

ANÁLISE DE CHURN

A G O S T O | 2 0 2 5



O que vamos explorar neste relatório?

O que é análise de churn?

Documentação & Metodologia

Análise Exploratória

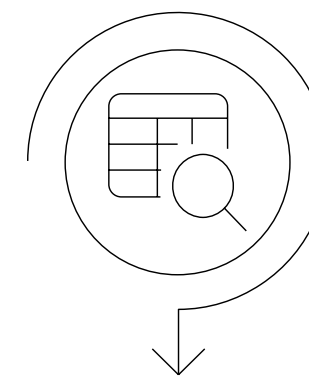
Modelagem Preditiva

Propostas & Próximos Passos

O que é Análise se churn?

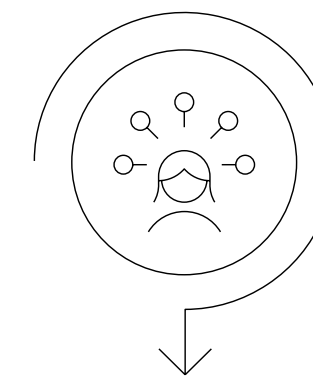
Análise de churn é o processo de identificação e entendimento do por que clientes deixam de comprar ou cancelam um serviço, seja em negócios de assinatura, e-commerce, brancos, telecom ou qualquer empresa que dependa de retenção para crescer.

- **Churn significa “perda de clientes” em um período** e é medido com a taxa ($\text{Clientes perdidos no período} \div \text{Clientes no início do período}$).
- **Reter clientes é mais barato** que conquistar novos (5 a 7 vezes menos custo)
- **Clientes fiéis gastam mais** e geram marketing boca a boca.
- **Reduzir churn estimula o aumento de receita recorrente** e valor do cliente (CLV).



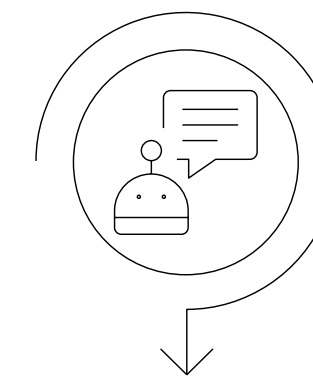
Coleta de Dados

Coletando informações dos clientes de várias fontes.



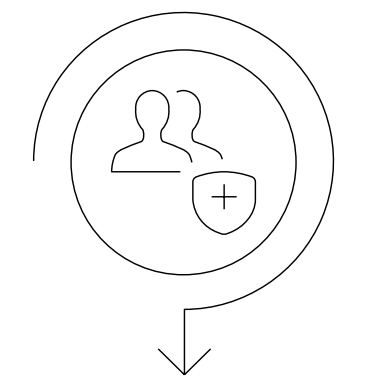
Exploração

Analizando dados para identificar padrões e insights.



Modelagem Preditiva

Aplicando técnicas para prever a probabilidade de churn.



Ações de Retenção

Implementando estratégias para prevenir o churn de clientes.

Documentação & Metodologia

- **Objetivo da Analise**

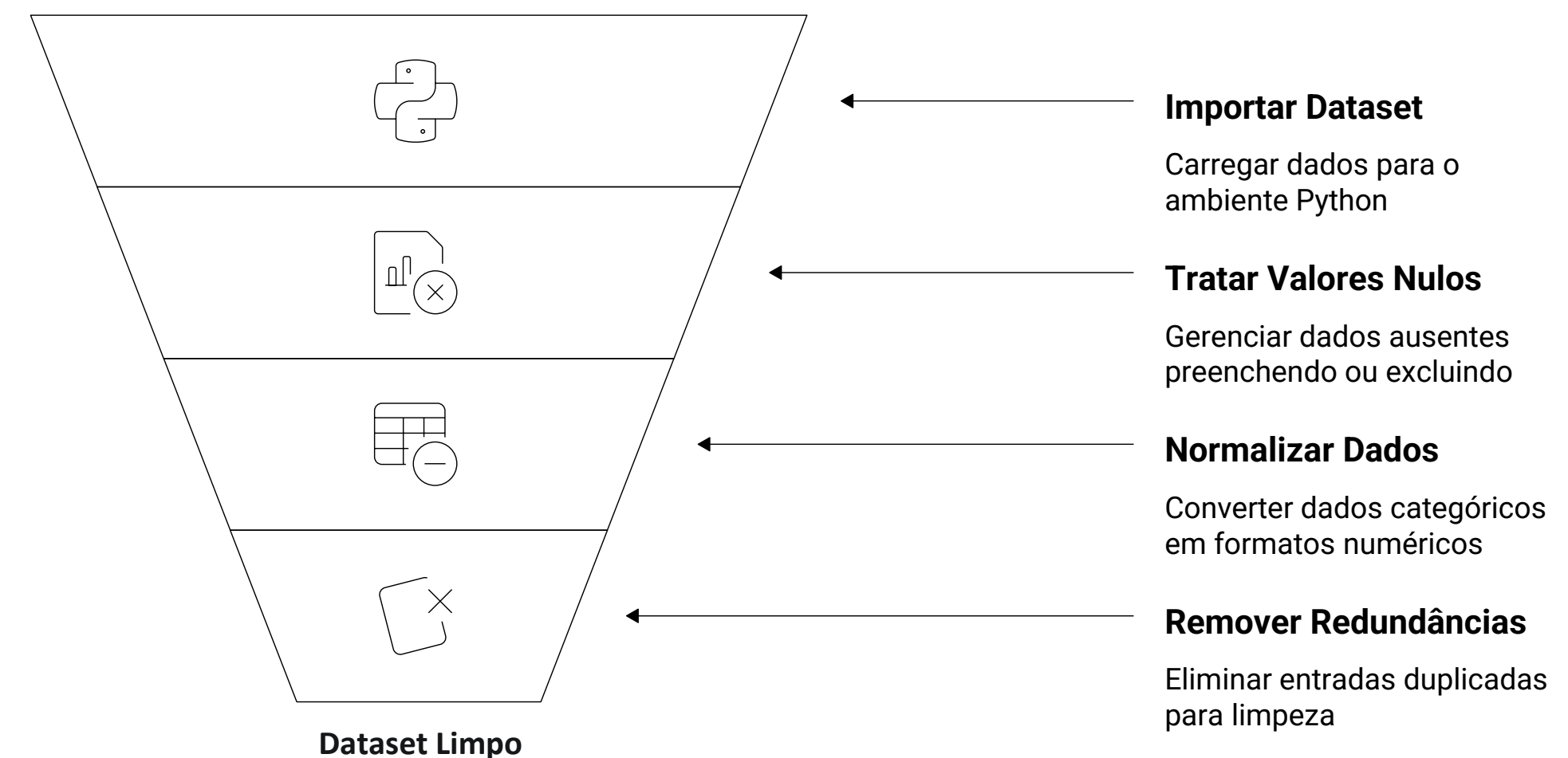
Compreender os principais fatores que influenciam o churn e desenvolver um modelo preditivo interpretável capaz de estimar a probabilidade de cancelamento a partir de variáveis históricas e comportamentais.

A base de dados utilizada é proveniente de uma empresa de telecomunicações iraniana, disponibilizada publicamente na comunidade de Machine Learning Keggles para fins de experimentação e estudo, sendo amplamente utilizada em grupos de análise de dados e machine learning para testes e demonstrações de metodologias analíticas.

