

# Estágio em Dados

**Itaú Unibanco**

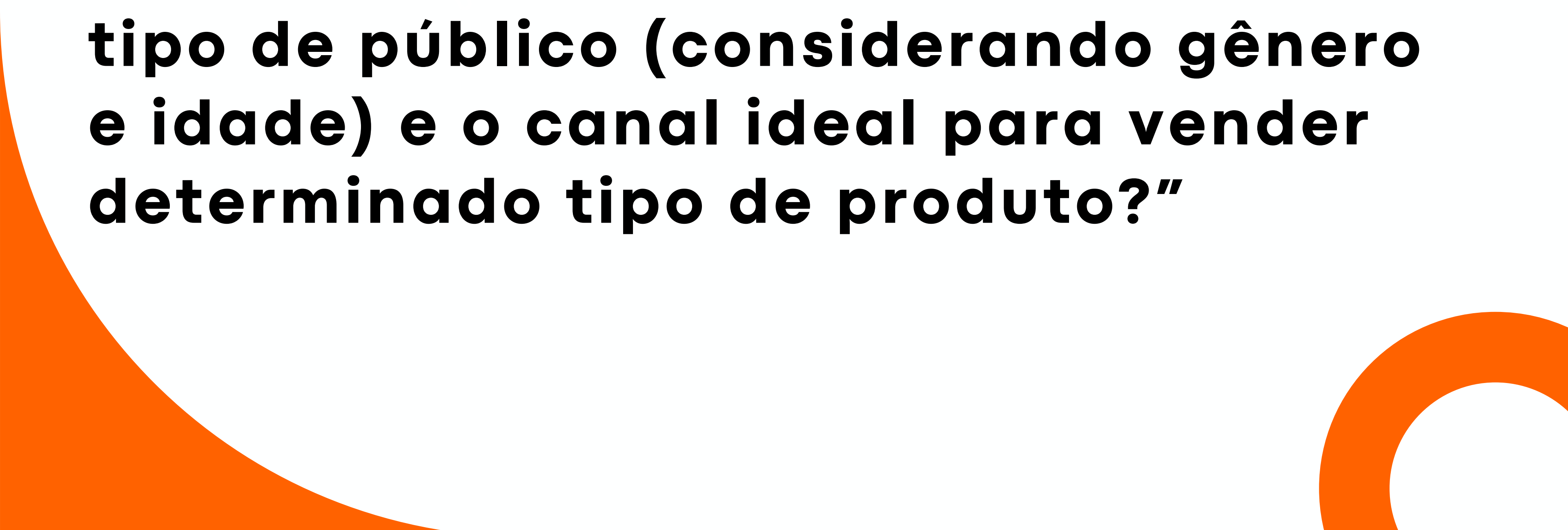


Presented By:  
Geraldo Vedrossi Neto



# Pergunta Norteadora

**“Analisando a base de dados, qual o tipo de público (considerando gênero e idade) e o canal ideal para vender determinado tipo de produto?”**



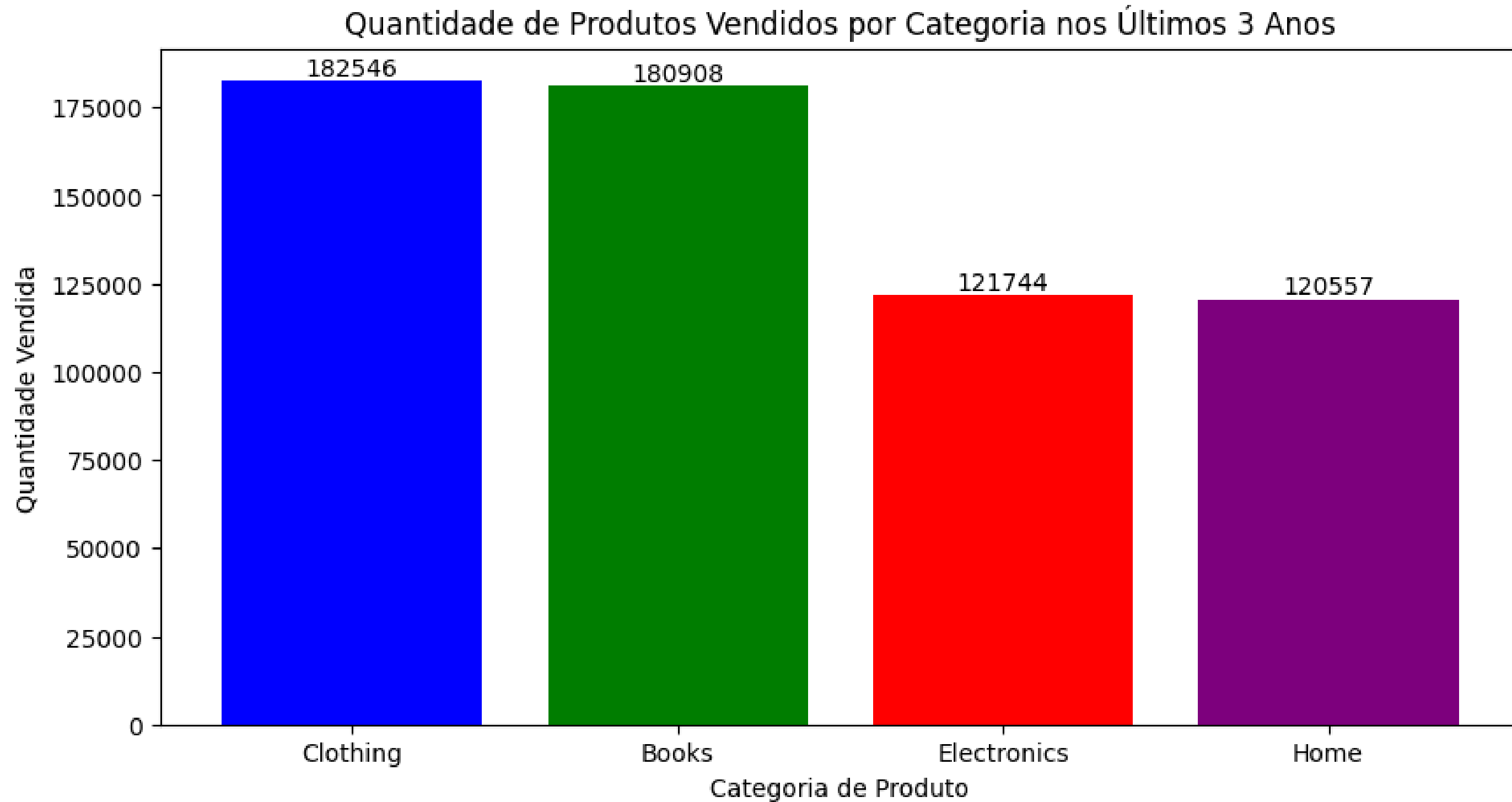
# DATASET

- |                          |               |
|--------------------------|---------------|
| 1. Customer ID           | 9. Gender     |
| 2. Purchase Date         | 10. Source    |
| 3. Product Category      | 11. Country   |
| 4. Product Price         | 12. State     |
| 5. Quantity              | 13. Latitude  |
| 6. Total Purchase Amount | 14. Longitude |
| 7. NPS                   |               |
| 8. Customer Age          |               |



