

PREDIÇÃO DE CHURN

Josiele Ferreira

VISÃO DE NEGÓCIO

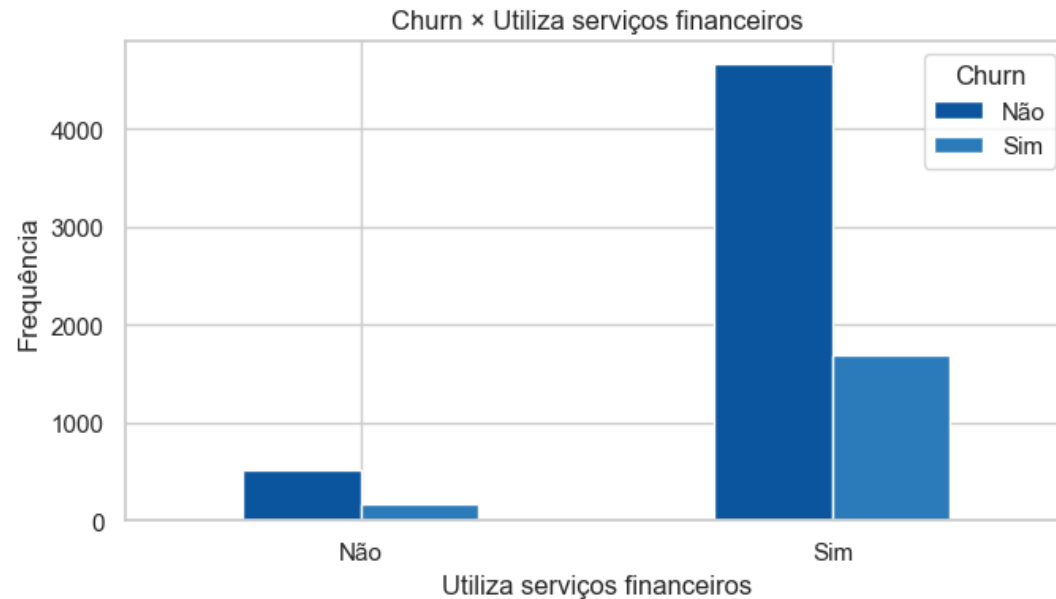
- O churn é um dos indicadores mais críticos para qualquer negócio que depende de relacionamentos recorrentes com clientes – como fintechs, telecom, SaaS, e-commerce, bancos e serviços de assinatura. Ele representa a perda de clientes em um determinado período, e tem impacto direto no crescimento sustentável da empresa;
- Machine Learning aplicado ao Churn é uma das aplicações mais estratégicas em Data Science, principalmente em negócios baseados em assinaturas, serviços financeiros, telecom, SaaS, varejo e e-commerce. A ideia é usar dados históricos de clientes para identificar padrões que indicam maior risco de cancelamento ou abandono e, assim, agir antes que o churn aconteça.
- Aquisição de novos clientes pode custar até **5x mais** do que retenção e assim, com o modelo preditivo permite agir proativamente para reduzir cancelamentos

DATASET

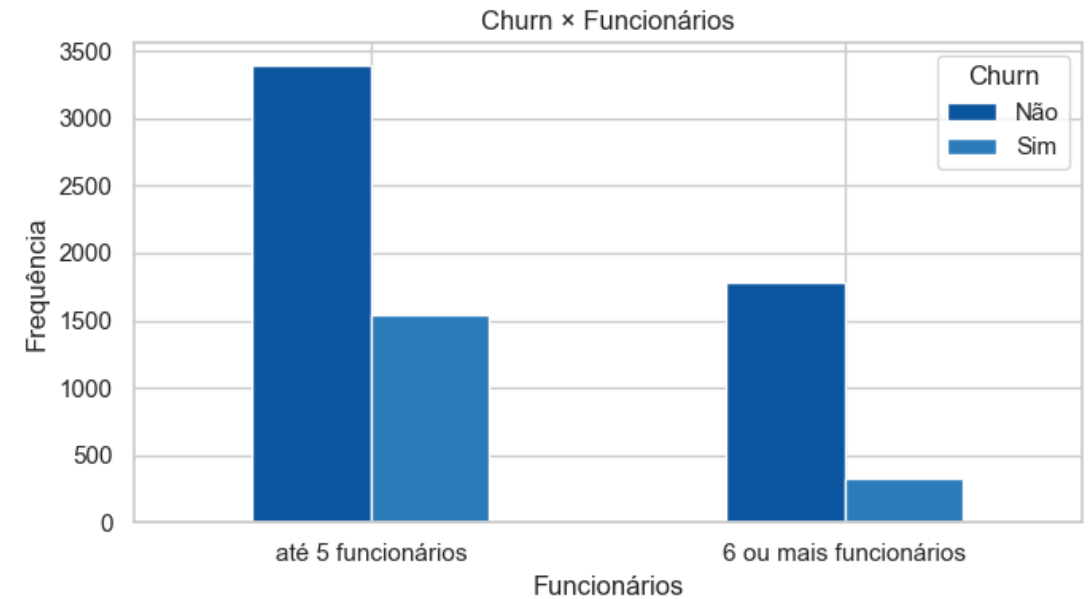
1. Dados: 7043
 2. Total de features: 22
 3. Target binário: Churn (Sim ou Não)
 4. Distribuição desbalanceada:
 - Não: 5.174
 - Sim: 1.869
- Exploração dos Dados: Tratamento e Limpeza dos Dados
 - Pre-processamento: Encoding de variáveis categóricas, Substituição de nulos por 0 ou mediana
 - Treinamento dos modelos: Regressão Logística e XGBoost
 - Métricas usadas: AUC ROC, Precision, F1-score, Recall, Acurácia
 - Modelo final: XGBoost
 - Feature importance

