

# Previsão de Churn Bancário — Banco Montes Claros

## Resumo Executivo

O Banco Montes Claros (contexto fictício) enfrenta o desafio do **churn** de clientes em seus produtos de crédito (cartão de crédito e empréstimos consignados). Churn representa a evasão de clientes, ou seja, clientes que cancelam serviços ou deixam de usar o cartão. A motivação para este projeto é clara: reter clientes existentes é muito mais barato e lucrativo do que adquirir novos <sup>1</sup>. Estudos de mercado indicam que conquistar um novo cliente pode custar de **5 a 25 vezes mais** do que manter um atual, e aumentar a retenção em apenas 5% pode elevar os lucros em **25% a 95%** <sup>2</sup>. Um churn elevado impacta diretamente a lucratividade e a saúde do negócio, pois representa perda de receita recorrente e aumenta despesas de aquisição para repor clientes perdidos <sup>3</sup> <sup>4</sup>. Portanto, minimizar a saída de clientes é **estrategicamente crítico**.

**Objetivo Geral:** Desenvolver um modelo preditivo de *churn* que identifique com antecedência quais clientes estão propensos a abandonar o banco, possibilitando ações proativas de retenção <sup>5</sup>.

### Objetivos Específicos:

- **Perfil e Drivers de Churn:** Analisar o perfil da base de clientes e entender quais fatores influenciam a saída (churn) <sup>6</sup>, descobrindo padrões de comportamento entre quem sai vs. permanece.
- **Modelos de Machine Learning:** Construir modelos de classificação supervisionada para prever churn, testando algoritmos como Regressão Logística, Árvores de Decisão, Random Forest, XGBoost, LightGBM, etc., comparando desempenho e adequação <sup>6</sup>.
- **Segmentação (Clustering):** Realizar análise de clusters não supervisionada (ex: K-Means) para segmentar os clientes em grupos com características similares, visando estratégias de atendimento/retensão diferenciadas <sup>7</sup>.
- **Lifetime Value (LTV):** Calcular um **proxy de Lifetime Value** para cada cliente ou segmento, estimando o valor de longo prazo de cada cliente para o banco, e integrar essa métrica ao risco de churn <sup>8</sup>. Assim, prioriza-se retenção dos clientes de maior valor.
- **Cenários de Negócio:** Construir simulações de cenários hipotéticos baseados nas previsões de churn e LTV – por exemplo, comparar impacto financeiro caso nenhuma ação seja tomada vs. implementação de ações de retenção – gerando insights acionáveis para suporte à decisão estratégica <sup>9</sup>.

Com esses objetivos, o projeto busca não apenas obter um modelo acurado, mas também **traduzir os resultados em recomendações práticas** que aumentem a retenção e maximizem o valor dos clientes para o banco <sup>4</sup>. Espera-se entregar ao Banco Montes Claros um **sistema de inteligência para retenção** de clientes, capaz de indicar *quem* pode sair, *por que* pode sair, segmentar *quem são* esses clientes e estimar *quanto vale a pena* retê-los, *como* retê-los e **qual o impacto financeiro** de agir ou não agir <sup>10</sup>. Os resultados esperados incluem a redução da taxa de churn nos próximos ciclos, aumento da efetividade das ações de fidelização e melhoria da rentabilidade, alinhados à estratégia de negócio de longo prazo.

# Sumário Técnico da Implementação

## Dados e Análise Exploratória

O dataset utilizado no projeto foi **BankChurners.csv** (fornecido em um case público), adaptado ao contexto do Banco Montes Claros. A base contém dados de **10.127 clientes** de cartão de crédito, com **1.627 casos de churn**, representando uma taxa de churn de aproximadamente **16%** <sup>11</sup>. Cada registro corresponde a um cliente, incluindo variáveis demográficas (idade, gênero, estado civil, educação, faixa de renda), informações de relacionamento financeiro (limite de crédito, quantidade de produtos no banco, tempo de relacionamento) e comportamento de uso do cartão (gastos, número de transações, frequência de uso, inatividade, contatos de atendimento, etc.) <sup>12</sup> <sup>13</sup>. A base não apresentou valores ausentes significativos, e as variáveis já estavam em formatos adequados; apenas foram realizadas conversões necessárias (por exemplo, a flag de churn foi mapeada para 0/1 e categorias foram renomeadas para português para facilitar a interpretação nos gráficos) <sup>14</sup>.

