

Análise Campanha de Marketing Bancário

Daniel Falcari

2025-01-30

Projeto de Análise de Campanha de Marketing Bancário

Bank Marketing Dataset

É um conjunto de dados, disponível no Kaggle. Ele se refere a uma campanha de marketing direto de uma instituição bancária, cujo objetivo era identificar clientes que poderiam se interessar por um produto financeiro específico: um depósito a prazo (deposit). O dataset foi utilizado analisar se um cliente subscreverá (ou não) o depósito a prazo com base em suas características demográficas.

Vamos usar o dataset do Kaggle: Bank Marketing <https://www.kaggle.com/datasets/janiobachmann/bank-marketing-dataset/data>

Abaixo, detalho as principais informações sobre o dataset.

Colunas e Valores Presentes

O dataset contém várias colunas (atributos) que descrevem características dos clientes, detalhes da campanha de marketing e informações contextuais. Aqui estão as colunas e seus significados:

Informações Demográficas dos Clientes: `age`: Idade do cliente (numérica).

`job`: Tipo de emprego (categórica: admin, blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, unknown).

`marital`: Estado civil (categórica: divorced, married, single).

`education`: Nível de educação (categórica: primary, secondary, tertiary, unknown).

`default`: Indica se o cliente tem crédito em inadimplência (categórica: no, yes, unknown).

`balance`: Saldo médio anual em euros (numérica).

`housing`: Indica se o cliente tem um empréstimo imobiliário (categórica: no, yes, unknown).

`loan`: Indica se o cliente tem um empréstimo pessoal (categórica: no, yes, unknown).

Informações Relacionadas à Campanha de Marketing: `contact`: Tipo de comunicação utilizada (categórica: cellular, telephone).

`day`: Último dia do mês em que o cliente foi contatado (numérica: 1 a 31).

`month`: Último mês em que o cliente foi contatado (categórica: jan, feb, mar, ..., nov, dec).

`duration`: Duração do último contato em segundos (numérica). (Nota: Este atributo não deve ser usado para treinamento, pois só é conhecido após o contato.)

campaign: Número de contatos realizados durante esta campanha para este cliente (numérica).

pdays: Número de dias desde o último contato de uma campanha anterior (numérica: -1 indica que o cliente não foi contatado anteriormente).

previous: Número de contatos realizados antes desta campanha (numérica).

poutcome: Resultado da campanha de marketing anterior (categórica: failure, nonexistent, success).

Variável Alvo (Target): deposit: Indica se o cliente subscreveu o depósito a prazo (categórica: yes, no). (Esta é a variável que se deseja prever.)

ASK - Perguntas Gerais Sobre a Campanha

Qual é a taxa de sucesso geral da campanha?

Qual é o perfil demográfico dos clientes que mais aderiram (idade, profissão, estado civil, nível de educação)?

PREPARE - IMPORTANDO E ENTENDENDO A BASE

```
# Verificando o caminho padrão do RStudio
getwd()
```

```
## [1] "C:/Users/LATITUDE/Documents/Linguagem_R"
```

```
# Definindo o caminho onde está salvo o banco de dados
# Obs: Inverter as barras ("\") para ("/") não ocorrer erro
setwd("C:/Users/LATITUDE/Documents/Linguagem_R/Dataset")
```

```
# Importando o dplyr e o tidyr
library(dplyr)
```

```
##
## Anexando pacote: 'dplyr'
```

```
## Os seguintes objetos são mascarados por 'package:stats':
##
##     filter, lag
```

```
## Os seguintes objetos são mascarados por 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
# Importando o ggplot2 e forcats
library(ggplot2)
library(forcats)
```

```
# Importando o dataset dos dados da campanha de marketing
dados_bank <- read.csv('bank.csv')
colnames(dados_bank)
```

```
## [1] "age"      "job"      "marital"  "education" "default"  "balance"
## [7] "housing"  "loan"     "contact"  "day"       "month"    "duration"
## [13] "campaign" "pdays"   "previous" "poutcome"  "deposit"
```

```
# Visualizando as primeiras linhas
head(dados_bank, 10)
```

```
##   age      job marital education default balance housing loan contact day
## 1  59   admin. married secondary    no   2343    yes  no unknown  5
## 2  56   admin. married secondary    no    45    no  no unknown  5
## 3  41 technician married secondary    no  1270    yes  no unknown  5
## 4  55  services married secondary    no  2476    yes  no unknown  5
## 5  54   admin. married tertiary    no   184    no  no unknown  5
## 6  42 management single tertiary    no    0    yes yes unknown  5
## 7  56 management married tertiary    no   830    yes yes unknown  6
## 8  60   retired divorced secondary    no   545    yes  no unknown  6
## 9  37 technician married secondary    no    1    yes  no unknown  6
## 10 28  services single secondary    no  5090    yes  no unknown  6
##   month duration campaign pdays previous poutcome deposit
## 1   may    1042         1    -1        0 unknown    yes
## 2   may    1467         1    -1        0 unknown    yes
## 3   may    1389         1    -1        0 unknown    yes
## 4   may     579         1    -1        0 unknown    yes
## 5   may     673         2    -1        0 unknown    yes
## 6   may     562         2    -1        0 unknown    yes
## 7   may    1201         1    -1        0 unknown    yes
## 8   may    1030         1    -1        0 unknown    yes
## 9   may     608         1    -1        0 unknown    yes
## 10  may    1297         3    -1        0 unknown    yes
```

```
# Verificando a estrutura desse dataset
str(dados_bank)
```

```
## 'data.frame': 11162 obs. of 17 variables:
## $ age : int 59 56 41 55 54 42 56 60 37 28 ...
## $ job : chr "admin." "admin." "technician" "services" ...
## $ marital : chr "married" "married" "married" "married" ...
## $ education: chr "secondary" "secondary" "secondary" "secondary" ...
## $ default : chr "no" "no" "no" "no" ...
## $ balance : int 2343 45 1270 2476 184 0 830 545 1 5090 ...
## $ housing : chr "yes" "no" "yes" "yes" ...
## $ loan : chr "no" "no" "no" "no" ...
## $ contact : chr "unknown" "unknown" "unknown" "unknown" ...
## $ day : int 5 5 5 5 5 5 6 6 6 6 ...
## $ month : chr "may" "may" "may" "may" ...
## $ duration : int 1042 1467 1389 579 673 562 1201 1030 608 1297 ...
## $ campaign : int 1 1 1 1 2 2 1 1 1 3 ...
## $ 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 : chr "unknown" "unknown" "unknown" "unknown" ...
## $ deposit : chr "yes" "yes" "yes" "yes" ...
```

```
# Verificando também a descrição estatística
summary(dados_bank)
```

```
##      age      job      marital      education
## Min.   :18.00 Length:11162 Length:11162 Length:11162
## 1st Qu.:32.00 Class :character Class :character Class :character
## Median :39.00 Mode  :character Mode  :character Mode  :character
## Mean   :41.23
## 3rd Qu.:49.00
## Max.   :95.00
##      default      balance      housing      loan
## Length:11162 Min.    :-6847 Length:11162 Length:11162
## Class :character 1st Qu.: 122 Class :character Class :character
## Mode  :character Median : 550 Mode  :character Mode  :character
##              Mean   : 1529
##              3rd Qu.: 1708
##              Max.   :81204
##      contact      day      month      duration
## Length:11162 Min.    : 1.00 Length:11162 Min.    : 2
## Class :character 1st Qu.: 8.00 Class :character 1st Qu.: 138
## Mode  :character Median :15.00 Mode  :character Median : 255
##              Mean   :15.66 Mean   : 372
##              3rd Qu.:22.00 3rd Qu.: 496
##              Max.   :31.00 Max.   :3881
##      campaign      pdays      previous      poutcome
## Min.    : 1.000 Min.    : -1.00 Min.    : 0.0000 Length:11162
## 1st Qu.: 1.000 1st Qu.: -1.00 1st Qu.: 0.0000 Class :character
## Median : 2.000 Median : -1.00 Median : 0.0000 Mode  :character
## Mean    : 2.508 Mean    : 51.33 Mean    : 0.8326
## 3rd Qu.: 3.000 3rd Qu.: 20.75 3rd Qu.: 1.0000
## Max.    :63.000 Max.    :854.00 Max.    :58.0000
##      deposit
## Length:11162
## Class :character
## Mode  :character
##
##
##
```

```
# Vamos verificar se existe algum valor vazio no banco de dados
any(is.na(dados_bank))
```

```
## [1] FALSE
```

```
# E quais são as colunas com valores vazios
colSums(is.na(dados_bank))
```

```
##      age      job      marital      education      default      balance      housing      loan
##      0      0      0      0      0      0      0      0
##      contact      day      month      duration      campaign      pdays      previous      poutcome
##      0      0      0      0      0      0      0      0
##      deposit
##      0
```

```
# Verificando se possui dados duplicados na base
duplicated_data <- dados_bank[duplicated(dados_bank), ]
print(duplicated_data)
```

```
## [1] age      job      marital  education default  balance  housing
## [8] loan     contact  day      month    duration campaign pdays
## [15] previous poutcome deposit
## <0 linhas> (ou row.names de comprimento 0)
```

PROCESS - MANIPULAÇÃO E LIMPEZA DOS DADOS

```
# Renomeando algumas colunas do dataset
dados_bank <- dados_bank %>%
  rename(
    type_contact = contact,
    housing_loan = housing,
    marital_status = marital
  )

# Verificando os valores distintos nessas colunas
distinct(dados_bank, job)
```

```
##           job
## 1      admin.
## 2  technician
## 3    services
## 4    management
## 5      retired
## 6 blue-collar
## 7    unemployed
## 8  entrepreneur
## 9    housemaid
## 10     unknown
## 11 self-employed
## 12      student
```

```
distinct(dados_bank, marital_status)
```

```
## marital_status
## 1      married
## 2       single
## 3     divorced
```

```
distinct(dados_bank, education)
```

```
## education
## 1 secondary
## 2 tertiary
## 3  primary
## 4   unknown
```

```
distinct(dados_bank, default)
```

```
## default
## 1      no
## 2      yes
```

```
distinct(dados_bank, housing_loan)
```

```
## housing_loan
## 1           yes
## 2           no
```

```
distinct(dados_bank, loan)
```

```
## loan
## 1   no
## 2   yes
```

```
distinct(dados_bank, type_contact)
```

```
## type_contact
## 1      unknown
## 2      cellular
## 3      telephone
```

```
distinct(dados_bank, poutcome)
```

```
## poutcome
## 1 unknown
## 2   other
## 3 failure
## 4 success
```

```
distinct(dados_bank, deposit)
```

```
## deposit
## 1      yes
## 2      no
```

```
# Verificando os valores em cada um desses rótulos
table(dados_bank$poutcome)
```

```
##
## failure   other success unknown
##    1228     537    1071    8326
```

```
table(dados_bank$job)
```

```
##
##      admin.  blue-collar  entrepreneur  housemaid  management
##      1334      1944      328      274      2566
##      retired self-employed  services  student  technician
##      778      405      923      360      1823
##      unemployed  unknown
##      357      70
```

```
# Agrupando na coluna poutcome "other" e "unknown" em "unknown" na coluna poutcome
dados_bank <- dados_bank %>%
  mutate(poutcome= ifelse(poutcome %in% c("other", "unknown"), "unknown", poutcome))

# Resumo estatístico da Distribuição demográfica em relação a coluna target
table(dados_bank$marital_status, dados_bank$deposit)
```

```
##
##           no  yes
## divorced  671  622
## married   3596 2755
## single    1606 1912
```

```
table(dados_bank$education, dados_bank$deposit)
```

```
##
##           no  yes
## primary    909  591
## secondary  3026 2450
## tertiary   1693 1996
## unknown    245  252
```

```
table(dados_bank$job, dados_bank$deposit)
```

```
##
##           no  yes
## admin.      703  631
## blue-collar 1236  708
## entrepreneur 205  123
## housemaid    165  109
## management   1265 1301
## retired      262  516
## self-employed 218  187
## services     554  369
## student      91  269
## technician   983  840
## unemployed   155  202
## unknown     36   34
```

ANALYZE - ANALISANDO E VISUALIZANDO INFORMAÇÕES

Respondendo a primeira pergunta: Qual é a taxa de sucesso geral da campanha? Para isso será realizado o cálculo da porcentagem dos clientes que realizaram e dos que não realizaram o depósito, apresentando o resultado em um gráfico de pizza.

```
# Obtendo os dados de porcentagem de quantos clientes fizeram o depósito a prazo
dados_pizza <- dados_bank %>%
  group_by(deposit) %>% # Agrupa os dados por "deposit"
  summarise(count = n()) %>% # Conta o número de ocorrências
  mutate(percentage = round(100 * count / sum(count), 1)) # Calcula a porcentagem

print(dados_pizza)
```

