

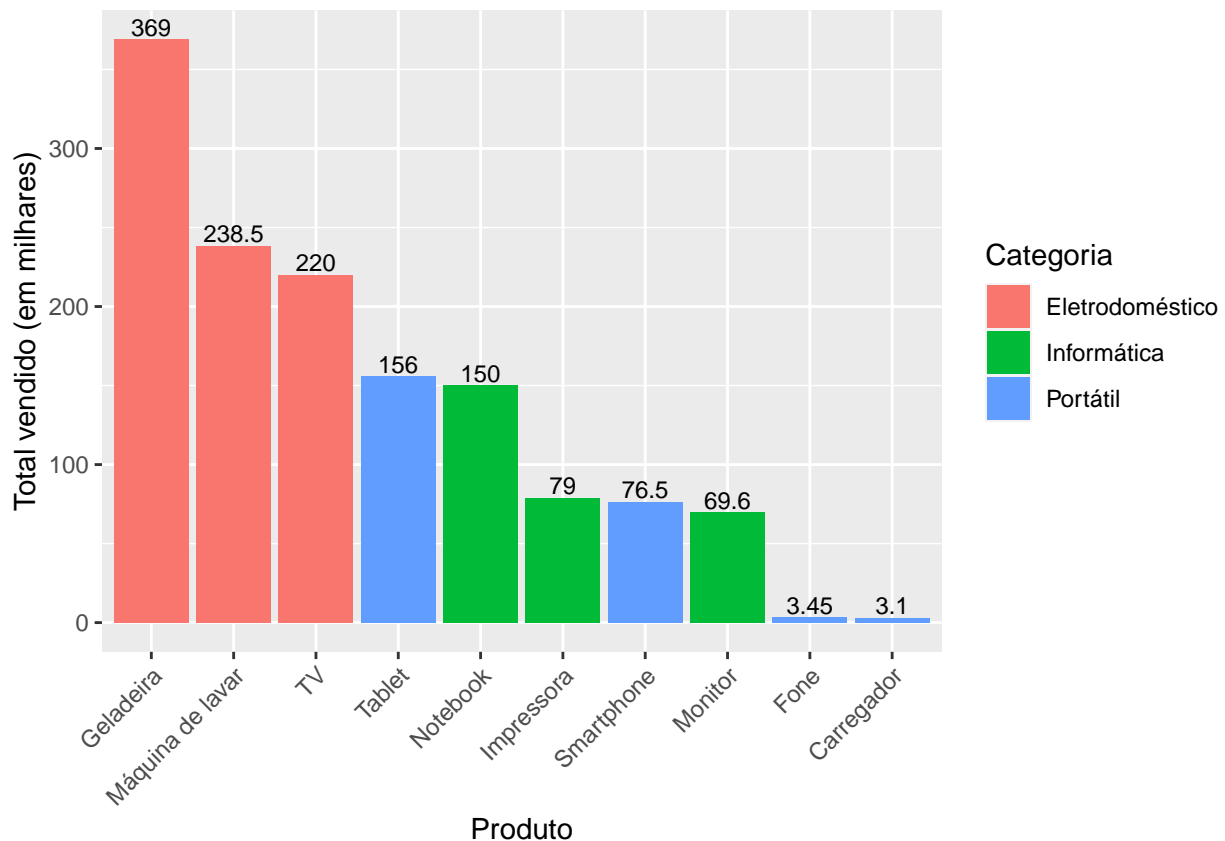
# Análise Descritiva

Carregando os pacotes necessários e fazendo a leitura dos dados.

```
library(ggplot2)
library(dplyr)
library(lubridate)
dados_clean = read.csv(
  'https://raw.githubusercontent.com/pedroww/Teste_Analytics_PedroCoelho/main/data_clean.csv'
)
```