# Roteiro de Análise

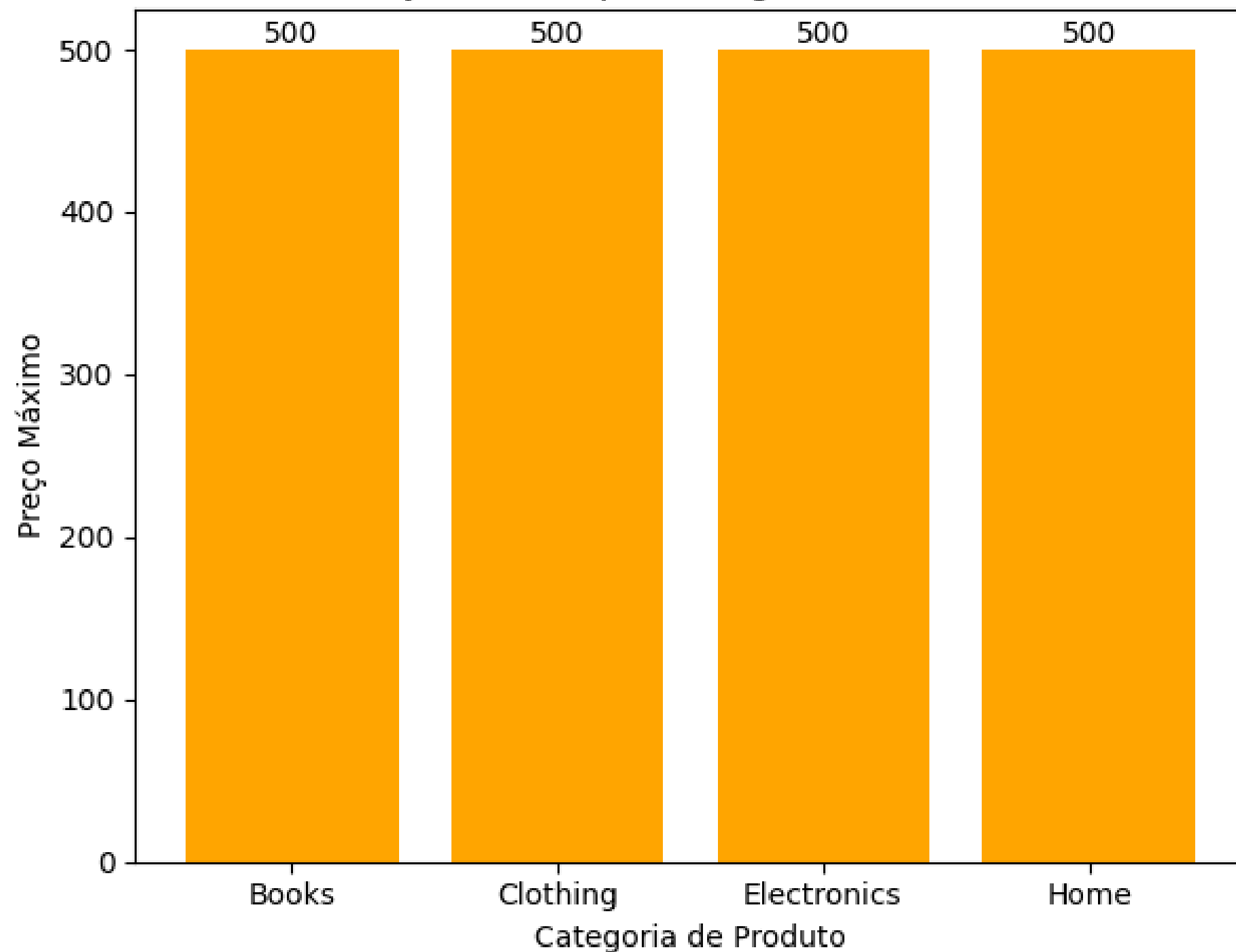
**Quais os produtos mais vendidos considerando os últimos 3 anos?**