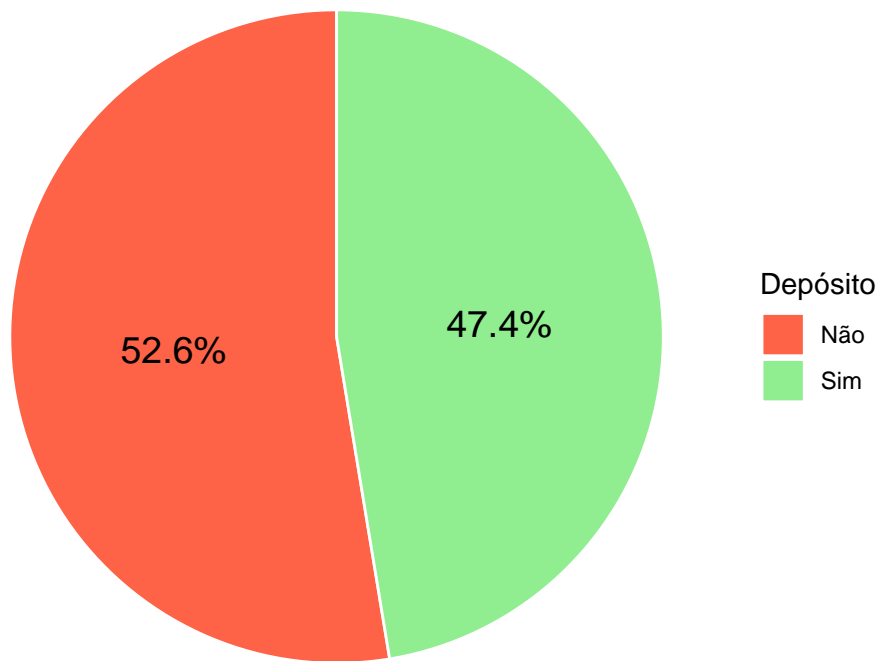
```
## # A tibble: 2 x 3
##   deposit count percentage
##   <chr>   <int>     <dbl>
## 1 no       5873       52.6
## 2 yes      5289       47.4
```

```
# Adicionar uma coluna para adicionar o símbolo de porcentagem
dados_pizza <- dados_pizza %>%
  mutate(label = paste0(percentage, "%"))

# Criar o gráfico de pizza para mostrar a porcentagem de quem fez ou não o depósito a prazo
grafico_pizza <- ggplot(dados_pizza, aes(x = "", y = percentage, fill = deposit)) +
  geom_bar(width = 1, stat = "identity", color = "white") +
  coord_polar("y", start = 0) + # Iniciar em 0 para uma apresentação mais limpa
  geom_text(aes(label = label), position = position_stack(vjust = 0.5), color = "black", size = 5) +
  theme_void() + # Tema minimalista sem eixos e grid
  scale_fill_manual(
    values = c("yes" = "lightgreen", "no" = "tomato"),
    name = "Depósito",
    labels = c("yes" = "Sim", "no" = "Não")
  ) + # Personalizar cores, nome da legenda e rótulos
  labs(title = "Depósitos a Prazo: Porcentagens") +
  theme(
    plot.title = element_text(hjust = 0.5)
  )

# Exibir o gráfico
print(grafico_pizza)
```


Depósitos a Prazo: Porcentagens



A taxa de sucesso da campanha baseada no dados foi de 47,4%

Analizando os dados como objetivo de responder a segunda pergunta

Qual é o perfil demográfico dos clientes que mais aderiram?

Analizando a Idade

```
# Estatísticas descritivas da idade  
summary(dados_bank$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##   18.00   32.00   39.00   41.23   49.00   95.00
```

```
# Média e desvio padrão  
sd(dados_bank$age)
```

```
## [1] 11.91337
```

```
# Filtrando os clientes que aderiram à campanha (deposit = 'yes')
dados_adquiridos <- dados_bank %>% filter(deposit == "yes")
```

```
# Filtrando os clientes que NÃO aderiram à campanha (deposit = 'no')
dados_no_adquiridos <- dados_bank %>% filter(deposit == "no")
```

```
# Estatísticas descritivas da idade dos que realizaram o depósito
summary(dados_adquiridos$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    18.00   31.00   38.00   41.67   50.00   95.00
```

```
# Média e desvio padrão dos que realizaram o depósito
mean(dados_adquiridos$age)
```

```
## [1] 41.67007
```

```
sd(dados_adquiridos$age)
```

```
## [1] 13.49778
```

```
# Calcular as porcentagens de depósitos a prazo por faixa etária
```

```
faixas_etarias <- dados_bank %>%
```

```
  mutate(faixa_etaria = cut(age, breaks = seq(17, 95, by = 6))) %>% # Ajuste os intervalos conforme ne
```

```
  group_by(faixa_etaria) %>%
```

```
  summarise(
```

```
    total = n(),
```

```
    depositaram = sum(deposit == "yes"),
```

```
    porcentagem = (depositaram / total) * 100
```

```
  ) %>%
```

```
  arrange(desc(porcentagem))
```

```
# Exibir os resultados
```

```
print(faixas_etarias)
```

```
## # A tibble: 13 x 4
```

```
##   faixa_etaria total depositaram porcentagem
```

```
##   <fct>          <int>      <int>      <dbl>
```

```
## 1 (89,95]         7         7        100
```

```
## 2 (65,71]       152        127      83.6
```

```
## 3 (71,77]       145        114      78.6
```

```
## 4 (77,83]        72         56      77.8
```

```
## 5 (17,23]       187        139      74.3
```

```
## 6 (59,65]       382        280      73.3
```

```
## 7 (83,89]        22         16      72.7
```

```
## 8 (23,29]      1364        789      57.8
```

```
## 9 (29,35]      2820       1261      44.7
```

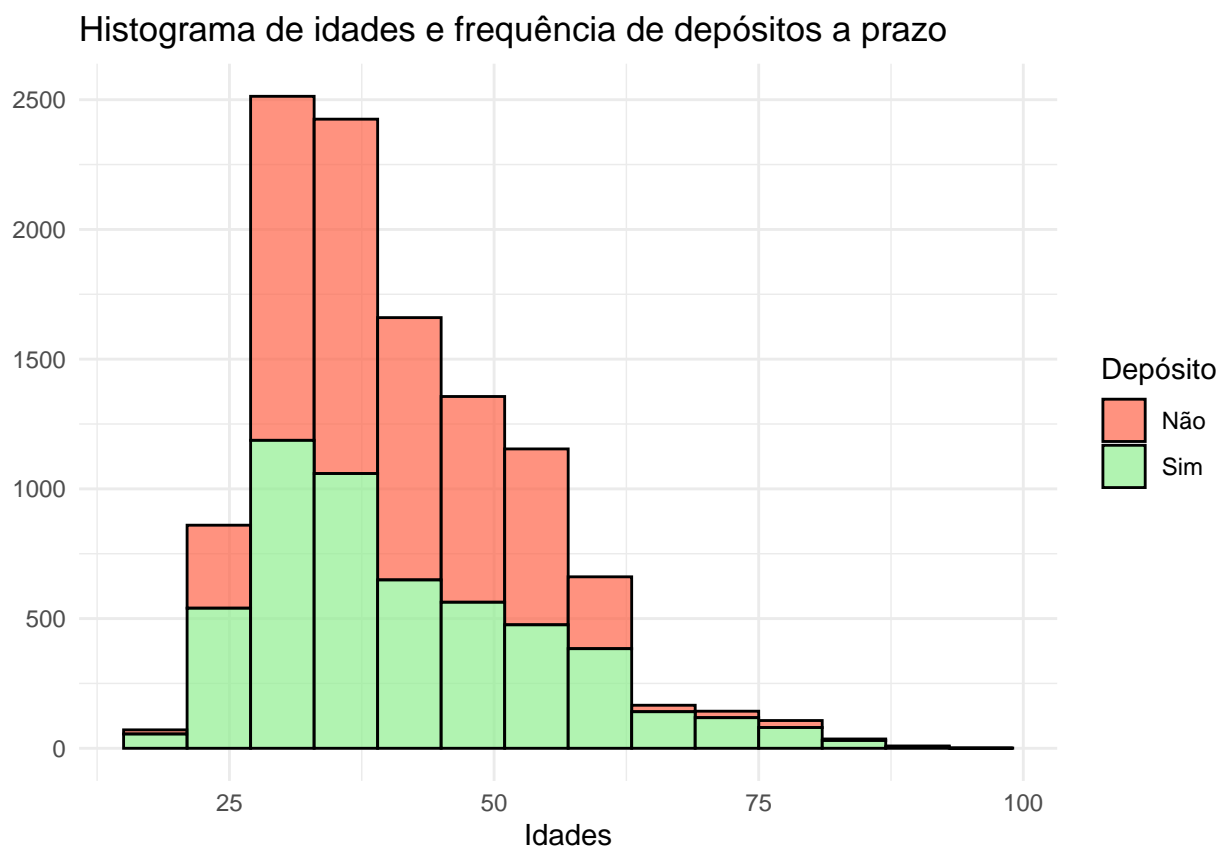
```
## 10 (53,59]     1088        466      42.8
```

```
## 11 (35,41]     2123        888      41.8
```

```
## 12 (41,47]     1548        644      41.6
```

```
## 13 (47,53]     1252        502      40.1
```

```
# Histograma de distribuição de idade separando por quem fez o depósito a prazo
ggplot(dados_bank, aes(x = age)) +
  geom_histogram(binwidth = 6, aes(fill = deposit), alpha = 0.7, color = "black") +
  labs(
    title = "Histograma de idades e frequência de depósitos a prazo",
    x = "Idades",
    y = NULL # Remove o rótulo do eixo Y
  ) +
  scale_fill_manual(
    values = c("yes" = "lightgreen", "no" = "tomato"),
    name = "Depósito",
    labels = c("yes" = "Sim", "no" = "Não")
  ) +
  theme_minimal()
```



Percepções das Análises em Relação à Idade

Insights

- **Clientes Idosos (60 a 95 anos):** Impressionantes 81% deste grupo realizaram depósitos a prazo
- **Jovens (18 a 23 anos):** Uma significativa parcela de 74,3% deste segmento fez depósitos a prazo
- **Jovens Adultos (24 a 29 anos):** Aproximadamente 57,8% dos jovens adultos optaram por depósitos a prazo
- **Adultos (36 a 59 anos):** Esta faixa etária apresentou o pior desempenho na campanha, com apenas 41,6% realizando depósitos a prazo.

Analisando as Profissões

```
# Tabela de frequências para análise das profissões
table(dados_bank$job)
```

```
##
##      admin.    blue-collar  entrepreneur    housemaid    management
##      1334      1944          328          274          2566
##      retired self-employed    services      student    technician
##      778        405          923          360          1823
##      unemployed      unknown
##      357            70
```

```
table(dados_adquiridos$job)
```

```
##
##      admin.    blue-collar  entrepreneur    housemaid    management
##      631      708          123          109          1301
##      retired self-employed    services      student    technician
##      516      187          369          269          840
##      unemployed      unknown
##      202            34
```

```
# Cálculo da porcentagem de cada profissão dos clientes com base no depósito a prazo (coluna target)
job_porcentagem <- dados_bank %>%
  group_by(job, deposit) %>%
  summarise(frequencia = n()) %>%
  mutate(percentual = frequencia / sum(frequencia) * 100)
```

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

```
print(job_porcentagem)
```

```
## # A tibble: 24 x 4
## # Groups:   job [12]
##   job      deposit frequencia percentual
##   <chr>      <chr>      <int>      <dbl>
## 1 admin.    no           703        52.7
## 2 admin.    yes          631        47.3
## 3 blue-collar no        1236        63.6
## 4 blue-collar yes         708        36.4
## 5 entrepreneur no         205        62.5
## 6 entrepreneur yes         123        37.5
## 7 housemaid no         165        60.2
## 8 housemaid yes         109        39.8
## 9 management no        1265        49.3
## 10 management yes         1301        50.7
## # i 14 more rows
```

