

# WALMART

INSIGHTS POR MEIO DE UM DATASET DE VENDAS



## OBJETIVO

Gerar insights e visualizar as tendências desses dados de maneira eficiente, traçando um perfil do cliente e suas tendências na hora da compra.

# Números Principais

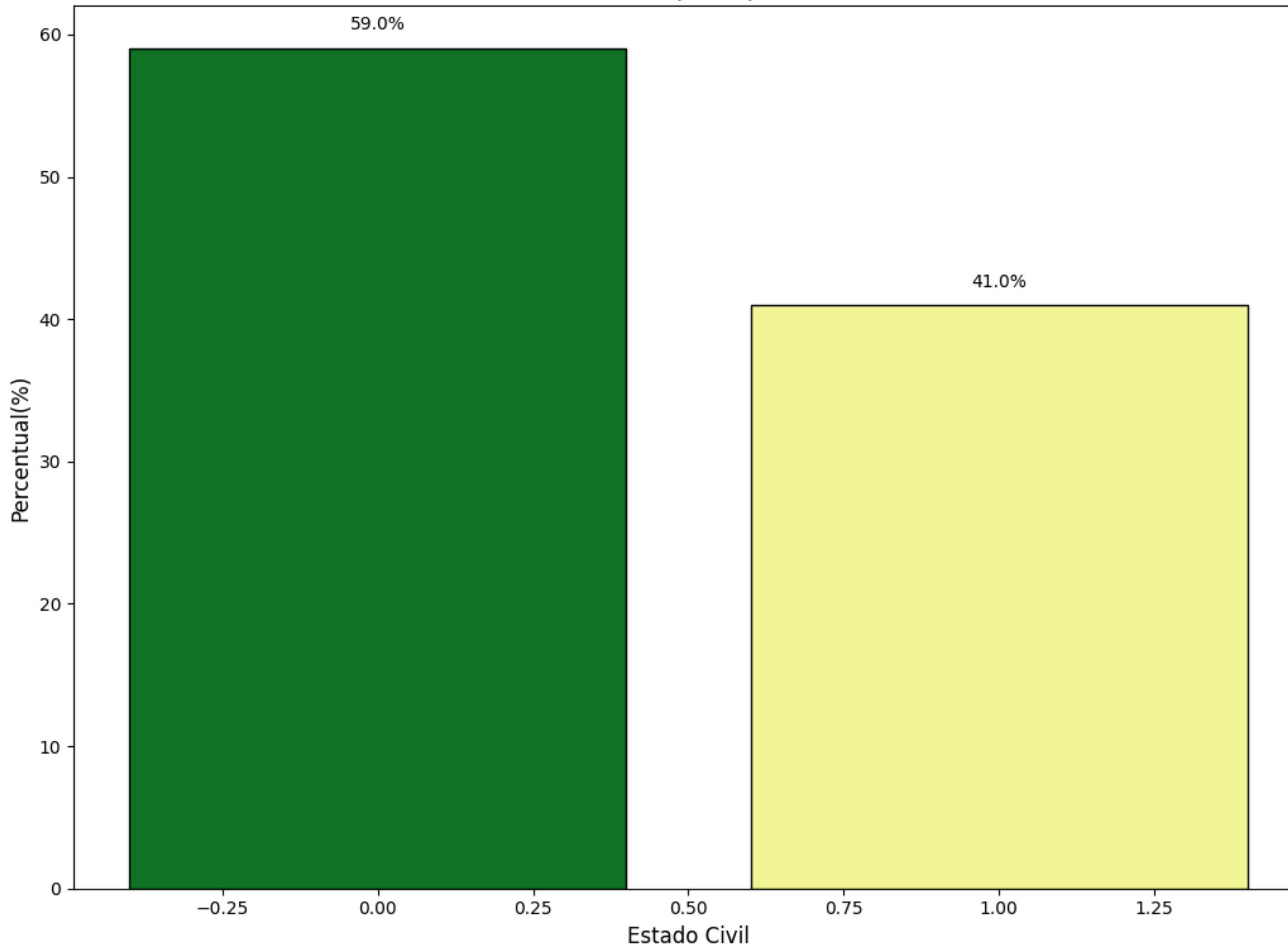
**QUANTIDADE DE CLIENTES ÚNICOS: 5.891**