- **Fonte:**

telecom_churn.csv

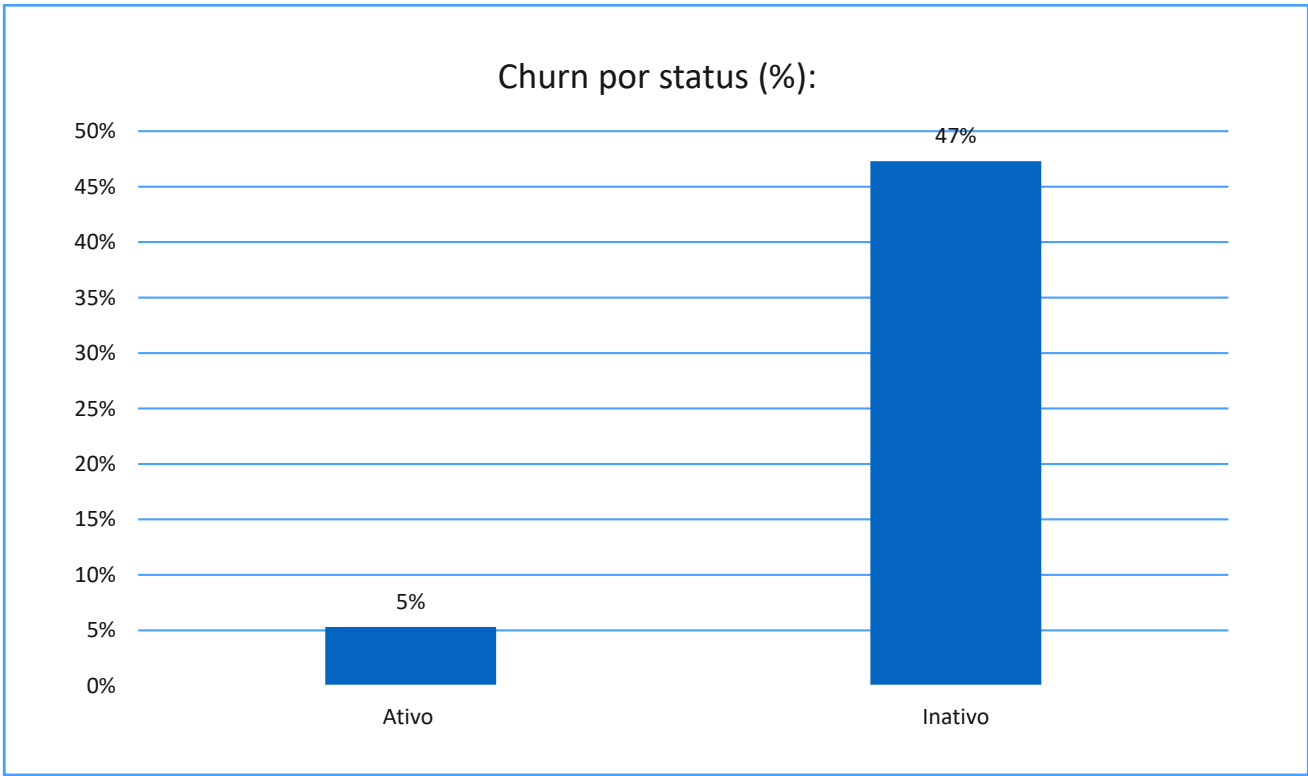
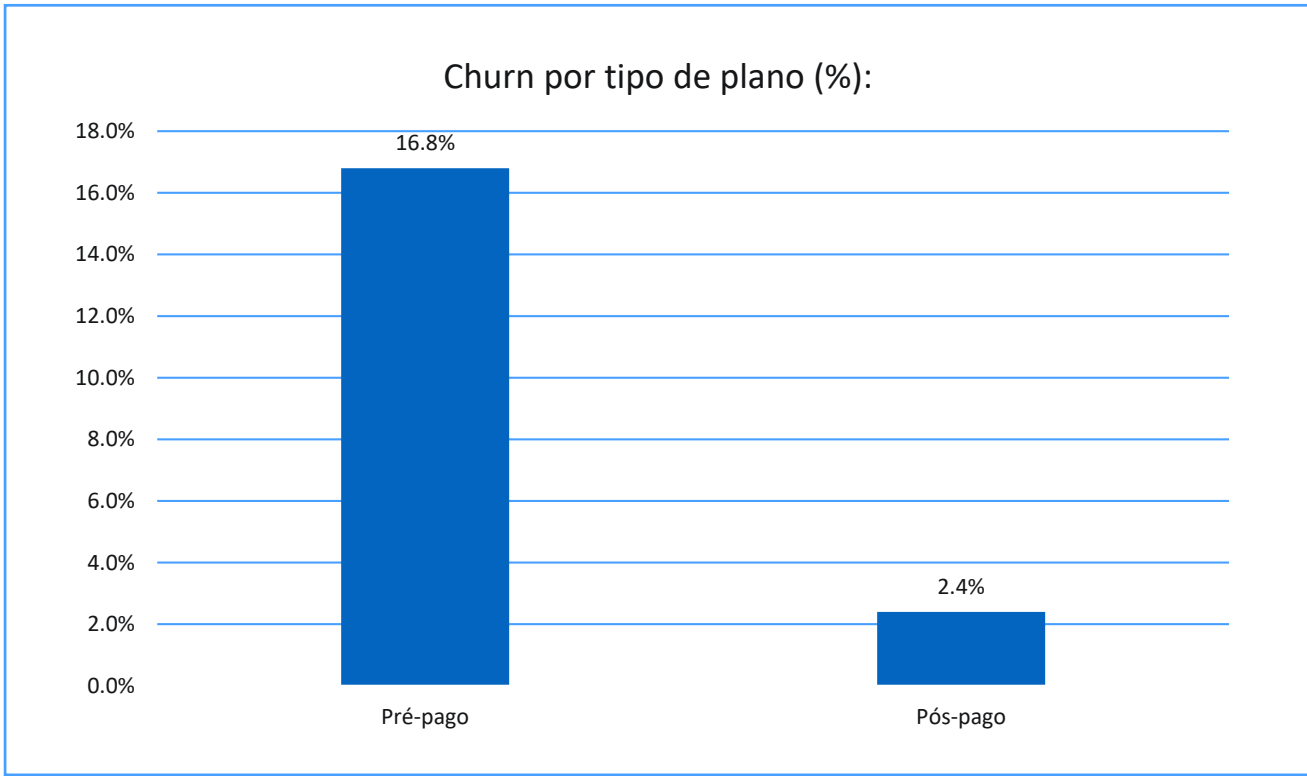
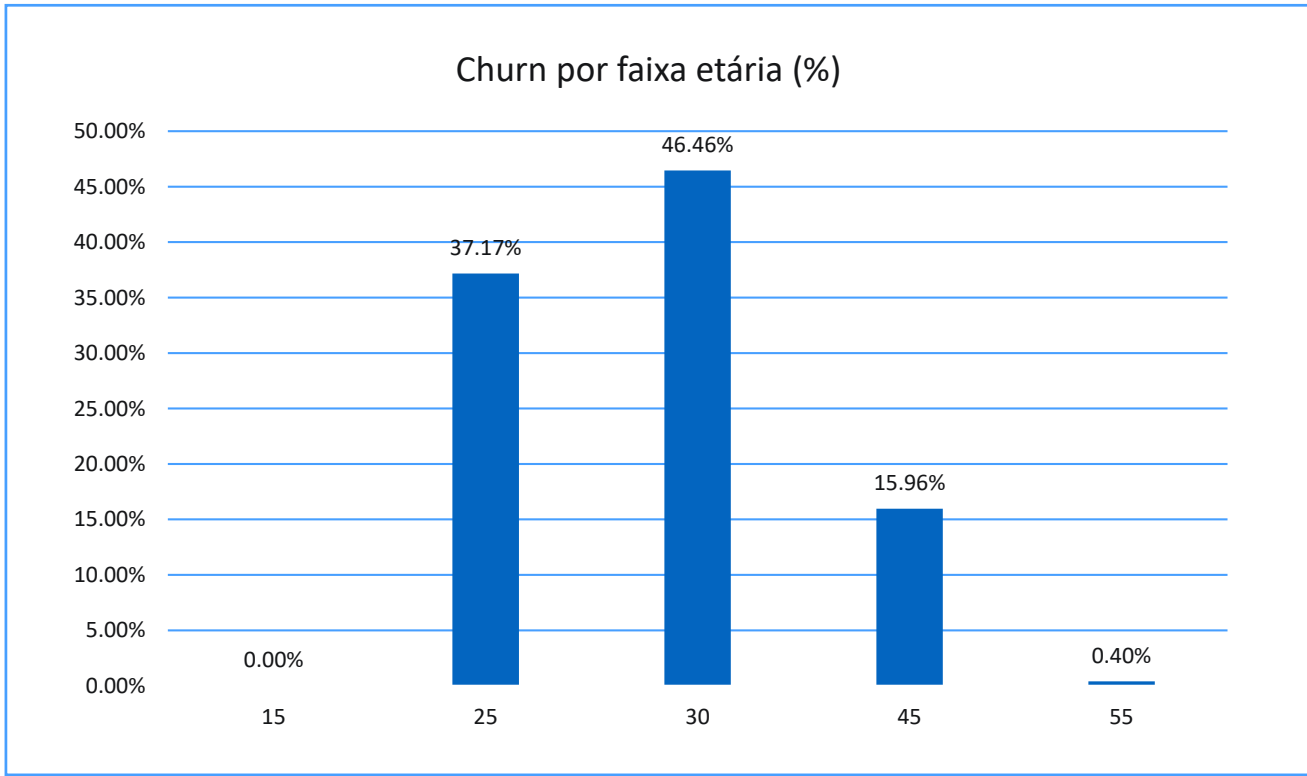
Metodologia de limpeza e transformação (ETL)



A N Á L I S E D E C H U R N

Análise Exploratória

Panorama Geral

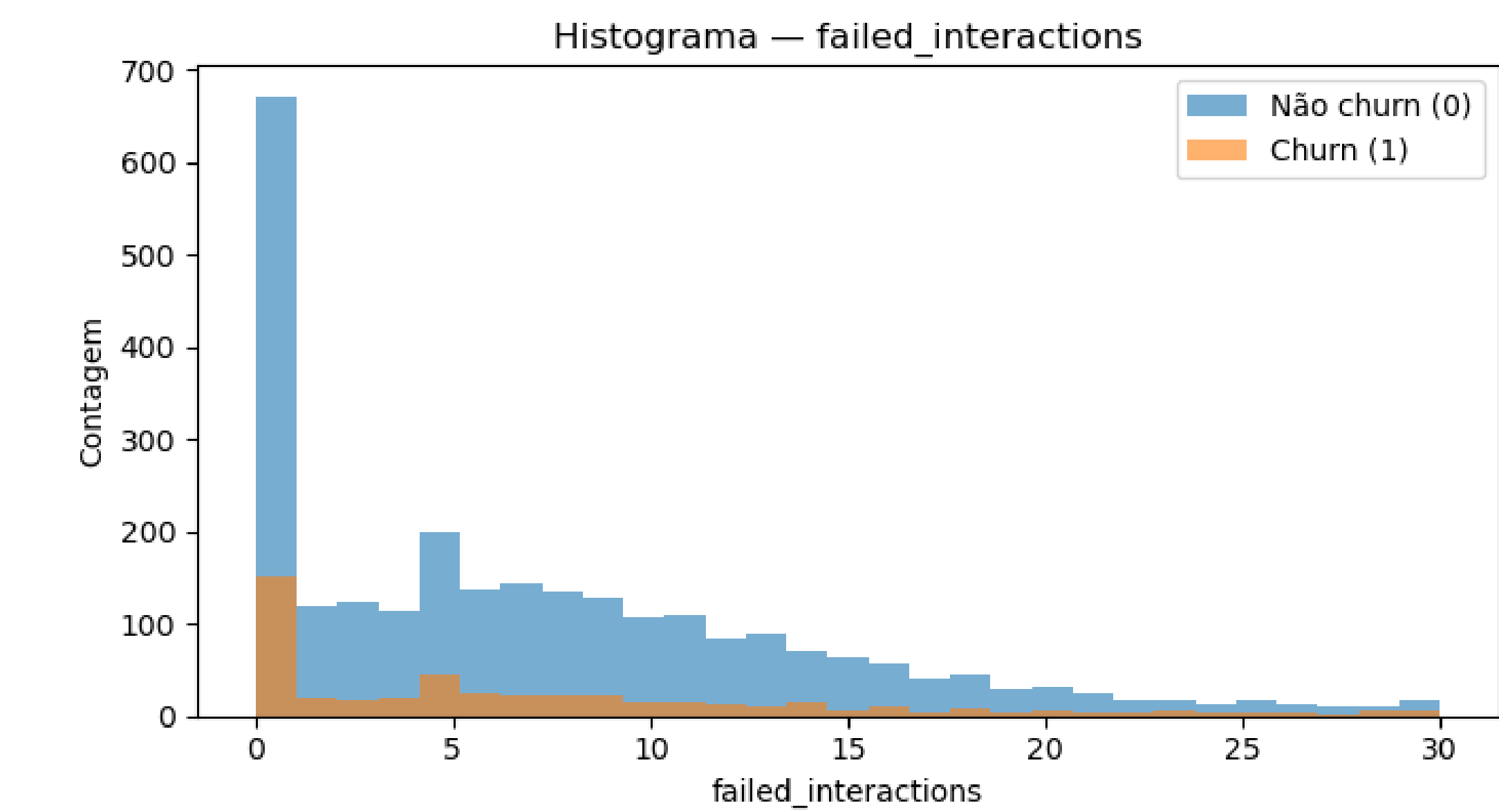


Perfil Inicial

Conseguimos observar um panorama inicial onde a maior incidência de churn se concentra nas faixas entre 21 e 40 anos, com pico aos 30, ocorrendo principalmente nos planos pré-pagos, nos perfis que se mostraram inativos

