

# Otimização de Campanhas de Marketing e Segmentação de Clientes

# Introdução e Objetivo

- Objetivo Principal: Otimizar a eficácia das campanhas de marketing para maximizar o ROI e aprofundar o conhecimento sobre a base de clientes.
- Problema de Negócio: Como identificar clientes mais propensos a responder a uma campanha e segmentar a base para ações mais direcionadas?

# Dados e Metodologia

- Dados Utilizados: Dataset de clientes com histórico de compras, informações demográficas e comportamento em campanhas anteriores.
- Metodologia Aplicada:
  - Entendimento e Pré-processamento de Dados
  - Análise Exploratória de Dados (EDA)
  - Segmentação de Clientes (Clustering)
  - Modelagem Preditiva de Resposta à Campanha
  - Otimização Financeira da Campanha

# Abordagem Técnica

- Ferramentas Utilizadas: Pandas, Scikit-learn, Matplotlib, Seaborn
- Apoio com IA: Utilizei o Gemini como ferramenta de apoio e colaboração. → Isso me permitiu focar na compreensão dos problemas de negócio e na interpretação dos resultados, elevando a complexidade e a abrangência da análise em um tempo mais eficiente.

# Etapa 1: Pré-processamento (Parte 1)

- Objetivo: Transformar dados brutos em formato limpo e estruturado.
- Ações:
  - Carregamento do dataset: 2240 linhas, 29 colunas
  - Preenchimento de 24 valores nulos em 'Income' com a mediana
  - Padronização de variáveis categóricas

# Etapa 1: Pré-processamento (Parte 2)

- Engenharia de Features Criadas:
  - Age (idade a partir do ano de nascimento)
  - Years\_Customer (tempo de relacionamento)
  - Total\_Spending (gasto total)
  - Total\_Purchases (número total de compras)
  - Children (número total de filhos)
  - HasChildren (flag binária)

## Etapa 1: Pré-processamento (Parte 3)

- Tratamento de outliers: → 3 linhas com idade > 90 anos removidas
- Remoção de colunas constantes ou irrelevantes
- Resultado: Base robusta, consistente e enriquecida com insights de negócio

## Etapa 2: EDA (Objetivo e Ações Iniciais)

- Objetivo:

Compreender características dos clientes, padrões de comportamento e relações entre variáveis.

- Ações:

- Histogramas para Age, Income, Total\_Spending, Recency
- Gráficos de barras para Education, Marital\_Status, Children



## Etapa 2: EDA (Parte 2)

- Análise de correlação: → Mapa de calor entre variáveis numéricas
- Gasto médio por categorias: → Education, Marital\_Status, HasChildren

# EDA: Análise Focada na Resposta

- Taxa de aceitação de campanhas anteriores
- Comparativo entre clientes que responderam e não responderam:

Métrica	Nãorespondeu	Respondeu
Renda	R\$	R\$
Média	55.342,30	72.620,40
Gasto	R\$	R\$
Médio	560,41	1.202,94
Recência	49,33dias	38,87dias

# EDA: Insights

- Clientes que responderam: → Mais ricos, gastam mais, compraram mais recentemente
- Identificação de:
  - Principais drivers de compra
  - Variáveis com maior potencial preditivo

## Etapa 3: Segmentação de Clientes

- Objetivo: Agrupar clientes com características e comportamentos similares
- Features utilizadas: Income, Total\_Spending, Recency, Age, Children, Years\_Customer, HasChildren
- Normalização com StandardScaler

# K-Means e Resultado dos Clusters

- Método do cotovelo  $\rightarrow K=4$
- Clusters formados com características distintas

# Cluster 0: Família Grande, Baixo Gasto

- Renda Média: R\$ 40.733,91
- Gasto Médio: R\$ 247,93
- Filhos: 1.33
- Recência: 52,27 dias
- Resposta à Campanha: 3,88%
- Perfil: Todos têm filhos, clientes antigos

# Cluster 1: Alto Valor e Engajamento

- Renda Média: R\$ 78.028,98
- Gasto Médio: R\$ 1.387,65
- Filhos: 0.43
- Recência: 39,76 dias
- Resposta: 22,33%
- Perfil: Alto gasto, menos filhos, mais engajados

## Cluster 2: Recentes, Baixo Gasto

- Renda Média: R\$ 73.322,25
- Gasto Médio: R\$ 244,58
- Filhos: 0.49
- Recência: 25,43 dias (mais recentes)
- Resposta: 2,19%
- Perfil: Alta renda, baixo gasto, baixíssima resposta



## Cluster 3: Renda Média, Gasto Moderado

- Renda Média: R\$ 51.179,94
- Gasto Médio: R\$ 467,31
- Filhos: 0.90
- Recência: 51,68 dias
- Resposta: 7,07%
- Perfil: Intermediário em quase todas as métricas

# Benefícios da Segmentação

- Estratégias de marketing personalizadas
- Ofertas direcionadas
- Otimização da comunicação e relacionamento

## Etapa 4: Modelagem Preditiva

- Objetivo: Prever probabilidade de resposta à campanha
- Dados preparados com One-Hot Encoding → `X_encoded` shape: (2237, 38)
- Treino/Teste: 80%/20%, com estratificação

# Modelo: Regressão Logística

- `class_weight = "balanced"`
- Geração de probabilidades para cada cliente

# Avaliação do Modelo

- Precision (classe 1): 0.44
- Recall (classe 1): 0.76
- F1-Score (classe 1): 0.56
- AUC da Curva ROC: 0.82 (Apresentar gráfico de ROC)

# Importância das Features

- Impacto Positivo:

- AcceptedCmp3 (~2.00)
- AcceptedCmp5 (~1.65)
- AcceptedCmp1 (~1.33)
- Years\_Customer (~1.22)
- AcceptedCmp4 (~1.19)

- Impacto Negativo:

- Marital\_Status\_Together (~-0.74)
- HasChildren (~-0.69)
- Teenhome (~-0.68)
- Marital\_Status\_Married (~-0.63)
- Education\_Basic (~-0.59)

## Etapa 5: Otimização Financeira

- Custo por contato: R\$ 3,00
- Receita por aceitação: R\$ 10,93
- Simulação para diferentes limiares

# Resultado da Otimização

- Lucro máximo: R\$ 233,86
- Limiar ótimo: 0.58
- Clientes contatados: 97
- Esperados respondentes: 48
- Taxa de sucesso esperada: 49,48%



# Comparação com Campanha Piloto

- Campanha Piloto: R\$ -3.046,00
- Novo resultado: R\$ 233,86
- Ganho total: R\$ 3.279,86

# Conclusões

- Conhecemos melhor os perfis de clientes
- Modelo preditivo com AUC = 0.82
- Transformação de prejuízo em lucro

# Próximos Passos

- Usar o modelo para selecionar os 97 clientes mais promissores
- Criar campanhas personalizadas por cluster
- Monitorar resultados e ajustar continuamente