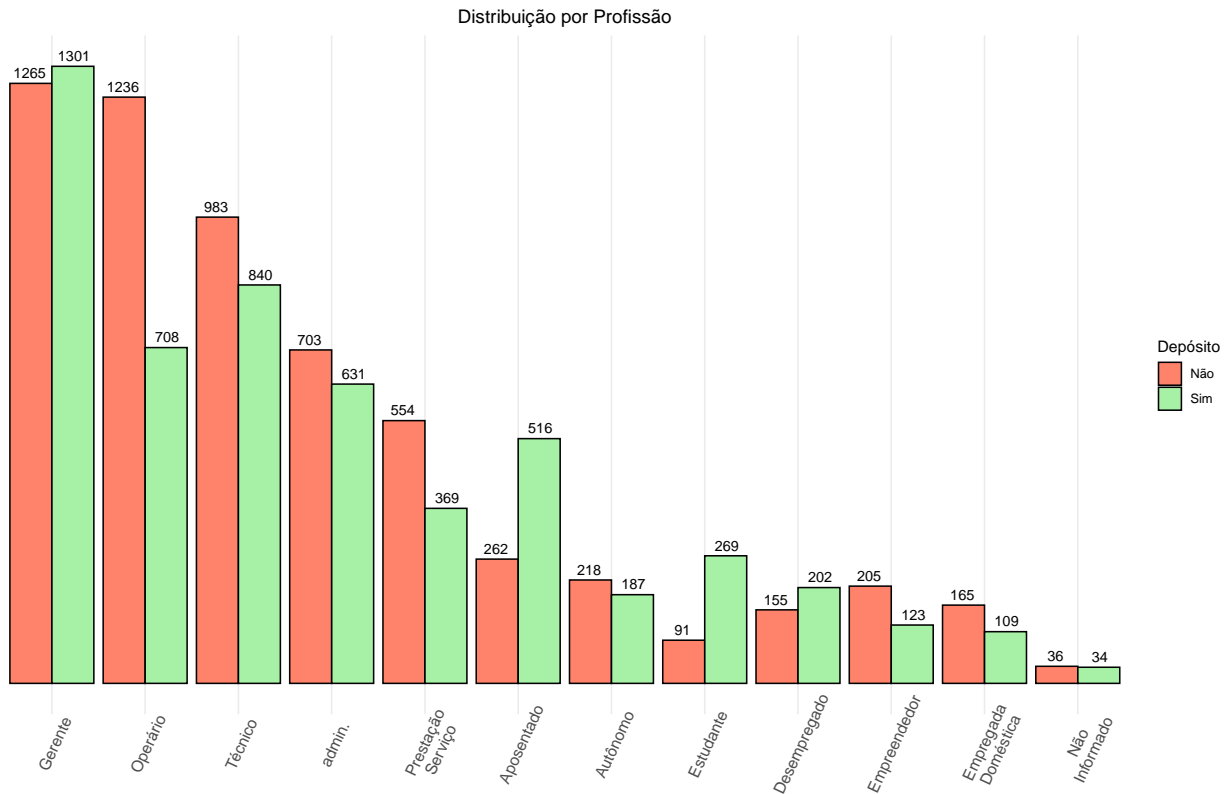
```

# Reorganizar a variável 'job' em ordem decrescente
dados_bank <- dados_bank %>%
  mutate(job = fct_infreq(job))

# Criar o gráfico de profissões dos clientes com o depósito a prazo (coluna target)
job_graf <- ggplot(dados_bank, aes(x = job, fill = deposit)) +
  geom_bar(position = position_dodge(width = 0.9), alpha = 0.8, color = "black") +
  geom_text(
    stat = "count",
    aes(label = ..count..),
    position = position_dodge(width = 0.95),
    vjust = -0.5,
    size = 3.5
  ) +
  labs(
    title = "Distribuição por Profissão",
    x = NULL,
    y = NULL # Remove o rótulo do eixo Y
  ) +
  scale_y_continuous(breaks = NULL) + # Remove os valores no eixo Y
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 65, size = 11, vjust = 0.8, hjust = 0.5), # Aumenta o tamanho e
    panel.grid.major.y = element_blank(), # Remove as linhas de grade horizontais
    axis.title.x = element_text(size = 14),
    plot.title = element_text(hjust = 0.5)
  ) +
  scale_fill_manual(
    values = c("yes" = "lightgreen", "no" = "tomato"),
    name = "Depósito",
    labels = c("yes" = "Sim", "no" = "Não")
  ) +
  scale_x_discrete(
    labels = c("admin" = "Administrador", "blue-collar" = "Operário",
      "entrepreneur" = "Empreendedor", "housemaid" = "Empregada\nDoméstica",
      "management" = "Gerente", "retired" = "Aposentado",
      "self-employed" = "Autônomo", "services" = "Prestação\nServiço",
      "student" = "Estudante", "technician" = "Técnico", "unemployed" = "Desempregado", "unknown" = "Não")
  )

print(job_graf)

```



```
# Filtrar apenas os casos onde deposit = "yes"
job_yes <- job_porcentagem %>%
  filter(deposit == "yes") %>%
  arrange(desc(percentual)) # Ordenar do maior para o menor percentual

print(job_yes)
```

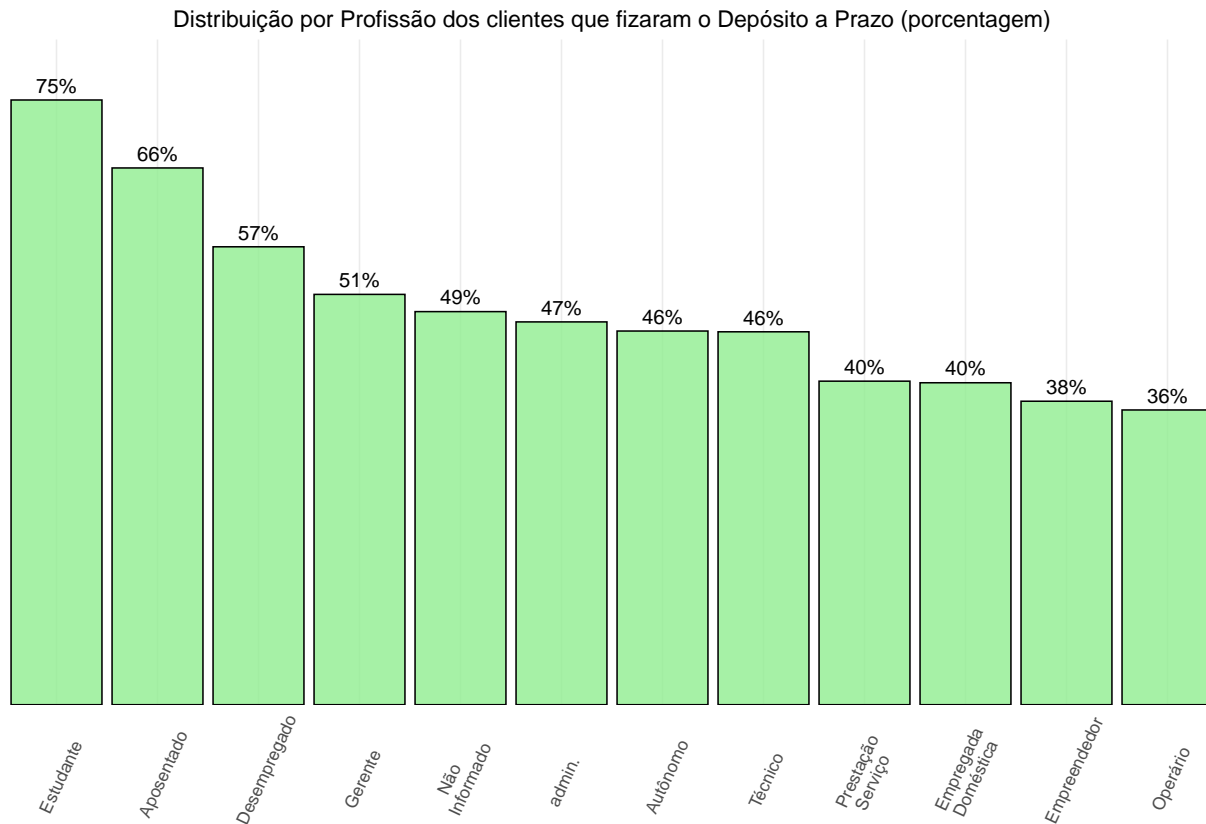
```
## # A tibble: 12 x 4
## # Groups:   job [12]
##   job      deposit frequencia percentual
##   <chr>      <chr>      <int>      <dbl>
## 1 student    yes          269        74.7
## 2 retired    yes          516        66.3
## 3 unemployed yes          202        56.6
## 4 management yes         1301        50.7
## 5 unknown    yes           34        48.6
## 6 admin.     yes          631        47.3
## 7 self-employed yes         187        46.2
## 8 technician yes          840        46.1
## 9 services   yes          369        40.0
## 10 housemaid yes          109        39.8
## 11 entrepreneur yes         123        37.5
## 12 blue-collar yes          708        36.4
```

```
# Criar o gráfico com porcentagens, relacionando as profissões dos clientes com o depósito a prazo (co
ggplot(job_yes, aes(x = reorder(job, -percentual), y = percentual)) +
```

```

geom_bar(stat = "identity", fill = "lightgreen", position = position_dodge(width = 0.9), alpha = 0.8,
geom_text(
  aes(label = paste0(round(percentual, 0), "%")),
  position = position_dodge(width = 0.95),
  vjust = -0.5,
  size = 5
) +
labs(
  title = "Distribuição por Profissão dos clientes que fizeram o Depósito a Prazo (porcentagem)",
  x = NULL,
  y = NULL,
) +
scale_y_continuous(breaks = NULL, expand = expansion(mult = c(0, 0.1))) + # Ajusta o espaço acima da
theme_minimal() +
theme(
  axis.text.x = element_text(angle = 65, size = 12, vjust = 0.3, hjust = 0.3), # Aumenta e ajusta ró
  panel.grid.major.y = element_blank(), # Remove as linhas do grid horizontal
  plot.title = element_text(hjust = 0.5, size = 16)
) +
scale_fill_manual(
  values = c("yes" = "lightgreen"),
  name = "Depósito",
  labels = c("yes" = "Sim")
) +
scale_x_discrete(
  labels = c("admin" = "Administrador", "blue-collar" = "Operário",
"entrepreneur" = "Empreendedor", "housemaid" = "Empregada\nDoméstica",
"management" = "Gerente", "retired" = "Aposentado",
"self-employed" = "Autônomo", "services" = "Prestação\nServiço",
"student" = "Estudante", "technician" = "Técnico", "unemployed" = "Desempregado", "unknown" = "Não\
)

```



Percepções das Análises em Relação às Profissões

Insights: Estudantes: Com uma alta taxa de conversão de 74,7%, embora representem apenas 3% da base de dados, os estudantes mostram um forte interesse em investir, mesmo sendo um grupo menor.

Aposentados: Apresentam uma taxa de conversão de 66,3%, representando 7% da base de dados. Esse grupo demonstra uma propensão significativa para investir em opções de longo prazo, possivelmente devido à estabilidade financeira adquirida ao longo dos anos.

Desempregados: Apesar de representarem 3% da base de dados, possuem uma conversão de 56,6%. Isso sugere que, mesmo sem uma fonte de renda estável, há um interesse em investir.

Gerentes: Com uma taxa de conversão de 50,7%, representam 23% da base de dados. Isso indica um bom interesse em depósitos a prazo.

Operários: Este grupo possui a menor taxa de conversão, 36,4%, e representa 17,5% da base de dados. Isso sugere menor acesso a interesse sobre investimentos, indicando a necessidade de estratégias específicas para aumentar a adesão.