PERFIL DE CLIENTES

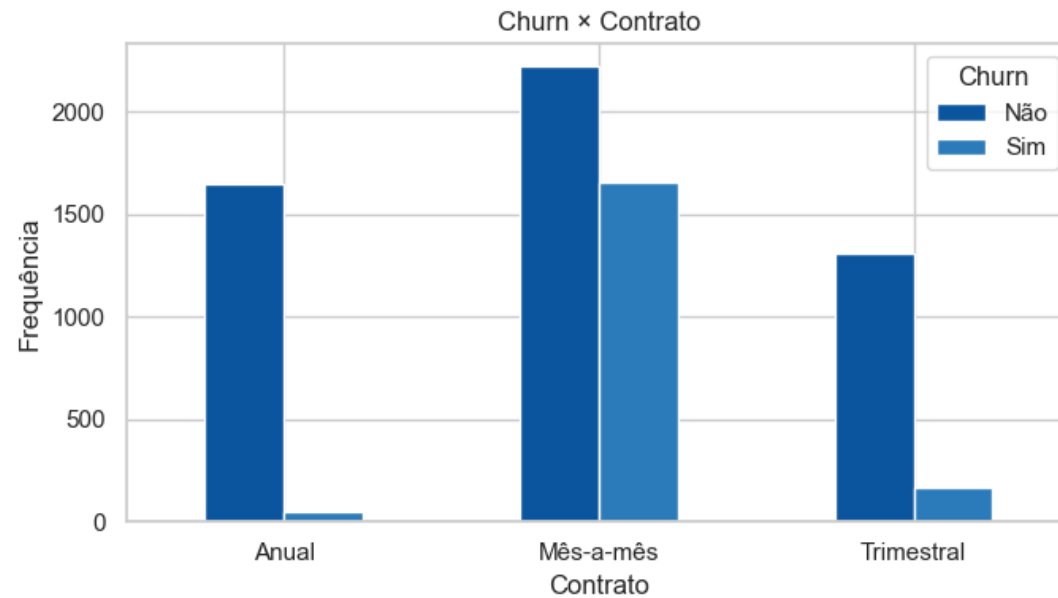
- Maioria utiliza **serviços financeiros** e possui contador;
- O uso de serviços financeiros não apresenta associação significativa com o churn ($p = 0,34$);
- Ou seja, esse fator não influencia na retenção.



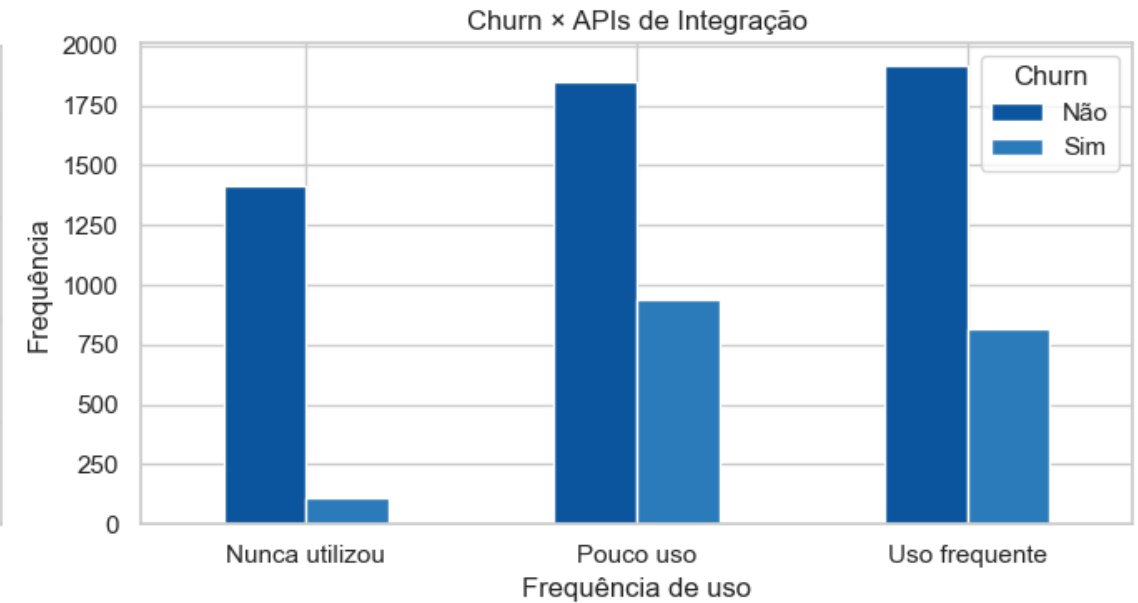
- Predomínio de **microe Pequenas empresas** (até 5 funcionários)
- Existe associação significativa entre o número de funcionários e o churn ($p < 0,001$);
- Empresas com mais de 5 funcionários têm menor probabilidade de cancelamento



