



iFood CRM Data Analyst Case

Visão Geral

Problemática

A empresa enfrenta o desafio de **melhorar significativamente** a rentabilidade de **suas campanhas de marketing direto**. Para isso, é necessário identificar o público certo, reduzir desperdícios de recursos e implementar uma abordagem baseada em dados que maximize o lucro e gere confiança nos métodos quantitativos.

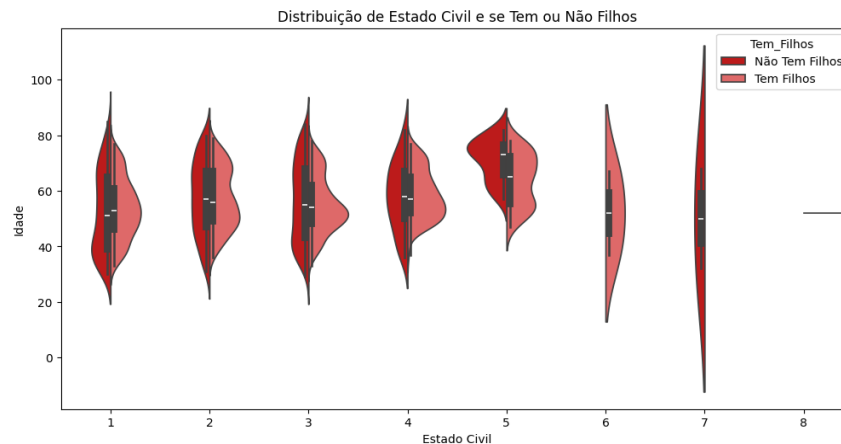
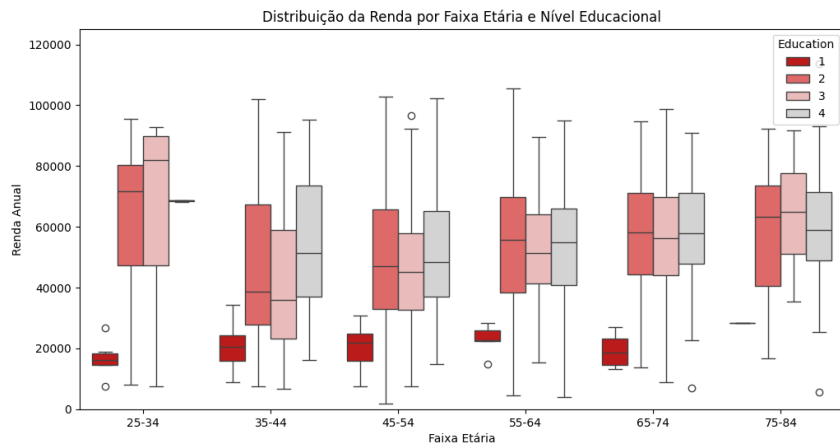
Objetivos

Objetivo Principal: Desenvolver um modelo preditivo que maximize o lucro da próxima campanha de marketing direto.

Objetivo Secundário: Fornecer insights detalhados sobre o perfil dos clientes que compraram o gadget, facilitando ações de marketing futuras.

Análise Exploratória dos Dados (**EDA**)