Analisando o Estado Civil

```
# Tabela de frequências para o estado civil
table(dados_bank$marital_status)
```

```
##
## divorced married single
##      1293      6351      3518
```



```
table(dados_adquiridos$marital_status)
```

```
##
## divorced   married    single
##         622      2755    1912
```

```
# Calcule a porcentagem de depósitos a prazo por estado civil
```

```
resultados_marital <- dados_bank %>%
  group_by(marital_status) %>%
  summarise(
    total = n(),
    depositaram = sum(deposit == "yes"),
    nao_depositaram = total - depositaram,
    porcentagem = (depositaram / total) * 100
  ) %>%
  arrange(desc(porcentagem))
```

```
# Exiba os resultados
```

```
print(resultados_marital)
```

```
## # A tibble: 3 x 5
```

```
##   marital_status total depositaram nao_depositaram porcentagem
##   <chr>          <int>      <int>          <int>      <dbl>
## 1 single         3518        1912          1606        54.3
## 2 divorced      1293         622           671        48.1
## 3 married       6351       2755          3596        43.4
```

```
# Reorganizar a variável 'marital_status' em ordem decrescente
```

```
dados_bank <- dados_bank %>%
  mutate(marital_status = fct_infreq(marital_status))
```

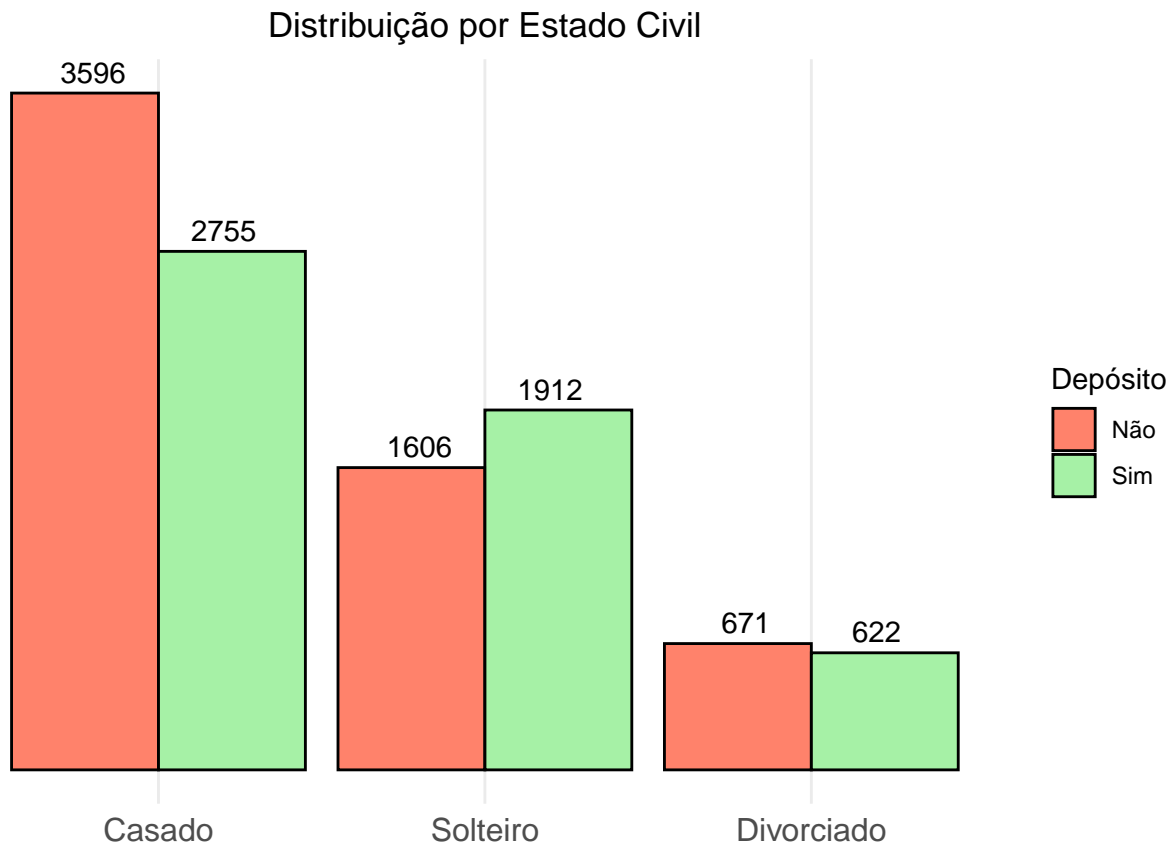
```
# Criar o gráfico com relacionando o estado civil com o depósito a prazo (coluna target)
```

```
ggplot(dados_bank, aes(x = marital_status, fill = deposit)) +
  geom_bar(position = position_dodge(width = 0.9), alpha = 0.8, color = "black") +
  geom_text(
    stat = "count",
    aes(label = ..count..),
    position = position_dodge(width = 0.8),
    vjust = -0.5,
    size = 4
  ) +
  labs(
    title = "Distribuição por Estado Civil",
    x = NULL,
    y = NULL # Remove o rótulo do eixo Y
  ) +
  scale_y_continuous(breaks = NULL) + # Remove os valores no eixo Y
  theme_minimal() +
  theme(
    axis.text.x = element_text(size = 12, vjust = 0.9, hjust = 0.5,),
    panel.grid.major.y = element_blank(), # Remove as linhas de grade horizontais
    axis.title.x = element_text(size = 14), # Aumenta o título do eixo X
  )
```

```

plot.title = element_text(hjust = 0.5)
) +
scale_fill_manual(
  values = c("yes" = "lightgreen", "no" = "tomato"),
  name = "Depósito",
  labels = c("yes" = "Sim", "no" = "Não")
) +
scale_x_discrete(
  labels = c("married" = "Casado", "single" = "Solteiro", "divorced" = "Divorciado")
)

```



```

# Calcular as porcentagens de quem fez e de quem não fez o depósito a prazo de acordo com o estado civil
marital_porcentagem <- dados_bank %>%
  group_by(marital_status, deposit) %>%
  summarise(frequencia = n()) %>%
  mutate(percentual = frequencia / sum(frequencia) * 100)

```

```

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

```

```

print(marital_porcentagem)

```

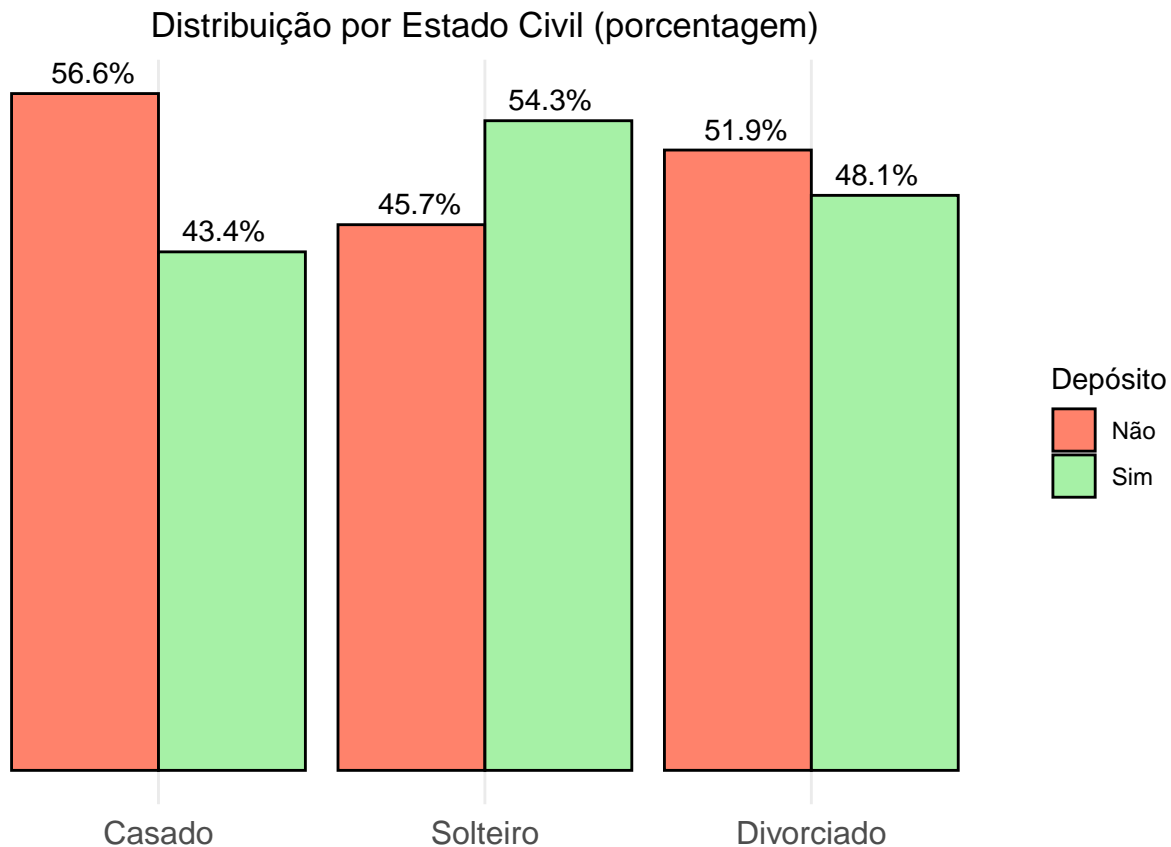
```

## # A tibble: 6 x 4
## # Groups:   marital_status [3]

```

```
## marital_status deposit frequencia percentual
## <fct>          <chr>          <int>      <dbl>
## 1 married      no             3596       56.6
## 2 married      yes             2755       43.4
## 3 single       no             1606       45.7
## 4 single       yes             1912       54.3
## 5 divorced     no              671       51.9
## 6 divorced     yes              622       48.1
```

```
# Criar o gráfico com porcentagens, relacionando o estado civil com o depósito a prazo (coluna target)
ggplot(marital_porcentagem, aes(x = marital_status, y = percentual, fill = deposit)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.9), alpha = 0.8, color = "black") +
  geom_text(
    aes(label = paste0(round(percentual, 1), "%")),
    position = position_dodge(width = 0.8),
    vjust = -0.5,
    size = 4
  ) +
  labs(
    title = "Distribuição por Estado Civil (porcentagem)",
    x = NULL,
    y = NULL
  ) +
  scale_y_continuous(breaks = NULL) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(size = 12, vjust = 0.5, hjust = 0.5), # Aumenta e ajusta rótulos do eix
    panel.grid.major.y = element_blank(), # Remove as linhas do grid horizontal
    plot.title = element_text(hjust = 0.5)
  ) +
  scale_fill_manual(
    values = c("yes" = "lightgreen", "no" = "tomato"),
    name = "Depósito",
    labels = c("yes" = "Sim", "no" = "Não")
  ) +
  scale_x_discrete(
    labels = c("married" = "Casado", "single" = "Solteiro", "divorced" = "Divorciado")
  )
```



Percepções das Análises em Relação ao Estado Civil

Insights: Solteiros: Apresentam a maior taxa de conversão de 54,3%, representando 31,5% da base de dados. Isso indica que este grupo está mais inclinado a investir, possivelmente devido a menos compromissos financeiros.

Divorciados: Com uma taxa de conversão de 48,1%, representam 11,6% da base de dados. Isso sugere que, apesar dos desafios financeiros que possam enfrentar, ainda estão dispostos a investir em depósitos a prazo.

Casados: Possuem a menor taxa de conversão de 43,4%, representando 56,9% da base de dados. Este grupo pode estar mais focado em outras prioridades financeiras, o que sugere a necessidade de estratégias específicas para aumentar o interesse em depósitos a prazo.

Analisando a Escolaridade

```
# Tabela de frequências para o nível educacional
table(dados_bank$education)
```

```
##
##  primary secondary tertiary unknown
##    1500      5476      3689      497
```

```
table(dados_adquiridos$education)
```

```
##
##   primary secondary tertiary   unknown
##      591      2450      1996      252
```

```
# Cálculo da porcentagem pela escolaridade dos clientes com base no depósito a prazo (coluna target)
education_porcentagem <- dados_bank %>%
  group_by(education, deposit) %>%
  summarise(frequencia = n()) %>%
  mutate(percentual = frequencia / sum(frequencia) * 100)
```

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

```
print(education_porcentagem)
```

```
## # A tibble: 8 x 4
## # Groups:   education [4]
##   education deposit frequencia percentual
##   <chr>      <chr>      <int>      <dbl>
## 1 primary    no           909        60.6
## 2 primary    yes          591        39.4
## 3 secondary no          3026        55.3
## 4 secondary yes          2450        44.7
## 5 tertiary  no          1693        45.9
## 6 tertiary  yes          1996        54.1
## 7 unknown   no           245        49.3
## 8 unknown   yes           252        50.7
```