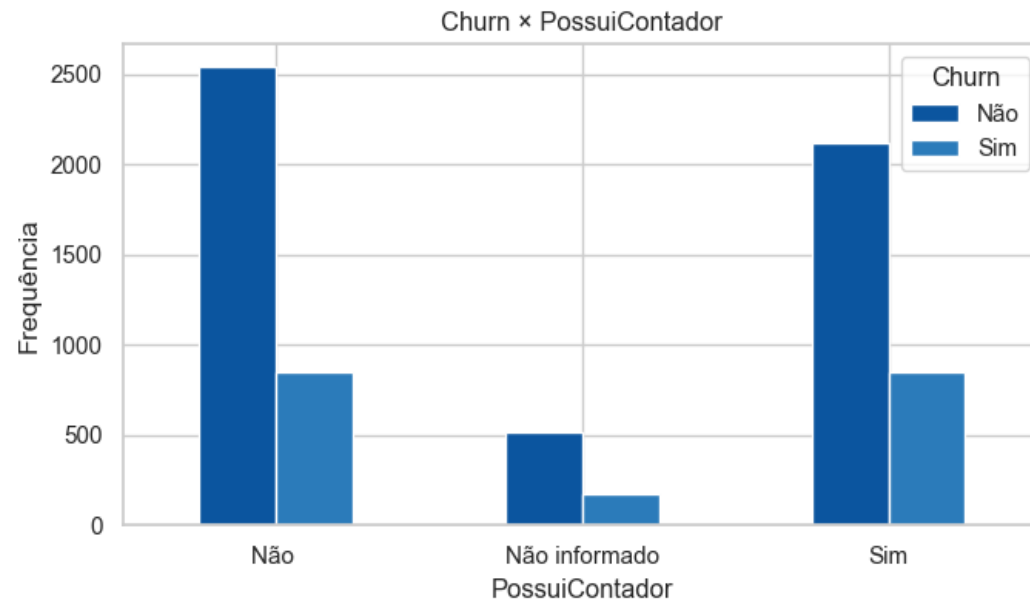
- **Contratos mensais predominam** (maior risco de churn);
- Há associação significativa entre o tipo de contrato e o churn ($p < 0,001$);
- Contratos anuais e trimestrais estão ligados a maior retenção, enquanto o mês a mês apresenta maior risco de cancelamento.



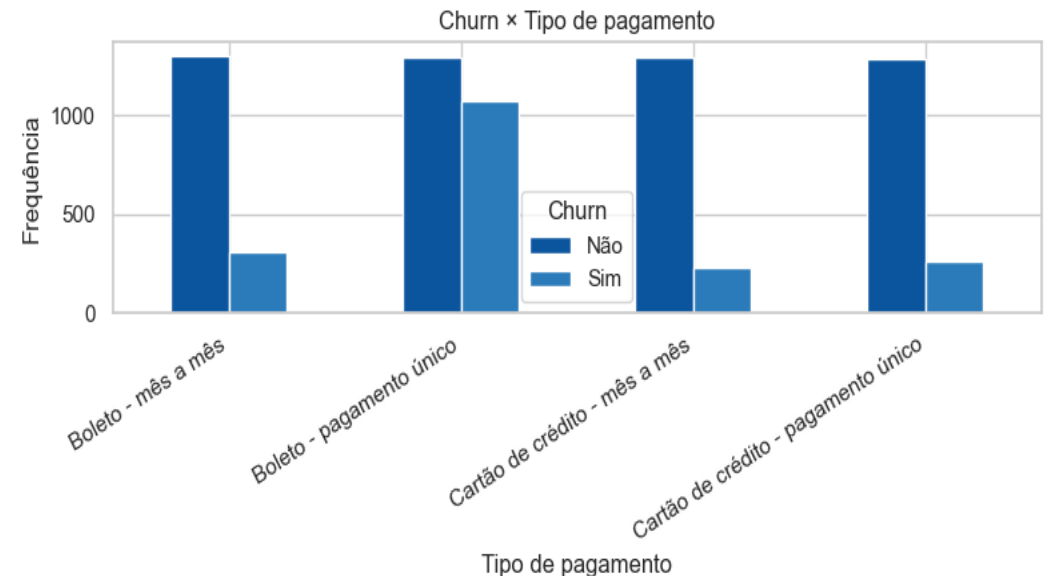
- Adoção desigual dos módulos: **vendas e APIs com baixo uso**
- Existe associação significativa entre o uso de APIs de integração e o churn ($p < 0,001$);
- O uso frequente está ligado à retenção, enquanto o pouco uso aumenta a chance de cancelamento.