**QUANTIDADE DE PRODUTOS: 3.631**

**QUANTIDADE DE OCUPAÇÕES: 21**

**QUANTIDADE DE CIDADES: 3**

Percentual de Compras por Estado Civil



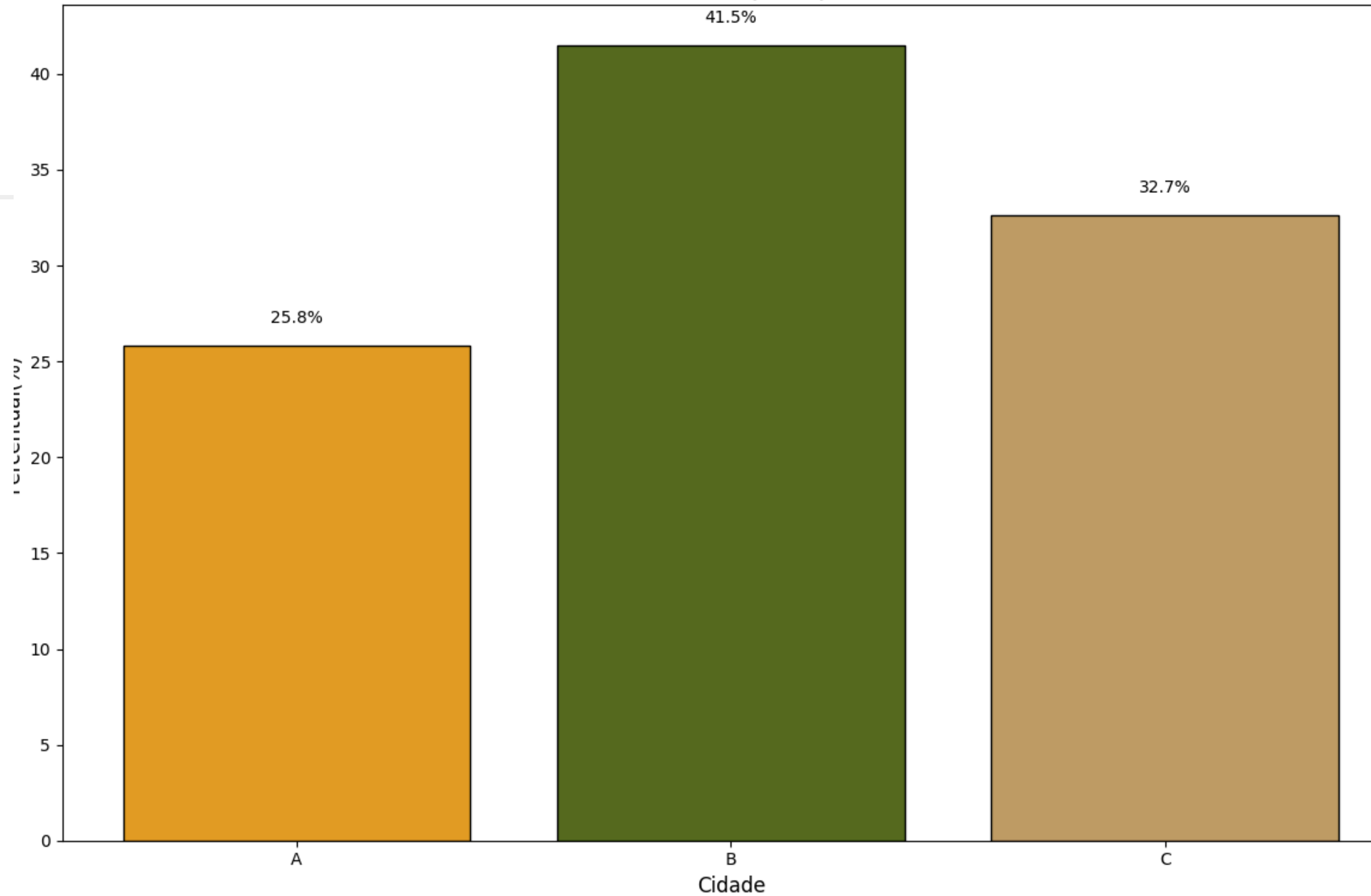
Estado Civil

- 0
- 1

Notamos que o público “0” tem o percentual de compras maior.

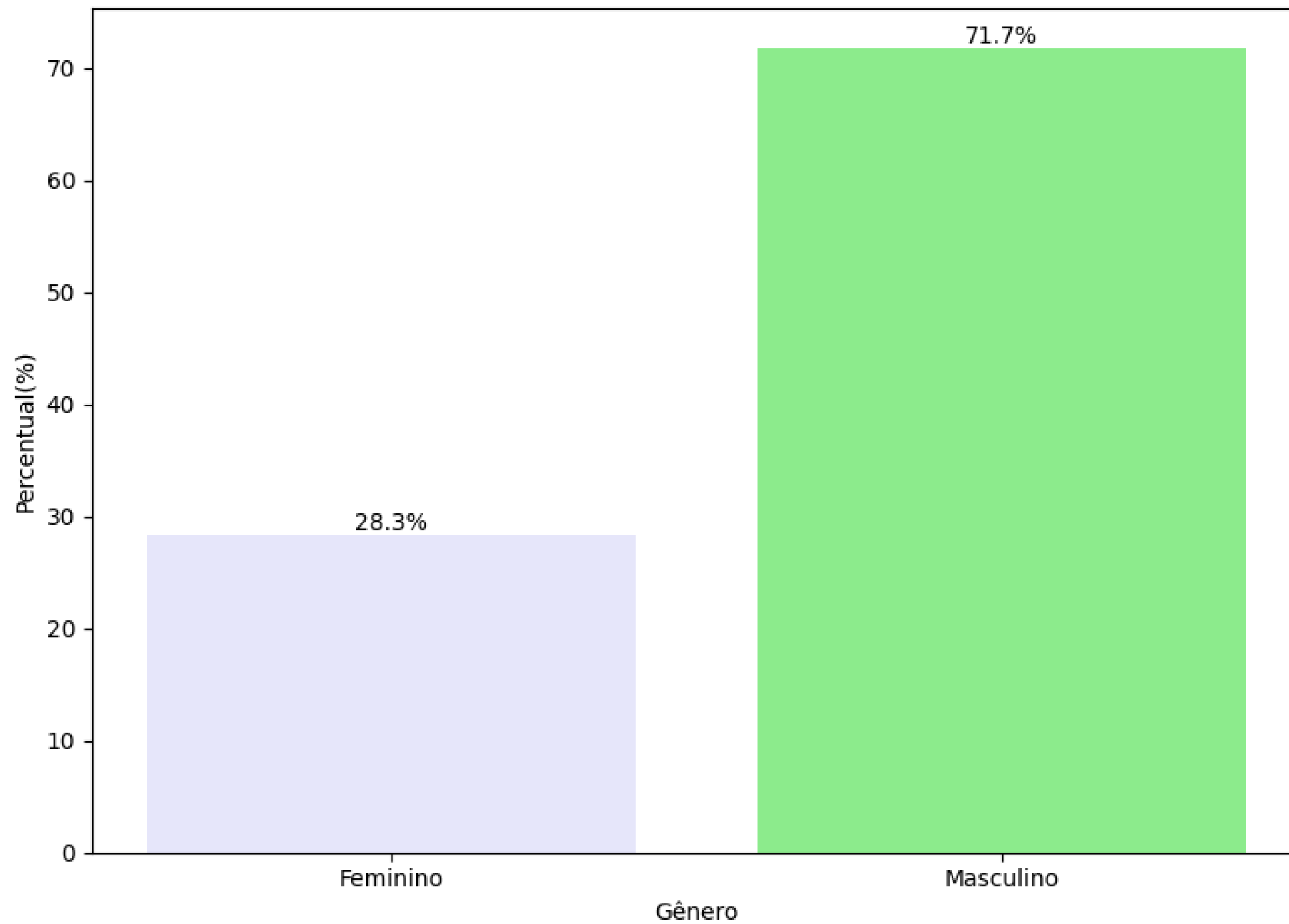
Esse público representa 58% da base total de clientes únicos.

