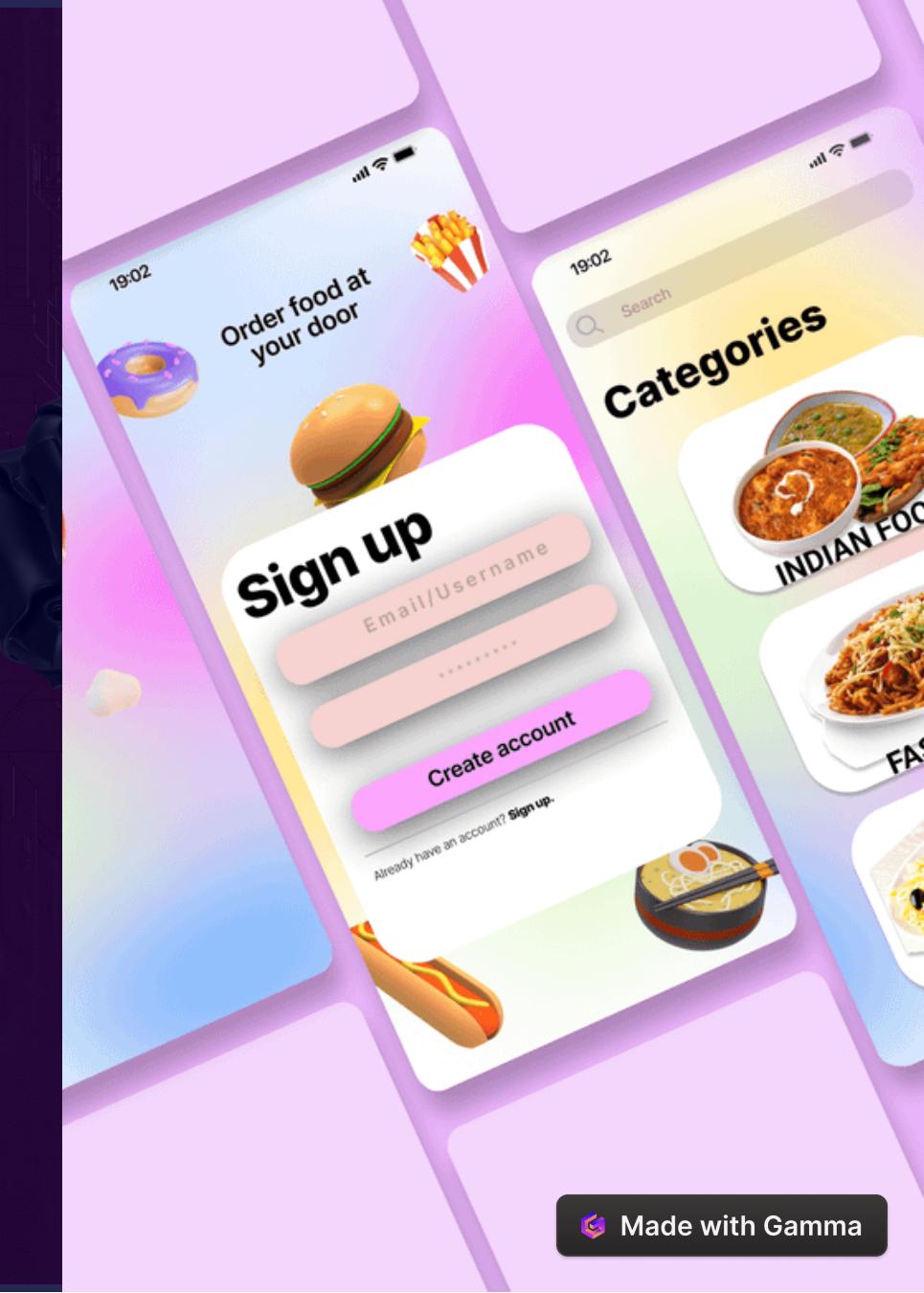


Churn em app de Delivery

Nessa apresentação, iremos analisar os dados de um aplicativo de delivery utilizando técnicas e ferramentas para análises de dados para responder questões importantes relacionadas ao Churn de Clientes e propor soluções eficazes.



by vinicius guirra



Regra de negócio

A área de CRM passou uma amostra de cerca de 10 mil clientes com suas respectivas informações de cadastro e transações nos próximos 4 meses a contar da data de extração usada como referência. Desta forma um cliente é classificado como Churn quando não transacionou nos últimos 30 dias. Nesta amostra vamos considerar que o último dia de transação foi 30/10/2019.