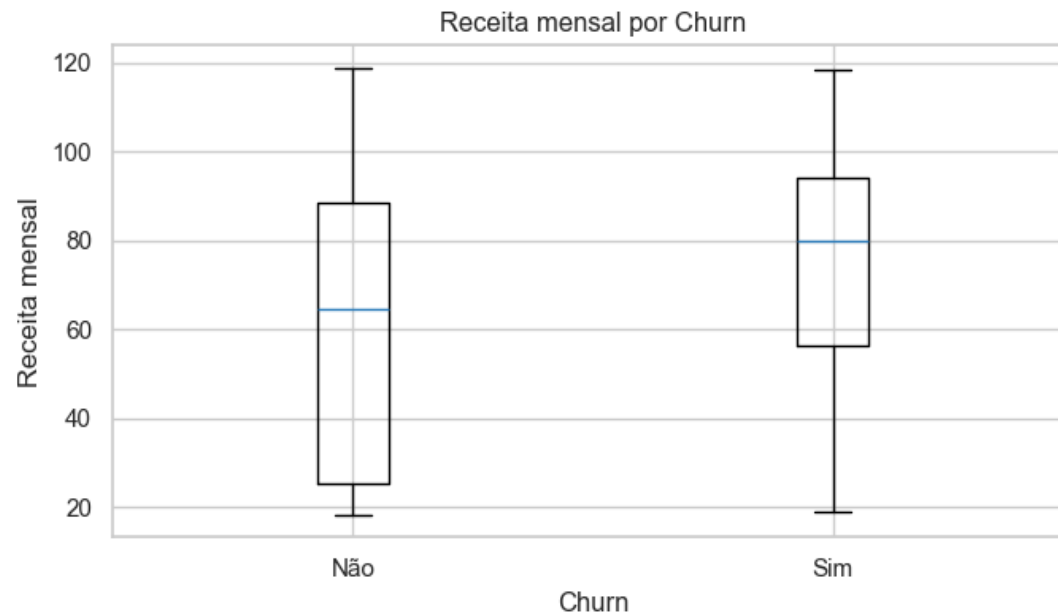
- **Possuir contador é um fator de retenção.** Empresas sem contador têm significativamente mais risco de churn, logo, devem ser priorizadas em ações preventivas.
- Empresas **sem contador** têm uma frequência proporcionalmente maior de churn e empresas **com contador** apresentam menor proporção de cancelamento.



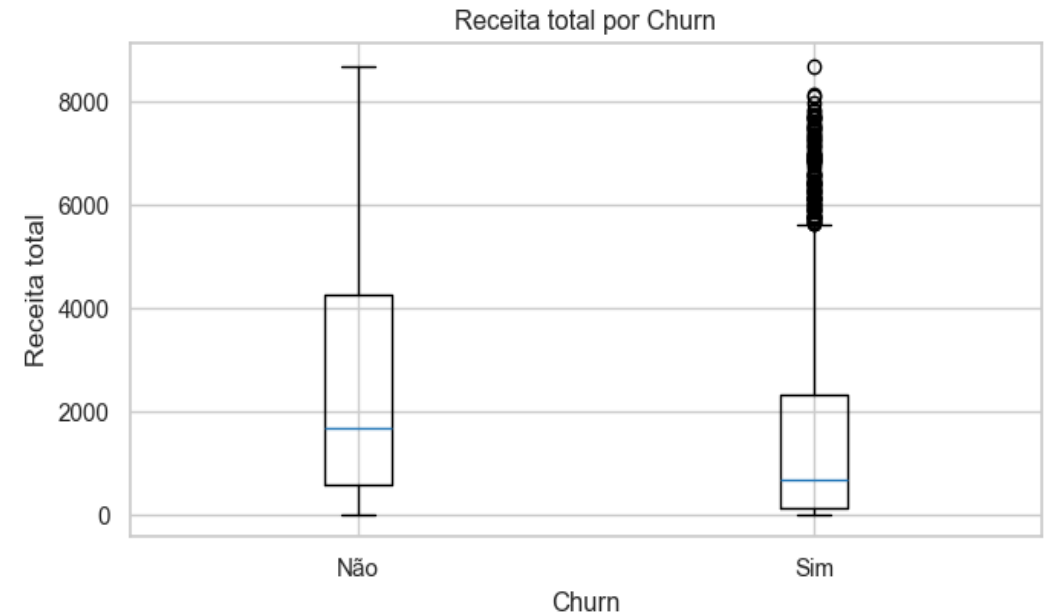
- **O método de pagamento é um dos fatores mais fortes de retenção.**
- Clientes em **boleto mês a mês** estão no grupo de maior risco, enquanto **cartão de crédito** (recorrente) e **pagamento único** são os métodos que mais favorecem a retenção;
- Há associação significativa entre o tipo de pagamento e o churn ($p < 0,001$).