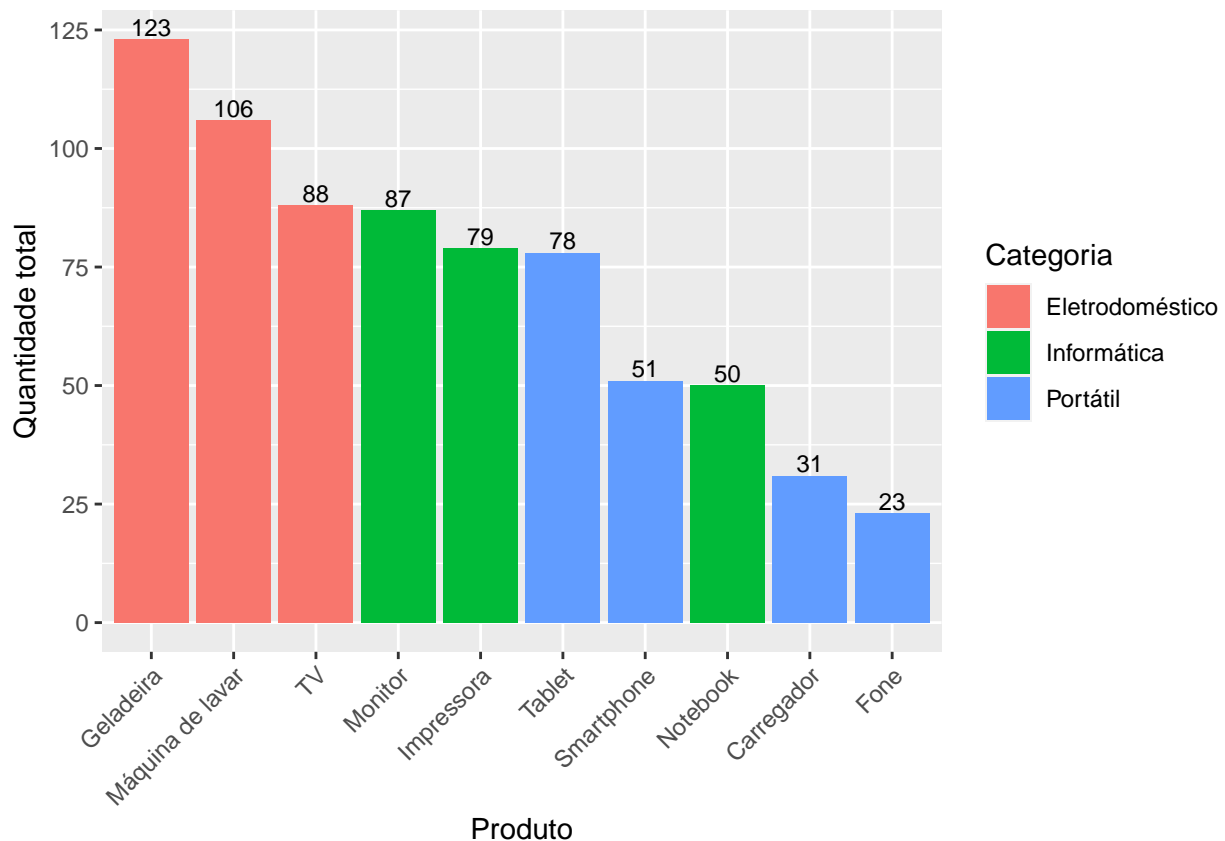
Criando a variável `Total_venda = Quantidade*Preço` e o gráfico do total de vendas de cada produto.

```
dados_clean$Venda_total = dados_clean$Quantidade*dados_clean$Preço
dados_clean |> group_by(Produto) |> summarise(Total_vendido = sum(Venda_total),
                                              Categoria = first(Categoria)) |>
  ggplot(aes(x = reorder(Produto, Total_vendido, decreasing = T), y = Total_vendido/1000,
             fill = Categoria)) + geom_bar(stat = 'identity') +
  labs(
    x = 'Produto',
    y = 'Total vendido (em milhares)'
  ) +
  geom_text(aes(label = paste0(Total_vendido/1000)),
            vjust = -.2, size = 3) +
  scale_x_discrete(guide = guide_axis(angle = 45))
```