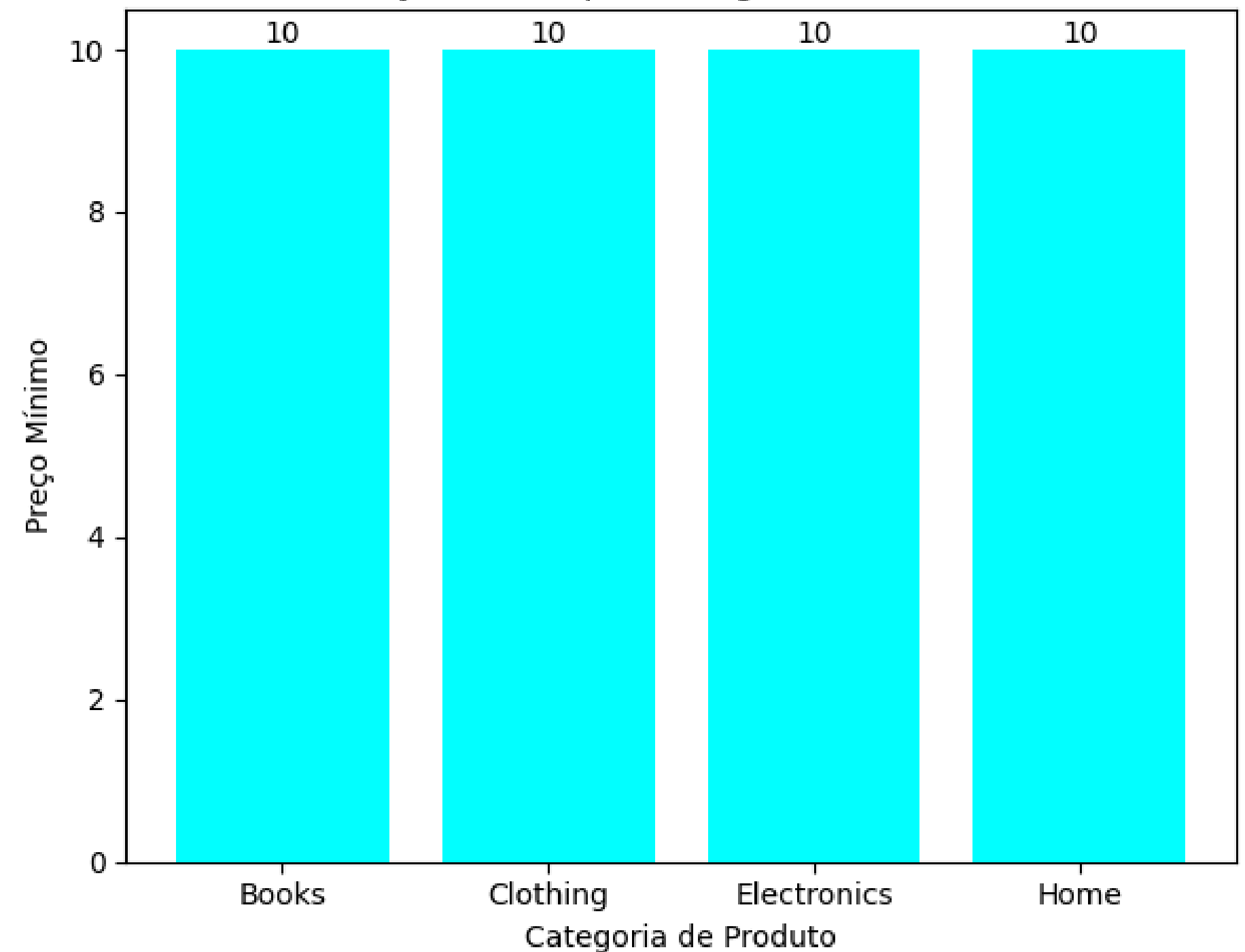
# Roteiro de Análise

**Qual o produto mais caro e o mais barato?**

Preço Máximo por Categoria de Produto

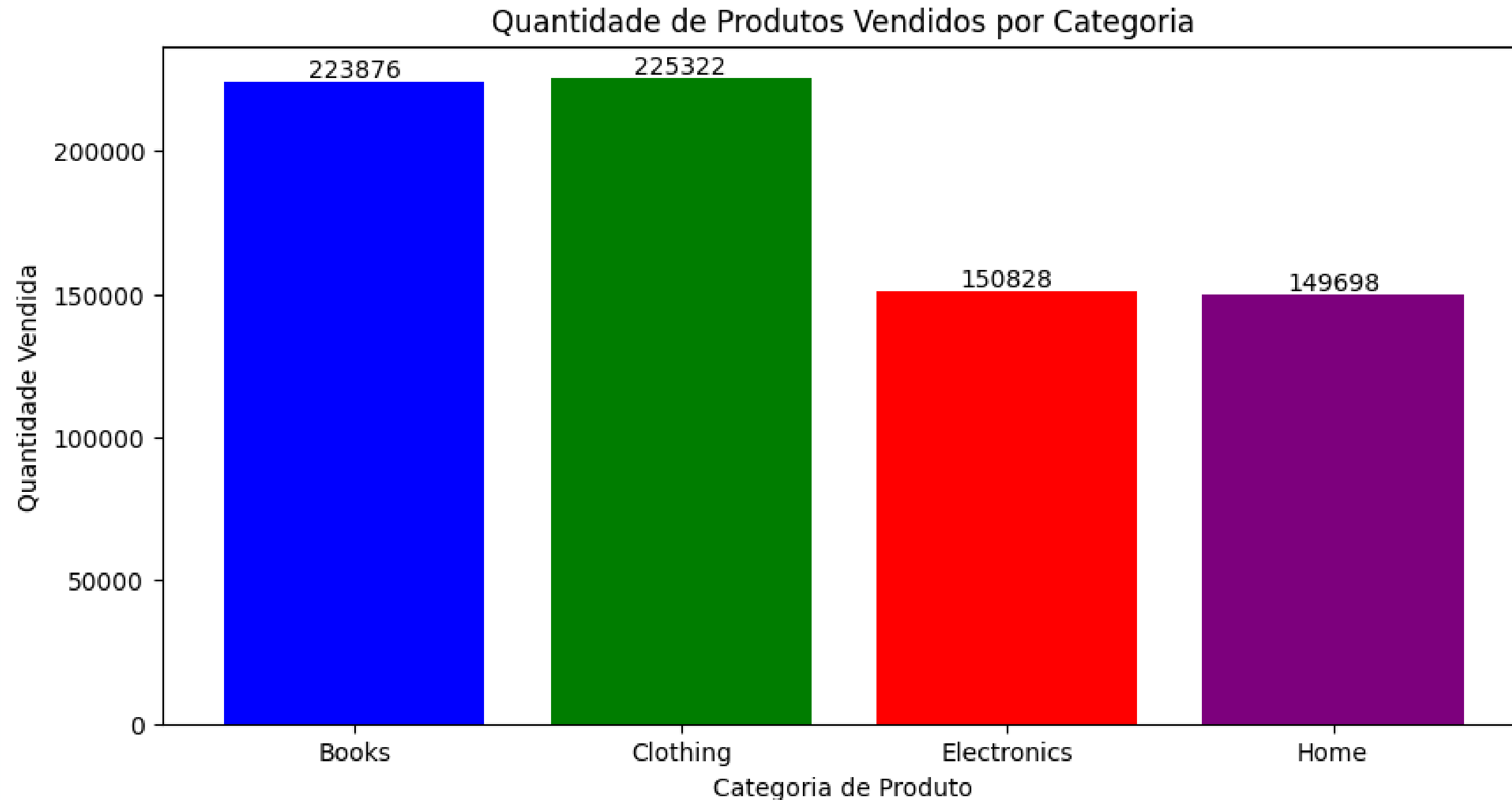