Entendimento do negócio

Segmentação de clientes

Segmentar os clientes com a probabilidade de darem Churn nos próximos 4 meses em relação à data de extração de referência.

1

Fatores de risco

Quais fatores de risco estão associados com o Churn de Clientes?

2

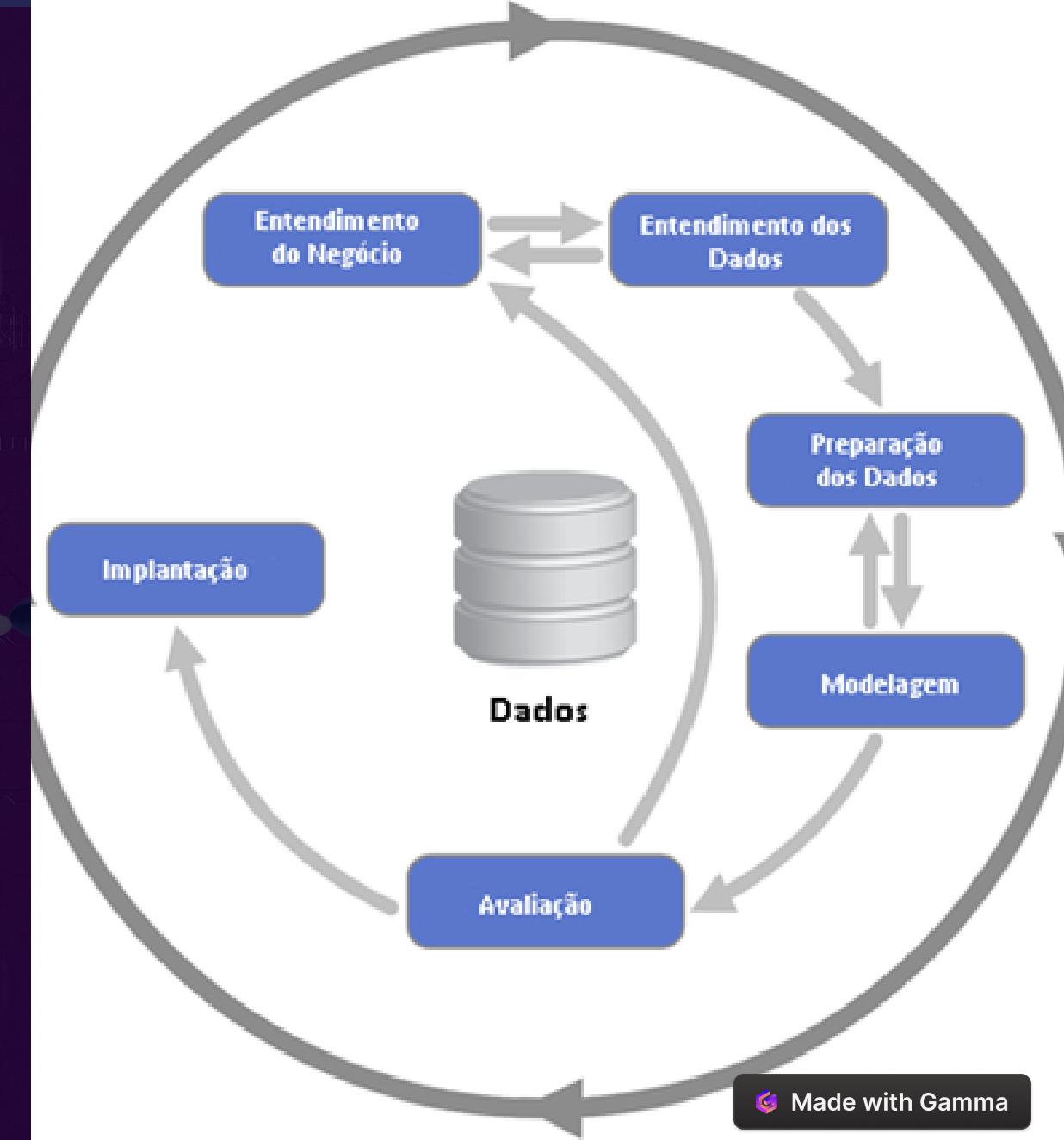
Plano de Ação

Quais os possíveis planos de ação que a empresa pode fazer para diminuir esse problema?

3



Metodologia Crispm



Workflow



Origem dos dados

- Extração dos dados do banco de dados da <https://www.preditiva.ai/>
- **Engenharia de variáveis:** Criação das colunas churn e dias sem transacionar seguindo a regra de negócio.



```
<alunos> medidas resumo <alunos> Script-8 <alunos> Script-10 <alunos> Script-11 <alunos> Script-12 ×
-- Adicionar coluna churn (1 para churn, 0 para não churn) e dias_sem_transacionar
SELECT
    fato_churn.*,
    dim_clientes.*,
    CASE
        WHEN fato_churn."DataUltimaTransacao" <= '2019-10-30' AND fato_churn."DataUltimaTransacao" >= '2019-09-30' THEN 1
        ELSE 0
    END AS churn,
    CASE
        WHEN fato_churn."DataUltimaTransacao" <= '2019-10-30' THEN
            EXTRACT(DAY FROM '2019-10-30'::timestamp - fato_churn."DataUltimaTransacao")
        ELSE
            NULL -- Ou outra lógica dependendo do que você preferir para lidar com datas inválidas
    END AS dias_sem_transacionar
FROM
    db_churn.fato_churn
INNER JOIN
    db_churn.dim_clientes ON fato_churn."ClientId" = dim_clientes."ClientId";
```

Análise exploratória

via colab: <https://colab.research.google.com/drive/1Kw09wHUX0s7UBql75cwJb19dptQsBPZ?usp=sharing>