Percentual de Compras por Cidade



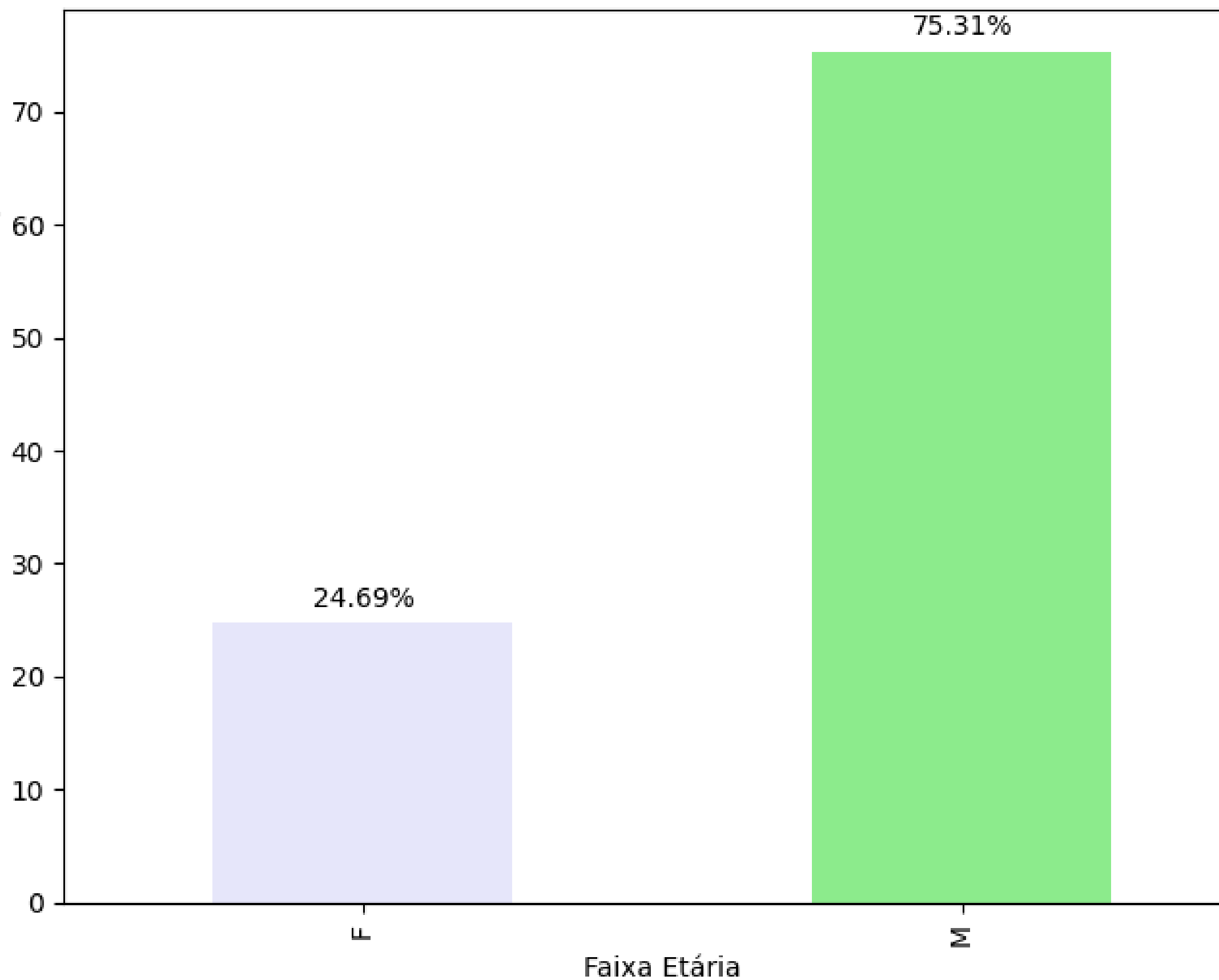
Nota-se que a cidade B possui o maior percentual de compras, enquanto a A possui o menor. Podemos criar campanhas de promoções e ofertas para a cidade A e rentabilizar mais ainda a cidade B e C.

Visão Geral do Público (Por Gênero)



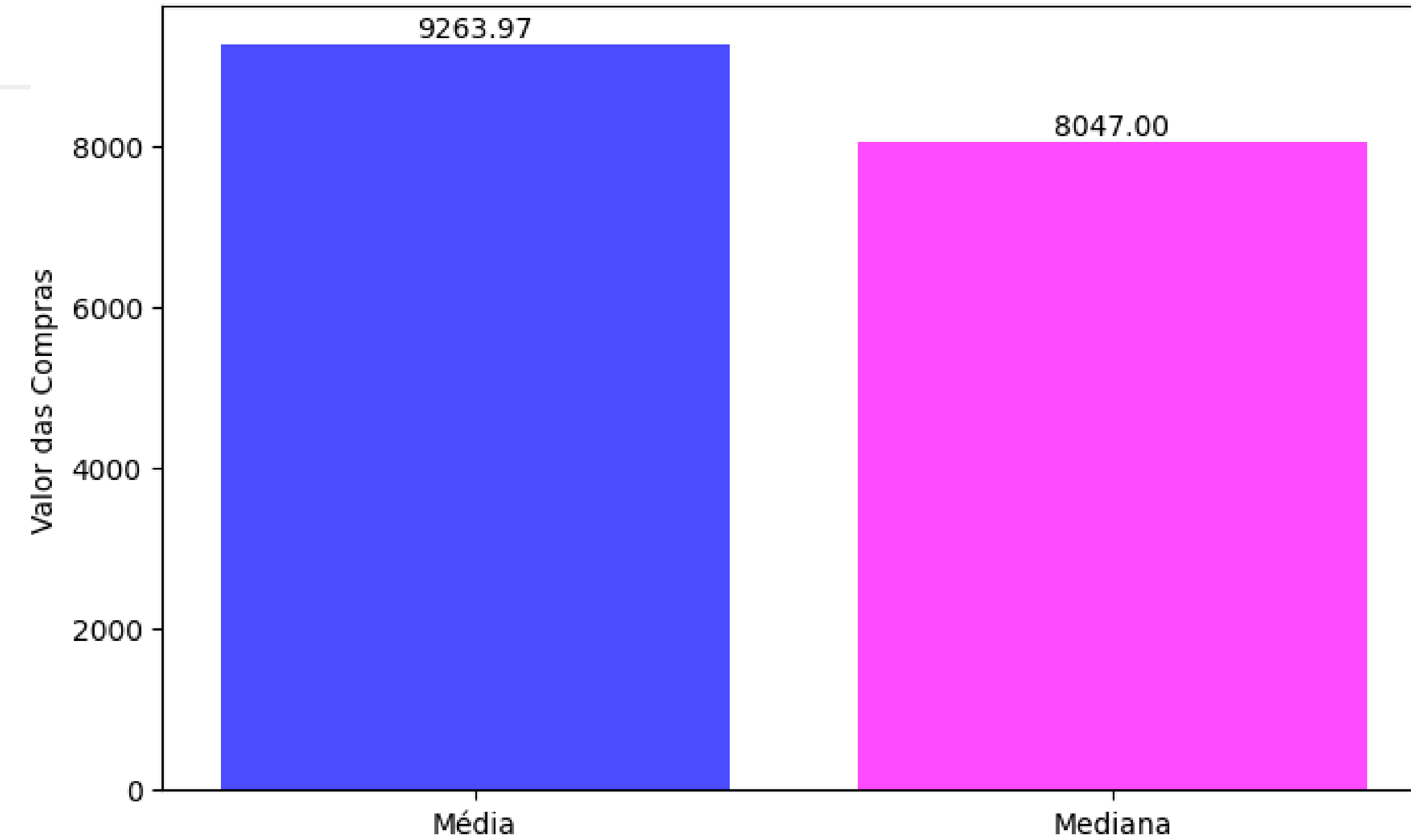
O público masculino representa 71% do público geral.

