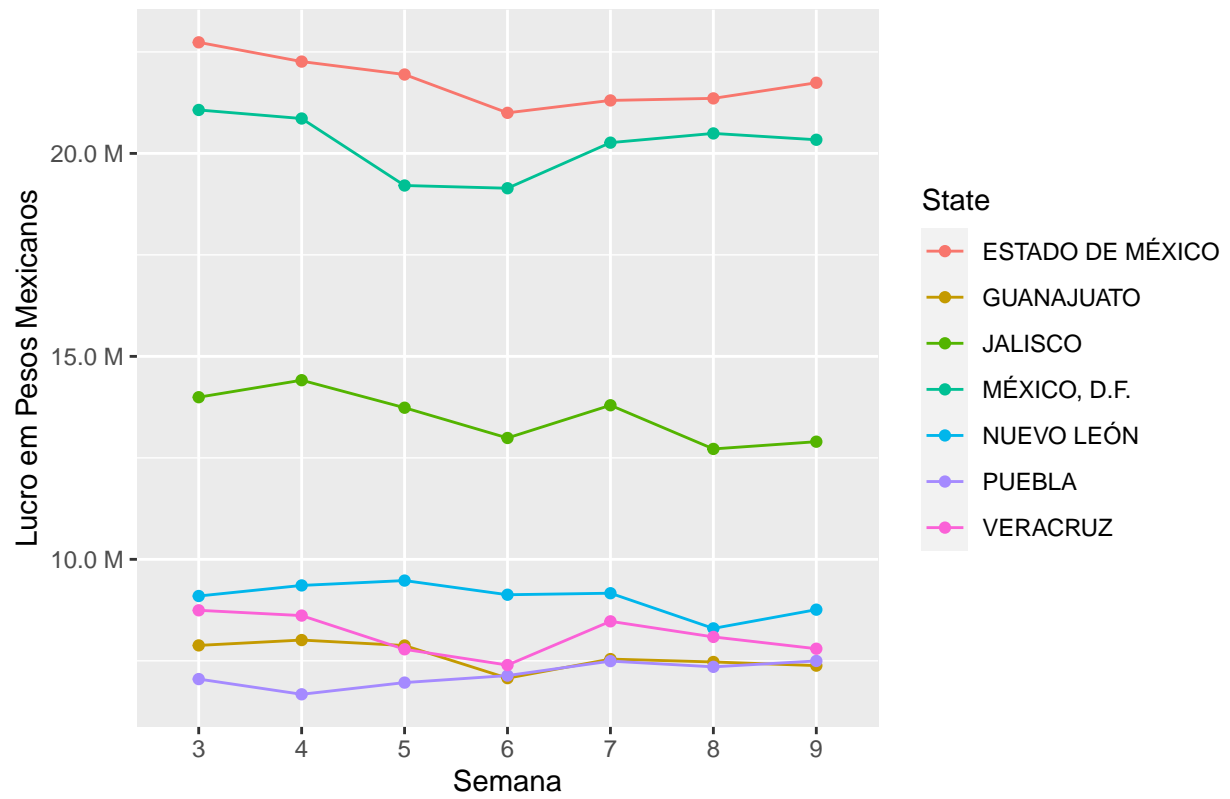
```
Semana_Vendas <- aggregate(Venta_hoy ~ Semana, dados, sum)
Semana_Vendas$Semana <- as.numeric(as.character(Semana_Vendas$Semana))
```