Quem são os clientes? | Dados sociodemográficos



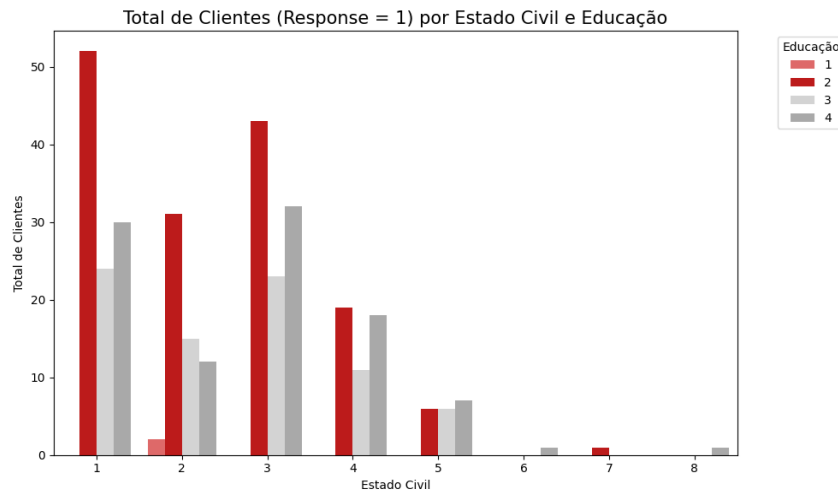
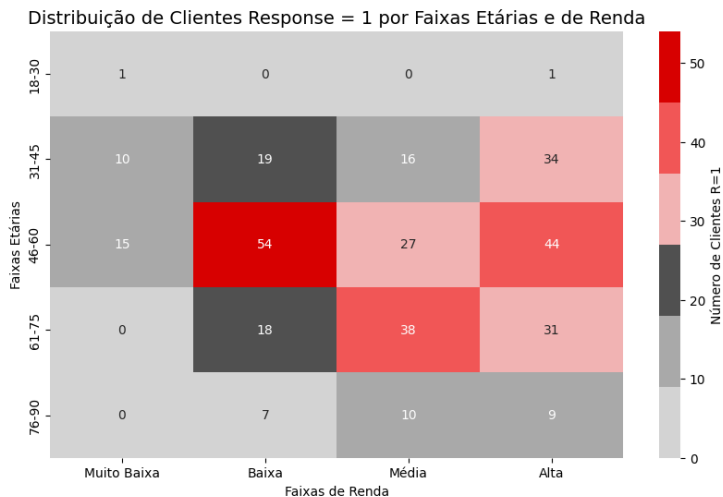
Education: Basic 1, Graduation 2, Master 3, PhD 4

Marital_Status: Single 1, Together 2, Married 3, Divorced 4, Widow 5, Alone 6, Absurd 7, YOLO 8

Distribuição de Idade, Níveis Educacionais, Renda, Filhos e Estado Civil | Todos que responderam a pesquisa

1. A faixa etária dos indivíduos varia de 29 a 85 anos, com a maioria concentrada entre **48 e 66 anos**. A renda mínima é de R\$ 1.730, e a máxima atinge R\$ 113.734, indicando uma grande desigualdade no rendimento.
2. Em termos de escolaridade, a maior **parte dos indivíduos possui Graduation (Nível 2)**, representando 50% da amostra, seguido por Master (Nível 3) e PhD (Nível 4). O Basic (Nível 1) é o nível educacional menos comum, com apenas 54 pessoas, evidenciando que a maioria tem formação educacional mais avançada.
3. A maioria das pessoas tem filhos, há ligeiramente mais filhos adolescentes do que filhos pequenos. O estado civil indica que a maioria está casada ou em união estável. Alguns indivíduos têm até dois filhos em cada faixa etária.

Quem são os clientes? | Dados sociodemográficos



Educação: Basic 1, Graduation 2, Master 3, PhD 4

Marital_Status: Single 1, Together 2, Married 3, Divorced 4, Widow 5, Alone 6, Absurd 7, YOLO 8

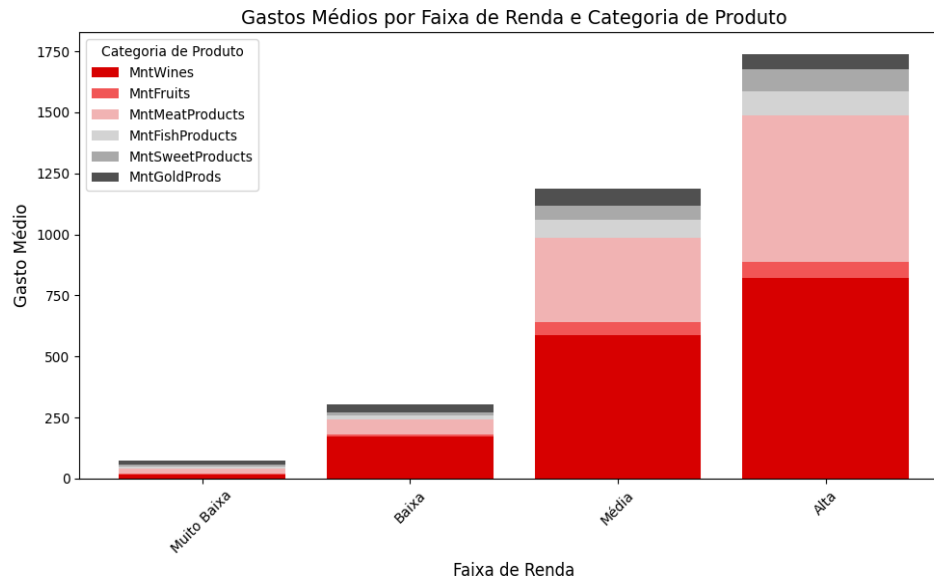
Distribuição de Idade, Níveis Educacionais, Renda, Filhos e Estado Civil | Aderiram a Campanha (Response = 1)