Preço Mínimo por Categoria de Produto



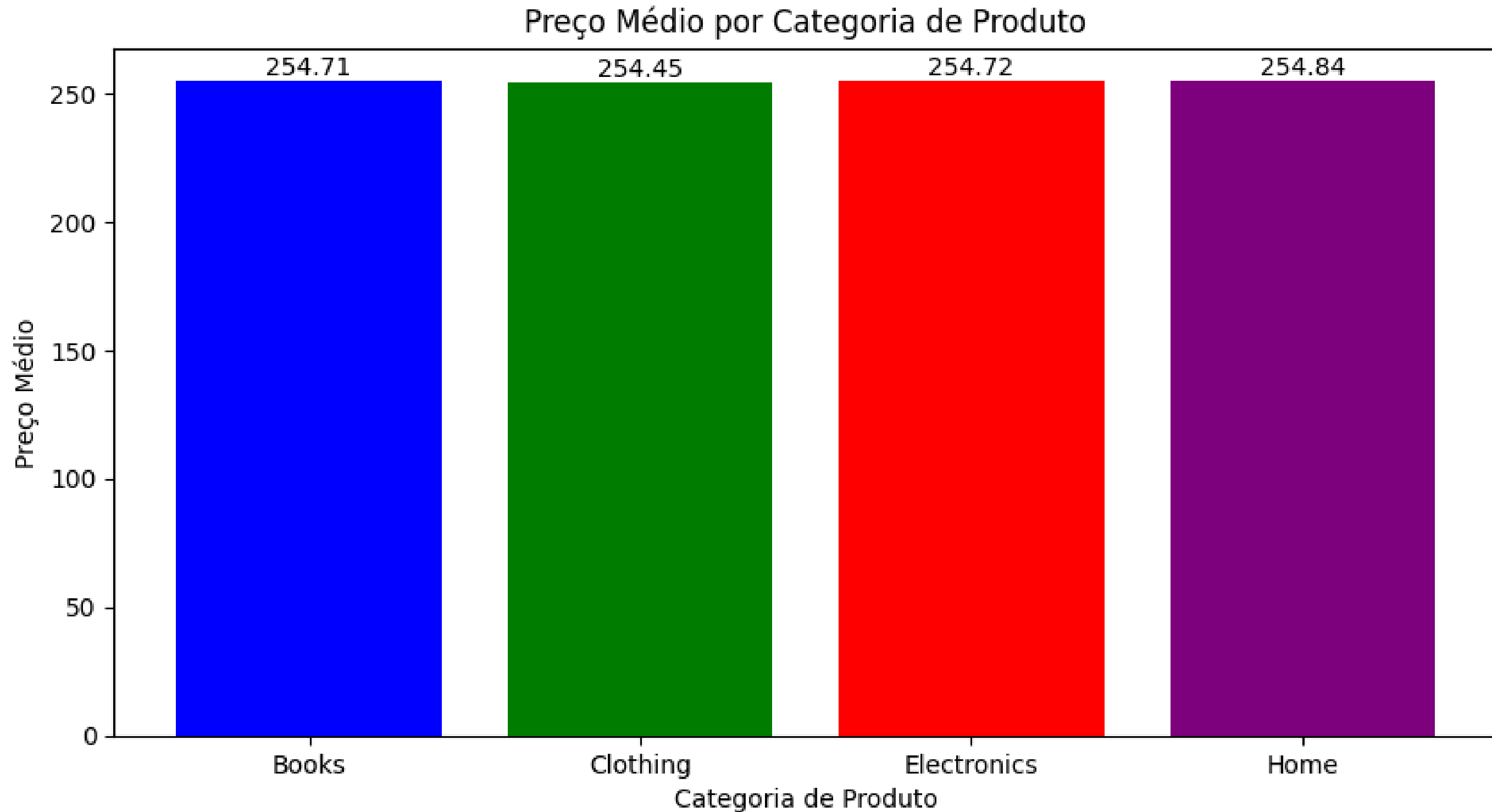
# Roteiro de Análise

**Qual a categoria mais vendida e a menos vendida?