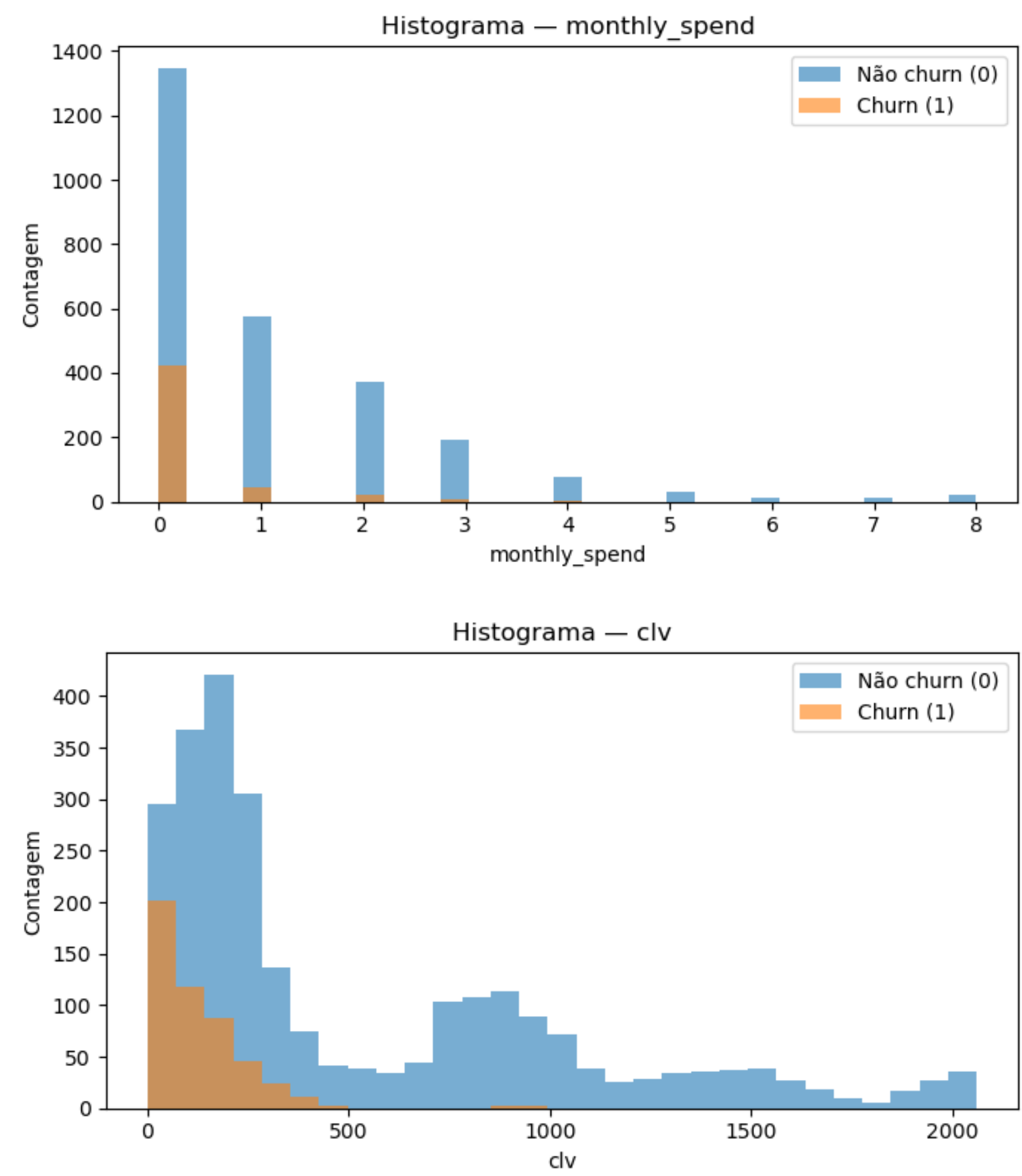
Churn

Silencioso



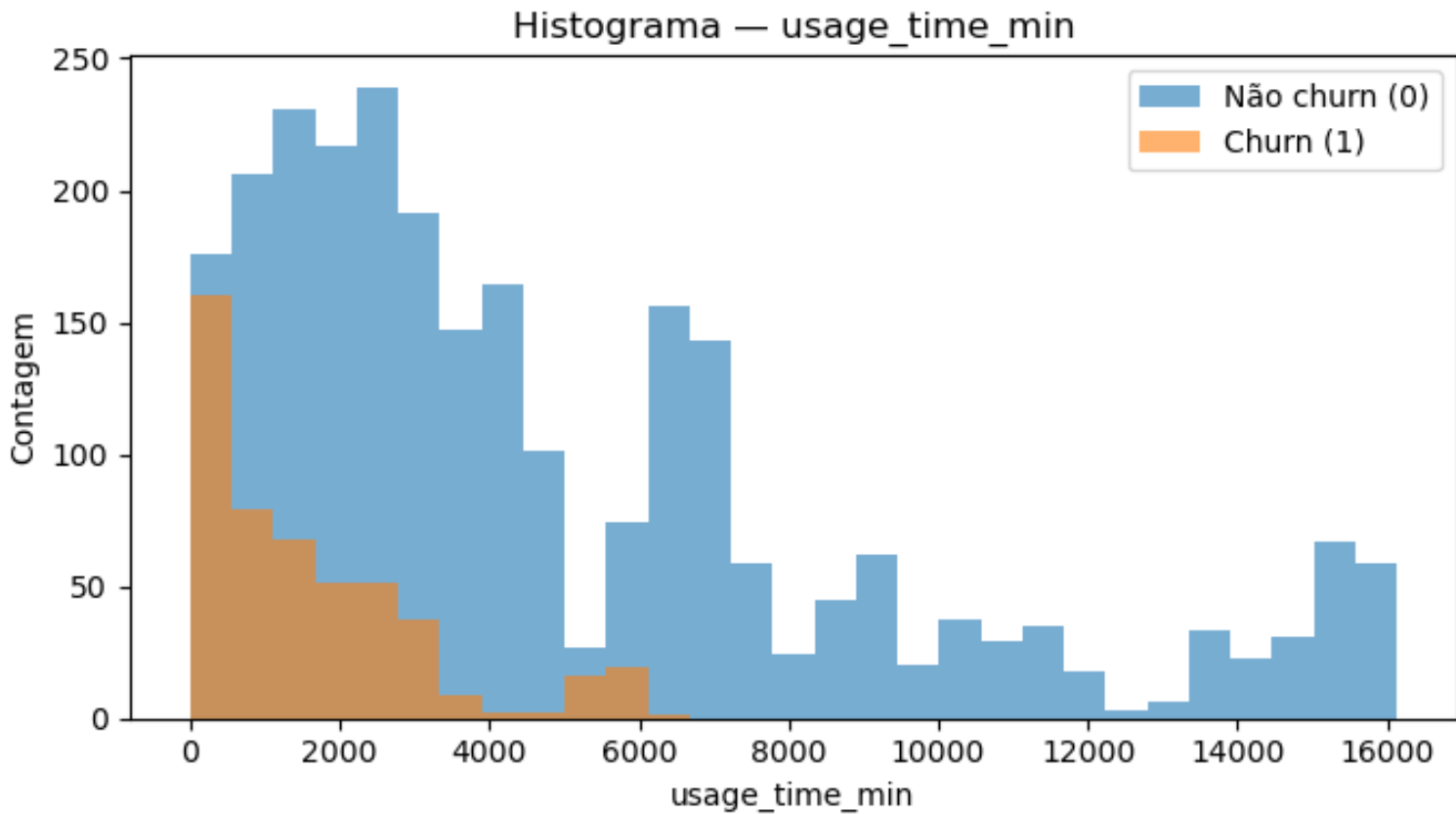
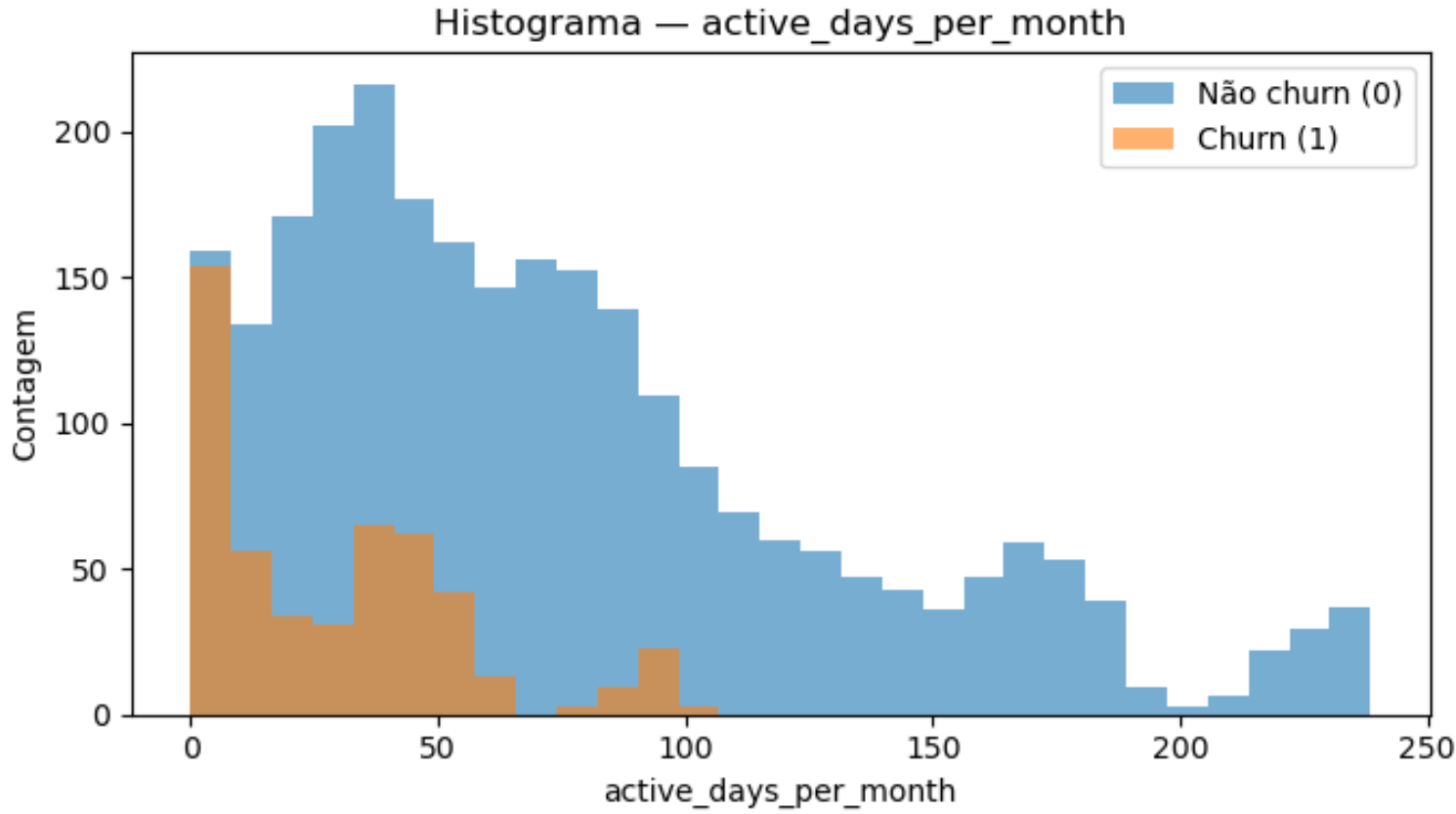
A cima podemos observar que há uma grande massa de clientes concentrada no ponto onde ainda não apresentaram interações falhas com a empresa, porém, neste mesmo ponto se apresenta um pico de churn, o que é contra intuitivo, uma vez que podemos pensar que o churn tende a aumentar com a frustração de contatos falhos.