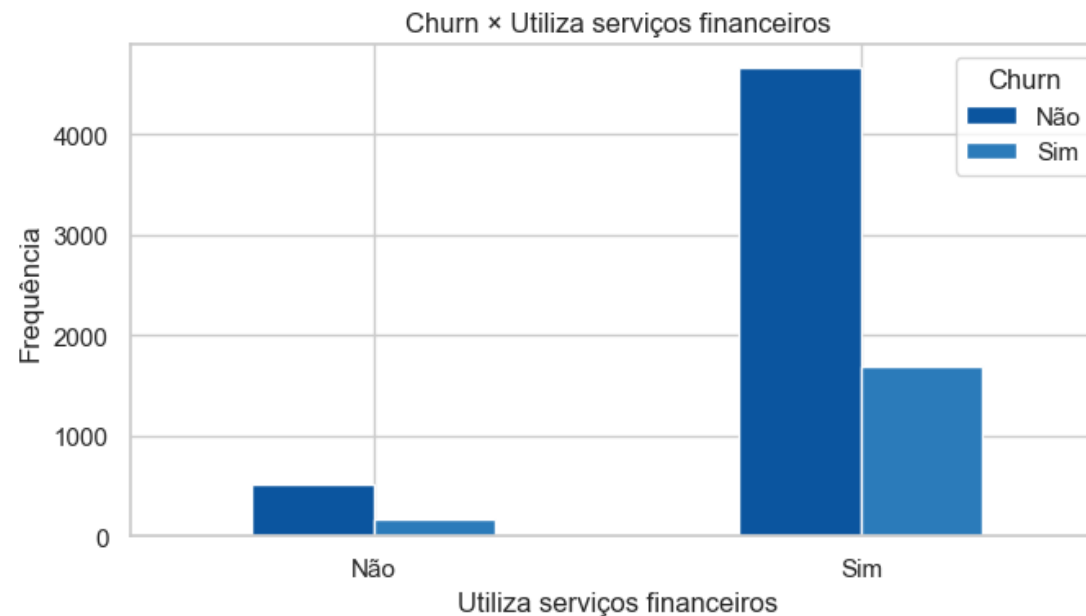
- **Clientes que pagam mais estão cancelando mais**, o que é um risco sério para o negócio. Essa variável é **altamente relevante** tanto para o modelo preditivo de churn quanto para definir **prioridades estratégicas de retenção**;
- Valor-p $\approx 3,31e-54$ (extremamente significativo), isso confirma que a diferença entre os grupos não é ao acaso.