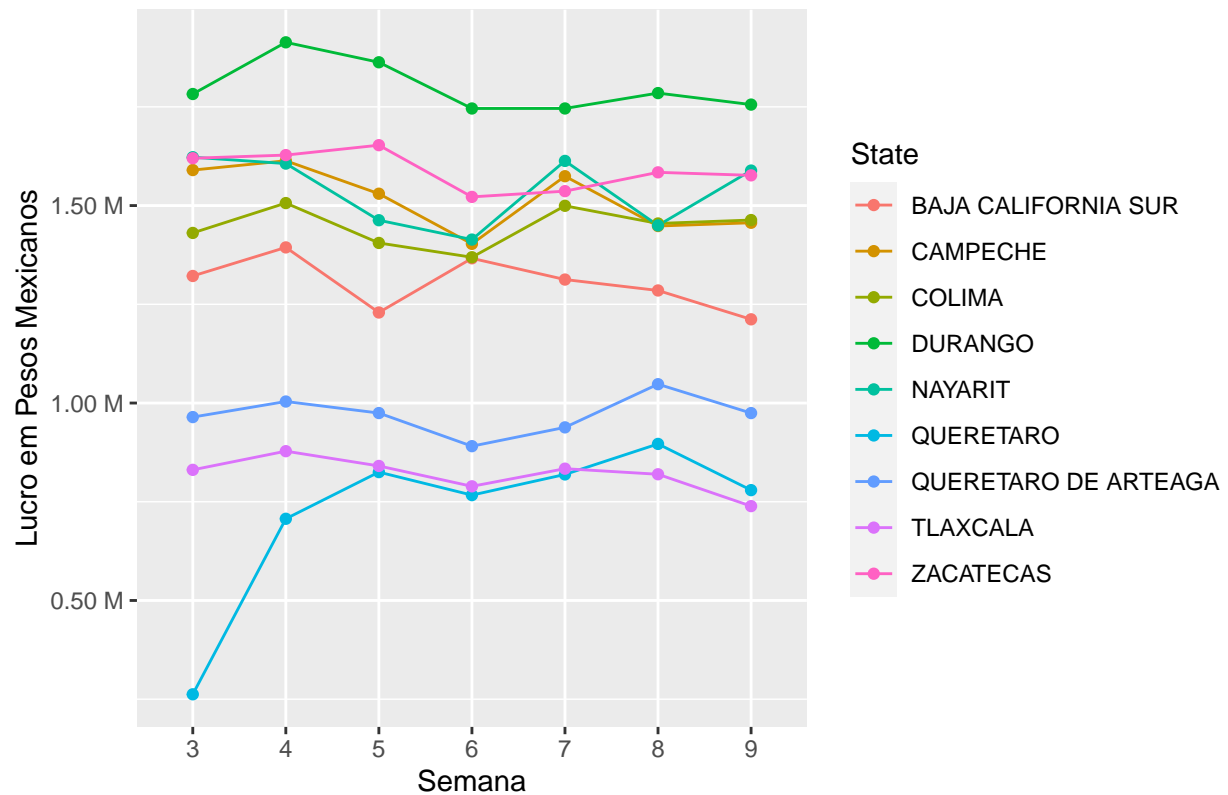
```
State_Vendas <- dados %>% group_by(State, Semana) %>%
  summarise(Profit = sum(Venta_hoy))
```

'summarise()' has grouped output by 'State'. You can override using the '.groups' argument.