1

Pré-análise

- Realizou-se uma análise inicial dos dados para entender a estrutura, tipos de variáveis, possíveis padrões e sua qualidade.

2

Transformação

- Verificação e tratamento de valores únicos e inconsistentes
- Verificação e tratamento de duplicadas
- Verificação e tratamento de valores nulos
- Transformação de tipos adequados dos dados

3

Carregamento

- Envio dos dados para pasta destino após tratamento.

Análise

Information Value

O "information value" é uma medida estatística que ajuda a identificar a importância das variáveis em um modelo preditivo. Ele é calculado com base na informação que cada variável fornece para prever uma determinada resposta. Quanto maior o valor do "information value", mais influente a variável é para a previsão.

Regressão Logística

A Regressão Logística é um modelo estatístico utilizado para prever a probabilidade de ocorrência de um evento binário, como sim/não, sucesso/fracasso, entre outros. Ele utiliza uma função logística para mapear a variável independente em uma probabilidade, e assim realizar a classificação. É amplamente utilizado em problemas de classificação, como previsão de churn de clientes, detecção de fraudes, entre outros.

Information Value

Variável	IV	Grau de Influência no Churn
Qte_Categorias	0,93	Muito forte
Idade	0,76	Muito forte
Estados	0,17	Médio
Programa_Fidelidade	0,15	Médio
Limite de crédito	0,10	Médio
Gênero	0,07	Fraca
Score de crédito	0,01	Fraca
Tempo Cliente	0,00	Fraca
Usa_Cartão_credito	0,00	Fraca
Soma_Pedidos_Acumulados	0,00	Fraca

Elaborado com microsoft Excel.