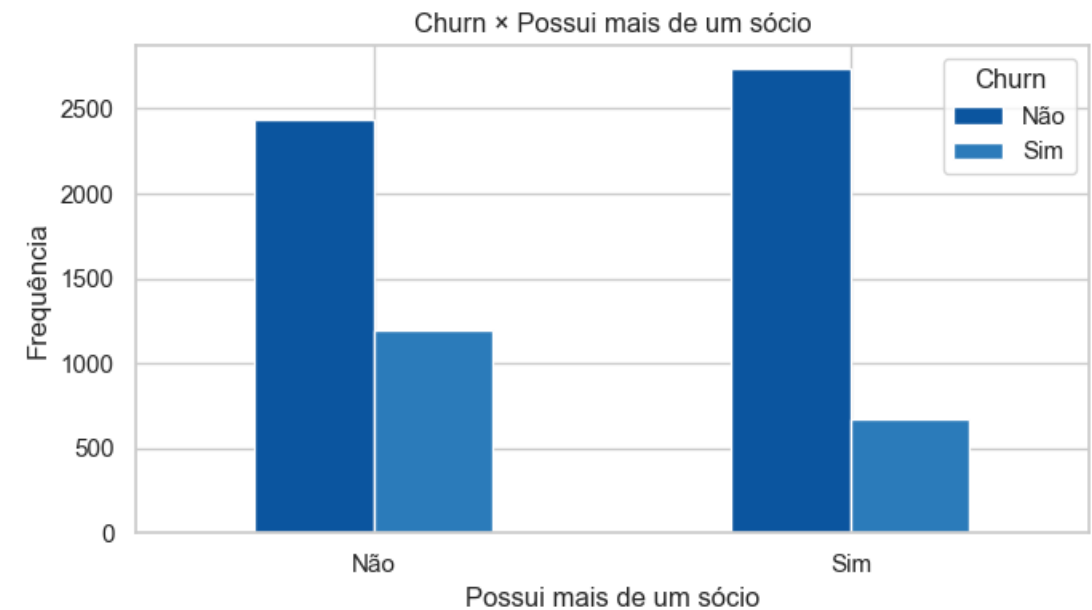
- Clientes que **não deram churn** apresentam **receita total acumulada muito maior**, isso acontece porque eles permanecem por mais tempo, acumulando pagamentos ao longo do ciclo de vida.
- Valor-p $\approx 1,99e-84$ (altamente significativo). Confirma que a diferença entre os grupos não é por acaso.



- O uso de serviços financeiros não tem impacto significativo na retenção. Ou seja, clientes que usam ou não usam serviços financeiros apresentam comportamento semelhante em relação ao churn;
- O uso de serviços financeiros não apresenta associação significativa com o churn ($p = 0,34$).



- Clientes com **mais de um sócio tendem a permanecer mais tempo**, enquanto empresas **com apenas um sócio estão mais propensas ao churn**.
- Há associação significativa entre ter mais de um sócio e o churn ($p < 0,001$).