Cientes de Baixo valor



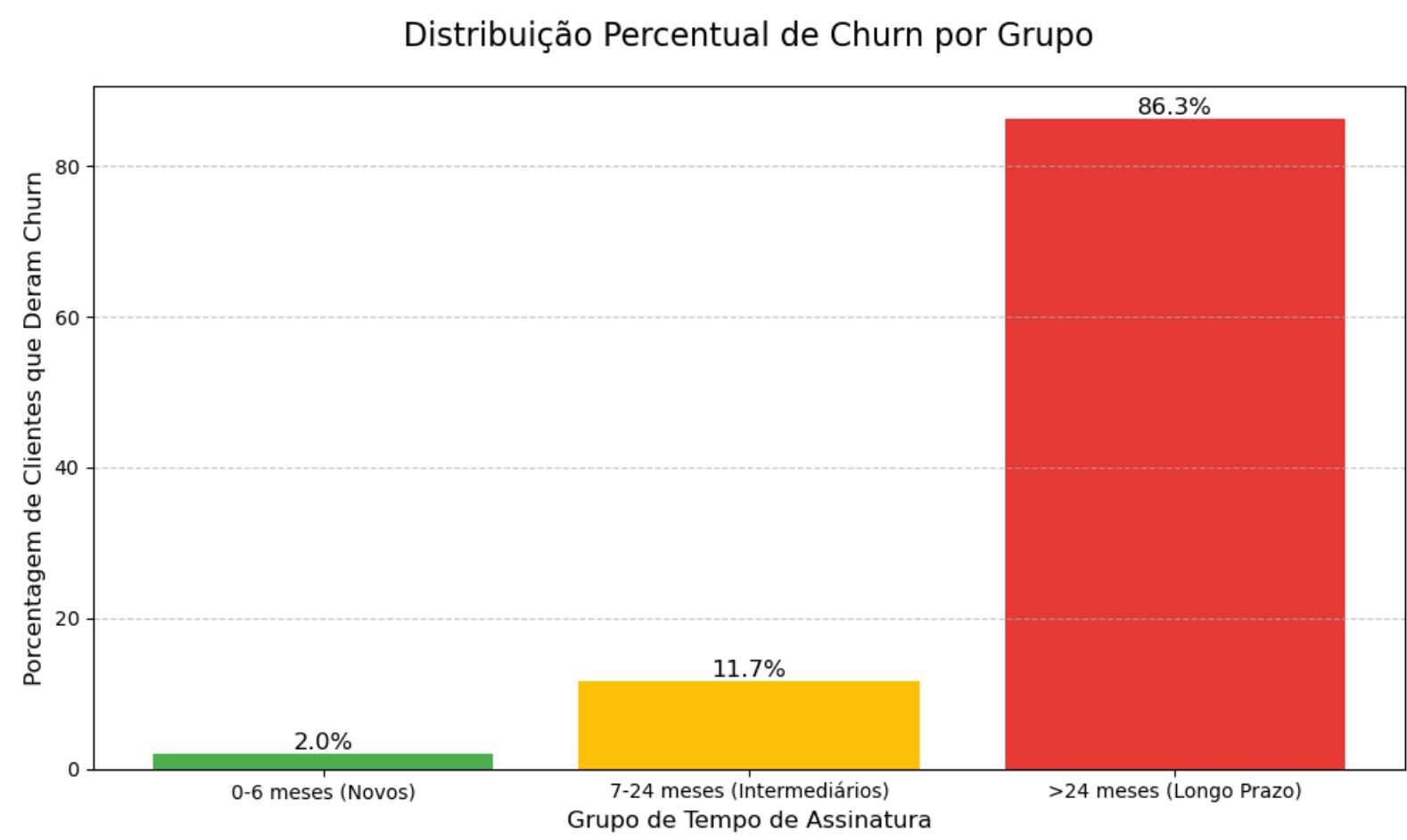
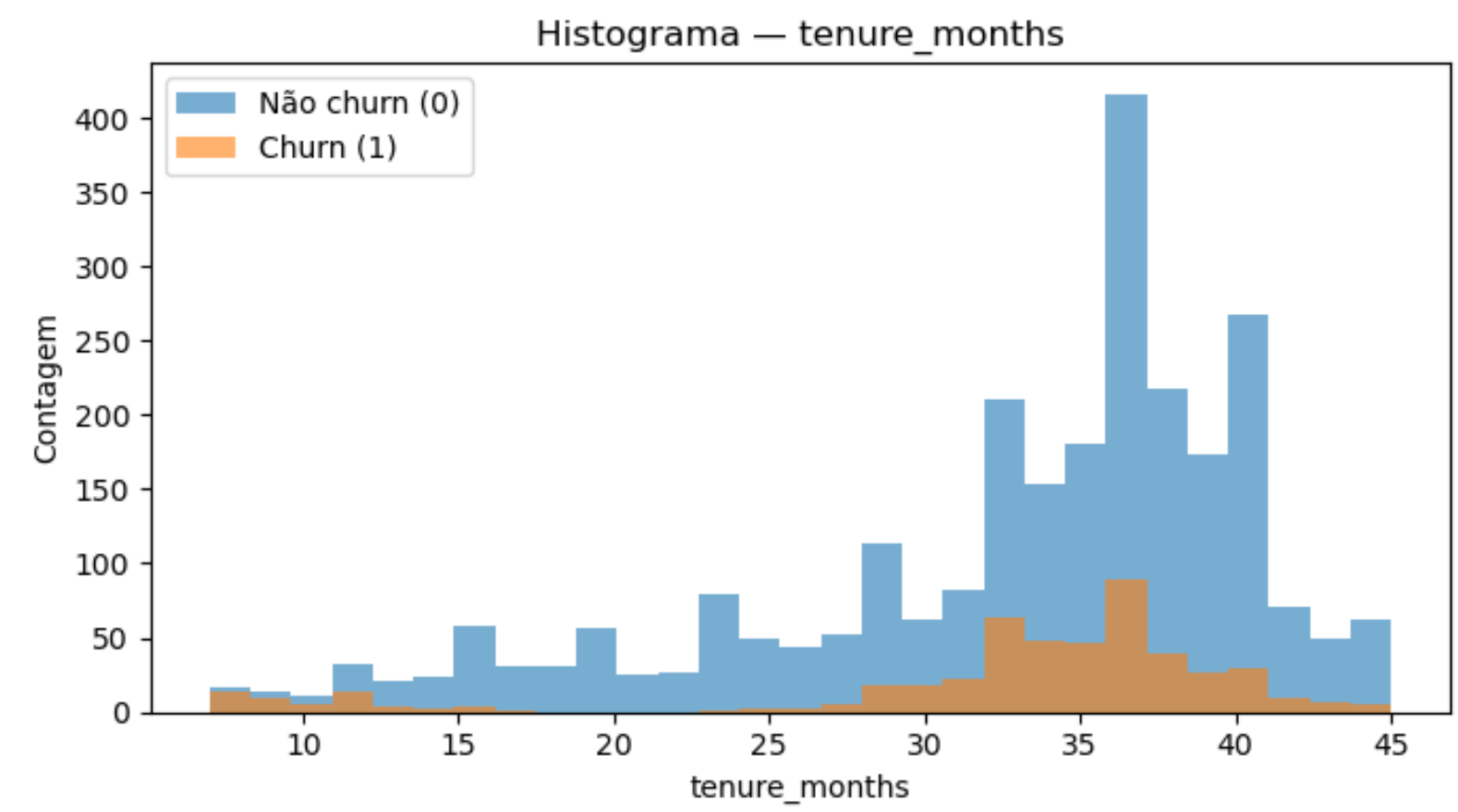
Quanto as tendências de churn relacionadas a gastos mensais e valor do cliente, podemos observar que há uma tendência de ocorrer com clientes de valor baixo, que não chegaram a gerar receita ou se tornaram valiosos ao longo do tempo