Lucro nos Estados Mais Lucrativos



Lucro nos Estados Menos Lucrativos

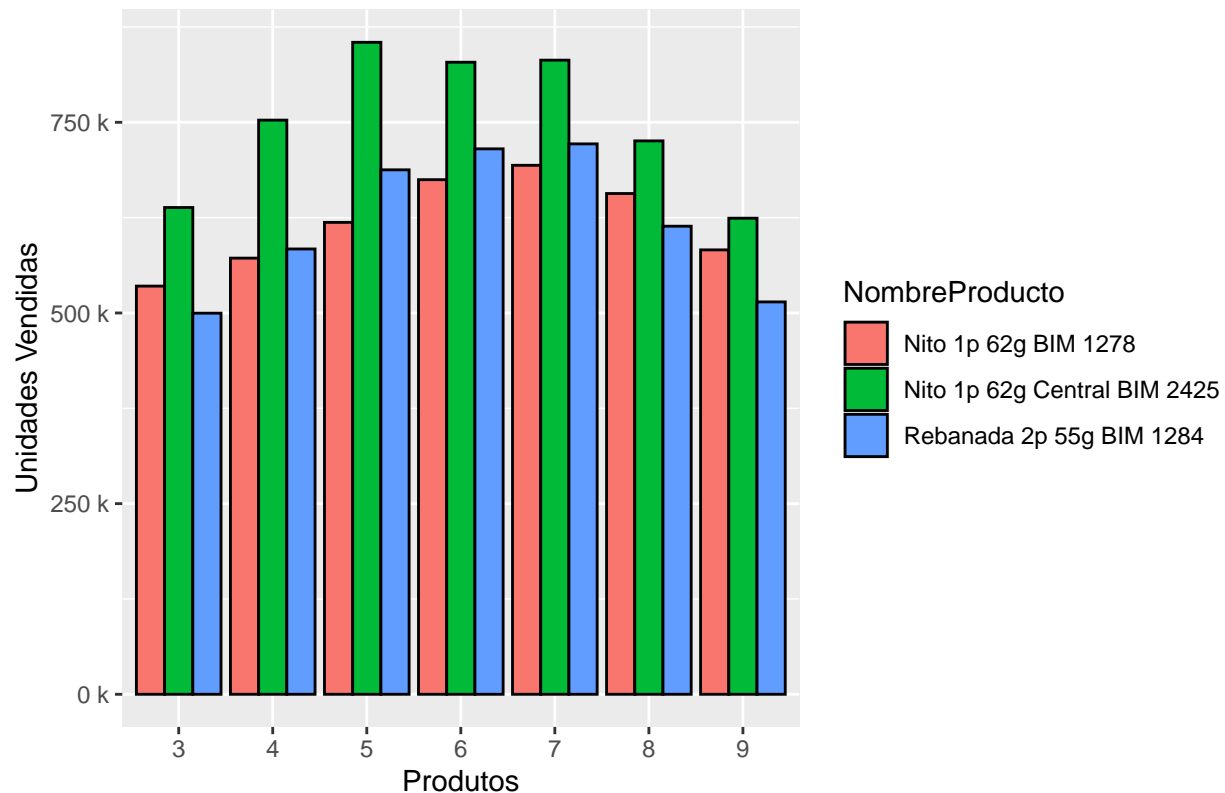


```
Produto_Vendas <- dados %>% group_by(NombreProducto, Semana) %>%
  summarise(Units = sum(Venta_uni_hoy))
```

'summarise()' has grouped output by 'NombreProducto'. You can override using the '.groups' argument.

```
Mais_Vendidos <- c("Nito 1p 62g Central BIM 2425", "Rebanada 2p 55g BIM 1284",
  "Nito 1p 62g BIM 1278")
```

Desempenho dos Produtos Mais Vendidos



```
Menos_Vendidos <- subset(Produto_Vendas$NombreProducto, Produto_Vendas$Units==0)
Menos_Vendidos <- unique(Menos_Vendidos)
Menos_Vendidos <- as.character(Menos_Vendidos)
```

Abaixo, você confere a lista de produtos que não venderam por ao menos uma semana.

```
## [1] "Animalitos Karamelo 200g MTA LAR 43261"
## [2] "Barra Multigrano Nuez 204g 6p TAB BIM 46299"
## [3] "Barri Pop 25g MP BIM 36453"
## [4] "Barri Pop 25g MP Tira BIM 36456"
## [5] "Barri Pop 25g Tira BIM 36258"
## [6] "Barritas Fresa Salvado 6p 67g Cj MTA MLA 43119"
## [7] "Barritas FresaPina Cj Exhib 5400g MLA 36175"
## [8] "Barritas Pina 67g Prom MLA 3303"
## [9] "Barritas Pina Kc 12p 768g MLA 5716"
## [10] "Besos de Nuez 120g CU GBI 35249"
## [11] "Bigotes Crema Pastelera 60g MTA TR 36911"
## [12] "Bimbollos Extra 6p 480g Prom MTA BIM 30706"
## [13] "Bimbollos Extra 6p 480g Prom TNB BIM 35294"
## [14] "Bimbunuelos con Canela 2p 33g MTA BIM 32301"
## [15] "Bollos BK 5in 6p 489g TAB SL 46214"
## [16] "Bolsita Surtido Rico 160g Tira CU GBI 33359"
## [17] "Bolsita Teddy Bear 350g GBI 33693"
## [18] "Bombonete 55g MLA 36060"
## [19] "Bran Frut con Avena 48g MTA BIM 35802"
```