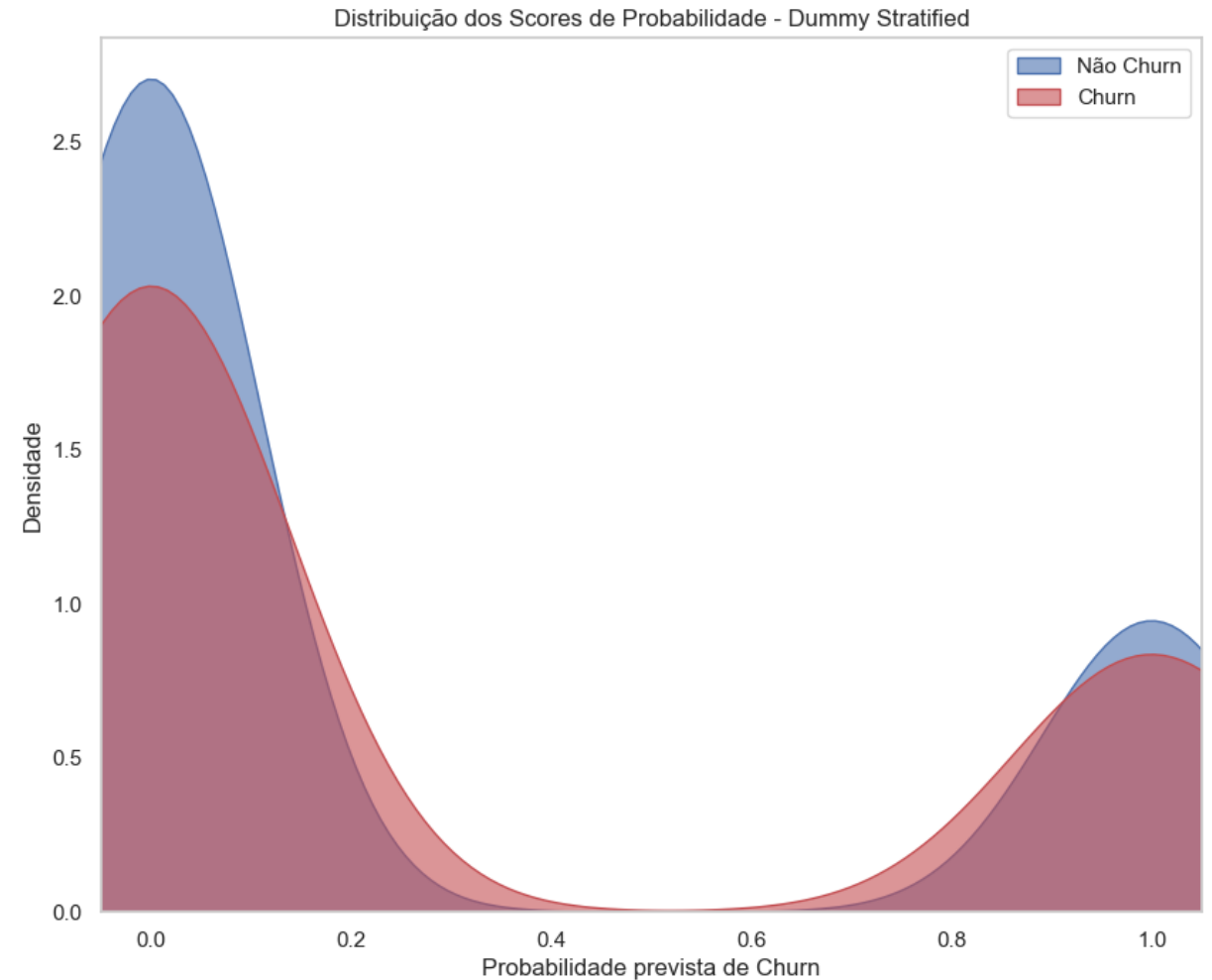


INTERPRETAÇÃO DOS MODELOS

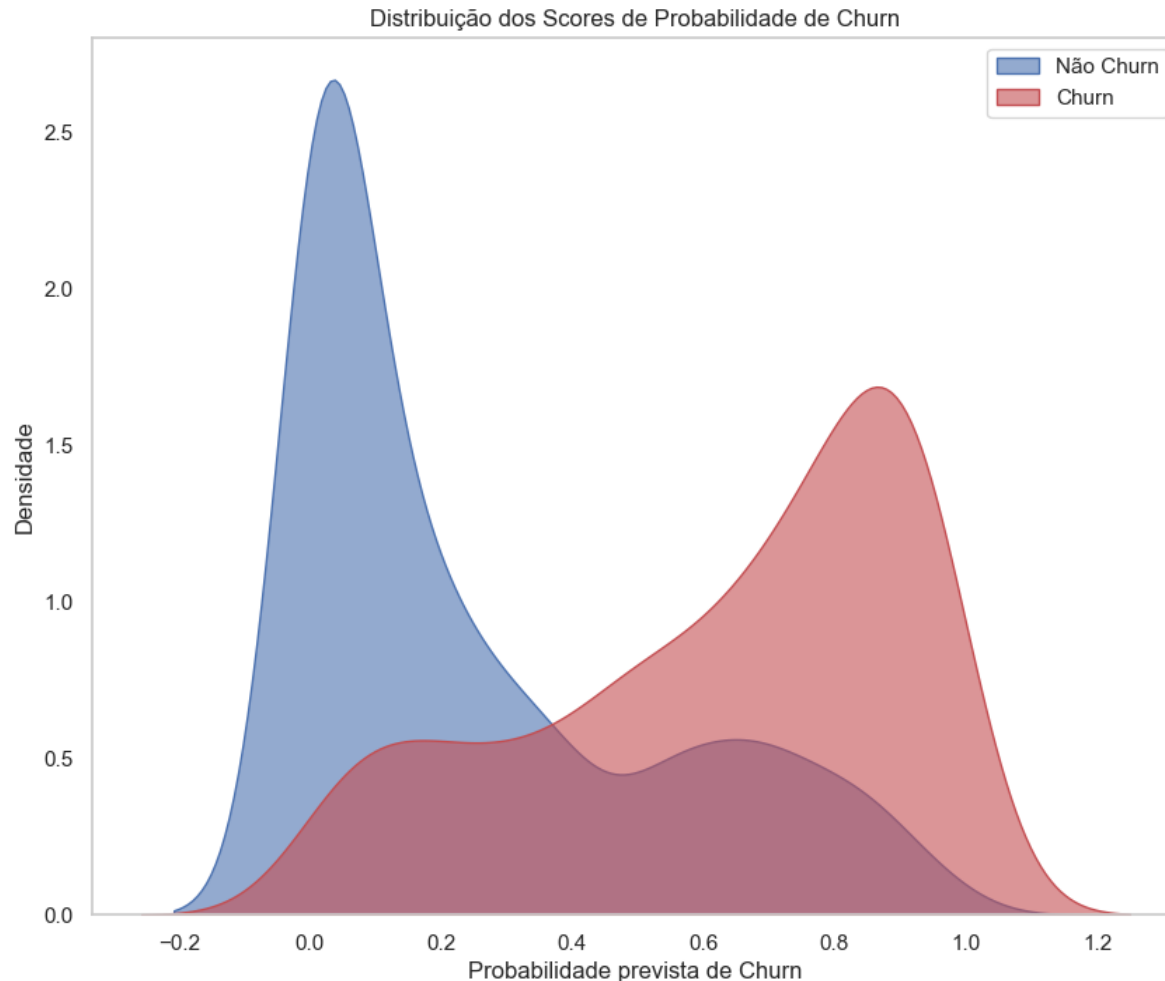
- O **baseline errava na identificação dos churners**, capturando menos de 1/3 dos casos.
- O **modelo XGBoost detecta 70% dos churns reais**, mesmo sacrificando um pouco da precisão (aceita alguns falsos positivos).
- Isso significa que a empresa pode **atuar preventivamente em clientes de alto risco**, reduzindo perdas e aumentando o LTV.
- Esse contraste é **muito poderoso para executivos**, porque mostra que:
 - Sem modelo → praticamente aleatório.
 - Com modelo → ganho real em **antecipar e agir sobre cancelamentos**.