1. A maioria dos clientes pertence ao Estado Civil 1 (single), especialmente com Educação 2 (Graduation).
2. Baixa adesão no Estado Civil 2 com Educação 1: Apenas 2 clientes, indicando menor engajamento desse grupo.
3. Faixa etária 46-60 anos é o destaque em quase todas as faixas de renda. Com destaque para clientes com a renda "Baixa"
4. Potencial em renda alta: Clientes 31-45 anos com renda "Alta" têm bom engajamento e poder aquisitivo.

Quais são os comportamentos de Compra? | Produtos

Relação Renda & Gastos nas categorias de produtos

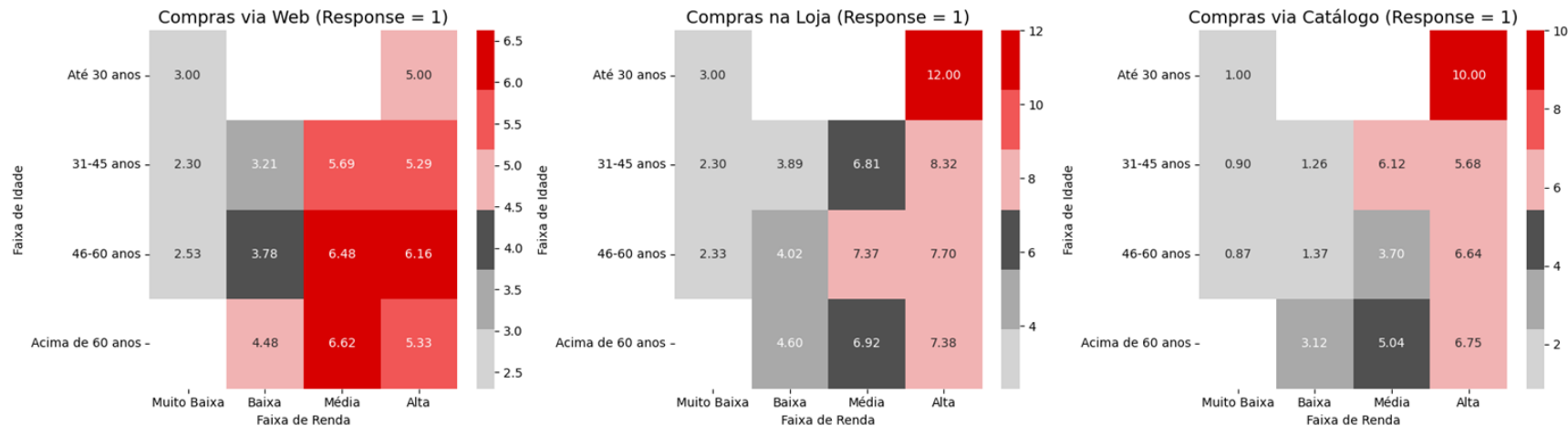
1. Existe uma tendência clara de **aumento nos gastos à medida que a faixa de renda sobe**, especialmente nas categorias de vinhos, carne e produtos "gold". Por exemplo, clientes na faixa de renda Muito Baixa gastam em média 11,10 em vinhos, enquanto na faixa Alta, o gasto sobe para 673,64.
2. Os clientes com maior renda tendem a gastar mais com produtos "gold" (72,53), indicando que a renda está fortemente associada à compra de itens mais exclusivos.
3. Os clientes da faixa Média têm uma distribuição de gastos mais equilibrada entre diferentes categorias, com destaque para os gastos com carnes (204,4) e produtos de peixe (48,24), o que pode indicar uma maior diversificação nas compras à medida que a renda se torna mais estável.