Em termos de perfil geral, a idade média dos clientes é ~46 anos (concentrada entre 40-55 anos), a maioria é casada (~54%) seguida de solteiros (~39%), e quase metade dos clientes possui renda anual abaixo de R\$40K (faixa de renda predominante), enquanto apenas ~7% têm renda > R\$120K <sup>15</sup> <sup>16</sup>. A taxa global de churn de ~16% indica que aproximadamente **1 em cada 6 clientes sai por ano**, o que é significativo: perder 16% da base anualmente pode comprometer crescimento e lucro, dependendo do valor desses clientes e do custo para adquiri-los novamente <sup>17</sup>.

**Padrões iniciais:** Foi observado que **clientes de meia-idade** apresentaram churn ligeiramente maior que muito jovens ou idosos, e intrigantemente clientes de renda e escolaridade mais altas tiveram churn relativamente **maior** do que grupos de renda média <sup>18</sup>. Isso sugere que o banco pode não estar atendendo plenamente as expectativas de clientes de perfil *premium*. Em contraste, características de uso mostraram sinais claros: clientes com **menos transações e longos períodos de inatividade** tendem a churnar mais, enquanto aqueles com **maior engajamento** (uso frequente do cartão) e relacionamento amplo com o banco (vários produtos) tendem a permanecer <sup>19</sup>. Clientes que contatam muito o SAC (indicador potencial de insatisfação) também apresentaram propensão maior de churn, reforçando que **insatisfação e baixo uso são drivers-chave** da evasão <sup>19</sup> <sup>20</sup>. Essas hipóteses foram exploradas quantitativamente na modelagem.

## Engenharia de Atributos (*Feature Engineering*)

Para melhorar a capacidade preditiva e interpretativa, realizamos extensa engenharia de atributos, criando variáveis derivadas a partir dos dados brutos. Entre as novas variáveis destacam-se <sup>21</sup>:

- **Ticket\_Medio**: gasto médio por transação (indicador de valor transacional médio do cliente).
- **Gasto\_Medio\_Mensal**: média mensal de gastos no cartão, calculada a partir do total anual de transações dividido pelo número de meses de relacionamento (ou 12 meses).
- **Rotativo\_Ratio**: razão de saldo rotativo/utilização de crédito, ex.: **Total\_Revolving\_Bal / Credit\_Limit**, indicando dependência de crédito rotativo (dívida no cartão) <sup>22</sup>.
- **Score\_Relacionamento**: índice composto de engajamento com o banco, combinando número de produtos (**Total\_Relationship\_Count**) e tempo de casa (**Months\_on\_book**), para refletir profundidade do relacionamento.
- **Caiu\_Valor** e **Caiu\_Transacoes**: indicadores binários de queda de uso, sinalizando clientes que reduziram significativamente o **valor gasto** ou **número de transações** do Q4 para Q1 (baseado nas variáveis de variação **Total\_Amt\_Chng\_Q4\_Q1** e **Total\_Ct\_Chng\_Q4\_Q1**). Esses flags identificam perda de engajamento recente.

- **Inativo\_3m**: outro indicador binário criado para marcar clientes **muito inativos** (true se **Months\_Inactive\_12\_mon**  $\geq 3$  meses sem uso em 12 meses) <sup>22</sup>. Também foi considerado um indicador para contatos excessivos (ex.: **Contatos\_12m\_Alto** se **Contacts\_Count\_12\_mon** acima de certo limiar).
- **Faixa\_Idade** e **Renda\_Class**: categorização de idade (e.g. <30, 30-50, >50 anos) e de renda (baixa, média, alta) para capturar efeitos não lineares dessas variáveis.
- **LTV\_Proxy**: cálculo customizado de **Lifetime Value proxy**, descrito adiante, estimando o valor relativo de cada cliente.