Percentual de Compras por Gênero



O público masculino possui o percentual de compra maior.

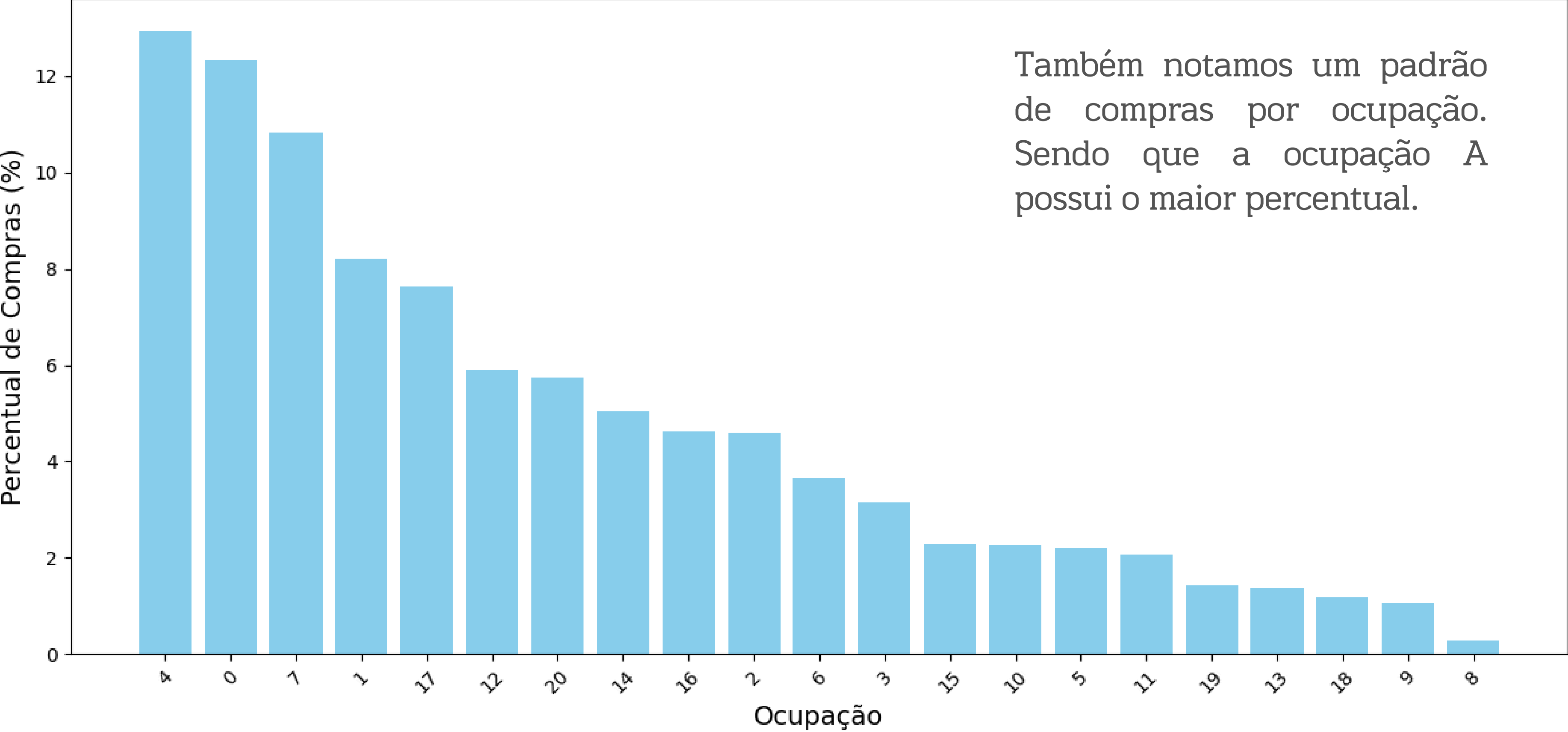
## Média e Mediana de Compras



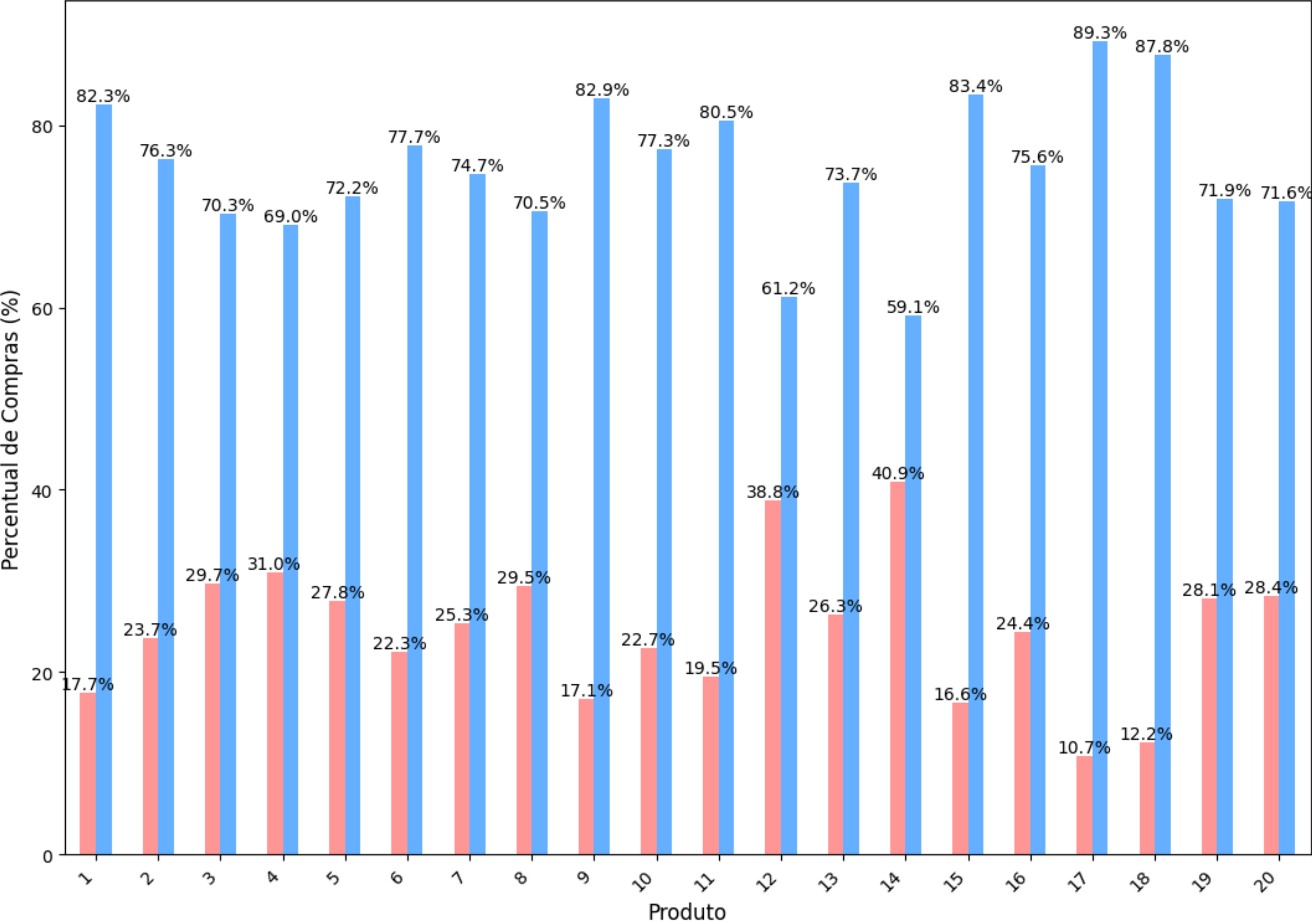
Valores gerais das compras por usuário.