Baixo Uso



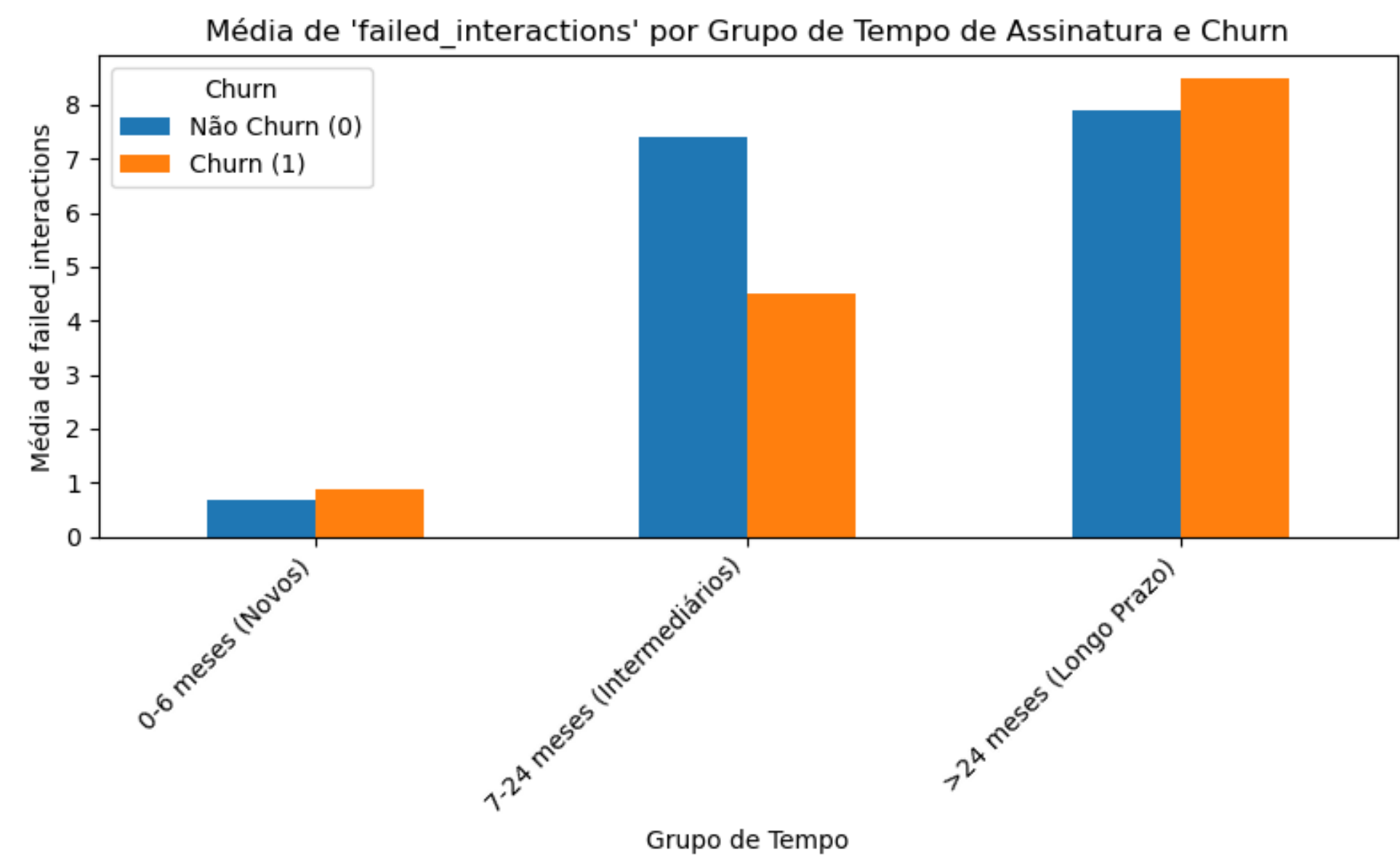
Outro sinal de churn aparece em clientes com menor tempo de uso e menor quantidade de dias ativos por mês, apontando novamente para um churn concentrado em um perfil de cliente de baixo engajamento.

Crise de Fidelidade



Este talvez seja o sinal mais drástico, pois, embora os indicadores anteriores pudessem sugerir um abandono de clientes de baixo valor e baixo engajamento, dando a ideia de que seriam clientes com pouco tempo de uso da plataforma, porém, o maior volume se concentra em clientes que já possuem histórico de 3 anos ou mais, sugerindo uma possível erosão fundamental do valor do produto, algo que deve ser investigada com atenção e urgência!.

Crise de Fidelidade



A média de interações falhas com a plataforma confirmam a crise de fidelidade, **indicando que problemas de interação na plataforma têm um impacto maior na retenção dos clientes mais fiéis**. A empresa precisa prestar atenção quando um cliente de longo prazo começa a ter problemas técnicos.

Com a análise exploratória conseguimos identificar 2 perfis distintos de clientes churners (os de baixo valor e engajamento, e os clientes fiéis) dando origem a 2 problemas distintos:

1- Churn no Início da Jornada

2- Crise de fidelidade

Problema 1

Churn no início

Hipótese 1

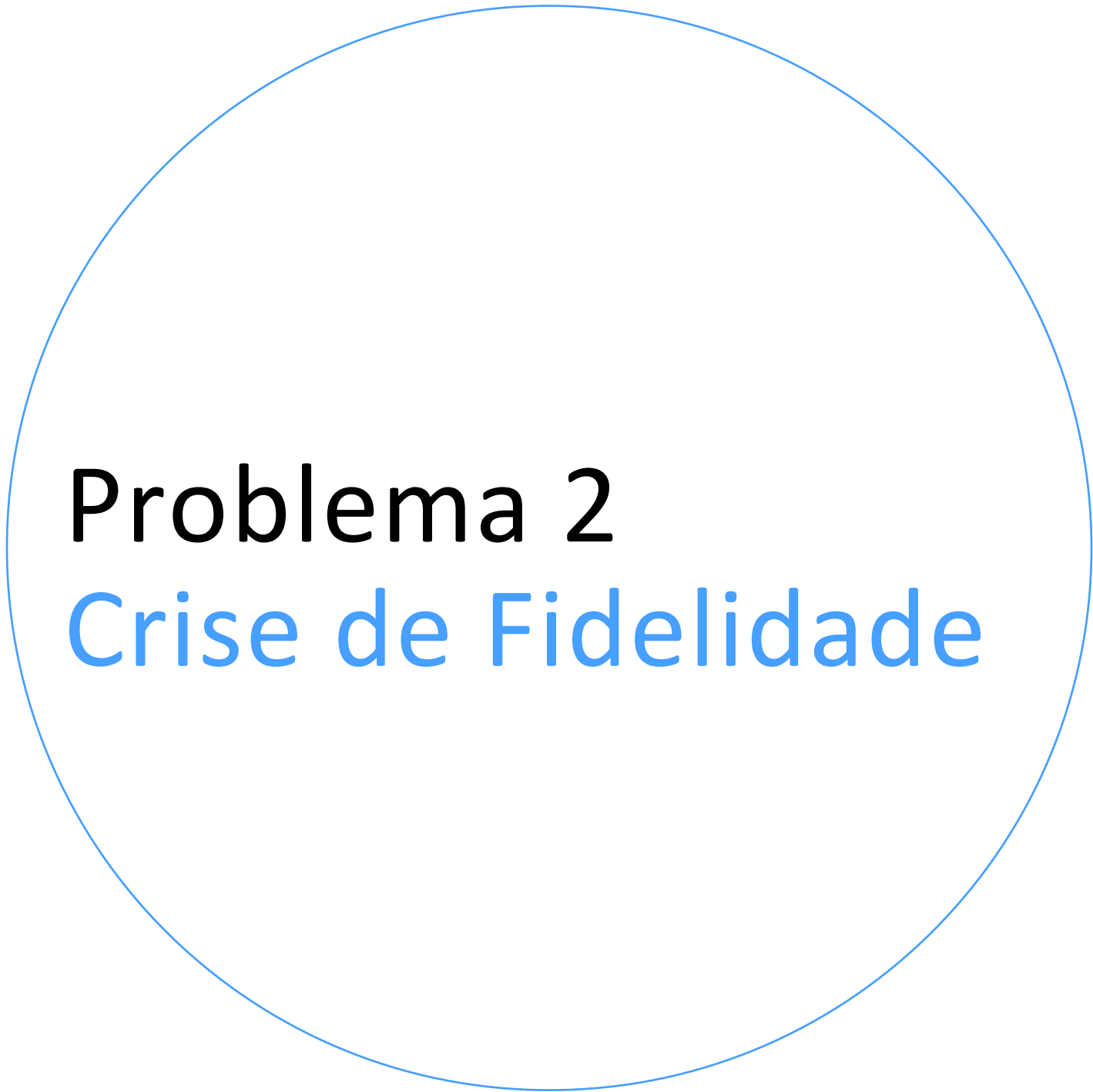
Problemas de Onboarding e Primeira Experiência: O cliente se cadastra, se frustra com a experiência inicial e decide cancelar antes de sequer ter um problema que o leve a abrir um ticket ou interagir com o suporte. Exemplos: O produto é muito difícil de usar, a interface é confusa, ou o cliente não consegue encontrar o valor prometido logo de cara.

Hipótese 2

Expectativas Não Alinhadas: A mensagem de marketing pode estar vendendo algo que o produto real não entrega. O cliente chega, percebe que não é bem o que ele esperava, e vai embora. A (interação falha) nunca acontece porque a falha não é técnica, mas sim de alinhamento de proposta de valor.

Hipótese 3

Fatores de Mercado: O cliente pode ter encontrado uma oferta melhor de um concorrente, ou seu contexto pessoal mudou (ex: não precisa mais do serviço). Esses fatores não têm relação com a sua experiência direta com o produto e, portanto, não gerariam uma "interação falha".



Problema 2

Crise de Fidelidade

Hipótese 1

Problema de Valor e Inovação: O churn acentuado em clientes de longa data podem sugerir uma perda na percepção de entrega de valor, funcionalidade e eficiência do produto.

Hipótese 2

Problemas de Preço e Custo: O churn pode ter sido estimulado com o aumento do preço de assinatura, no qual o cliente não percebeu um aumento correspondente no valor entregue, ou então, o cliente pode ter encontrado um concorrente que oferecesse uma solução similar com melhor relação custo-benefício.

Hipótese 3

Problemas de Relacionamento: A empresa pode ter parado de se comunicar ou engajar ativamente com o cliente, ou o cliente enfrentou problemas recorrentes e experiências de suporte ruim