IV - Score de Crédito

Score de crédito	Frequência de churn	IV
350-449	32%	0,01
450-549	21%	0,00
550-649	21%	0,00
650-749	19%	0,00
750-850	19%	0,00
Total Geral	20%	-
IV TOTAL		0,01

A faixa de 350-449 apresenta a maior frequência de churn (32%) com 12% de diferença em relação a total, mas contribui pouco para a capacidade do modelo ($IV = 0,01$) o que sugere que, apesar da alta taxa de churn, a influência dessa variável específica no modelo é limitada e por ser uma variável adquirida de um birô de crédito, vale discutir sua compra futura.

IV - Estado

Estados	Frequência de churn	IV
Minas Gerais	32%	0,12
Rio de Janeiro	16%	0,01
São Paulo	16%	0,04
Total Geral	20%	-
IV TOTAL		0,17

Minas Gerais apresenta a maior taxa de churn, atingindo 32%, indicando uma possível área de atenção para estratégias de retenção. O Valor do IV para Minas Gerais é notavelmente alto, representando 0,12 do IV total de 0,17.

IV - Idade

Idade	Frequência de churn	IV
18-27	7%	0,10
28-37	9%	0,24
38-47	24%	0,02
48-57	54%	0,35
58-67	45%	0,06
>68	17%	0,00
Total Geral	20%	-
IV TOTAL		0,76

Clientes com idades entre 48-57 anos e 58-67 anos demonstram as maiores taxas de churn, 54% e 45%, respectivamente. Além disso, o Valor do IV Total é significativamente alto, atingindo 0,76, indicando que a idade exerce uma forte influência na predição de churn.

IV - Qte de Categorias

Qte_Categorias	Frequência de churn	IV
1	27%	0,09
2	7%	0,41
3	82%	0,29
4	97%	0,14
Total Geral	20%	-
IV TOTAL		0,93

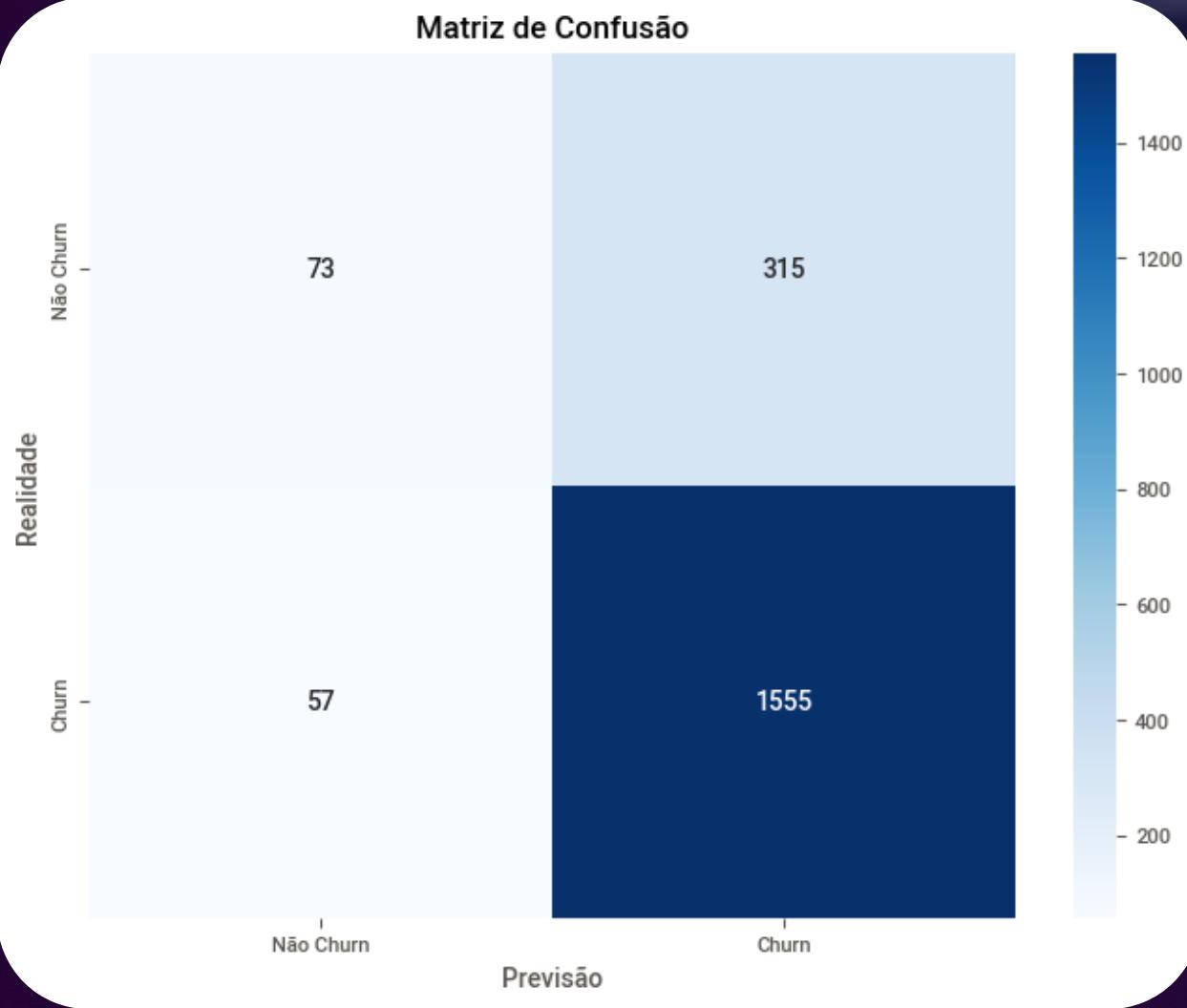
Clientes que pedem em 4 categorias apresentam uma taxa de churn expressivamente alta, atingindo 97%. A variável contribui significativamente para o IV Total, alcançando 0,93. Porém a variável é inconclusiva pois vale entender o que cada categoria significa.