Income: Muito Baixa (0–25.000), Baixa (25.001–50.000), Média (50.001–75.000), Alta (Acima de 75.000)

*Valores Anuais

Quais são os comportamentos de Compra? | Canais de Compra



Relação entre idade e renda por canal de venda (web, loja, catálogo) | Aderiram a Campanha (Response = 1)

1. Em geral, **pessoas com maior renda tendem a comprar mais** tanto pela web quanto nas lojas físicas, com uma presença significativa de **compras via catálogo** nas faixas de renda "Alta".
2. As **compras online** são mais populares **em pessoas com maior renda (média e alta)**, tendo consumidores de **baixa renda preferindo realizar compras em lojas físicas**.
3. As faixas etárias mais jovens (Até 30 anos) tendem a ter um comportamento mais expressivo nas compras por web e loja, enquanto as **faixas etárias mais velhas** (Acima de 60 anos) **possuem compras mais equilibradas**, com um aumento nas compras via catálogo nas faixas de renda mais altas.

Segmentação de Clientes - **Clusterização**

Segmentação de Clientes | Clusterização

O objetivo desta etapa é realizar a **segmentação de clientes** para identificar grupos com características semelhantes, permitindo ações de marketing direcionadas e maximizando o potencial de sucesso da campanha do gadget. A segmentação foi construída com base na Análise Exploratória de Dados (EDA).

Clusterização:

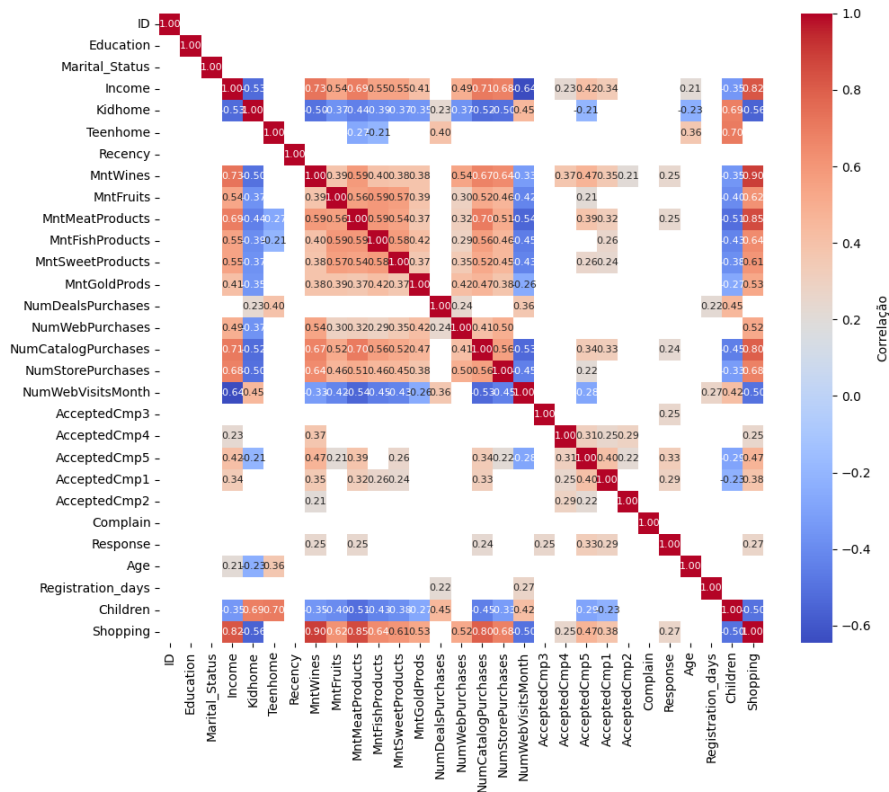
- Utilizou-se algoritmos como **K-Means** para agrupar clientes com características semelhantes. Além disso, foi aplicada a técnica de **Análise de Componentes Principais (PCA)** para redução de dimensionalidade.