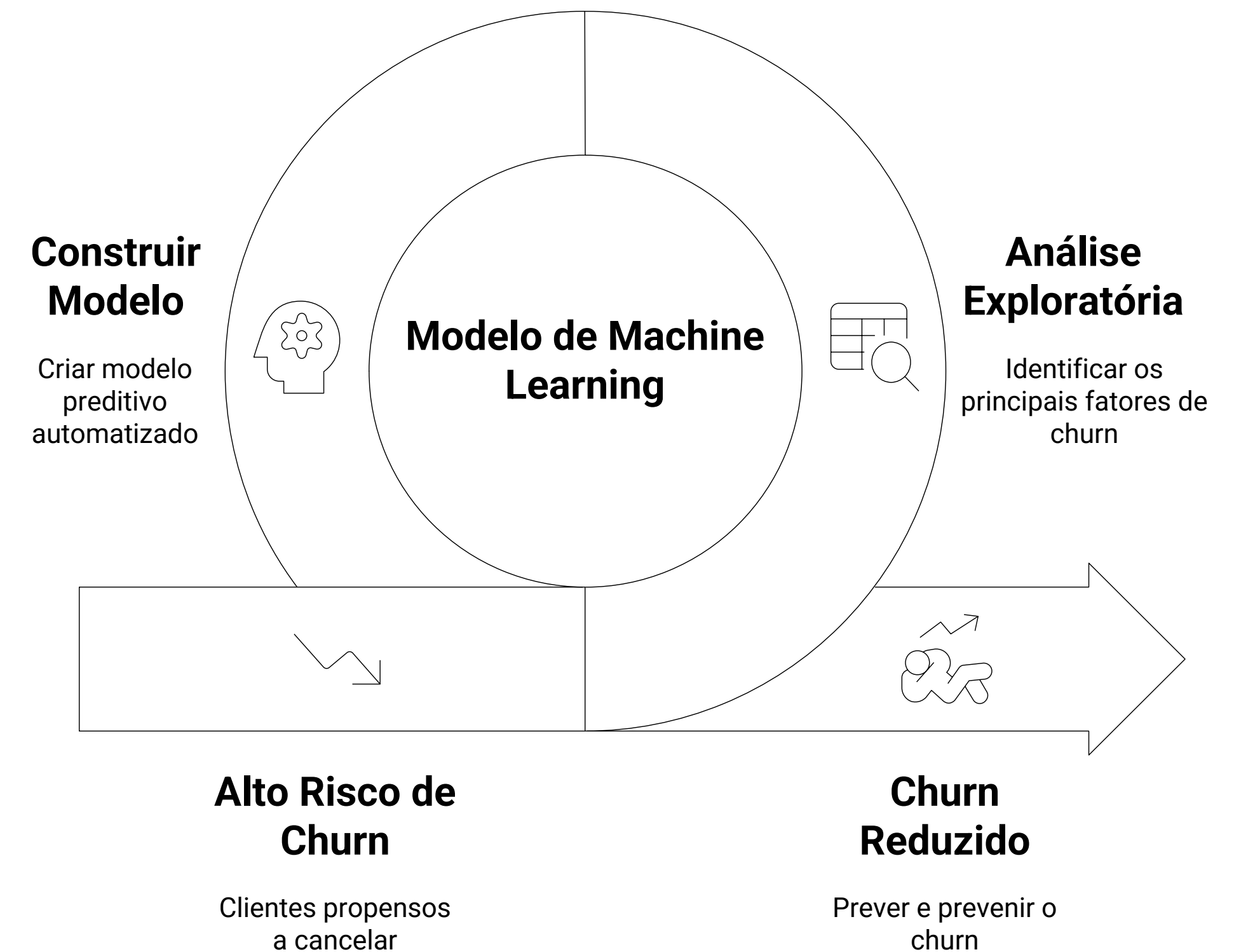
Hipótese 3

Fatores Externos e Concorrência: Possibilidade de forte entrada de um concorrente com uma proposta de valor superior, ou então uma mudança no contexto pessoal e profissional do cliente que tornou o serviço desnecessário para ele.

A N Á L I S E D E C H U R N

Modelagem Preditiva

Com esses insights da análise exploratória, o próximo passo foi construir um modelo de Machine Learning capaz de prever, de forma automatizada, quais clientes têm maior risco de cancelamento.



Modelo de Regressão Logística

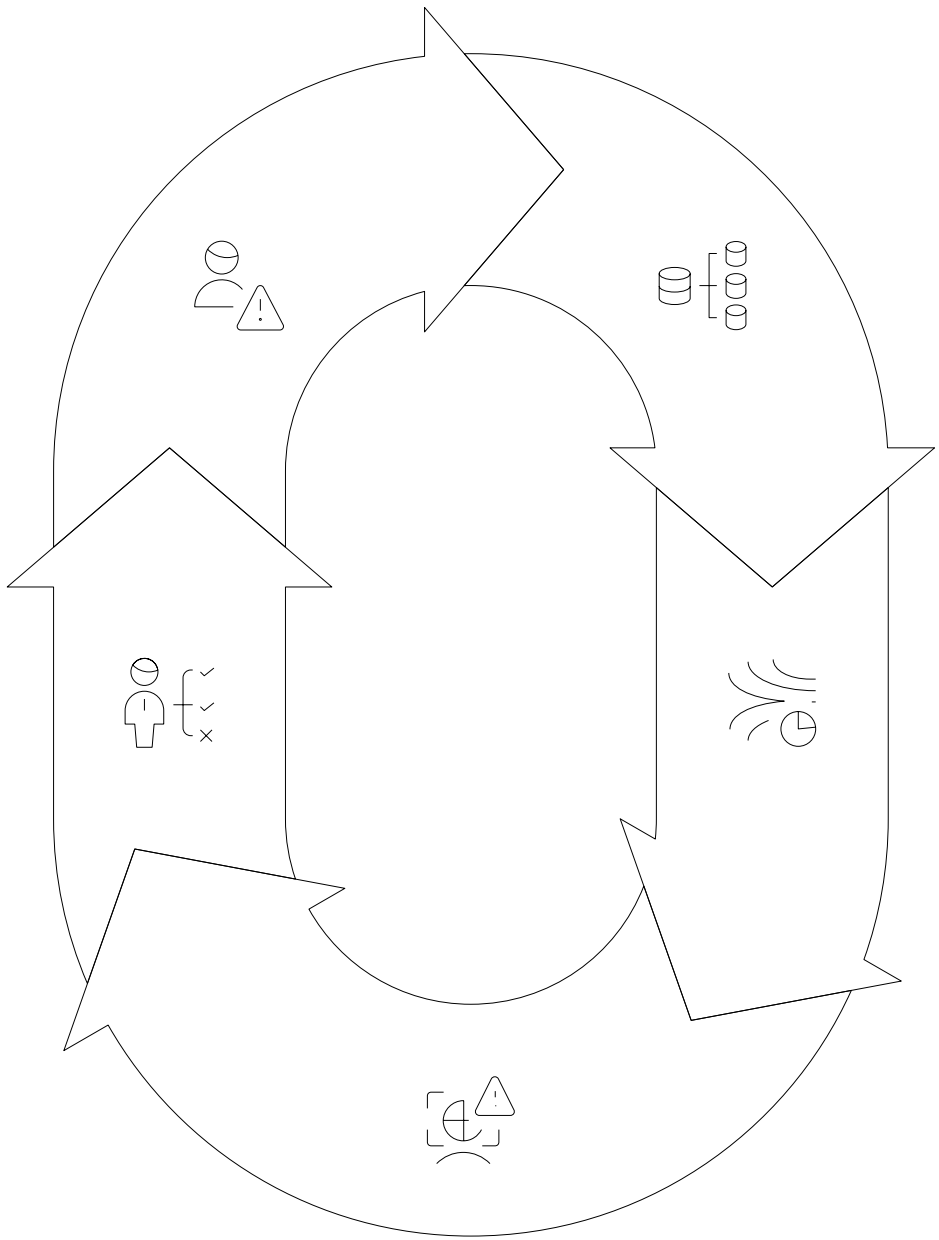
Utilizamos um modelo de **Regressão Logística**, que é um método estatístico robusto para prever um resultado binário, como "sim" ou "não". No nosso caso, o modelo é treinado para responder à pergunta: **"Este cliente tem uma alta probabilidade de churn?"**

O modelo analisa as características de um cliente (como tempo de uso e reclamações de suporte) e calcula um **score de risco** que varia de 0% a 100%. Quanto maior a porcentagem, maior a probabilidade de o cliente cancelar.

Para garantir que o modelo fosse o mais eficaz possível na identificação de clientes em risco (a classe Churn), realizamos uma otimização com um limite de decisão (threshold) recalibrado.

Identificar Clientes em Risco
Marcar clientes com alta probabilidade de churn

Otimizar Limite de Decisão
Ajustar para melhor precisão



Coletar Dados do Cliente
Reunir informações relevantes do cliente

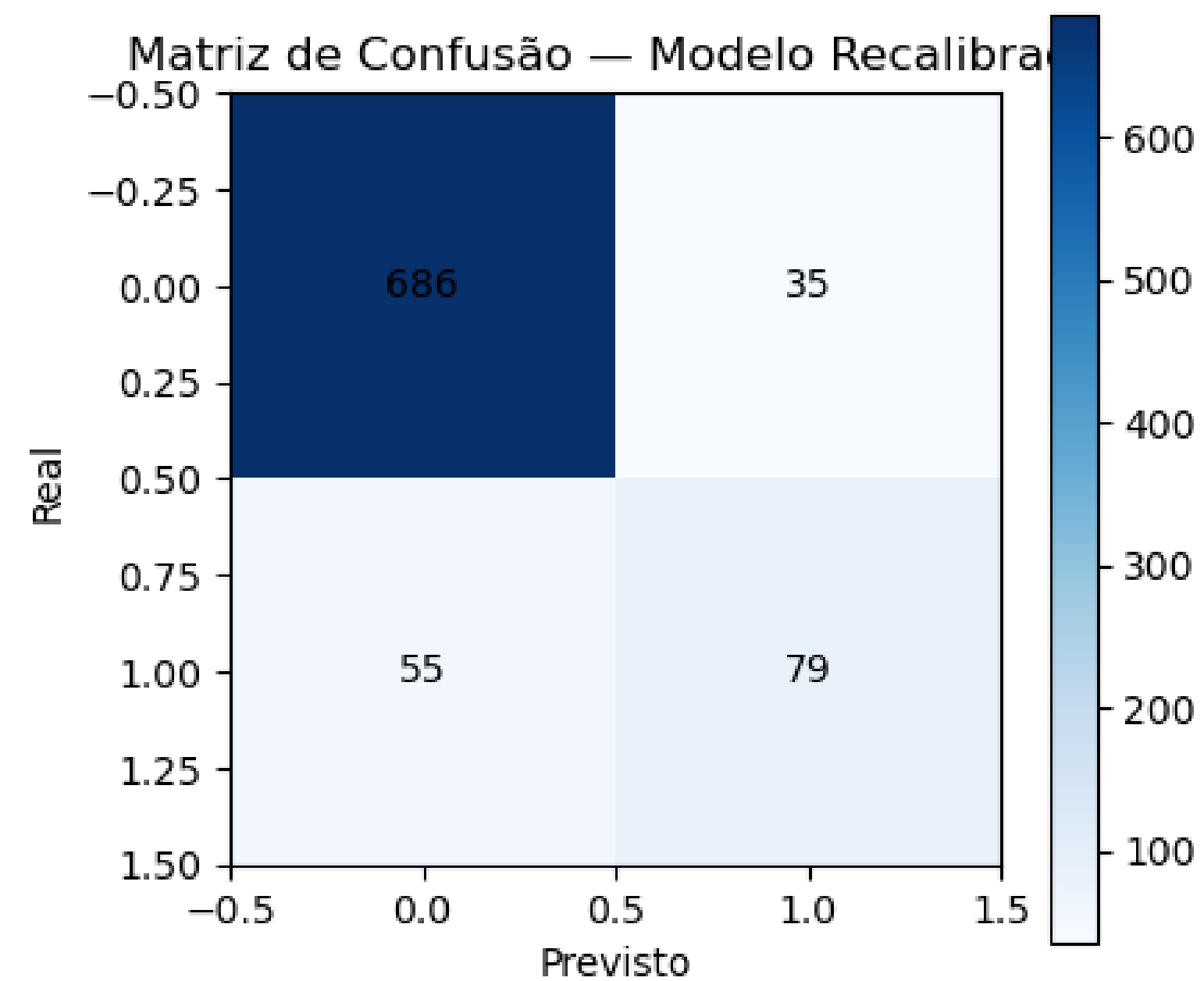
Analisar Características do Cliente
Avaliar dados para insights

Calcular Score de Risco
Determinar a probabilidade de churn

Matriz De Confusão

A matriz de confusão é um resumo do desempenho do modelo de classificação.