[20] "Bran Frut Mango 48g BIM 30218"
 ## [21] "Bran Frut Mango 48g MTB BIM 30265"
 ## [22] "Burritos Sincro 170g CU LON 53"
 ## [23] "Butter Milk Bread 680g ORO 9756"
 ## [24] "Cajilla Galletones Naturales 1140g CU TR 36791"
 ## [25] "Canelitas Navidad 60g MTA MLA 37177"
 ## [26] "Canelitas Tubo 33p 180g MTB MLA 32361"
 ## [27] "Choco Roles 2p 80g TAB MLA 46240"
 ## [28] "Chocochips 100 cal 8P 22g GBI 48303"
 ## [29] "Clasicas de Regalo Sabor Chocol 216g GBI 36041"
 ## [30] "Combo Salma mas Levite1360g Prom CUB SAN 36927"
 ## [31] "Deliciosas Betunadas 180g CU LAR 33657"
 ## [32] "Deliciosas Chochitos 180g LAR 35127"
 ## [33] "Deliciosas Chochitos 180g MTB LAR 35128"
 ## [34] "Deliciosas Chochitos 180g Prom MTA LAR 35761"
 ## [35] "Deliciosas Chocochispas 280g SP MTA LAR 35704"
 ## [36] "Deliciosas Chocolate 180g SP LAR 35124"
 ## [37] "Deliciosas Chocolate 180g SP MTA LAR 35125"
 ## [38] "Dona 1p 26g Tira BIM 49928"
 ## [39] "Dona 1p 26g Tira MTA BIM 36932"
 ## [40] "Dona 1p 26g Tira TAB BIM 36933"
 ## [41] "Doraditas 4p TR 1641"
 ## [42] "Duo Barrita FsaPna 110g Prom MTA MLA 35558"
 ## [43] "Duo Bran Fresa Pina 96g SP BIM 35724"
 ## [44] "Duo BranFrut Fsa mas 3 Avena 93g BIM 35822"
 ## [45] "Duo Plativolos 148g MLA 32363"
 ## [46] "Duo Plativolos 148g MTA MLA 32364"
 ## [47] "Empanizador Hojuela 185g MTA BIM 48695"
 ## [48] "Galleta Almenuez 57g SP Tira TR 37222"
 ## [49] "Galleta Almenuez Bajio 57g CR2 TR 36735"
 ## [50] "Galleta Almenuez Bajio 57g Tira MTB TR 36310"
 ## [51] "Galleta Arandaleta 57g SP Tira TR 37220"
 ## [52] "Galleta Cereleta 57g Tira MTB TR 36006"
 ## [53] "Galleta Cereleta MG 57g CR2 TR 36732"
 ## [54] "Galleta Confeti Prueb 5p 65g Tira TR 36291"
 ## [55] "Galletas Panditas 3p 33g Tira MLA 35754"
 ## [56] "Galleton Chocolate 57g Tira MTB TR 36000"
 ## [57] "Gansi Premio 1p 50g MLA 31013"
 ## [58] "Lata Corazon Mixta 10Mayo Gen 200g GBI 36240"
 ## [59] "Lata Cuadrad Mix Navidad 200g CU GBI 36963"
 ## [60] "Lors Tubo 18p 148g SP MTB MLA 31469"
 ## [61] "Mantecadas 1p 33g BIM 40507"
 ## [62] "Mantecadas 1p 33g MTB BIM 40820"
 ## [63] "Mantecadas Vain 16p 500g Prom Cj BIM 36806"
 ## [64] "Mantecadas Vain 6p 188g Det Prom MTA BIM 44950"
 ## [65] "Mi Gansito 5p 125g ME TAB MLA 32954"
 ## [66] "Mini Doraditas 4p 60g MTB TR 33048"
 ## [67] "Mini Gansito Pack 3p 75g MTB MLA 36128"
 ## [68] "Mini Rocko 13g CU MLA 37185"
 ## [69] "Orejas 4p 51g Tira TR 36924"
 ## [70] "Orejitas 5g GBI 49552"
 ## [71] "Pack Galleta Leche Gansito 262g CU MLA 36169"
 ## [72] "Pan Blanco Rainbo 28Reb MTA BIM 36153"
 ## [73] "Pan Molido 240g 14pct mas MTA BIM 43278"

```