Análise dos Segmentos

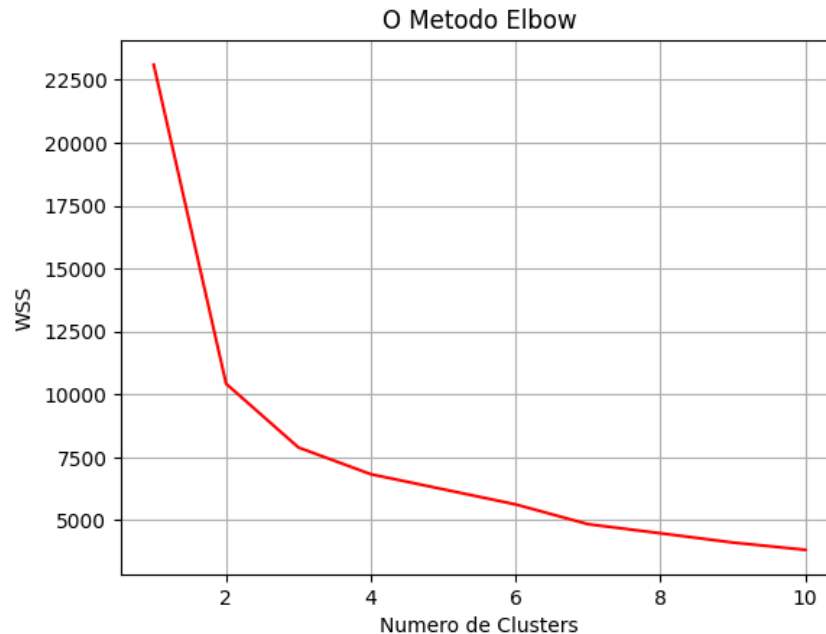
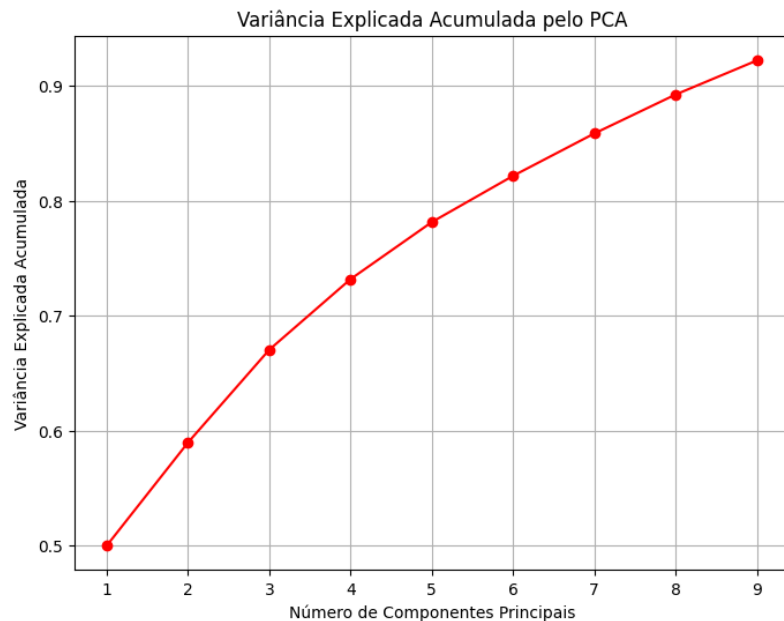
- Cada segmento foi descrito destacando suas principais características (ex.: clientes de alta renda, famílias numerosas de média renda).
- Foi analisada a relação de cada grupo com a **probabilidade de compra**, identificando os segmentos mais promissores para a campanha.

Validação dos Segmentos

- A coerência dos grupos foi avaliada por métricas de clusterização (como silhueta) e validação cruzada.