IV - Programa de Fidelidade

Programa_Fidelidade	Frequência de churn	IV
0	27%	0,07
1	14%	0,08
Total Geral	20%	
	IV TOTAL	0,15
	IV TOTAL	0,93

A adesão ao programa de fidelidade exerce uma influência significativa na redução da probabilidade de churn, destacando a eficácia desse programa como estratégia de retenção.

Regressão Logística



Relatório de Classificação:

	precision	recall	f1-score	support
0	0.56	0.19	0.28	388
1	0.83	0.96	0.89	1612
accuracy			0.81	2000
macro avg	0.70	0.58	0.59	2000
weighted avg	0.78	0.81	0.77	2000

Acurácia do Modelo: 0.82

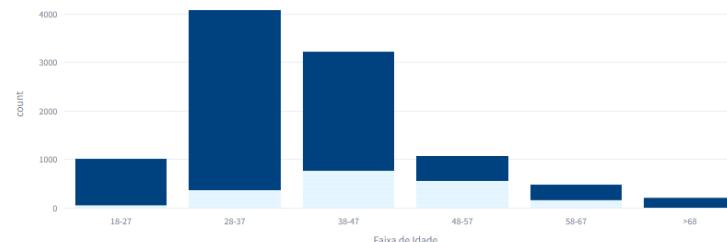
- Em termos simples, o modelo acertou aproximadamente 82% das previsões.
- A matriz de confusão resume o desempenho do modelo. Por exemplo, ele previu corretamente 82 casos de "Não Churn" e 1555 casos de "Churn". No entanto, também fez algumas previsões incorretas.
- **Precisão (precision):** Aproximadamente 59% dos casos previstos como "Não Churn" eram realmente "Não Churn". Cerca de 84% dos casos previstos como "Churn" eram realmente "Churn".
- **Revocação (recall):** Cerca de 21% dos casos reais de "Não Churn" foram corretamente previstos. Para "Churn", aproximadamente 96% dos casos reais foram corretamente previstos.
- **F1-score:** Uma média ponderada entre precisão e recall. Para "Não Churn", é cerca de 31%. Para "Churn", é bastante alto, aproximadamente 90%.
- **Acurácia (accuracy):** O modelo acertou aproximadamente 82% das previsões em geral.
- **Macro avg e weighted avg:** São médias ponderadas de precisão, recall e f1-score para todas as classes.