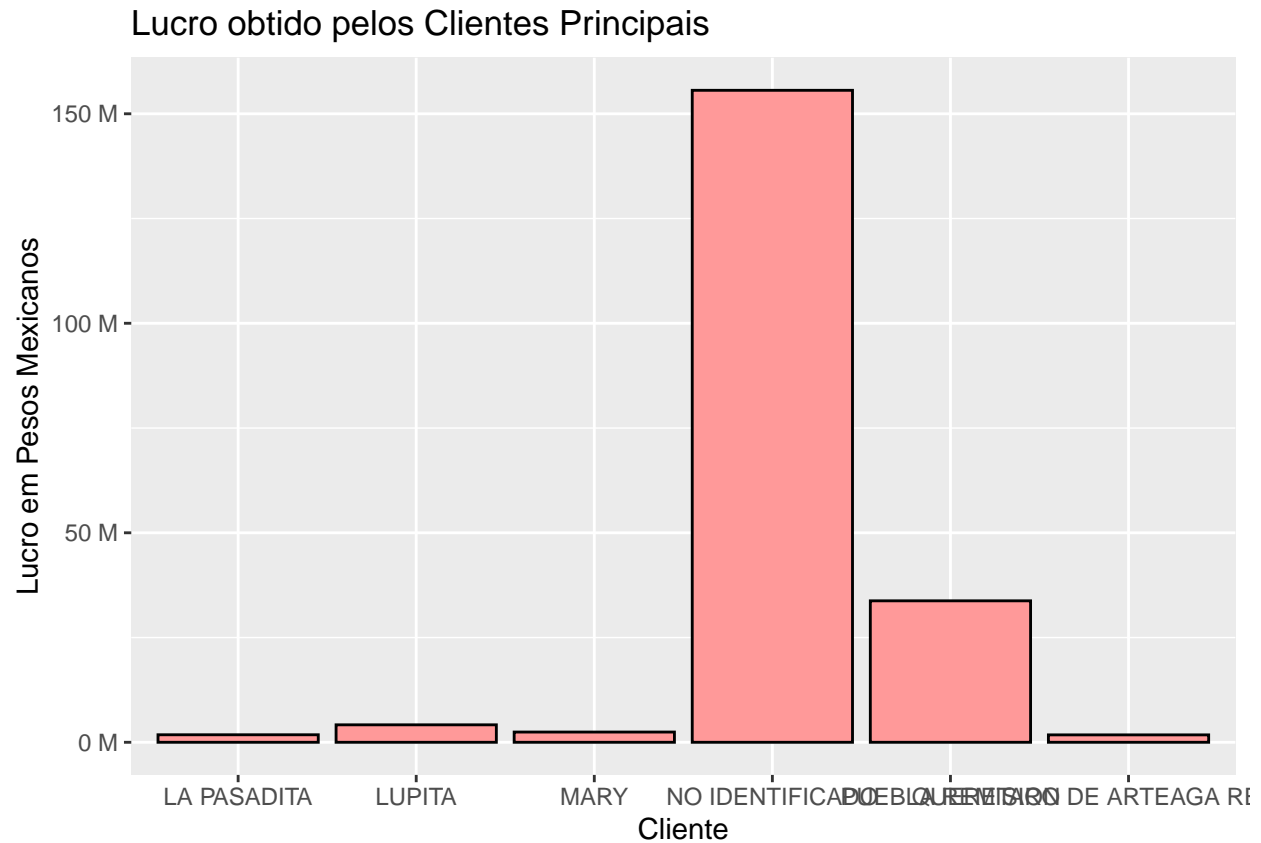
## [74] "Pan Multigrano 540g TNB BIM 35851"
## [75] "Pan Multigrano 610g MTA BIM 37131"
## [76] "Pan Multigrano Linaza 610g TNB BIM 37137"
## [77] "Pan Tost Silueta 8 Cereales 250g CU BIM 35442"
## [78] "Pan Tostado con Fibra 270g CU BIM 35441"
## [79] "Pan Tostado Tradicional Indi 180g CU BIM 35444"
## [80] "Panera Polvoron 14p 65g TR 4946"
## [81] "Panetone Suandy 750g SUA 37191"
## [82] "Principe Crunch 8p 60g Cj CU MLA 33004"
## [83] "Principe Crunch 8p 60g Cj MTA MLA 33353"
## [84] "Principe mas Navigalleta42 5g Cj MTA MLA 37232"
## [85] "Saladas 175g CU LAR 32590"
## [86] "Salmas 18g CJM SAN 32015"
## [87] "Salmas 18g CU SAN 35095"
## [88] "Salmas Duo 36g TBE SAN 33286"
## [89] "Sandwich Blanco Jamon 90g Prom LON 32220"
## [90] "Sandwich Integral PPlus 112g MTA LON 31670"
## [91] "Sandwichera Epix Futbol 72g BIM 31799"
## [92] "Spirrones 25g CU BAR 31861"
## [93] "Sponch Navidad 120g Cj MTA MLA 43291"
## [94] "Suavicremas Fresa 158g MTA MLA 34548"
## [95] "Suavicremas Fresa KC 480g MLA 5718"
## [96] "Submarinos Chocolate 3p 105g MTA MLA 30010"
## [97] "Super Pan Bco Ajonjoli 680g SP MTA WON 43381"
## [98] "Surtido Fino 750g GBI 48000"
## [99] "Thins Multigrano 340g TAB ORO 32885"
## [100] "Tortilla Hna RB 10p 260g Prom 2p DH 35438"
## [101] "Tortillinas 10p 255g Prom MTA TR 41935"
## [102] "Tortillinas 12p 310g MTA NTE TR 40606"
## [103] "Tortillinas 12p 310g Prom MTA TR 49781"
## [104] "Tortillinas 12p 310g Prom TR 30370"
## [105] "Tortillinas 15p 390g MTA TR 37639"
## [106] "Tortillinas 22p 570g NTE MTA TR 49495"
## [107] "Tortillinas 22p 570g Prom TR 30505"
## [108] "Tostachos Monchis 75g Tira BAR 32426"
## [109] "Tostada Kodyz 20p 350g TBE KOD 30746"
## [110] "Tostada Plana 15p 175g TAB MR 46088"
## [111] "Tostadita Mediana 184g CJM SAN 31191"
## [112] "Tostado Blanco 180g TAB WON 49989"
## [113] "Totopos Redondo 350g MTA MR 43097"
## [114] "Triki Intenso 20g Cj CU MLA 35862"
## [115] "Triki Intenso 8p 80g CU MLA 35840"
## [116] "Triki Trakes 8p 69g Prom CU MLA 36529"
## [117] "Triki Trakes 8p 69g Prom MTB MLA 40930"
## [118] "TuinKids Naranja 2p 50g MTA MLA 36224"
## [119] "Twin Pack Thins Multig 16p 680g MTA ORO 36289"
## [120] "Wonder 100pct Ajonjoli 680g SP MTA WON 43359"
## [121] "Wonder 100pct Ajonjoli 680g SP WON 4468"
## [122] "Wonder 100pct gde 680g SP MTA WON 43339"