Percentual de Compras por Ocupação

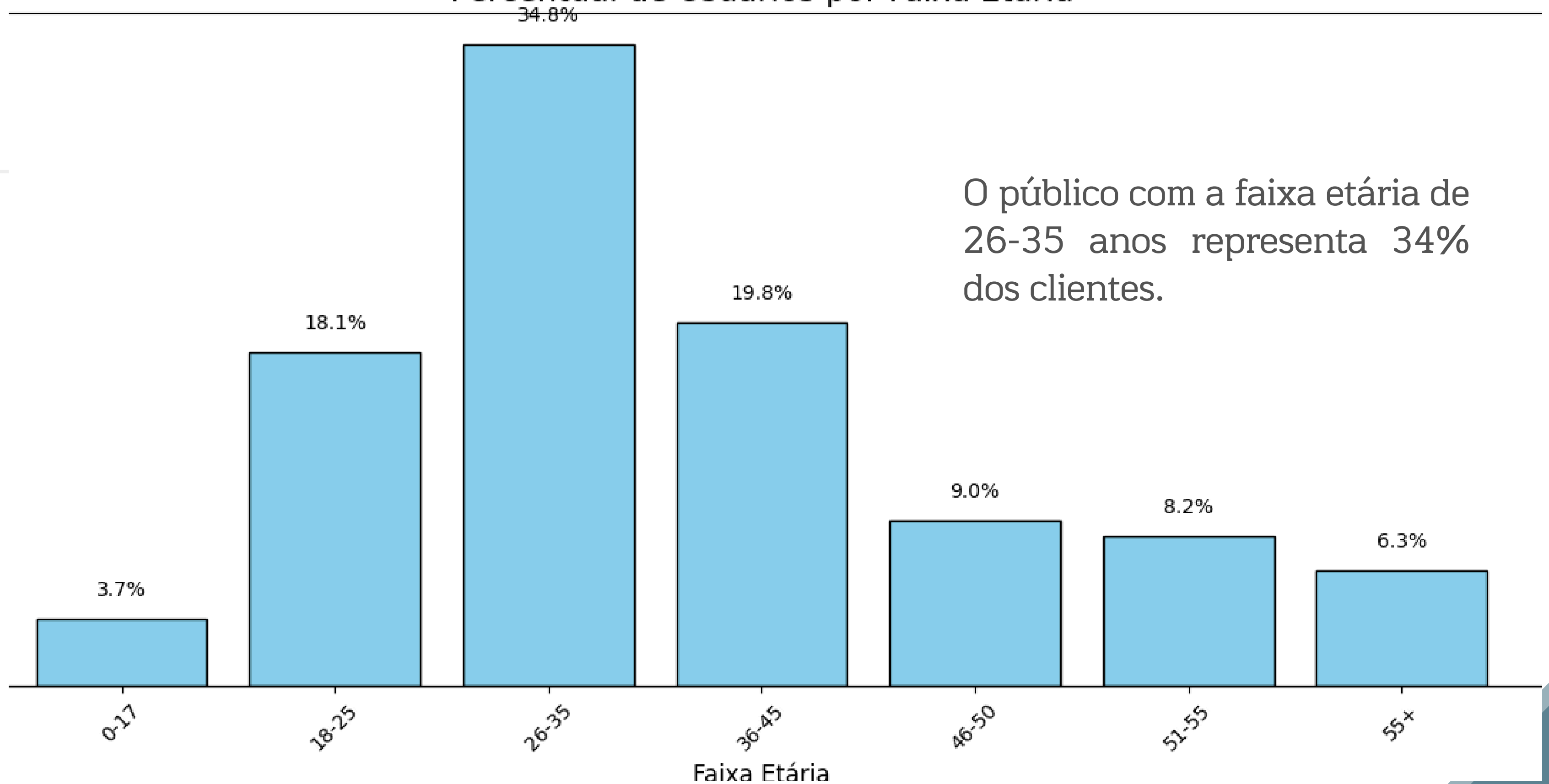


Percentual de Compras por Produto e Gênero

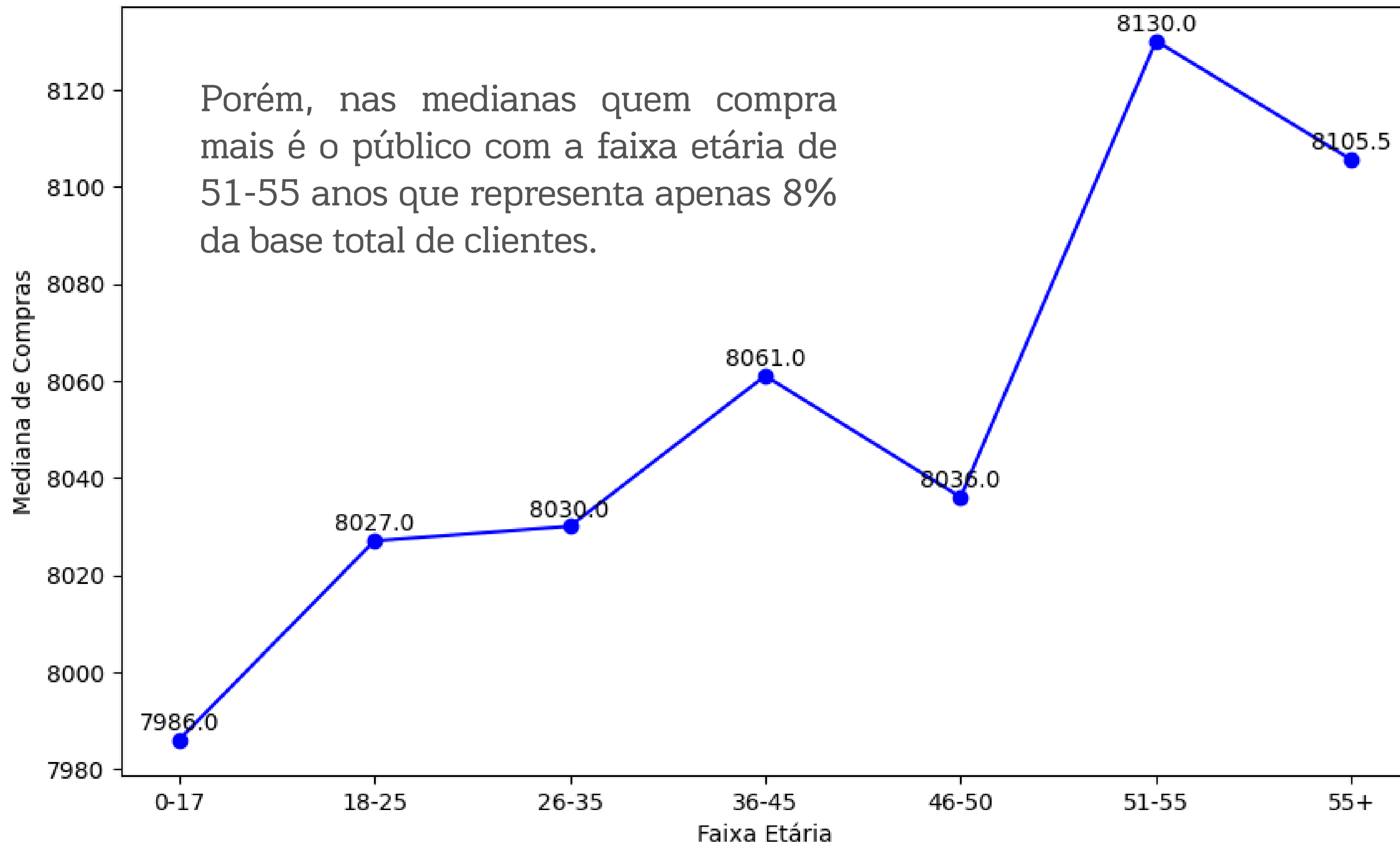


Aqui podemos ver que o público masculino compra mais o produto 17, enquanto o público feminino compra o produto 14. Podemos adaptar campanhas de marketing com bases nessas informações.