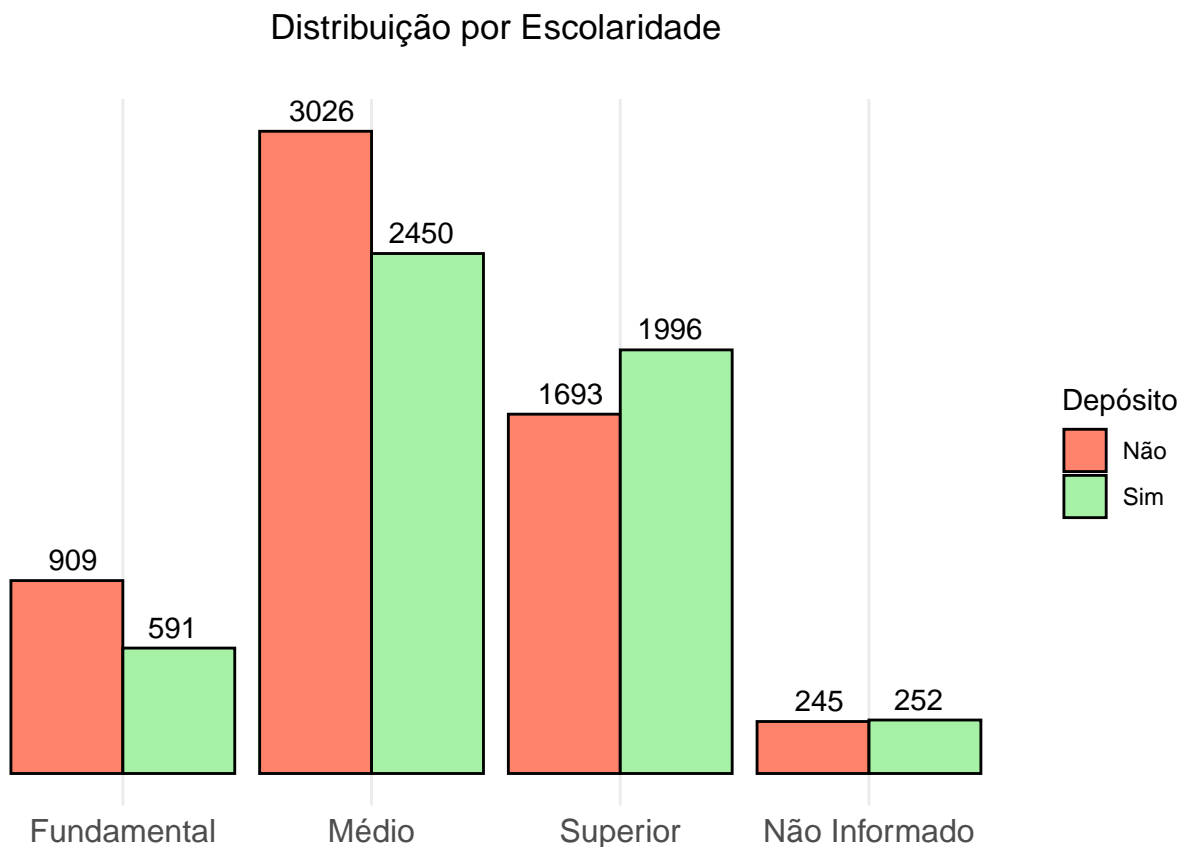
```
# Reorganizar a variável 'education' em ordem decrescente
# dados_bank <- dados_bank %>%
#   mutate(education = fct_infreq(education))
```

```
# Criar o gráfico da escolaridade dos clientes com base no depósito a prazo (coluna target)
ggplot(dados_bank, aes(x = education, fill = deposit)) +
  geom_bar(position = position_dodge(width = 0.9), alpha = 0.8, color = "black") +
  geom_text(
    stat = "count",
    aes(label = ..count..),
    position = position_dodge(width = 0.8),
    vjust = -0.5,
    size = 4
  ) +
  labs(
    title = "Distribuição por Escolaridade\n",
    x = NULL,
    y = NULL # Remove o rótulo do eixo Y
  ) +
  scale_y_continuous(breaks = NULL) + # Remove os valores no eixo Y
  theme_minimal() +
```

```

theme(
  axis.text.x = element_text(size = 12, vjust = 0.9, hjust = 0.5,), # Aumenta o tamanho e ajusta pos
  panel.grid.major.y = element_blank(), # Remove as linhas de grade horizontais
  axis.title.x = element_text(size = 14),
  plot.title = element_text(hjust = 0.5)
) +
scale_fill_manual(
  values = c("yes" = "lightgreen", "no" = "tomato"),
  name = "Depósito",
  labels = c("yes" = "Sim", "no" = "Não")
) +
scale_x_discrete(
  labels = c("unknown" = "Não Informado", "primary" = "Fundamental",
    "secondary" = "Médio", "tertiary" = "Superior")
)

```



```

# Criar o gráfico com porcentagens, relacionando a escolaridade dos clientes com base no depósito a pr
ggplot(education_porcentagem, aes(x = education, y = percentual, fill = deposit)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.9), alpha = 0.8, color = "black") +
  geom_text(
    aes(label = paste0(round(percentual, 1), "%")),
    position = position_dodge(width = 0.8),
    vjust = -0.5,
    size = 4
  ) +

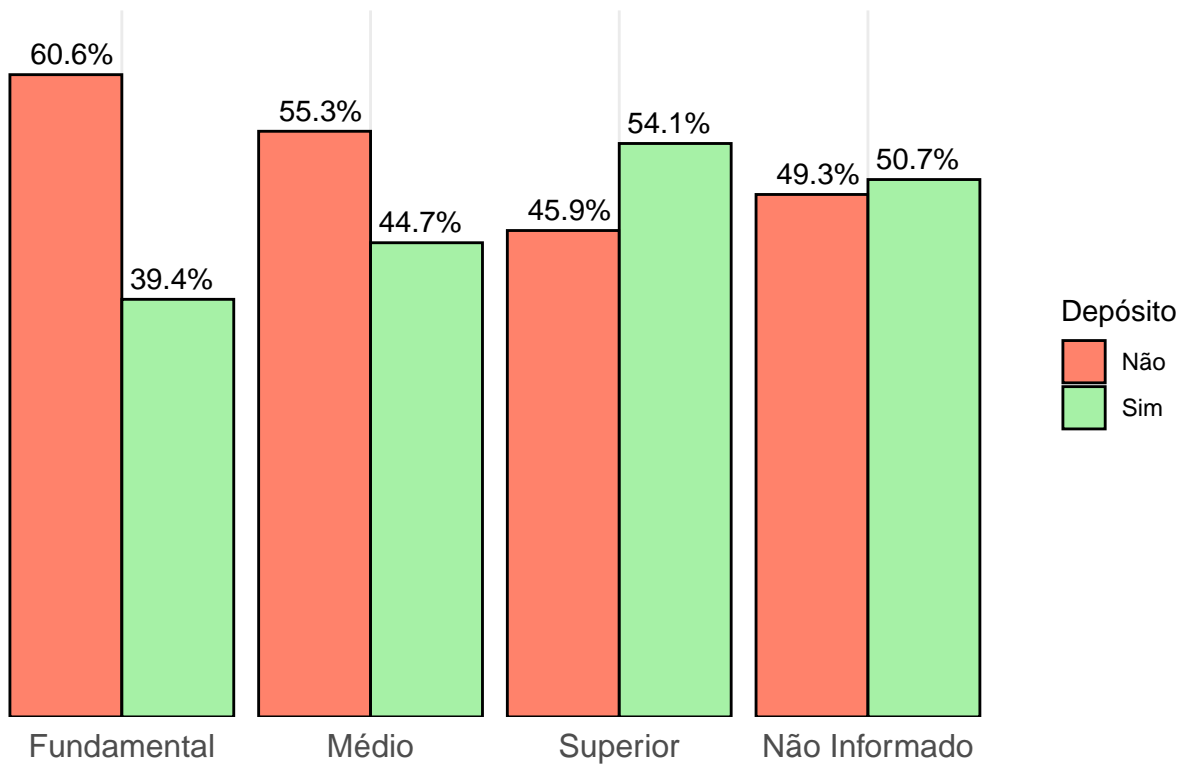
```

```

labs(
  title = "Distribuição por Escolaridade (porcentagem)\n",
  x = NULL,
  y = NULL,
) +
scale_y_continuous(breaks = NULL, expand = expansion(mult = c(0, 0.1))) + # Ajusta o espaço acima da
theme_minimal() +
theme(
  axis.text.x = element_text(size = 12, vjust = 0.5, hjust = 0.5), # Aumenta e ajusta rótulos do eix
  panel.grid.major.y = element_blank(), # Remove as linhas do grid horizontal
  plot.title = element_text(hjust = 0.5)
) +
scale_fill_manual(
  values = c("yes" = "lightgreen", "no" = "tomato"),
  name = "Depósito",
  labels = c("yes" = "Sim", "no" = "Não")
) +
scale_x_discrete(
  labels = c("unknown" = "Não Informado", "primary" = "Fundamental",
    "secondary" = "Médio", "tertiary" = "Superior")
)

```

Distribuição por Escolaridade (porcentagem)



Percepções das Análises em Relação à Escolaridade

Insights: Ensino Superior (Tertiary): Clientes com nível de escolaridade superior apresentaram uma taxa de conversão de 54,1%. Isso indica que quanto maior o nível de escolaridade, maior a propensão para investir em depósitos a prazo.

Ensino Médio (Secondary): Com uma taxa de conversão de 44,7%, os clientes com ensino médio também demonstraram um interesse significativo em depósitos a prazo, embora menor do que aqueles com ensino superior.

Ensino Fundamental (Primary): Este grupo apresentou uma taxa de conversão de 39,4%. Apresentando o menor percentual entre os níveis de escolaridade conhecidos.

Desconhecido (Unknown): Clientes com escolaridade não especificada tiveram uma taxa de conversão de 50,7%, o que é relativamente alta e destaca a importância de entender melhor este grupo para direcionar estratégias de marketing.

Analisando Saldo Bancário Médio Anual

```
# Analisando os dados estatísticos do saldo bancário médio anual
summary(dados_bank$balance)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -6847     122     550    1529    1708    81204
```

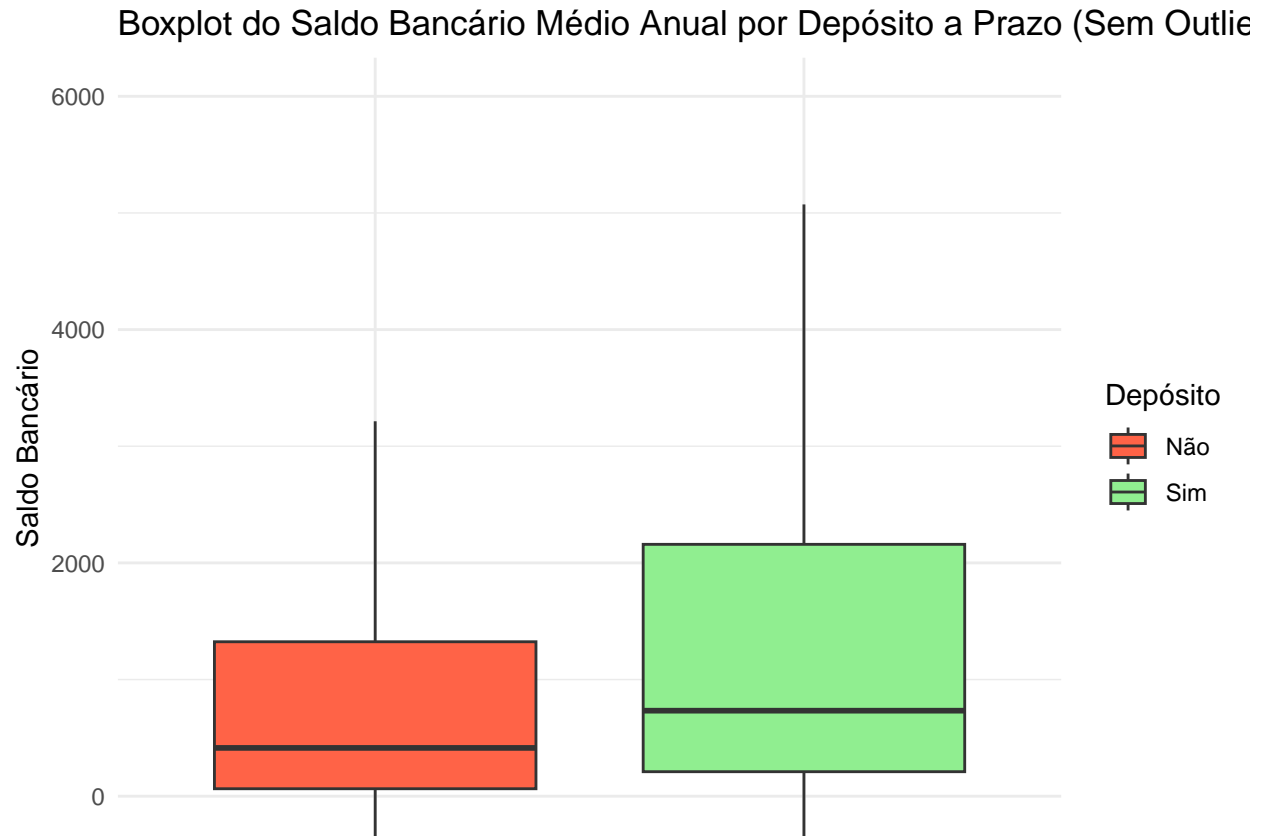
```
# Obter resumo estatístico do saldo bancário médio anual para depósitos "yes"
summary(dados_adquiridos$balance)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -3058     210     733    1804    2159    81204
```

```
# Obter resumo estatístico do saldo bancário médio anual para depósitos "no"
summary(dados_no_adquiridos$balance)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -6847      64     414    1280    1324   66653
```

```
# boxplot do saldo bancário médio anual em relação a realização ou não do depósito a prazo
ggplot(dados_bank, aes(x = deposit, y = balance, fill = deposit)) +
  geom_boxplot(outlier.shape = NA) + # Remove os outliers para melhor visualização
  coord_cartesian(ylim = c(quantile(dados_bank$balance, 0.05), quantile(dados_bank$balance, 0.95))) +
  labs(title = "Boxplot do Saldo Bancário Médio Anual por Depósito a Prazo (Sem Outliers)",
       x = NULL,
       y = "Saldo Bancário") +
  scale_fill_manual(
    values = c("tomato", "lightgreen"),
    name = "Depósito",
    labels = c("yes" = "Sim", "no" = "Não")
  ) +
  scale_x_discrete(labels = NULL) +
  theme_minimal()
```

Percepções das Análises em Relação ao Saldo Bancário Médio Anual

Insights: Clientes com Saldos Bancários Mais Altos: Têm uma maior probabilidade de aderir ao depósito a prazo, sugerindo que aqueles com mais recursos financeiros são mais propensos a investir.