```

```

Cliente_Vendas <- aggregate(Venta_hoy ~ NombreCliente, datos, sum)

```

Como pode ser observado acima, a maioria dos clientes não possuem cadastro na empresa.

Etapa 4: Construindo Modelos Preditivos

```
# Criando Datasets de treino e teste
```

```
library(caret)
```

```
## Loading required package: lattice
```

```
sample2 <- sample(1:nrow(dados), 350000)
sample3 <- sample(350001:nrow(dados), 150000)
treino <- dados[sample2,]
teste <- dados[sample3,]
```

```
# Criando o Modelo
```

```
modelo_v1 <- train(Demanda_uni_equil ~ Venta_hoy + Venta_uni_hoy, data = treino, method = 'lm')
summary(modelo_v1)
```

```
##
## Call:
## lm(formula = .outcome ~ ., data = dat)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -974.04    0.01     0.03     0.07    39.88
##
## Coefficients:
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  1.197e-02  3.506e-03   3.414 0.000641 ***
## Venta_hoy    1.857e-05  3.046e-06   6.097 1.08e-09 ***
## Venta_uni_hoy 9.867e-01  1.621e-04 6086.951 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.96 on 349997 degrees of freedom
## Multiple R-squared:  0.9913, Adjusted R-squared:  0.9913
## F-statistic: 1.993e+07 on 2 and 349997 DF,  p-value: < 2.2e-16
```

```
# Importância das variáveis
```

```
varImp(modelo_v1)
```

```
## lm variable importance
```

```
##
```

```
##              Overall
```

```
## Venta_uni_hoy    100
```

```
## Venta_hoy        0
```

```
# Realizando a previsão
```

```
previsao <- predict(modelo_v1, teste)
```

```
previsao <- round(previsao)
```