Visualização com Streamilt

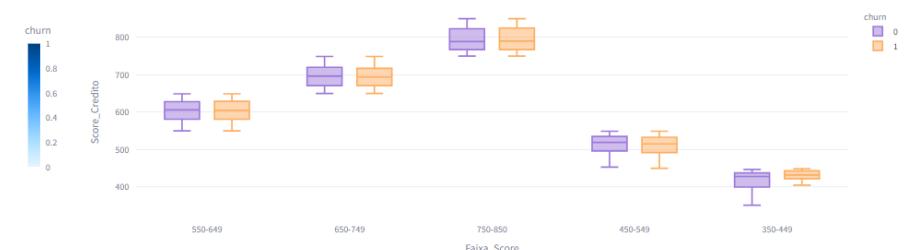
Análise de Churn

por Vinícius Guirra

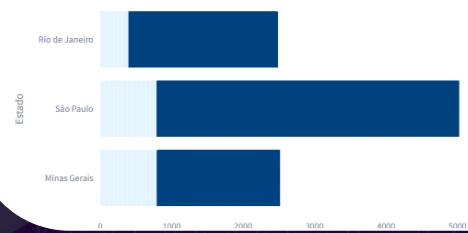
Churn x Idade



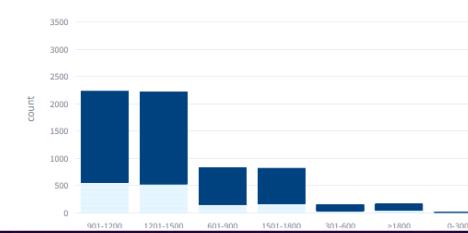
Churn x Score de Crédito



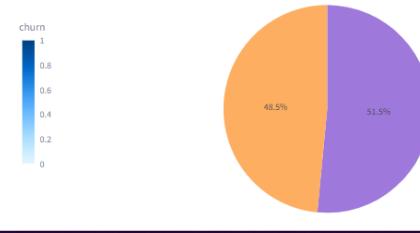
Churn x Barras de Estado



Churn x Limite_Credito_Mercado



Churn x Programa Fidelidade



Made with Gamma

Possíveis planos de ação

Variável	O que fazer ?	Porquê?	Quem?	Como fazer?	Qual custo?	Quando?	Onde?
Score de crédito	Desenvolver promoções para Score 350-449	Influência significativa na predição de churn.	Equipe de Retenção de Clientes	Ofertas personalizadas, descontos.	Orçamento de marketing para promoções especiais.	Próximos 3 meses	Em todo o mercado
Estados	Campanhas específicas para clientes em Minas Gerais	Minas Gerais tem a maior taxa de churn e alto IV.	Equipe de Marketing	Promoções locais, coleta de feedbacks específicos.	Orçamento de marketing para campanhas locais.	Próximos 3 meses	Em todo o mercado
Idade	Estratégias focadas em clientes de 48-57 e 58-67 anos	Essas faixas etárias têm as maiores taxas de churn e alto IV.	Equipe de Retenção de Clientes	Benefícios personalizados, programas de fidelidade direcionados.	Desenvolvimento e promoção de benefícios exclusivos.	Próximos 3 meses	Em todo o mercado
Qte_Categorias	Incentivos para clientes que pedem nas 4 categorias	97% de churn nessa categoria é alto IV.	Equipe de Marketing, Equipe de Produto	Programas de pontos, descontos progressivos	Desenvolvimento de sistema de pontos e descontos.	Próximos 3 meses	Em todo o mercado
Programa_Fidelidade	Reforçar o programa de fidelidade	Clientes participantes têm menor churn e contribuem para o IV.	Equipe de Retenção de Clientes	Recompensas, promoções exclusivas para participantes.	Marketing para promover o programa.	Próximos 3 meses	Em todo o mercado

Conclusão

Reducir o churn de clientes é uma tarefa desafiadora, mas com uma análise cuidadosa dos dados, segmentação eficiente e estratégias de retenção personalizadas, é possível alcançar resultados significativos. Acompanhar e atualizar constantemente as estratégias é essencial para manter o sucesso a longo prazo.