PCA e Elbow | Dimensionalidade e Número de Clusters

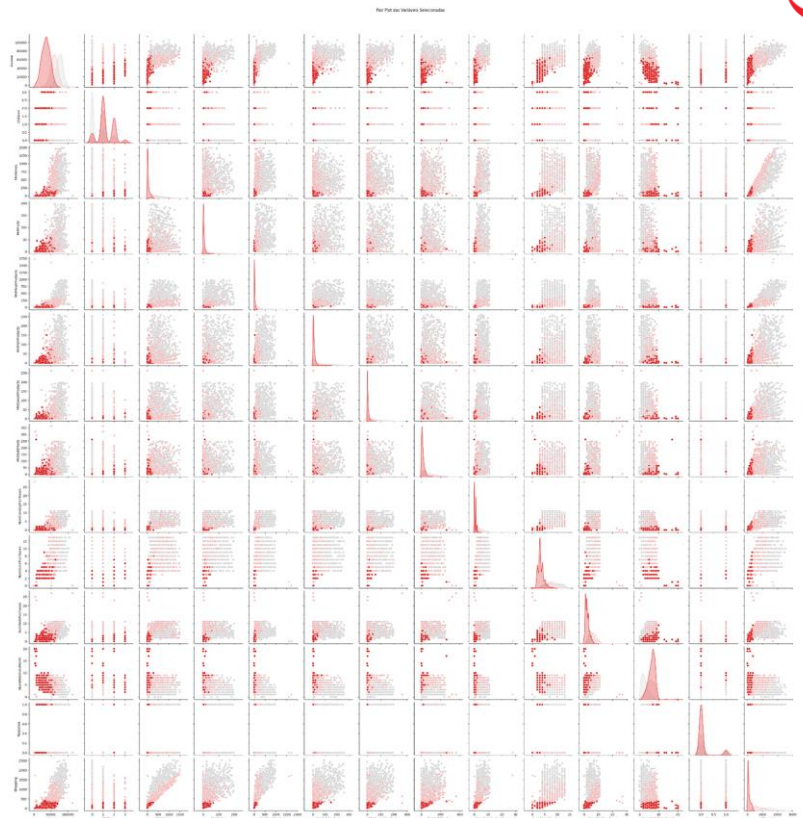


Validação dos Segmentos | Clusters

Cluster 0: Renda baixa, com filhos e pouco a moderado consumo de produtos. Possui o maior número de clientes e taxa de conversão relativamente baixa (**9,2%**), o que sugere que o engajamento com esse grupo é menor.

Cluster 1: Clientes de renda alta, sem filhos, com interesse específico em produtos de alta qualidade, como vinhos e carne. Apresenta a maior taxa de conversão de clientes (**29.4%**)

Cluster 2: Clientes renda moderada, com filhos, maior gasto em vinhos, carne e doces tem uma taxa de conversão intermediária (**12,3%**)



Modelo Preditivo - **Classificação**

Modelo Preditivo | **Classificação**

O objetivo desta etapa é prever a probabilidade de resposta dos clientes a campanhas de marketing, auxiliando na tomada de decisões estratégicas, utilizaremos a segmentação de clientes para contribuir como atributo do modelo.

Processo de Segmentação e Modelagem:

- **Divisão dos Dados:** Separação dos dados em 80% para treino e 20% para teste.
- **Pesos de Amostras:** Atribuição de pesos aos clusters para priorizar certos grupos no treinamento do modelo.
- **Normalização:** Aplicação do StandardScaler para padronizar os dados de treino e teste.

Classificação e Avaliação:

- **Modelos de Classificação:** Teste de diferentes modelos, como Random Forest, SVM, Logistic Regression, entre outros.
- **Avaliação de Desempenho:** Treinamento com pesos de amostras, cálculo de acurácia e uso de validação cruzada para medir a performance dos modelos.