Mediana do Saldo: A mediana confirma que clientes com saldos mais altos são mais propensos a aderir ao depósito a prazo.

Quartis:

- **Q1 (25% dos Clientes):** 25% dos clientes que aderiram ao depósito têm saldo acima de 210, enquanto apenas 25% dos que não aderiram têm saldo acima de 64. Isso demonstra uma clara tendência de clientes com saldos mais altos serem mais propensos a investir.
- **Q3 (75% dos Clientes):** 75% dos clientes que aderiram ao depósito têm saldo abaixo de 2.159, enquanto 75% dos que não aderiram têm saldo abaixo de 1.324. Isso reforça que saldos mais altos estão associados a uma maior probabilidade de adesão.

Conclusão Geral: Saldos mais altos estão fortemente associados à adesão ao depósito a prazo, enquanto clientes com saldos mais baixos ou negativos têm menor probabilidade de aderir.

Analisando Clientes Inadimplentes e Adimplentes

```
# Tabela dos clientes que Inadimplêntes
table(dados_bank$default)
```

```
##
##    no    yes
## 10994   168
```

```
table(dados_adquiridos$default)
```

```
##
##    no    yes
## 5237    52
```

```
# Calcule a porcentagem de depósitos a prazo por Inadimplência
```

```
resultados_default <- dados_bank %>%
  group_by(default) %>%
  summarise(
    total = n(),
    depositaram = sum(deposit == "yes"),
    nao_depositaram = total - depositaram,
    porcentagem = (depositaram / total) * 100
  ) %>%
  arrange(desc(porcentagem))
```

```
# Exiba os resultados
print(resultados_default)
```

```
## # A tibble: 2 x 5
##   default total depositaram nao_depositaram porcentagem
##   <chr>   <int>      <int>          <int>      <dbl>
## 1 no     10994      5237           5757      47.6
## 2 yes     168       52            116      31.0
```

```
# Reorganizar a variável 'default' em ordem decrescente
```

```
dados_bank <- dados_bank %>%
  mutate(default = fct_infreq(default))
```

```
# Cálculo da porcentagem de inadimplência ou não dos clientes com o depósito a prazo (coluna target)
```

```
default_porcentagem <- dados_bank %>%
  group_by(default, deposit) %>%
  summarise(frequencia = n()) %>%
  mutate(percentual = frequencia / sum(frequencia) * 100)
```

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

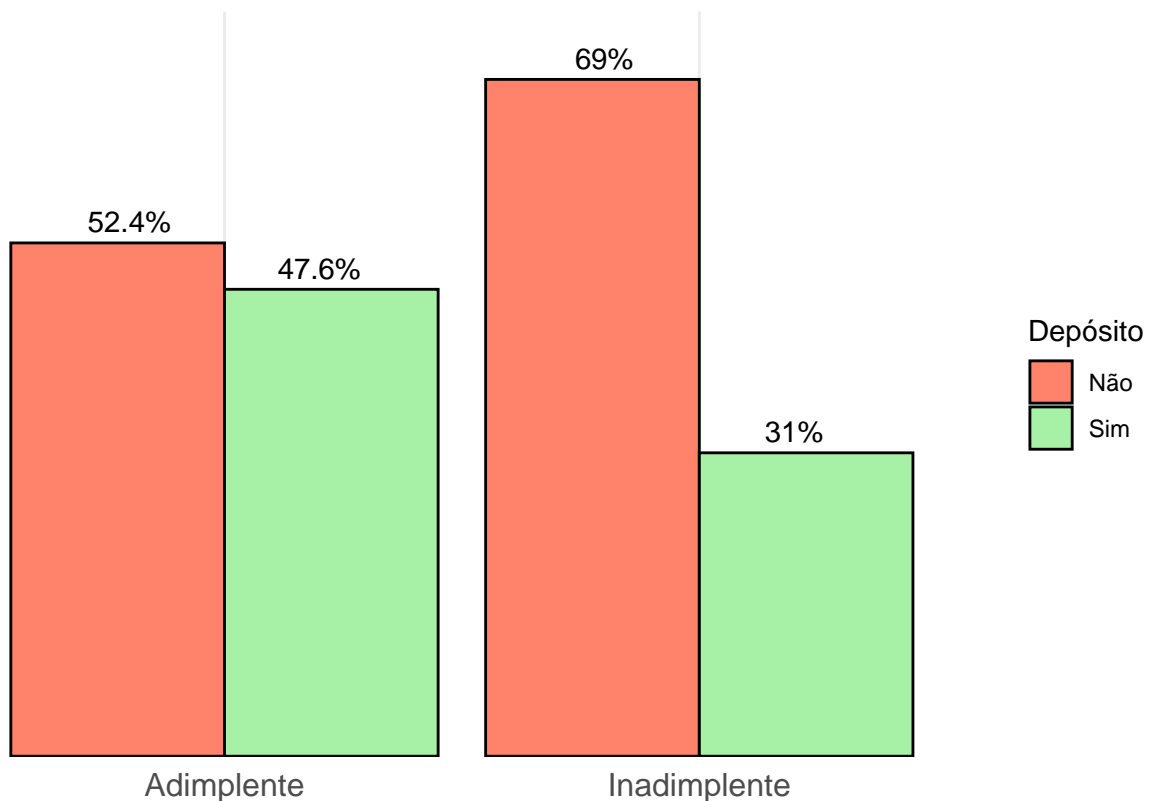
```
print(default_porcentagem)
```

```
## # A tibble: 4 x 4
## # Groups:   default [2]
```

```
## default deposit frequencia percentual
## <fct> <chr> <int> <dbl>
## 1 no no 5757 52.4
## 2 no yes 5237 47.6
## 3 yes no 116 69.0
## 4 yes yes 52 31.0
```

```
# Criar o gráfico com porcentagens, relacionando a inadimplência ou não dos clientes com o depósito a
ggplot(default_porcentagem, aes(x = default, y = percentual, fill = deposit)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.9), alpha = 0.8, color = "black") +
  geom_text(
    aes(label = paste0(round(percentual, 1), "%")),
    position = position_dodge(width = 0.8),
    vjust = -0.5,
    size = 4
  ) +
  labs(
    title = "Distribuição por Inadimplência e Adimplente (porcentagem)",
    x = NULL,
    y = NULL,
  ) +
  scale_y_continuous(breaks = NULL, expand = expansion(mult = c(0, 0.1))) + # Ajusta o espaço acima da
  theme_minimal() +
  theme(
    axis.text.x = element_text(size = 12, vjust = 0.5, hjust = 0.5), # Aumenta e ajusta rótulos do eix
    panel.grid.major.y = element_blank(), # Remove as linhas do grid horizontal
    plot.title = element_text(hjust = 0.5)
  ) +
  scale_fill_manual(
    values = c("yes" = "lightgreen", "no" = "tomato"),
    name = "Depósito",
    labels = c("yes" = "Sim", "no" = "Não")
  ) +
  scale_x_discrete(
    labels = c("no" = "Adimplente", "yes" = "Inadimplente")
  )
```

Distribuição por Inadimplência e Adimplente (porcentagem)



Percepções das Análises em Relação à Inadimplência

Insights: Clientes Adimplentes: Os clientes que não possuem histórico de inadimplência têm uma taxa de adesão ao depósito a prazo de 47,6%. Isso sugere que clientes com um bom histórico financeiro são mais propensos a investir em depósitos a prazo.

Clientes Inadimplentes: Aqueles com histórico de inadimplência apresentam uma taxa de adesão ao depósito a prazo significativamente menor, de apenas 31,0%. Isso indica que clientes com problemas financeiros são menos propensos a investir..

Conclusão Geral: O histórico de inadimplência é um fator importante na decisão de investir. Clientes com um bom histórico financeiro tendem a ser mais confiantes e dispostos a investir, enquanto aqueles com histórico de inadimplência enfrentam maiores barreiras.

Analisando Clientes que possuem ou não Empréstimo Imobiliário

```
# Calcule a porcentagem de depósitos a prazo por clientes que possuem ou não empréstimo imobiliário
resultados_housing <- dados_bank %>%
  group_by(housing_loan) %>%
  summarise(
    total = n(),
    depositaram = sum(deposit == "yes"),
    nao_depositaram = total - depositaram,
    porcentagem = (depositaram / total) * 100
```

```
) %>%
  arrange(desc(porcentagem))
```

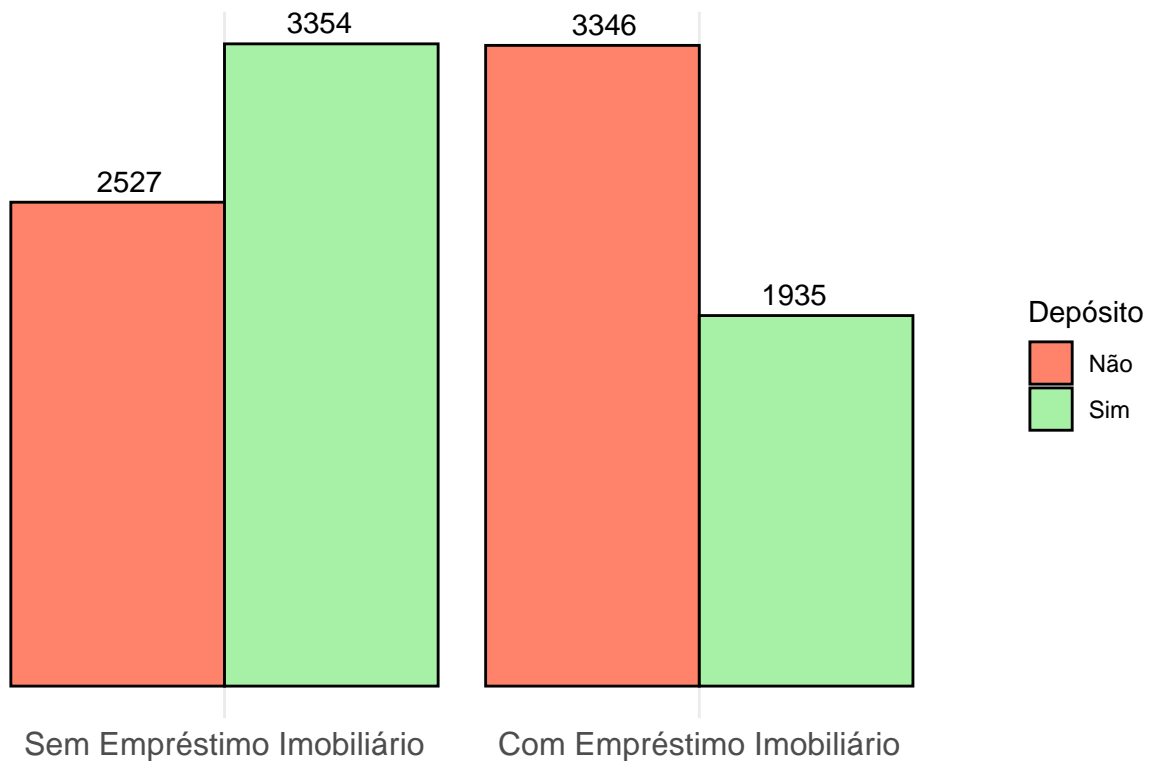
```
# Exiba os resultados
print(resultados_housing)
```

```
## # A tibble: 2 x 5
##   housing_loan total depositaram nao_depositaram porcentagem
##   <chr>          <int>          <int>          <int>          <dbl>
## 1 no            5881            3354            2527            57.0
## 2 yes           5281            1935            3346            36.6
```

```
# Reorganizar a variável 'housing_loan' em ordem decrescente
dados_bank <- dados_bank %>%
  mutate(housing_loan = fct_infreq(housing_loan))
```

```
# Criar o gráfico relacionando clientes que possuem ou não empréstimo imobiliário com o depósito a prazo
ggplot(dados_bank, aes(x = housing_loan, fill = deposit)) +
  geom_bar(position = position_dodge(width = 0.9), alpha = 0.8, color = "black") +
  geom_text(
    stat = "count",
    aes(label = ..count..),
    position = position_dodge(width = 0.8),
    vjust = -0.5,
    size = 4
  ) +
  labs(
    title = "Distribuição por Empréstimo Imobiliário\n",
    x = NULL,
    y = NULL # Remove o rótulo do eixo Y
  ) +
  scale_y_continuous(breaks = NULL) + # Remove os valores no eixo Y
  theme_minimal() +
  theme(
    axis.text.x = element_text(size = 12, vjust = 0.9, hjust = 0.5), # Aumenta o tamanho e ajusta pos
    panel.grid.major.y = element_blank(), # Remove as linhas de grade horizontais
    axis.title.x = element_text(size = 14), # Aumenta o título do eixo X, se necessário
    plot.title = element_text(hjust = 0.5)
  ) +
  scale_fill_manual(
    values = c("yes" = "lightgreen", "no" = "tomato"),
    name = "Depósito",
    labels = c("yes" = "Sim", "no" = "Não")
  ) +
  scale_x_discrete(
    labels = c("no" = "Sem Empréstimo Imobiliário", "yes" = "Com Empréstimo Imobiliário")
  )
```