Ela compara o que o modelo **previu** com o que realmente aconteceu (**Real**).



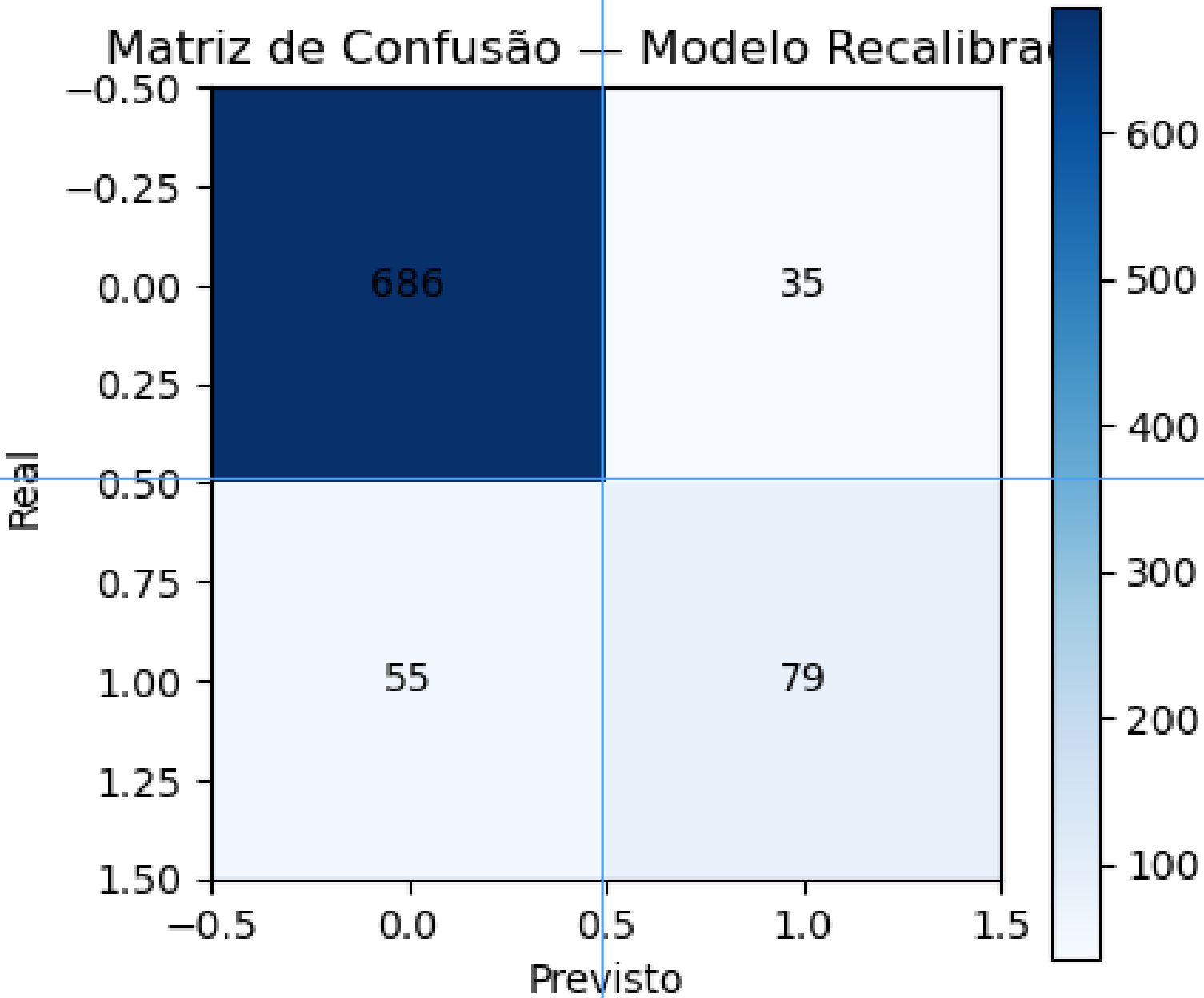
VERDADEIROS NEGATIVOS

O modelo **previu** que o cliente não daria churn e o cliente **realmente não deu churn**.

FALSOS POSITIVOS

O modelo **previu** que o cliente daria churn, mas o cliente **na verdade não deu churn**.

- Estes são os "alarmes falsos". Cada falso positivo representa tempo da equipe de Sucesso do Cliente gasto com um cliente que não estava em risco.



FALSOS NEGATIVOS

O modelo **previu** que o cliente não daria churn, mas o cliente **na verdade deu churn**.

Cada falso negativo é um cliente que o **modelo não conseguiu identificar como de risco e que acabou cancelando**. O principal objetivo é minimizar este número para não perder clientes de alto valor.

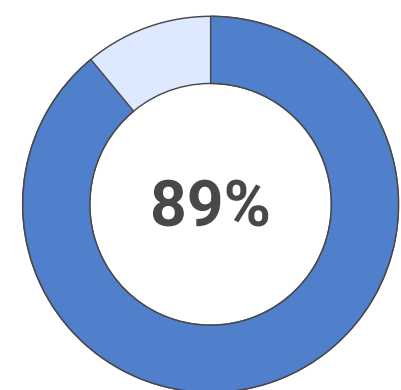
VERDADEIROS POSITIVOS

O modelo **previu** que o cliente daria churn e o cliente **realmente deu churn**.

Cada verdadeiro positivo é um cliente que o modelo identificou corretamente como de risco, dando à sua equipe a chance de agir e tentar retê-lo.

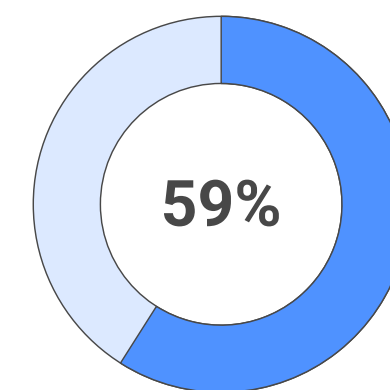
Performance

Do Modelo



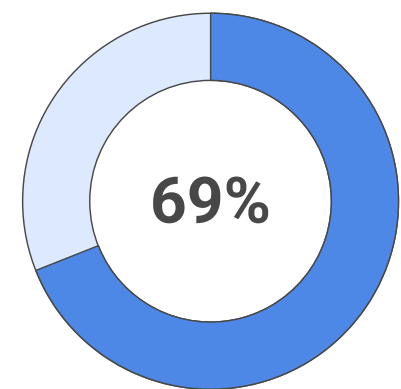
Acurácia

A acurácia mostra a performance geral do modelo, ou seja, **a porcentagem de previsões que ele acertou no total**. Um valor de 89% é um ótimo positivo para a previsão.



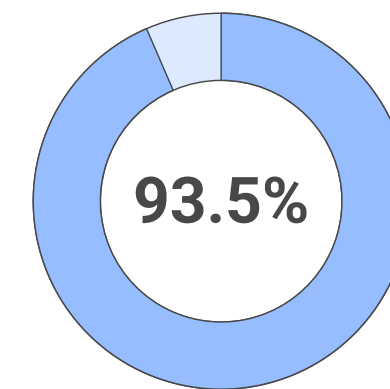
Recall)

O recall responde à pergunta: "De todos os clientes que realmente deram churn, quantos o modelo conseguiu identificar?". **Um recall de 59% indica que o modelo é capaz de "pegar" a maioria dos clientes que estão prestes a cancelar.**



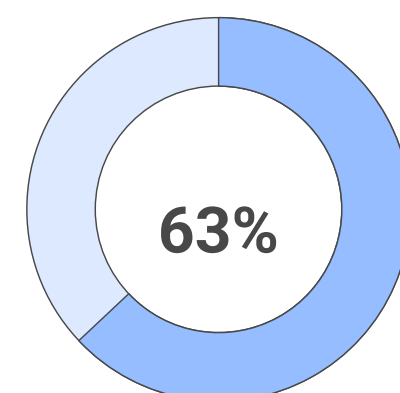
Precisão

A precisão responde à pergunta: "Dos clientes que o modelo previu como Churn, quantos realmente deram churn?". **Um valor de 69% é sólido e significa que, para cada 10 clientes sinalizados como de risco, cerca de 7 realmente estão em alto risco.**



AUC-ROC

A área sob a curva (AUC-ROC) é uma das melhores métricas para avaliar modelos de classificação. **Um valor de 0.935 (em uma escala de 0 a 1) mostra que o modelo é muito bom em distinguir entre clientes de churn e não-churn, superando em muito a aleatoriedade.**