**



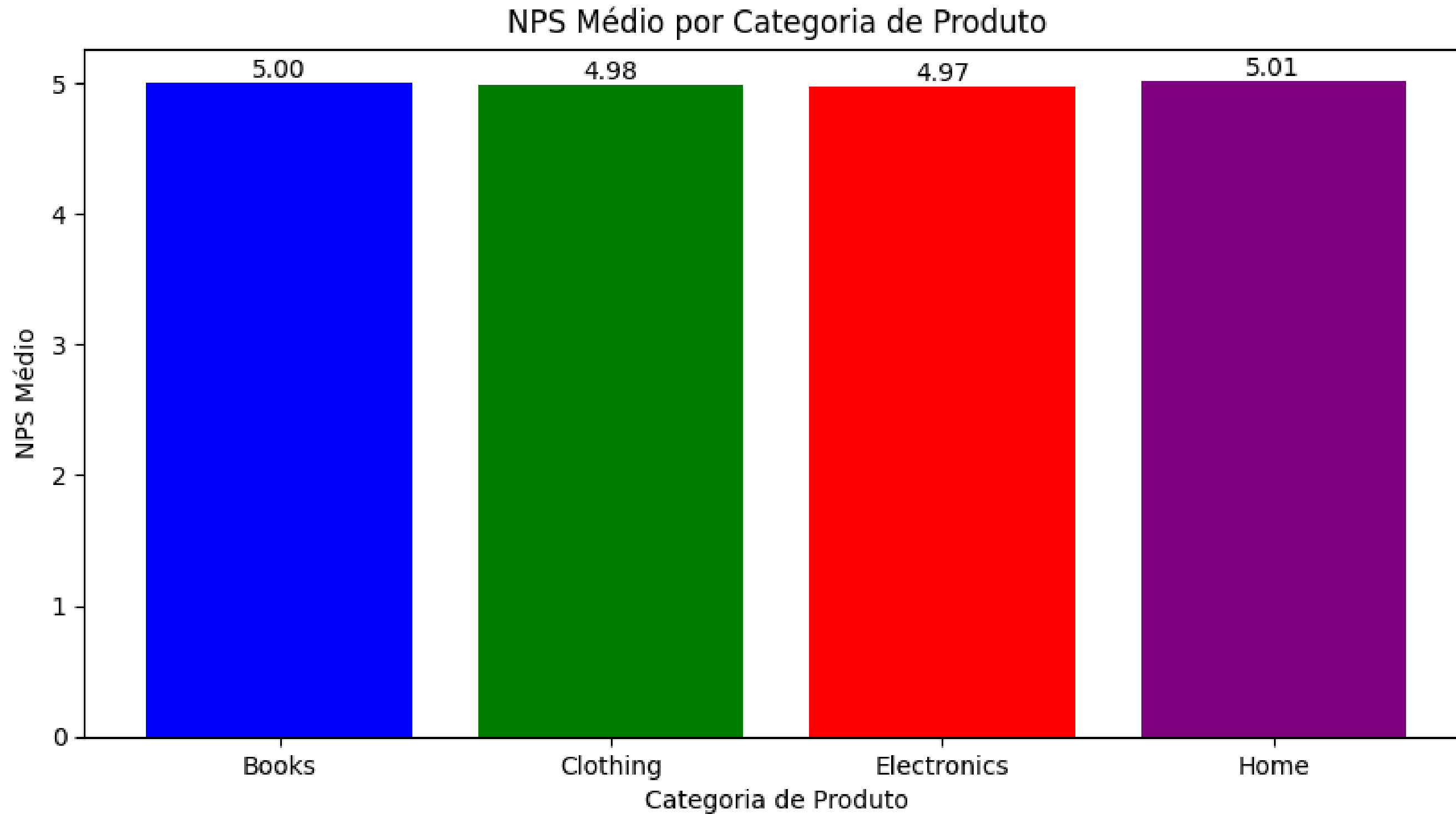
# Roteiro de Análise

**Qual a categoria mais cara e a menos cara?**



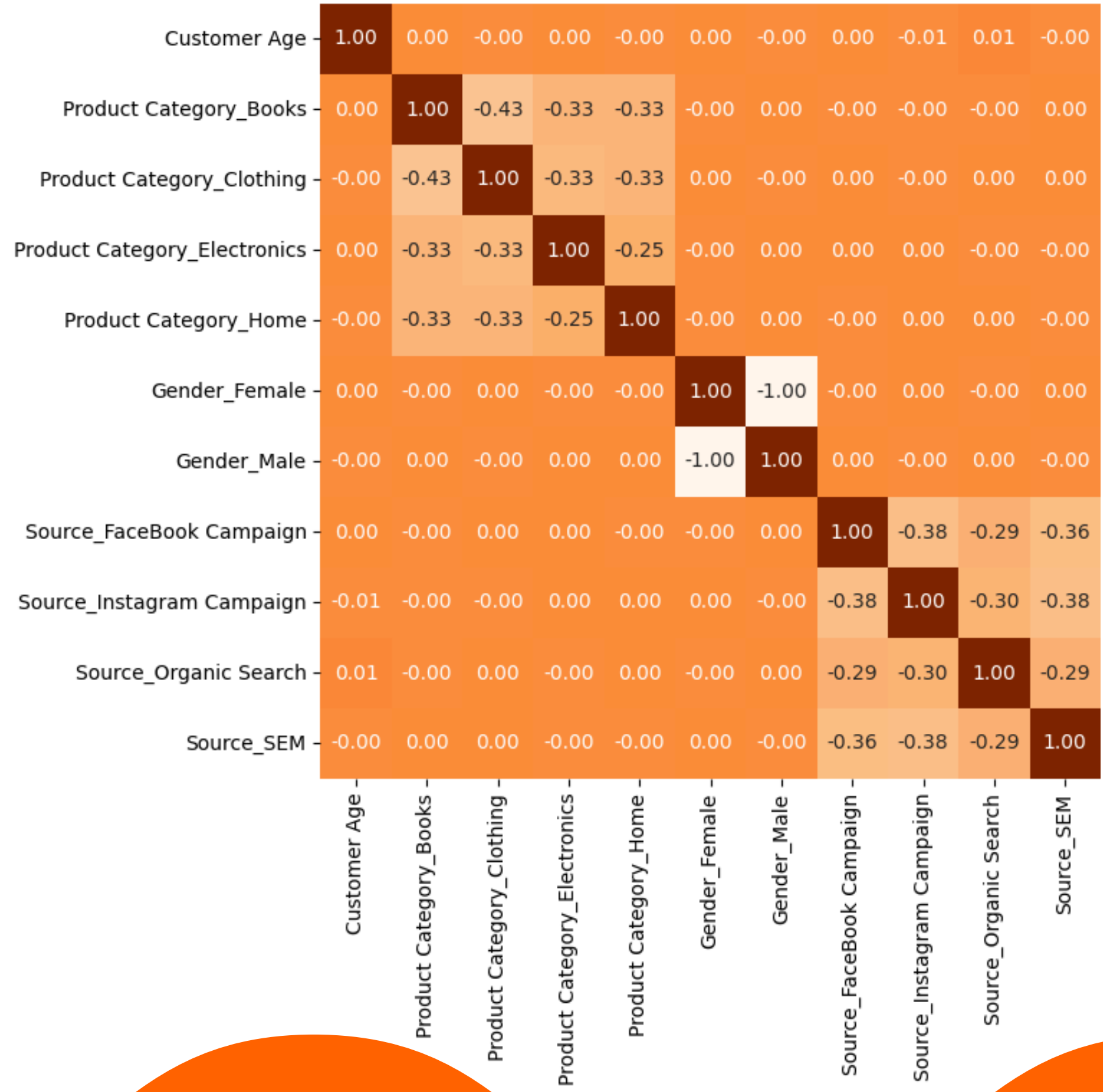