Distribuição por Empréstimo Imobiliário



```
# Calcular as porcentagens relacionando clientes que possuem ou não empréstimo imobiliário com o depós
housing_porcentagem <- dados_bank %>%
  group_by(housing_loan, deposit) %>%
  summarise(frequencia = n()) %>%
  mutate(percentual = frequencia / sum(frequencia) * 100)
```

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

```
print(housing_porcentagem)
```

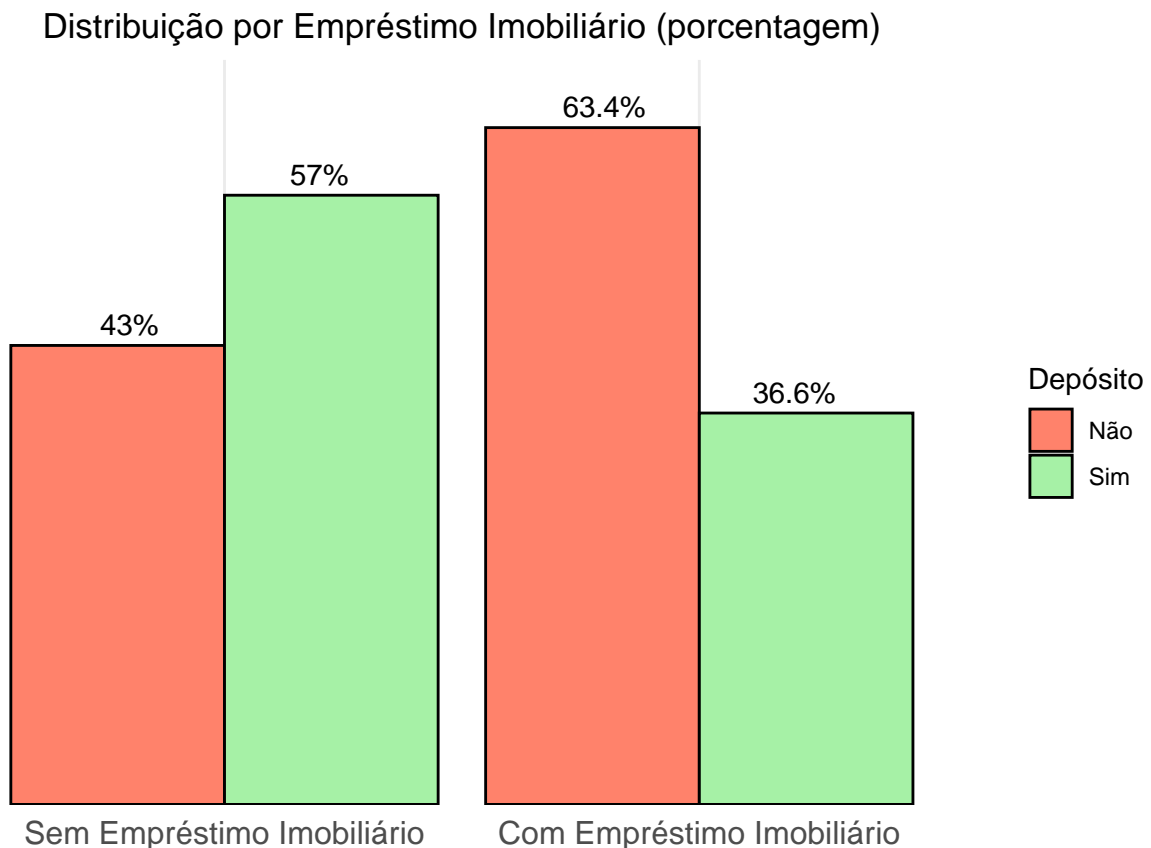
```
## # A tibble: 4 x 4
## # Groups:   housing_loan [2]
##   housing_loan deposit frequencia percentual
##   <fct>         <chr>      <int>      <dbl>
## 1 no          no         2527       43.0
## 2 no          yes        3354       57.0
## 3 yes         no         3346       63.4
## 4 yes         yes        1935       36.6
```

```
# Criar o gráfico com porcentagens relacionando clientes que possuem ou não empréstimo imobiliário com
ggplot(housing_porcentagem, aes(x = housing_loan, y = percentual, fill = deposit)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.9), alpha = 0.8, color = "black") +
  geom_text(
```

```

aes(label = paste0(round(percentual, 1), "%")),
position = position_dodge(width = 0.8),
vjust = -0.5,
size = 4
) +
labs(
  title = "Distribuição por Empréstimo Imobiliário (porcentagem)",
  x = NULL,
  y = NULL,
) +
scale_y_continuous(breaks = NULL, expand = expansion(mult = c(0, 0.1))) + # Ajusta o espaço acima da
theme_minimal() +
theme(
  axis.text.x = element_text(size = 12, vjust = 0.5, hjust = 0.5), # Aumenta e ajusta rótulos do eix
  panel.grid.major.y = element_blank(), # Remove as linhas do grid horizontal
  plot.title = element_text(hjust = 0.5)
) +
scale_fill_manual(
  values = c("yes" = "lightgreen", "no" = "tomato"),
  name = "Depósito",
  labels = c("yes" = "Sim", "no" = "Não")
) +
scale_x_discrete(
  labels = c("no" = "Sem Empréstimo Imobiliário", "yes" = "Com Empréstimo Imobiliário")
)

```



Percepções das Análises em Relação ao Crédito Imobiliário

Insights: Clientes sem Crédito Imobiliário: Este grupo apresenta uma taxa de adesão ao depósito a prazo de 57,0%. Isso sugere que clientes sem compromissos financeiros relacionados a imóveis têm maior disponibilidade para investir em opções de longo prazo.

Clientes com Crédito Imobiliário: Com uma taxa de adesão ao depósito a prazo de apenas 36,6%, este grupo é menos propenso a investir. .

Conclusão Geral: Clientes sem crédito imobiliário são mais propensos a investir, enquanto aqueles com crédito imobiliário tendem a evitar novos compromissos financeiros devido às suas obrigações existentes.

Analisando Clientes que possuem ou não Empréstimo Pessoal

```
# Calcular as porcentagens relacionando clientes que possuem ou não empréstimo pessoal com o depósito a
loan_porcentagem <- dados_bank %>%
  group_by(loan, deposit) %>%
  summarise(frequencia = n()) %>%
  mutate(percentual = frequencia / sum(frequencia) * 100)
```

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

```
print(loan_porcentagem)
```

```
## # A tibble: 4 x 4
## # Groups:   loan [2]
##   loan deposit frequencia percentual
##   <chr> <chr>      <int>      <dbl>
## 1 no    no           4897         50.5
## 2 no    yes          4805         49.5
## 3 yes   no           976          66.8
## 4 yes   yes           484          33.2
```

```
# Reorganizar a variável 'loan' em ordem decrescente
dados_bank <- dados_bank %>%
  mutate(loan = fct_infreq(loan))

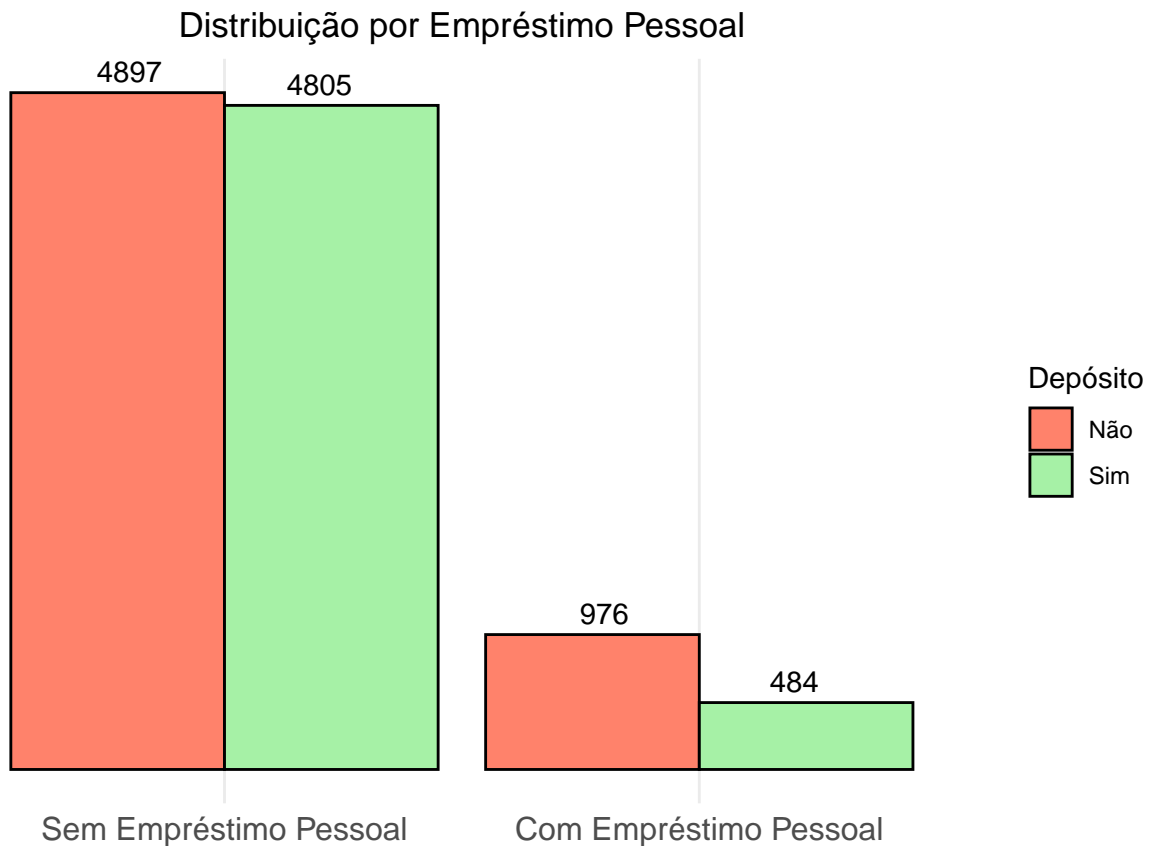
# Criar o gráfico relacionando clientes que possuem ou não empréstimo pessoal com o depósito a prazo (c
ggplot(dados_bank, aes(x = loan, fill = deposit)) +
  geom_bar(position = position_dodge(width = 0.9), alpha = 0.8, color = "black") +
  geom_text(
    stat = "count",
    aes(label = ..count..),
    position = position_dodge(width = 0.8),
    vjust = -0.5,
    size = 4
  ) +
  labs(
    title = "Distribuição por Empréstimo Pessoal",
    x = NULL,
    y = NULL # Remove o rótulo do eixo Y
```



```

) +
scale_y_continuous(breaks = NULL) + # Remove os valores no eixo Y
theme_minimal() +
theme(
  axis.text.x = element_text(size = 12, vjust = 0.9, hjust = 0.5,), # Aumenta o tamanho e ajusta pos
  panel.grid.major.y = element_blank(), # Remove as linhas de grade horizontais
  axis.title.x = element_text(size = 14), # Aumenta o título do eixo X, se necessário
  plot.title = element_text(hjust = 0.5)
) +
scale_fill_manual(
  values = c("yes" = "lightgreen", "no" = "tomato"),
  name = "Depósito",
  labels = c("yes" = "Sim", "no" = "Não")
) +
scale_x_discrete(
  labels = c("no" = "Sem Empréstimo Pessoal", "yes" = "Com Empréstimo Pessoal")
)