BASELINE

- Baseline (Dummy Stratified)
- **Accuracy:** 62% (reflete apenas o desbalanceamento da base)
- **Precision/Recall para Churn:** 0.29 / 0.29, quase aleatório
- **ROC AUC:** 0.52, praticamente igual ao acaso
- **Insight:** O baseline não diferencia efetivamente positivos de negativos, servindo apenas como referência mínima.



MODELO PREDITIVO



- Modelo XGBoost (com ajuste de classe)
- **Accuracy:** 76%
- **Precision (Churn):** 0.53
- **Recall (Churn):** 0.70 - Aumentou muito a capacidade de detectar churn real
- **Macro F1:** 0.71, Equilíbrio entre classes
- **ROC AUC:** 0.83, Boa capacidade de separação entre churn e não churn
- **Insight:** Clientes Não Churn concentram-se em probabilidades baixas (<0.2), enquanto clientes Churn têm distribuição mais espalhada (0.3–0.8). O modelo separa razoavelmente as classes, mas com sobreposição significativa.

PADRÕES DE CHURN

- Os padrões de churn apontam que **a ausência de contador, o uso de boleto recorrente e a estrutura societária limitada a um único sócio** aumentam a propensão ao cancelamento;
- Além disso, embora clientes de maior ticket mensal estejam mais expostos ao churn, **a retenção prolongada é o verdadeiro diferencial para geração de receita acumulada**. Esses achados oferecem subsídios claros para segmentação de clientes em risco e desenho de estratégias de retenção mais eficazes;
- O **modelo preditivo (XGBoost)** conseguiu capturar justamente esses padrões e entregar capacidade de detecção de churn **muito superior ao baseline** (ROC AUC 0.83 vs. 0.52);
- Taxa significativa de cancelamentos em contratos mensais;
- Baixo engajamento digital (pouco uso de módulos e integrações);
- Clientes menos conectados apresentam maior risco de churn;

CHURN MODEL DATA

Modelo / Métrica	Resultado	Insight
Baseline (Dummy Stratified)	Accuracy: 62% Precision/Recall (Churn): 0.29 / 0.29 ROC AUC: 0.52	Não captura padrões reais; apenas reflete o desbalanceamento da base. Serve como referência mínima.
XGBoost (ajuste de classe)	Accuracy: 76% Precision (Churn): 0.53 Recall (Churn): 0.70 Macro F1: 0.71 ROC AUC: 0.83	Grande melhoria na detecção de churn. Aumenta recall e capacidade de separação, permitindo ações preventivas mais eficazes.

EXPERIMENTOS

- Experimento 1:
 - O experimento avaliou o impacto de uma ação de tratamento sobre clientes de alto risco de churn. Em um teste A/B, o grupo de tratamento apresentou taxa de churn de **19,3%**, contra **25,8%** no grupo controle, representando uma **redução relativa de ~25%**. O resultado foi estatisticamente significativo ($p=0,0136$), confirmando que a intervenção foi eficaz em **reter clientes de alto risco** e trazendo impacto relevante para o negócio.
- Experimento 2:
 - O experimento avaliou a efetividade do modelo em ambiente real, quando o time de negócios decide quais ações aplicar. O churn no grupo tratado foi de **27,4%**, contra **26,4%** no grupo não tratado, indicando apenas uma redução modesta de **~3,8%**, sem significância estatística ($p=0,48$). Entre as ações testadas, **desconto** apresentou melhor resultado (25,8%), enquanto **contato** foi menos eficaz (29,7%). No geral, o impacto global foi pequeno, mas o teste demonstrou a **utilidade prática do modelo** e a capacidade da área de negócios em usar a informação de forma operacional.

RECOMENDAÇÕES ESTRATÉGICAS

- Foco em retenção de contratos mensais, migrar para planos longos
- Campanhas segmentadas para clientes com baixo engajamento
- Ações personalizadas com base no score de churn
- Investimento em Customer Success = aumento direto no LTV

OBRIGADA