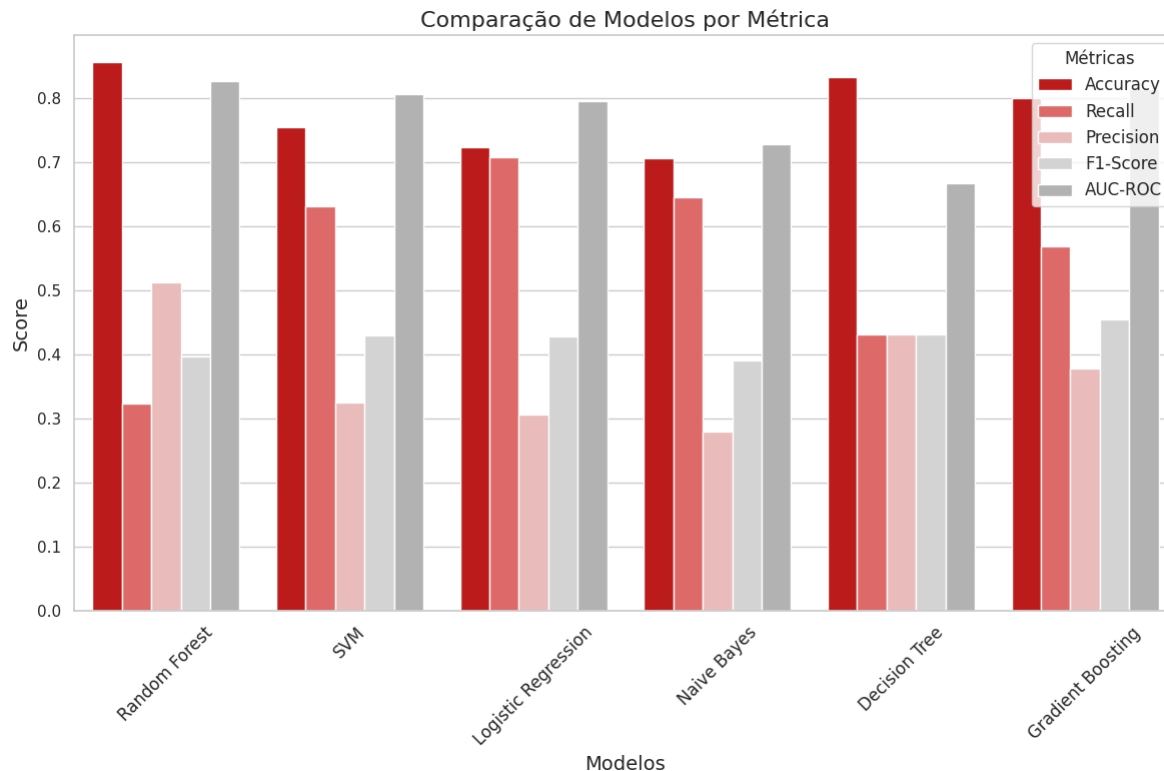
Otimização do Modelo Selecionado

- Otimizar os hiperparâmetros do modelo escolhido.

Modelos de Classificação | Random Forest

Random Forest

- Accuracy: 0.8161
- Recall: 0.6000
- Precision: 0.4105
- F1-Score: 0.4875
- AUC-ROC: 0.8213
- CV Accuracy: 0.9518



Taxa de Sucesso da Próxima Campanha | **Classificador**

Com base nas métricas fornecidas para o modelo de Random Forest (Accuracy: 0.8587, Recall: 0.60, Precision: 0.4105), podemos calcular as estimativas de receita para a próxima campanha de marketing se utilizar o modelo montado.

Estimativa de Clientes Selecionados

Para estimar quantos clientes o modelo identifica como potenciais compradores, utilizaremos a precisão do modelo.

- Total de clientes da campanha (baseado no número da campanha anterior): 2.240
- Clientes que realmente compram (segundo o recall de 60%): $60\% \text{ de } 2.240 = 551,3$ compradores (aproximadamente 1235 compradores).

Estimativa de Receita

- Com base na receita gerada por cada comprador (11 MU por cliente), podemos calcular a receita estimada:
- Receita estimada: $1235 \text{ compradores} \times 11 \text{ MU} = \mathbf{13.585 \text{ MU}}$.

A campanha anterior gerou 3.674 MU, enquanto a estimativa de receita para esta campanha, com base no modelo de Random Forest, é 13.585 MU, indicando um aumento significativo na receita projetada.