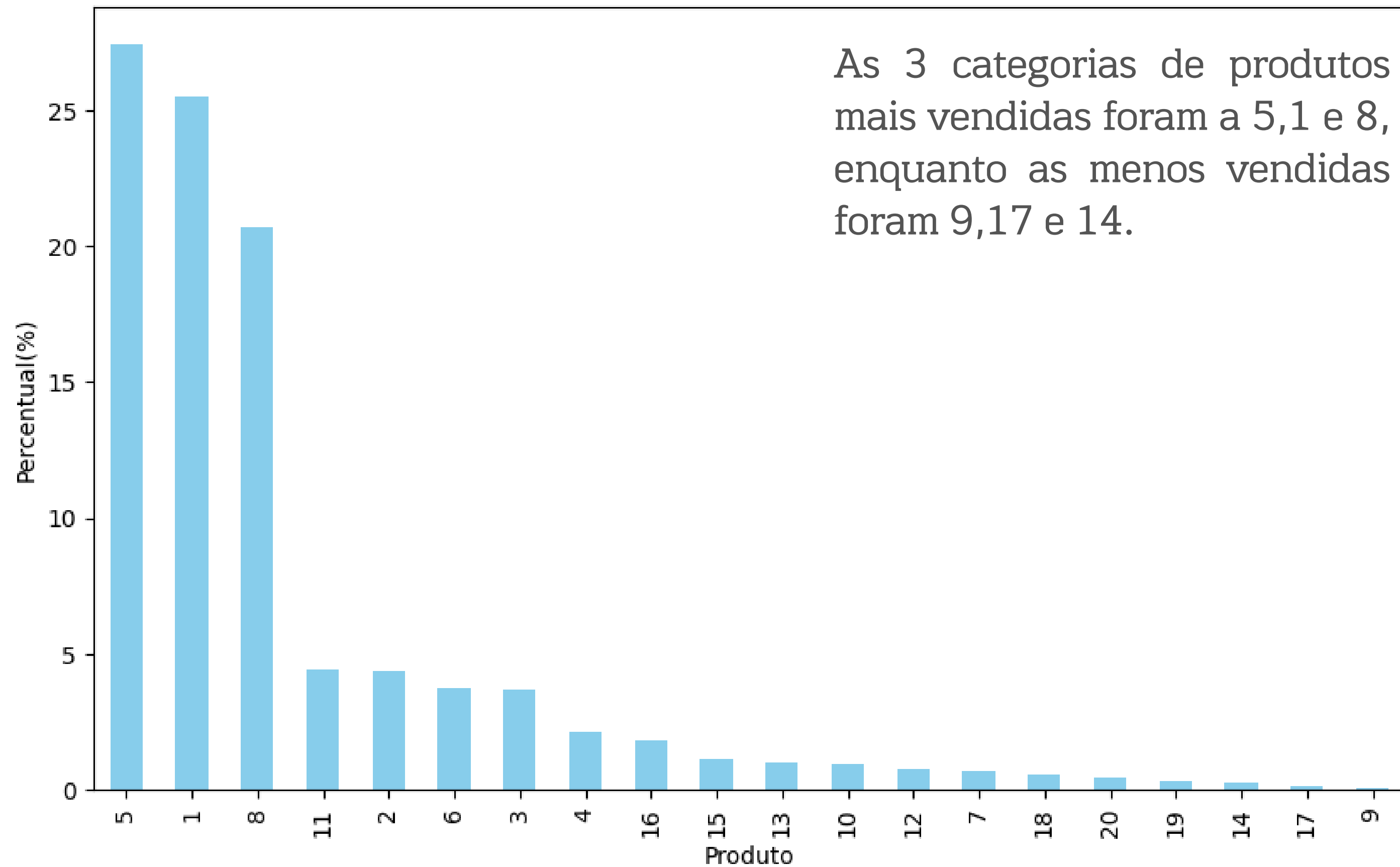
## Percentual de Usuários por Faixa Etária