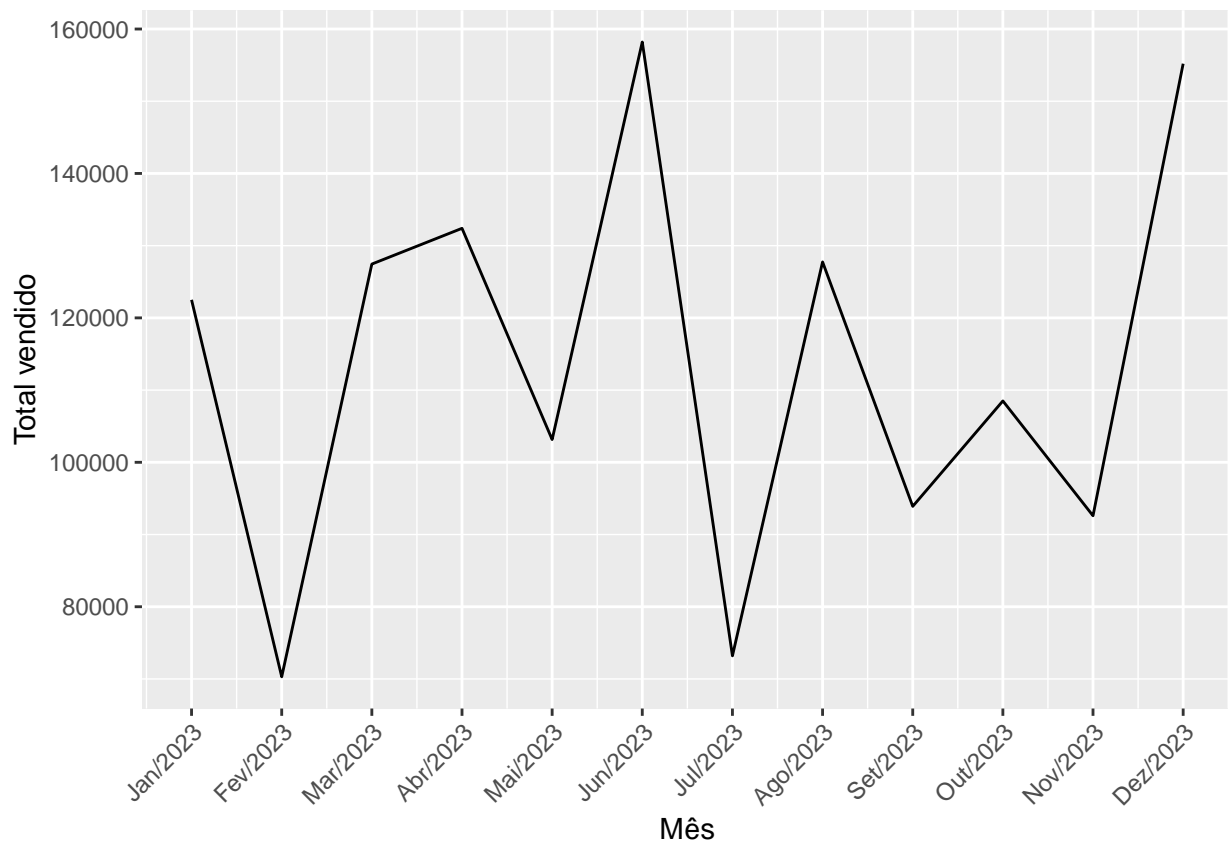
O produto com maior valor em vendas durante o período foi a geladeira, seguida da máquina de lavar e TV, os três da categoria eletrodoméstico. Esses produtos possuem valores individuais altos, o que ajuda a entender esses números. Para uma melhor análise, vamos analisar o gráfico da quantidade vendida por produto no período.

```
dados_clean |> group_by(Produto) |> summarise(Qtd_total = sum(Quantidade),
                                              Categoria = first(Categoria)) |>
ggplot(aes(x = reorder(Produto, Qtd_total, decreasing = T), y = Qtd_total,
           fill = Categoria)) + geom_bar(stat = 'identity') +
labs(
  x = 'Produto',
  y = 'Quantidade total'
) +
geom_text(aes(label = Qtd_total),
          vjust = -.2, size = 3) +
scale_x_discrete(guide = guide_axis(angle = 45))
```



Além de possuírem os maiores valores de vendas, os eletrodomésticos também foram os que mais venderam em quantidade, indicando que essa categoria parece realmente ser o ponto forte da loja.

```
dados_clean$Mes = month(dados_clean$Data)
nomes_meses = c('Jan/2023', 'Fev/2023', 'Mar/2023', 'Abr/2023',
                 'Mai/2023', 'Jun/2023', 'Jul/2023', 'Ago/2023',
                 'Set/2023', 'Out/2023', 'Nov/2023', 'Dez/2023')
dados_clean |> group_by(Mes) |> summarise(total_mes = sum(Venda_total)) |>
  ggplot(aes(x = Mes, y = total_mes)) + geom_line() +
  scale_x_continuous(
    breaks = seq_along(nomes_meses),
    labels = nomes_meses,
    guide = guide_axis(angle = 45)
  ) +
  labs(
    x = 'Mês',
    y = 'Total vendido'
  )
```



Os meses de fevereiro e julho foram aqueles com pior desempenho de vendas, enquanto janeiro e dezembro foram os meses com maiores vendas. No final/início do ano geralmente ocorre mais vendas devido às festividades, algo que justifica os números. Em junho ocorre algo parecido. Nos meses que se seguem a esses, há uma queda nas vendas.