```



```

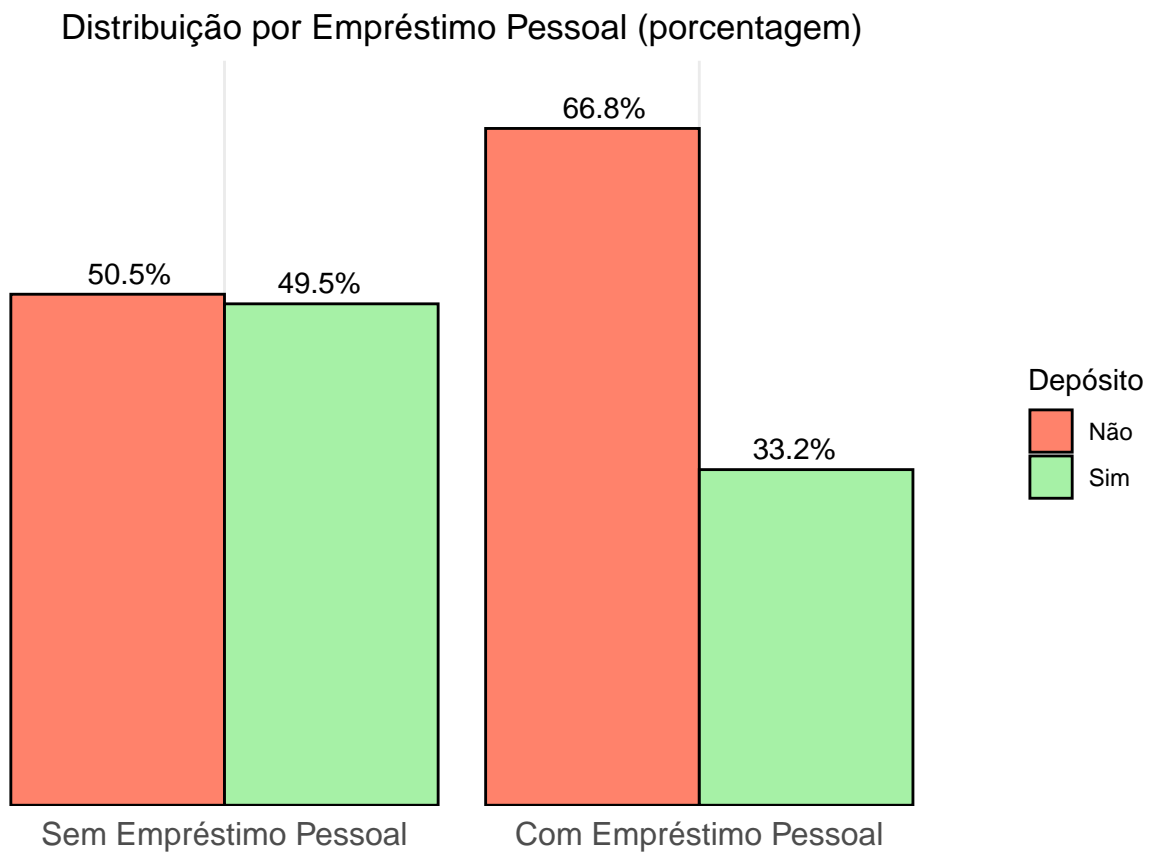
# Criar o gráfico com porcentagens relacionando clientes que possuem ou não empréstimo pessoal com o de
ggplot(loan_porcentagem, aes(x = loan, y = percentual, fill = deposit)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.9), alpha = 0.8, color = "black") +
  geom_text(
    aes(label = paste0(round(percentual, 1), "%")),
    position = position_dodge(width = 0.8),
    vjust = -0.5,

```

```

    size = 4
  ) +
  labs(
    title = "Distribuição por Empréstimo Pessoal (porcentagem)",
    x = NULL,
    y = NULL,
  ) +
  scale_y_continuous(breaks = NULL, expand = expansion(mult = c(0, 0.1))) + # Ajusta o espaço acima da
  theme_minimal() +
  theme(
    axis.text.x = element_text(size = 12, vjust = 0.5, hjust = 0.5), # Aumenta e ajusta rótulos do eix
    panel.grid.major.y = element_blank(), # Remove as linhas do grid horizontal
    plot.title = element_text(hjust = 0.5)
  ) +
  scale_fill_manual(
    values = c("yes" = "lightgreen", "no" = "tomato"),
    name = "Depósito",
    labels = c("yes" = "Sim", "no" = "Não")
  ) +
  scale_x_discrete(
    labels = c("no" = "Sem Empréstimo Pessoal", "yes" = "Com Empréstimo Pessoal")
  )
)

```



Percepções das Análises em Relação ao Empréstimo Pessoal

Insights: Clientes sem Empréstimo Pessoal: Este grupo apresenta uma taxa de adesão ao depósito a prazo de 49,5%.

Clientes com Empréstimo Pessoal: Com uma taxa de adesão ao depósito a prazo significativamente menor de 33,2%..

Conclusão Geral: Clientes sem empréstimo pessoal são mais propensos a investir em depósitos a prazo, enquanto aqueles com empréstimo pessoal tendem a evitar novos compromissos financeiros devido às suas obrigações existentes.

Analizando o cruzamento de informações de Estado Civil e Escolaridade

```
# Calcular as porcentagens relacionando clientes que possuem ou não empréstimo pessoal com o depósito a
```

```
# Filtrar os valores da coluna education que não são "unknown"
dados_education_filtrado <- dados_bank %>%
  filter(education != "unknown")
```

```
porcentagem_teste <- dados_education_filtrado %>%
  group_by(marital_status, education, deposit) %>%
  summarise(frequencia = n()) %>%
  mutate(percentual = frequencia / sum(frequencia) * 100)
```

```
## 'summarise()' has grouped output by 'marital_status', 'education'. You can
## override using the '.groups' argument.
```

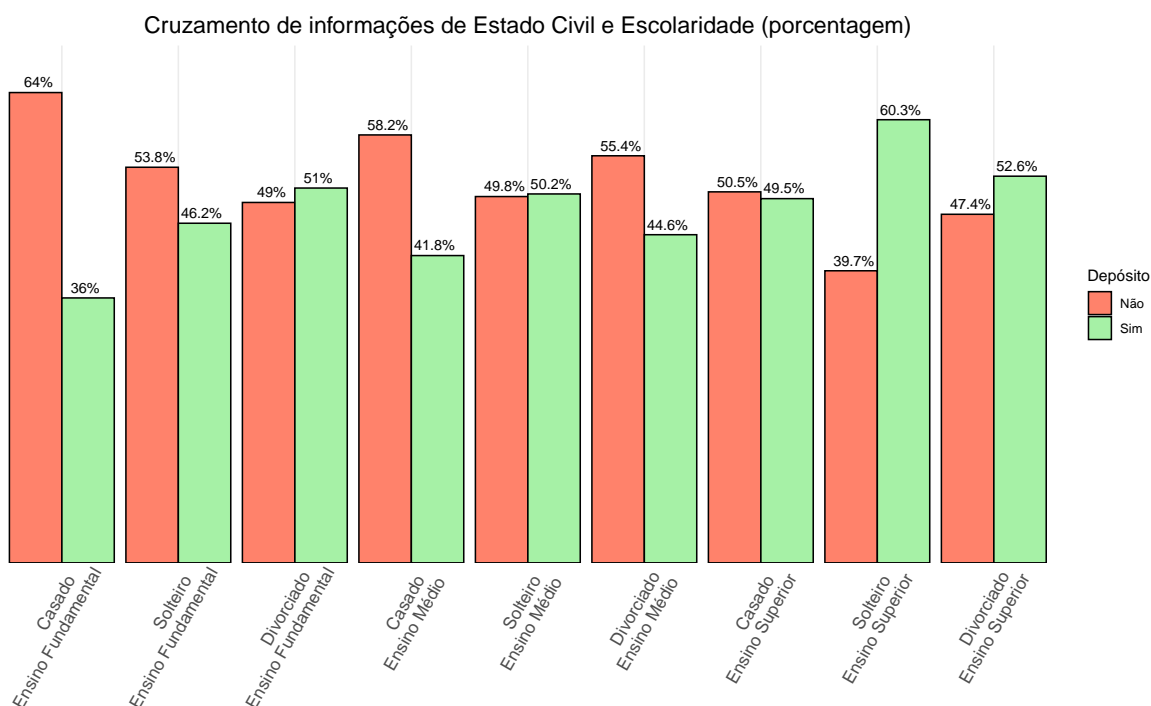
```
print(porcentagem_teste)
```

```
## # A tibble: 18 x 5
## # Groups:   marital_status, education [9]
##   marital_status education deposit frequencia percentual
##   <fct>          <chr>    <chr>      <int>      <dbl>
## 1 married      primary    no         703        64.0
## 2 married      primary    yes        396        36.0
## 3 married      secondary no        1816       58.2
## 4 married      secondary yes        1304       41.8
## 5 married      tertiary  no         930       50.5
## 6 married      tertiary  yes         913       49.5
## 7 single       primary    no         106       53.8
## 8 single       primary    yes          91       46.2
## 9 single       secondary no         849       49.8
## 10 single      secondary yes         855       50.2
## 11 single      tertiary  no         580       39.7
## 12 single      tertiary  yes         880       60.3
## 13 divorced    primary    no          100       49.0
## 14 divorced    primary    yes          104       51.0
## 15 divorced    secondary no         361       55.4
## 16 divorced    secondary yes         291       44.6
## 17 divorced    tertiary  no          183       47.4
## 18 divorced    tertiary  yes         203       52.6
```

```

# Criar o gráfico com porcentagens relacionando clientes que possuem ou não empréstimo pessoal com o de
ggplot(porcentagem_teste, aes(x = interaction(marital_status, education), y = percentual, fill = deposi
geom_bar(stat = "identity", position = position_dodge(width = 0.9), alpha = 0.8, color = "black") +
geom_text(
  aes(label = paste0(round(percentual, 1), "%")),
  position = position_dodge(width = 0.8),
  vjust = -0.5,
  size = 3.5
) +
labs(
  title = "Cruzamento de informações de Estado Civil e Escolaridade (porcentagem)",
  x = NULL,
  y = NULL,
) +
scale_y_continuous(breaks = NULL, expand = expansion(mult = c(0, 0.1))) +
theme_minimal() +
theme(
  axis.text.x = element_text(angle = 60, size = 12, vjust = 0.9, hjust = 0.8), # Aumenta e ajusta ró
  panel.grid.major.y = element_blank(), # Remove as linhas do grid horizontal
  plot.title = element_text(size = 16, hjust = 0.5),
  plot.margin = unit(c(1, 1, 2, 1), "cm") # Ajusta as margens do gráfico
) +
scale_fill_manual(
  values = c("yes" = "lightgreen", "no" = "tomato"),
  name = "Depósito",
  labels = c("yes" = "Sim", "no" = "Não")
) +
scale_x_discrete(
  labels = c("married.secondary" = "Casado\nEnsino Médio", "single.secondary" = "Solteiro\nEnsino Méd
  "divorced.secondary" = "Divorciado\nEnsino Médio", "married.tertiary" = "Casado\nEnsino Superior",
  "single.tertiary" = "Solteiro\nEnsino Superior", "divorced.tertiary" = "Divorciado\nEnsino Superior
  "married.primary" = "Casado\nEnsino Fundamental", "single.primary" = "Solteiro\nEnsino Fundamental"
  "divorced.primary" = "Divorciado\nEnsino Fundamental")
)

```



Percepções das Análises do cruzamento de informações de Estado Civil e Escolaridade

Insights: Combinação de Escolaridade e Estado Civil

- A combinação de solteiros e ensino superior resulta na maior taxa de adesão ao depósito a prazo.
- Casados com ensino fundamental são os menos propensos a investir.
- A interação entre escolaridade e estado civil reforça os padrões observados. Clientes solteiros e com maior escolaridade têm maior probabilidade de aderir ao depósito, enquanto casados com menor escolaridade são os menos propensos.

ACT - RECOMENDAÇÕES COM BASE NA ANÁLISE DOS DADOS

Perfil do Cliente Ideal para Depósito a Prazo

Com base na análise dos dados, identificamos o perfil do cliente com maior propensão a aderir ao depósito a prazo. Esse perfil pode ser utilizado para direcionar as campanhas de marketing e maximizar a eficácia da próxima estratégia. Abaixo, destacamos as características do cliente ideal:

Idade

- Clientes Idosos (60 a 95 anos):
- Jovens (18 a 23 anos):

Profissão

- Estudantes e Aposentados:

Estado Civil

- Solteiros:

Escolaridade

- Ensino Superior:

Saldo Bancário

- Clientes com Saldos Mais Altos:

Histórico Financeiro

- Clientes Adimplentes, que não possuem Empréstimos pessoais e imobiliários

Recomendações para a Próxima Campanha

Foco nos Grupos com Maior Propensão:

- Priorize os clientes que com o perfil recomendado pela análise dos dados.
- Desenvolva mensagens que ressaltem a segurança, a rentabilidade e a facilidade de acesso ao depósito a prazo.
- Personalização da abordagem para os diferentes perfis de clientes, por exemplo jovens e idosos.

Educação Financeira:

- Implemente programas de educação financeira para aumentar a compreensão e o interesse de clientes com menor propensão a investir, como casados e clientes com ensino médio ou fundamental.