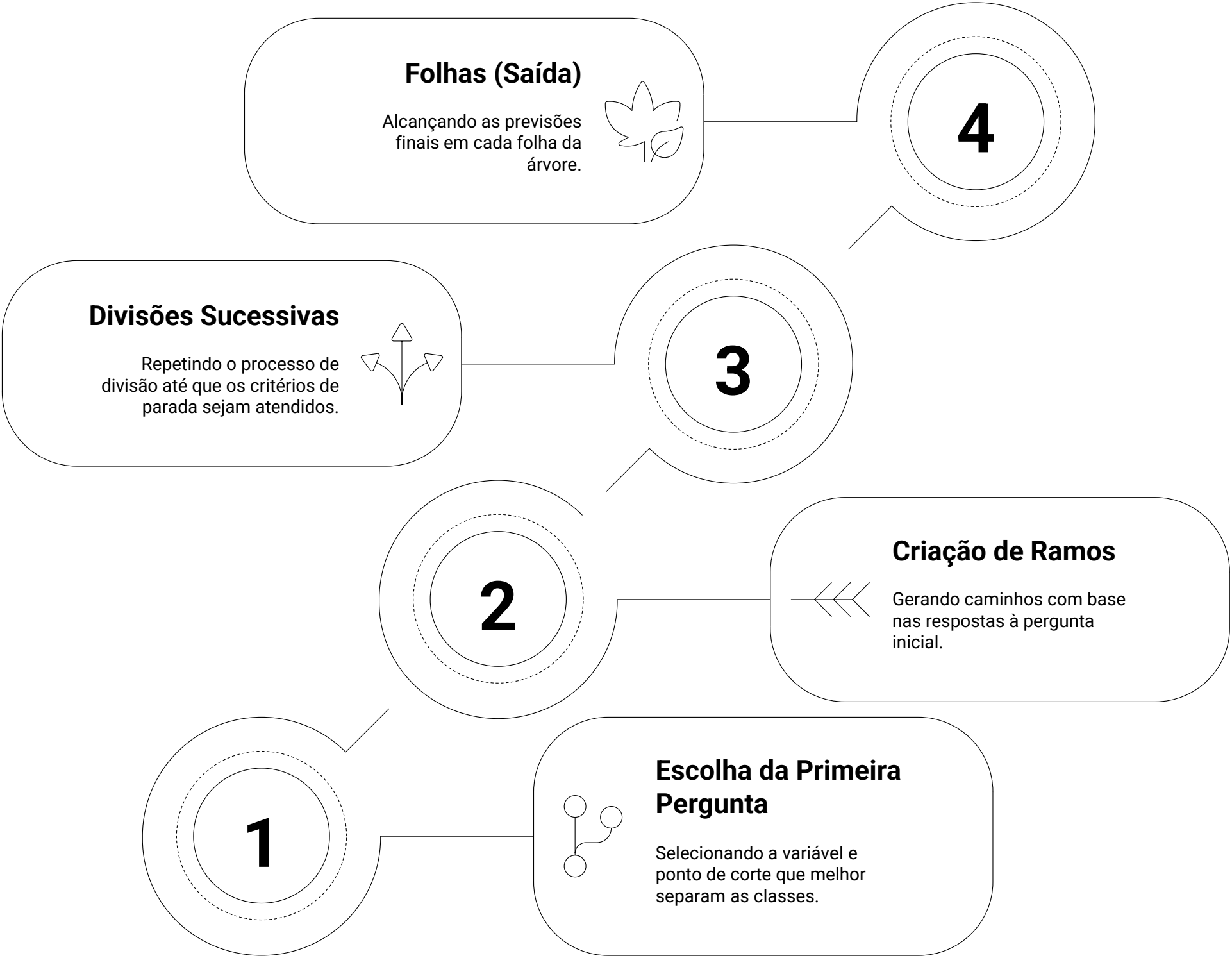


F1-score

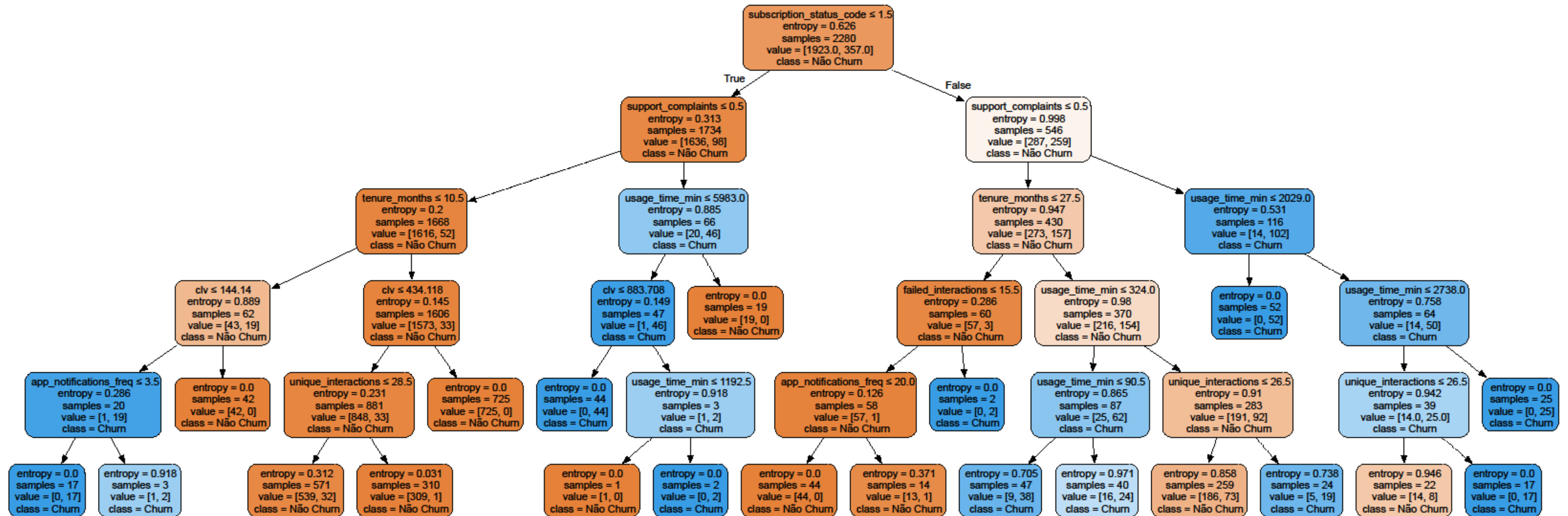
O F1-score sinaliza o balanço entre precisão e recall em um único número. **Um F1-score de 0.637 mostra um bom balanço**, indicando que o modelo é confiável tanto para identificar clientes em risco quanto para evitar falsos alarmes.

Modelo de Arvore de Decisões

Para ter uma melhor visão de modelos preditivos, realizamos o teste com um segundo modelo chamado Árvore de decisões, que cria ramificações utilizando padrões históricos em suas previsões, alertando a possibilidade de churn baseadas em interações que o cliente vai realizando com o serviço (ex: uma reclamação ou tempo de inatividade), se ramificando com cada interação e por fim criando probabilidades deste cliente se tornar um churn.

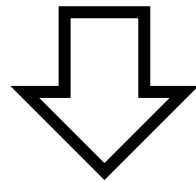


A N Á L I S E D E C H U R N



Comparação De Modelos

Modelo	Acurácia	Precisão	Recall	F1-score	AUC-ROC
Regressão Logística	~90%	~81%	~48%	~0.60	~0.93
Árvore de Decisão	~85%	~81%	~61%	~0.69	(não usado)



No quesito Recall, a **Árvore de Decisões** apresentou uma performance superior, significando que ela tem a capacidade de identificar 13% a mais de clientes em risco de Churn em comparação com a **Regressão Logística**.

Os modelos preditivos são ferramentas que precisam de constante supervisão e refinamentos
Com o tempo, a ideia é otimizar o equilíbrio entre precisão e recall, evitando falsos alarmes e perdas de oportunidades reais.

A N Á L I S E D E C H U R N

Propostas &
Próximos Passos

Automatização

De Score de Risco

Implementar os modelos preditivos é um passo importante para prever e evitar churn, porém não há a necessidade de optar apenas por um caminho, e sim utilizar as dois modelos simultaneamente:

- **Regressão:** usada como ranking probabilístico.
- **Árvore:** usado para explicação de segmentação (saber o por que o cliente está em risco e qual ação tomar).

Prioridade Máxima: Clientes com probabilidade de churn acima de um limiar definido (ex: 0.8). Eles são os alvos de maior risco.

Prioridade Média: Clientes com probabilidade de churn entre 0.5 e 0.8. Eles exigem atenção, mas o risco é moderado.

Baixa Prioridade: Clientes com probabilidade de churn abaixo de 0.5. O modelo os considera seguros e o time de CS pode focar em outros grupos.

Pesquisa sobre a Crise de Fidelidade

Para tratarmos o maior problema da plataforma, vamos realizar uma pesquisa qualitativa para identificar os motivos que levam clientes de longo prazo a darem churn, identificando gatilhos não capturados pelos dados quantitativos e validando as hipóteses levantadas pela análise da jornada.

Público-Alvo da Entrevista: Clientes que cancelaram a assinatura e que tinham um tempo de assinatura (tenure_months) superior a 24 meses.

Amostragem: De 10 a 15 ex-clientes para garantir uma amostra diversificada.

Pesquisa sobre a Crise de Fidelidade

ROTEIRO DE ENTREVISTA

Introdução:

- "Olá, [Nome do Cliente]. Agradeço por seu tempo. Nós notamos que você cancelou a sua assinatura recentemente e, como você foi um cliente fiel por um bom tempo, sua experiência é muito importante para nós. Poderia nos contar um pouco sobre sua jornada com nosso serviço, desde quando começou a usar?"

Fase Inicial:

- "No início, o que te fez escolher nosso serviço e o que você mais gostava nele?"
- "Você se lembra de algum momento em que você sentiu que o serviço realmente te ajudou ou superou suas expectativas?"

Fase de Transição (A Crise):

- "Em algum momento, sua percepção sobre o serviço começou a mudar? O que aconteceu?"
- "Houve algum evento, falha ou problema que te fez começar a pensar em cancelar?"
- "Você chegou a entrar em contato com o nosso suporte por esses problemas? Como foi a experiência?"

Fase Final (O Churn):

- "Qual foi o principal fator que te levou a tomar a decisão de cancelar? Foi um evento único ou uma soma de pequenos problemas?"
- "Se pudéssemos ter feito algo para evitar sua saída, o que seria?"
- "O que você está usando agora para suprir essa necessidade? O que o concorrente oferece que nós não oferecíamos?"

Pesquisa sobre a Crise de Fidelidade

ANÁLISE DOS RESULTADOS

Codificação Qualitativa:

Após as entrevistas, o time de CX deve transcrever as respostas e "codificá-las" por tema. Ex: "Problemas Técnicos Recorrentes", "Falta de Novos Recursos", "Custo-benefício", "Atendimento Lento".

Identificação de Pontos de Dores:

A partir dos códigos, identifique os temas mais recorrentes. Isso revelará os pontos de dor mais comuns entre os clientes fiéis.

Criação de Personas de Churn:

Desenvolva perfis fictícios que representem os principais grupos de clientes que deram churn, com base nas suas histórias.

Pesquisa sobre a Crise de Fidelidade

ENTREGÁVEIS

Relatório de Insights:

Após a análise das entrevistas será apresentado um relatório com as conclusões e as principais citações dos clientes para trazer o valor humano aos dados.

Recomendação de produto

O maior objetivo com as entrevistas é extrair as dores e as necessidades dos clientes para que possamos apresentar uma lista de recomendações para o time de produto, focando em melhorias que resolvam as dores identificadas.

Cronograma De Pesquisa

SETEMBRO

[illegible]

Obrigado!