# Roteiro de Análise

**Qual o produto com melhor e pior NPS?**



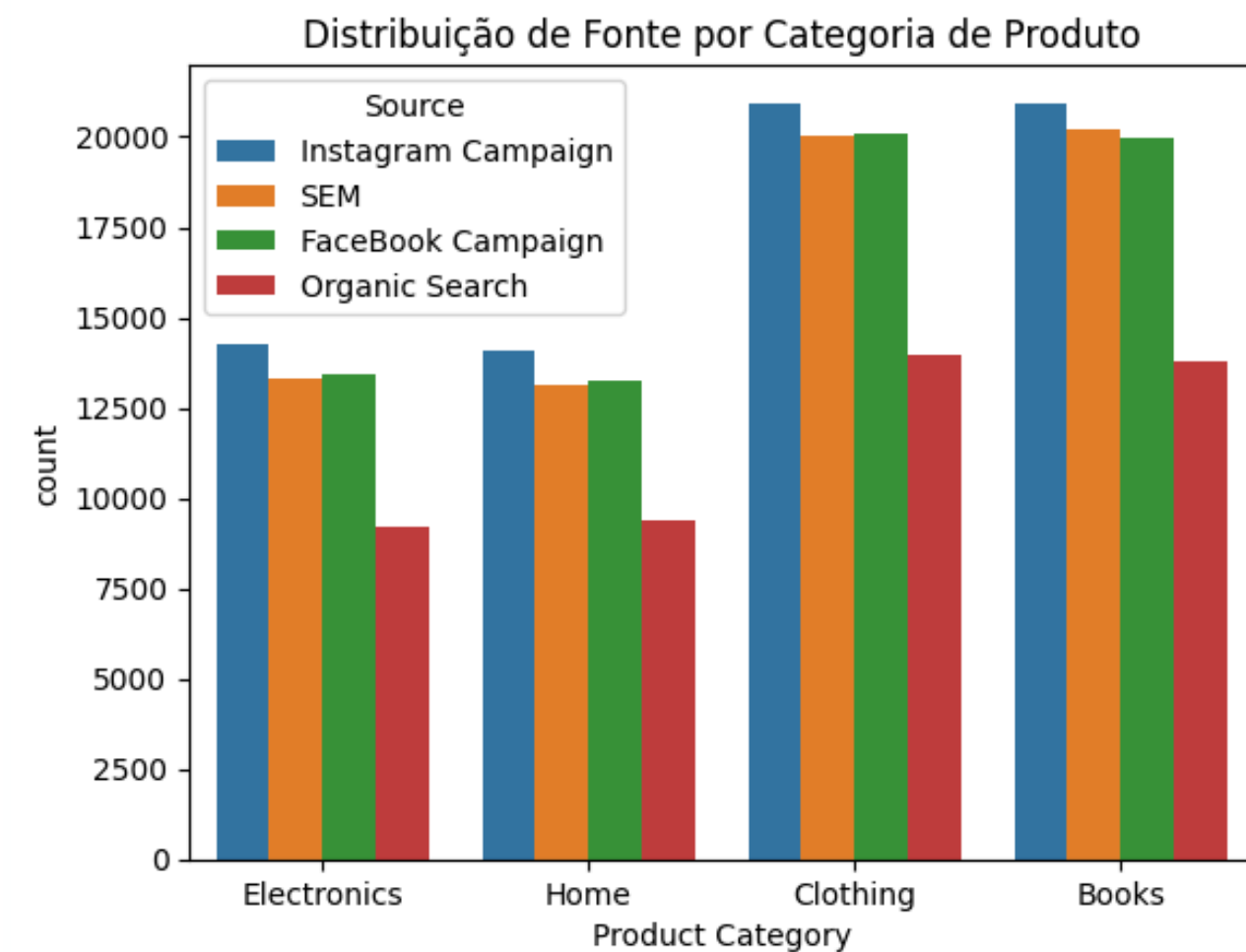
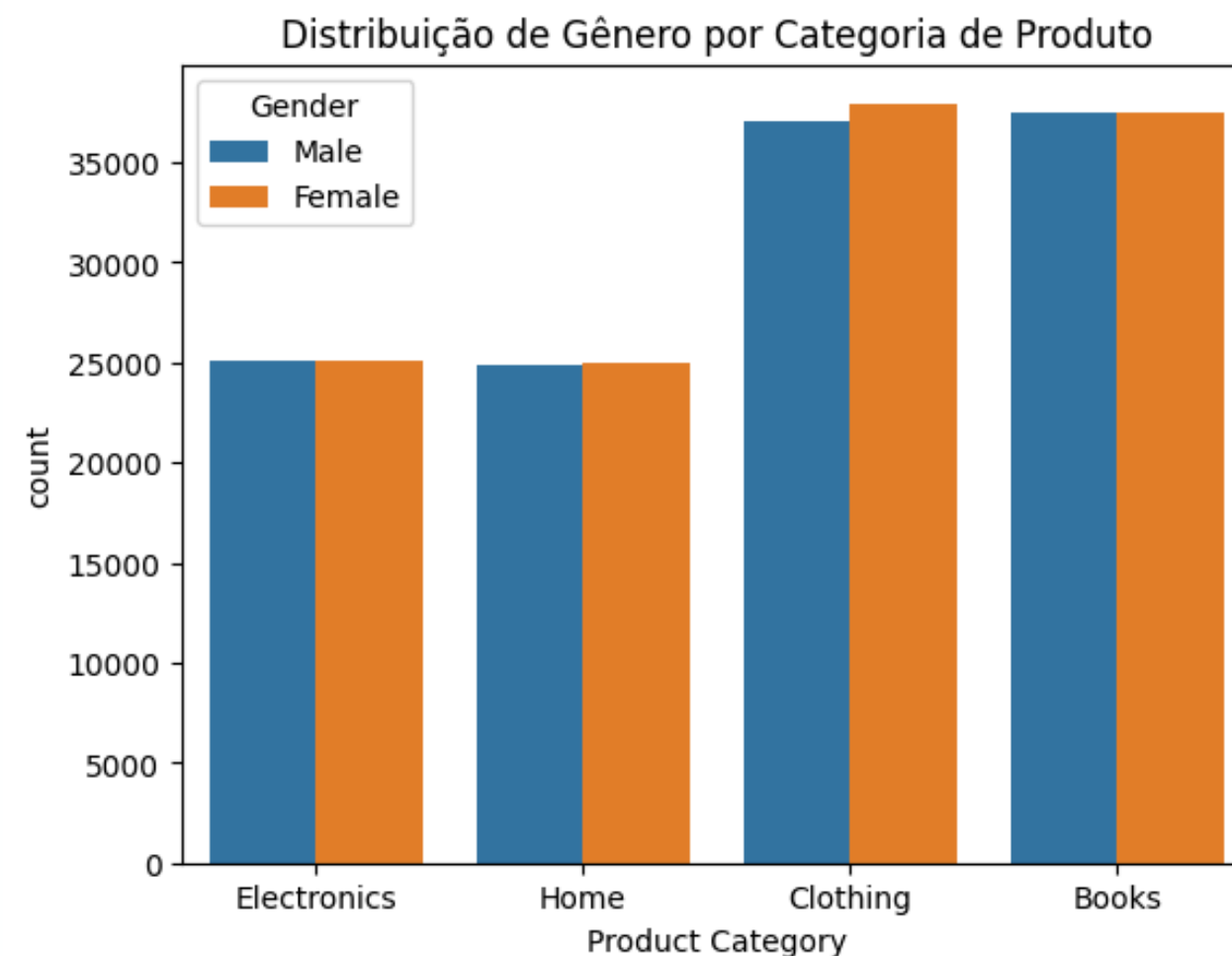
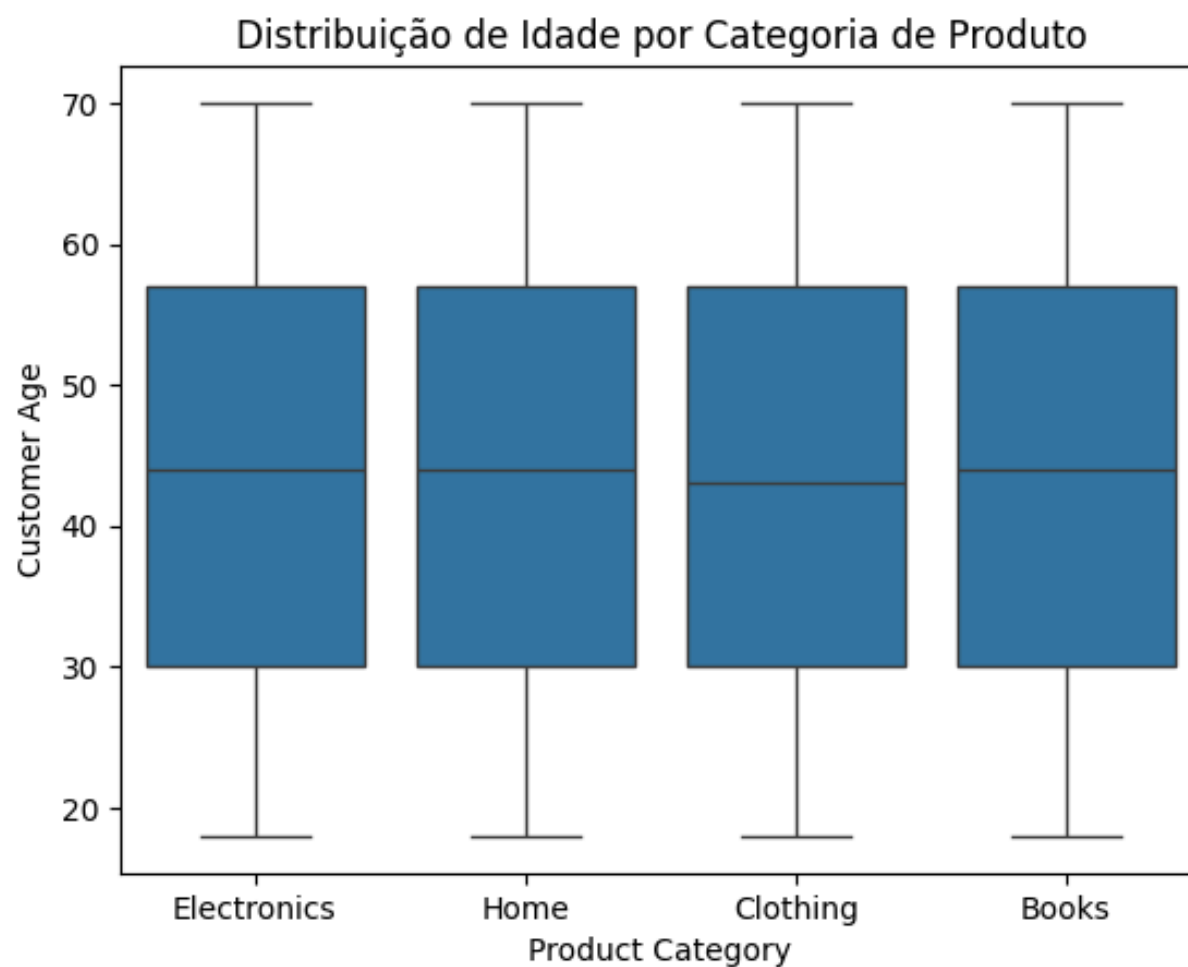