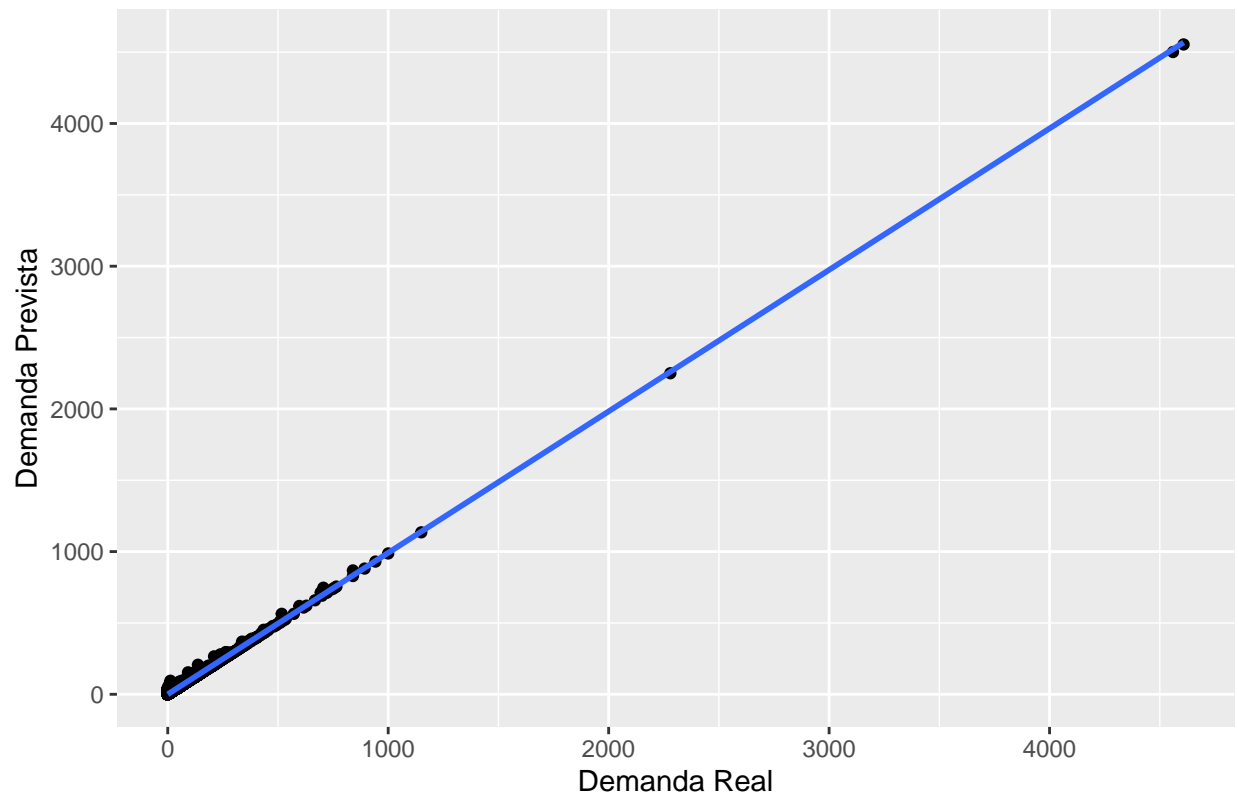
```
# Precisão do modelo
```

```
mean(previsao==teste$Demanda_uni_equil)
```

```
## [1] 0.9447733
```

```
## 'geom_smooth()' using formula 'y ~ x'
```

Modelo de Regressão Linear



Fim

www.github.com/henryjordan