Essas variáveis adicionadas enriqueceram a representação do cliente, permitindo aos modelos capturar padrões mais complexos. Por exemplo, **Ticket\_Medio** e **Rotativo\_Ratio** ajudam a distinguir estilos de uso (cliente que faz poucas transações de alto valor vs. muitas transações pequenas; cliente que utiliza todo limite vs. paga fatura integral), enquanto **Score\_Relacionamento** resume a fidelidade geral. Os indicadores **Caiu\_Valor/Transacoes** antecipam tendência de abandono, marcando clientes que diminuíram seu uso recentemente – possivelmente *signal vermelho* para churn. Em suma, a engenharia de atributos buscou explicitar comportamentos de risco e valor, **combinando variáveis brutas em insights mais acionáveis**.

## Modelagem Preditiva Supervisionada

Preparamos o dataset para modelagem, realizando a padronização necessária e divisão em treino/teste estratificada (mantenedora da proporção de ~16% churn). Avaliamos **cinco algoritmos de classificação** principais: **Regressão Logística**, **Árvore de Decisão**, **Random Forest**, **XGBoost** e **LightGBM** <sup>6</sup>. Como a classe churn é minoritária, utilizamos técnicas como ajuste de pesos/classe e avaliação por métricas apropriadas (AUC, Recall, Precision, F1) além da acurácia, garantindo sensibilidade ao desbalanceamento <sup>23</sup>.

**Desempenho dos Modelos:** Os resultados mostraram que os métodos de ensemble baseados em árvores (**Random Forest**, **XGBoost**, **LightGBM**) superaram os modelos mais simples (logística e árvore única) em poder preditivo <sup>24</sup>. A tabela abaixo resume as principais métricas obtidas:

Modelo	ROC AUC (aprox.)	Desempenho e Observações (Classe Churn)
<b>Reg. Logística</b>	~0.84 – 0.92	<i>Baseline</i> interpretável. Alto <i>recall</i> (~82%) de churn, mas gerou muitos falsos positivos (precisão ~53%) <sup>25</sup> <sup>24</sup> . Coeficientes indicaram impactos claros das variáveis, útil para insights.
<b>Árvore Decisão</b>	~0.80	Modelo simples e interpretável, porém com menor capacidade preditiva. <i>Recall</i> ~61% apenas <sup>26</sup> – tendência a classificar como não-churn (erro de omissão de churners).
<b>Random Forest</b>	~0.94	Melhorou a detecção de churn ( <i>recall</i> ~68%) mantendo boa precisão (~85%) <sup>27</sup> . Acurácia ~93%. Indícios de overfitting controlados via validação cruzada.
<b>XGBoost</b>	~0.96	Destaque em equilíbrio: capturou >80% dos churners ( <i>recall</i> ~82%) com precisão ~71% <sup>25</sup> . F1 ~76%. Alto <i>AUC</i> e <i>balanced accuracy</i> (~0.88). Modelo robusto, porém mais "agressivo" (sensível a churn).

Modelo	ROC AUC (aprox.)	Desempenho e Observações (Classe Churn)
<b>LightGBM</b>	~0.96	Desempenho equivalente ao XGBoost (recall ~80%, precisão ~72%) <sup>28</sup> . Diferenças sutis; também atingiu F1 ~76%. Modelo mais leve, facilitando rapidez de pontuação.