# Mapa de Calor - Correlação



# Gerando Hipóteses

- 1º Não há diferença na distribuição de idades por categoria de produto
- 2º Não há diferença na distribuição de gênero por categoria de produto
- 3º Não há diferença na distribuição de canais por categoria de produto



# Gerando Hipóteses - Idade

## **Hipótese Nula ( $H_0$ ):**

**Não há relação significativa entre as variáveis independentes (categoria do produto) e a variável dependente (idade do cliente).**

## **Hipótese Alternativa ( $H_1$ ):**

**Existe uma relação significativa entre as variáveis independentes (categoria do produto) e a variável dependente (idade do cliente).**

# Testando Hipóteses - Idade

	Coeficiente	Erro Padrão	Estatística t	Valor p	Intervalo de Confiança (2.5%)	Intervalo de Confiança (97.5%)
Const	1,23192E+11	1,44029E+11	0,855329142	0,392369778	-1,591E+11	4,05484E+11
Books	-1,23192E+11	1,44029E+11	-0,855329142	0,392369778	-4,05484E+11	1,591E+11
Clothing	-1,23192E+11	1,44029E+11	-0,855329142	0,392369778	-4,05484E+11	1,591E+11
Electronics	-1,23192E+11	1,44029E+11	-0,855329142	0,392369778	-4,05484E+11	1,591E+11
Home	-1,23192E+11	1,44029E+11	-0,855329142	0,392369778	-4,05484E+11	1,591E+11

# Gerando Hipóteses - Gênero

## **Hipótese Nula ( $H_0$ ):**

**Não há relação significativa entre as variáveis independentes (categoria do produto) e a variável dependente (gênero do cliente).**

## **Hipótese Alternativa ( $H_1$ ):**

**Existe uma relação significativa entre as variáveis independentes (categoria do produto) e a variável dependente (gênero do cliente).**