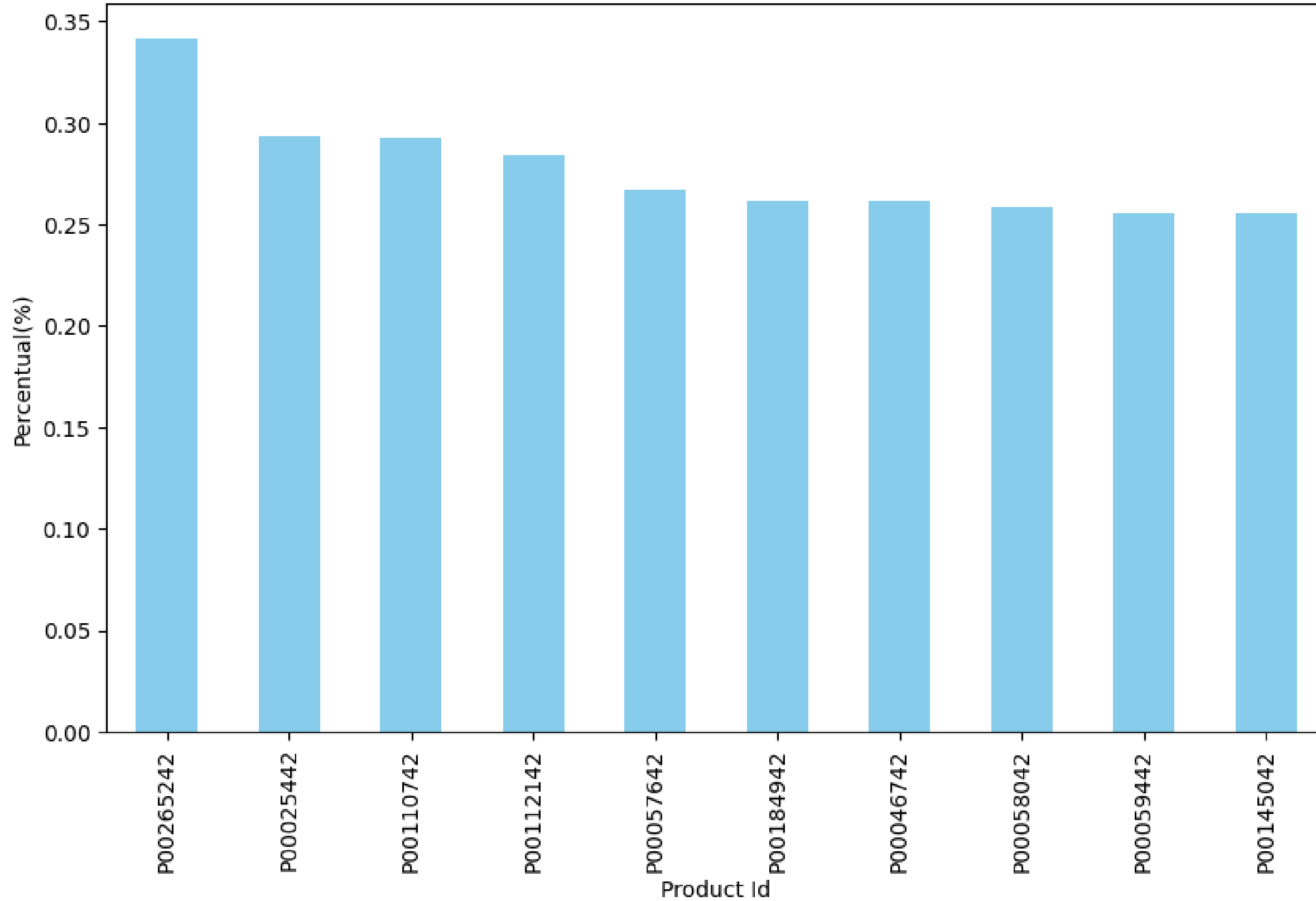
## Mediana de Compras por Faixa Etária



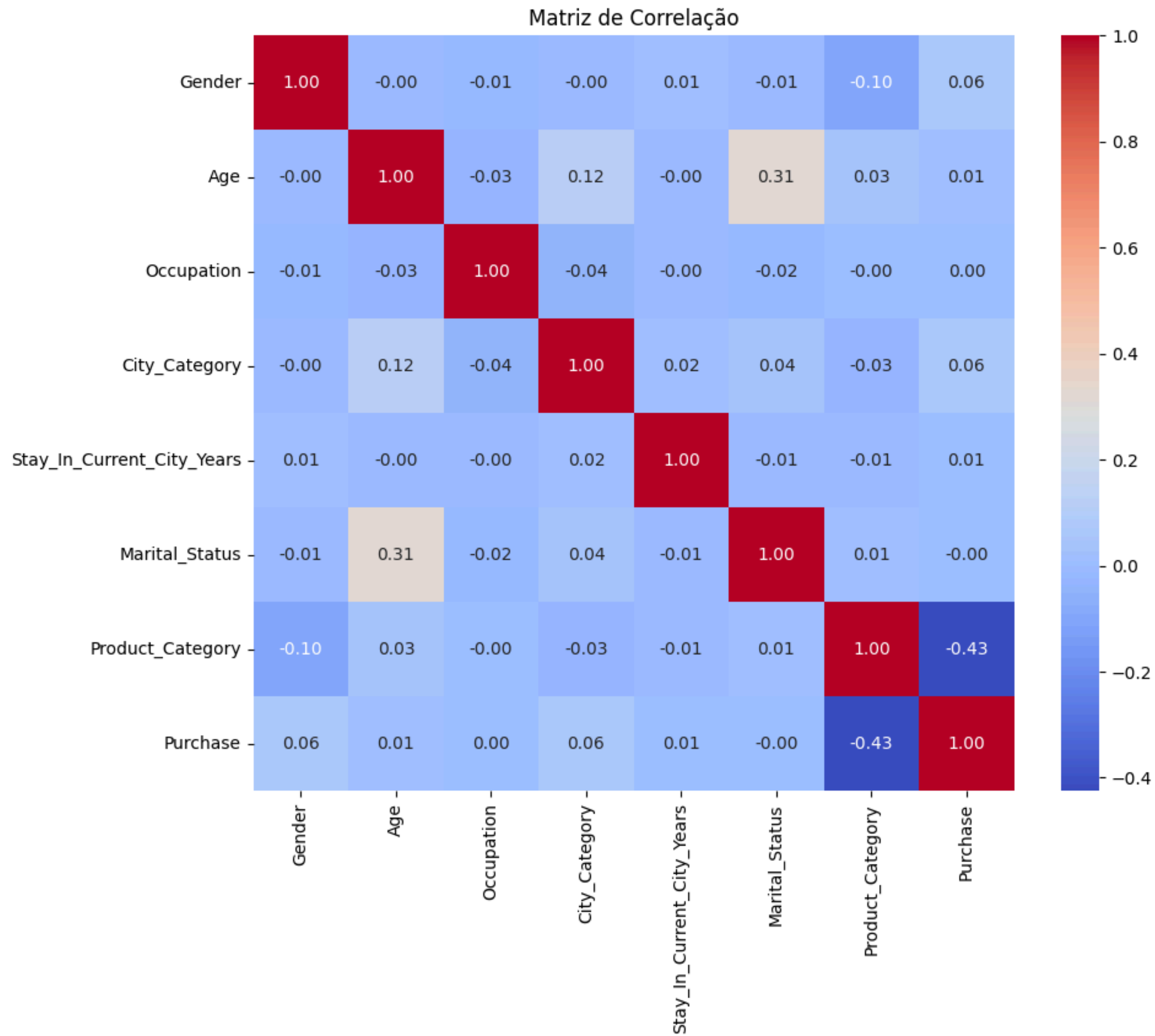
Percentual de Compras por Categoria de Produto



Percentual Product Id Mais Vendidos (Top 10)




Top 10 dos produtos mais vendidos.



Utilizaremos a matriz de correlação para ver a correlação entre as variáveis notamos que a maior correlação é entre 'age' e 'marital\_status', usaremos essa informação para treinar os modelos de Machine Learning, mas podemos ver que há uma baixa correlação entre as variáveis o que pode atrapalhar no treinamento dos modelos.

Modelo	MSE	MAE	R² Score	Observações
Árvore de Decisão	2.212	3.445	6.413	Bom desempenho, pode ser superado por modelos mais robustos
Regressão Linear	2.212	3.445	6.413	Mesmo desempenho da Árvore de Decisão, simplicidade pode ser uma vantagem
Gradient Boosting	2.293	3.663	6.282	Levemente pior que os outros, mas ainda competitivo
Random Forest	2.184	3.436	6.458	Melhor desempenho geral, captura complexidade
Regressão Ridge	5.043	5.426	1.822	Desempenho ruim, não captura bem os padrões




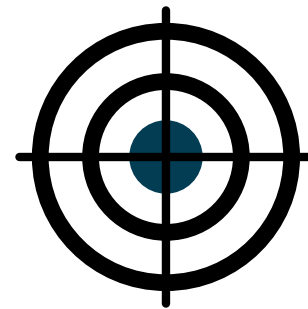


Melhor Modelo: O Random Forest teve o melhor desempenho com os menores erros (MSE e MAE) e o maior  $R^2$ , se mostrando que o mais eficaz para o conjunto de dados.

Piores Modelos: Tanto a Regressão Ridge quanto a Regressão Lasso tiveram desempenhos fracos, sugerindo que elas não são adequadas para este problema.

Modelos Satisfatórios: Árvore de Decisão e Regressão Linear tiveram desempenhos praticamente iguais e razoavelmente bons, embora não tão eficazes quanto o Random Forest.





Obrigado.