# Testando Hipóteses – Gênero

	Coeficiente	Erro Padrão	Estatística t	Valor p	Intervalo de Confiança (2.5%)	Intervalo de Confiança (97.5%)
Const	-2599281139	4678601946	-0,555567917	0,578506766	-11769216847	6570654570
Books	2599281139	4678601946	0,555567917	0,578506766	-6570654569	11769216848
Clothing	2599281139	4678601946	0,555567917	0,578506766	-6570654569	11769216848
Electronics	2599281139	4678601946	0,555567917	0,578506766	-6570654569	11769216848
Home	2599281139	4678601946	0,555567917	0,578506766	-6570654569	11769216848

# Gerando Hipóteses - Canal

## **Hipótese Nula ( $H_0$ ):**

**Não há relação significativa entre as variáveis independentes (categoria do produto) e a variável dependente (canal do cliente).**

## **Hipótese Alternativa ( $H_1$ ):**

**Existe uma relação significativa entre as variáveis independentes (categoria do produto) e a variável dependente (canal do cliente).**

# Testando Hipóteses - Canal

	Coeficiente	Erro Padrão	Estatística t	Valor p	Intervalo de Confiança (2.5%)	Intervalo de Confiança (97.5%)
Const	4458849005	4139451354	1,077159416	0,281410092	-3654365845	12572063854
Books	-4458849004	4139451354	-1,077159416	0,281410092	-12572063854	3654365845
Clothing	-4458849004	4139451354	-1,077159416	0,281410092	-12572063854	3654365845
Electronics	-4458849004	4139451354	-1,077159416	0,281410092	-12572063854	3654365845
Home	-4458849005	4139451354	-1,077159416	0,281410092	-12572063854	3654365845



# Comprovando Hipóteses


## Resultado:

Os valores **p** associados aos coeficientes indicam a probabilidade de observar um efeito igual ou mais extremo do que o observado, assumindo que a **Hipótese Nula( $H_0$ )** é verdadeira (ou seja, não há relação significativa entre a variável independente e a variável dependente).



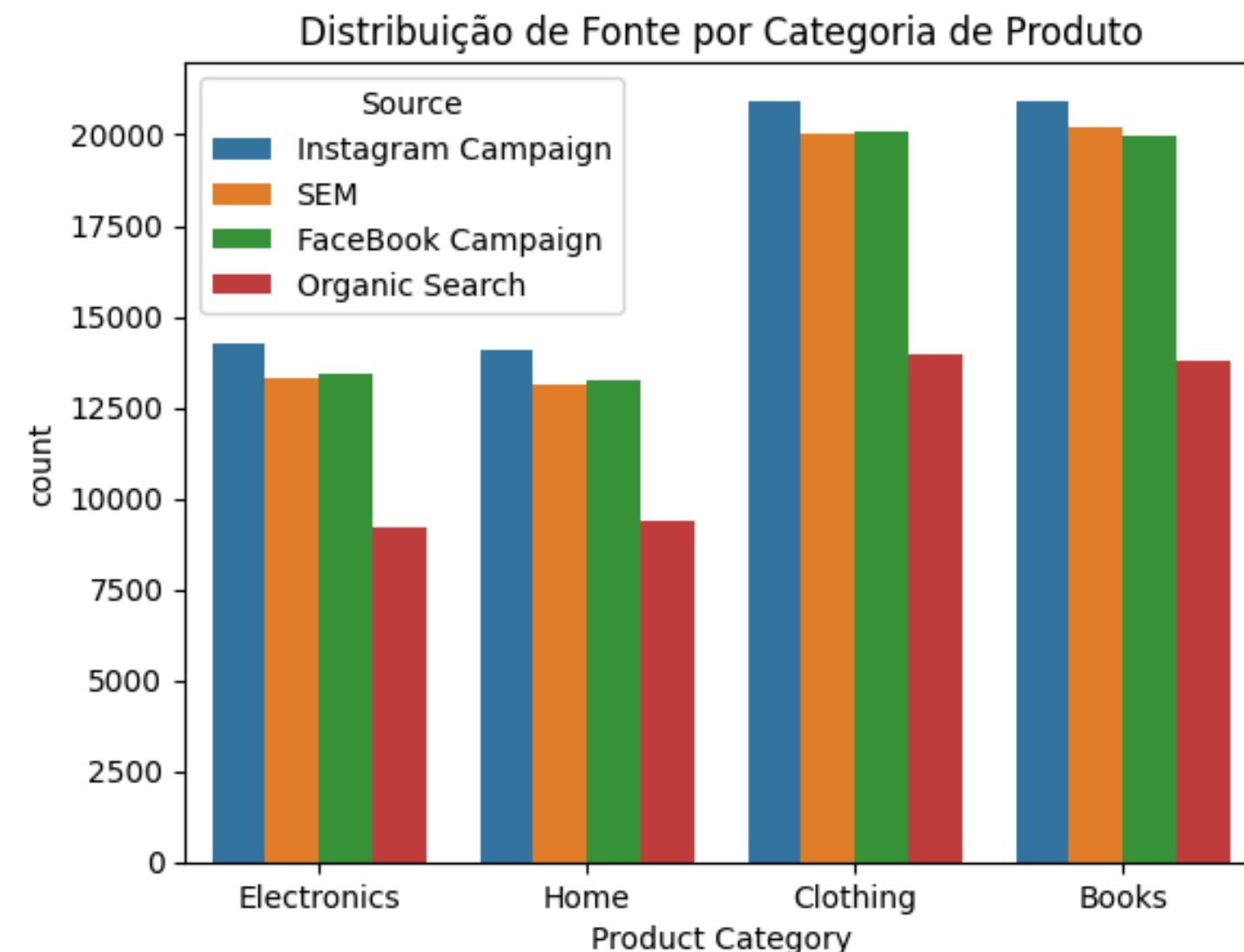
# CONCLUSÃO

**“Qual o tipo de público (considerando gênero e idade) e o canal ideal para vender determinado tipo de produto?”**

- Não existe correlação entre o **tipo de público** e **tipo de produto**.
  - Enquanto a distribuição dos **canais** se mantém a mesma para **tipo de público** e **tipo de produto**.
- 

# CONCLUSÃO

- Logo, todo **tipo de produto** pode ser vendido para qualquer **perfil de cliente**.
- E o **canal** recomendado será a **campanha do Instagram**.





Muito

Obrigado!

Geraldo Vedrossi Neto