Como era esperado, os ensembles obtiveram **maior AUC e recall** de churn, características desejáveis no contexto (identificar o máximo de churners possível)<sup>24</sup>. Em contrapartida, a Regressão Logística – embora menos acurada – forneceu interpretabilidade valiosa: seus coeficientes evidenciaram relações entre variáveis e churn, servindo de *check* de consistência com o conhecimento de negócio<sup>29</sup>. Por exemplo, na logística verificou-se que cada mês adicional de inatividade aumentava substancialmente a odds de churn, enquanto clientes com mais produtos tinham odds menores – alinhado às expectativas.

**Modelo Final:** Optamos por utilizar o **XGBoost** treinado como modelo final de previsão de churn em produção, devido ao seu excelente desempenho e eficiência computacional (predição de novos clientes em milissegundos)<sup>30</sup>. O modelo foi salvo como artefato (`modelo_churn_xgb.pkl`) e integrado ao pipeline do projeto. O **LightGBM** teve performance similar, de modo que poderia ser uma alternativa viável; porém, seguimos com XGBoost dado seu ligeiro melhor recall e maturidade em nossos testes.

Para avaliar a performance final do modelo selecionado, analisamos a **matriz de confusão** no conjunto de teste, confirmando que o modelo consegue identificar cerca de 4 em cada 5 churners (bom **recall**), mantendo baixo o alarmismo (taxa de falso positivo controlada, precisão ~70%)<sup>25</sup><sup>31</sup>. A acurácia global atingiu ~92%, mas este número por si só é menos relevante devido ao desbalanceamento de classes – preferimos enfatizar AUC e recall do churn nas análises<sup>32</sup>. Abaixo visualizamos a importância das variáveis no modelo final:

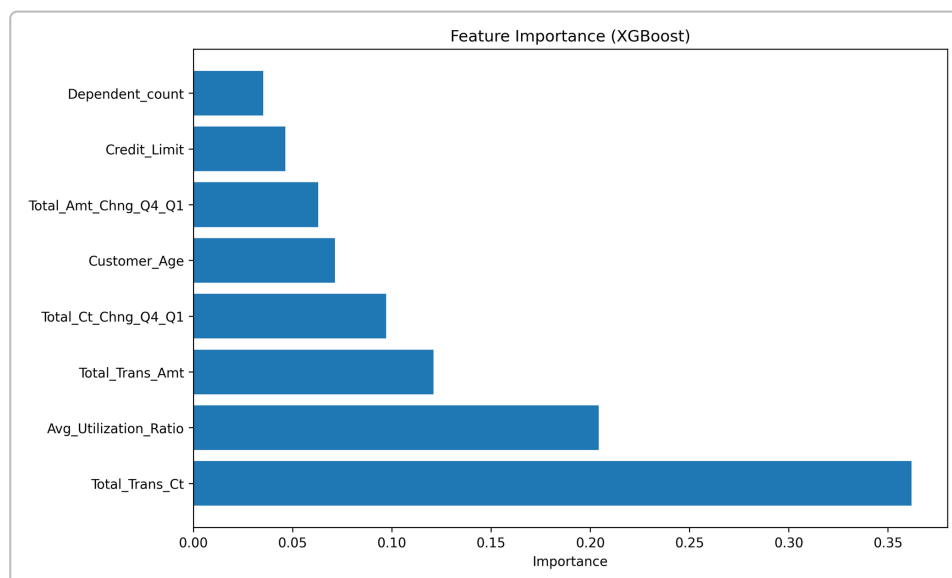


Gráfico de importância de variáveis segundo o modelo XGBoost. As barras representam a contribuição relativa de cada variável na decisão do modelo. Observa-se que o número total de transações nos últimos 12 meses (*Total\_Trans\_Ct*) despontou como o fator mais influente, seguido da taxa média de utilização de limite (*Avg\_Utilization\_Ratio*) e do valor total transacionado (*Total\_Trans\_Amt*). Variáveis que indicam mudança de comportamento, como a variação no número de transações (*Total\_Ct\_Chng\_Q4\_Q1*), e atributos demográficos como idade do cliente, também surgem com importância considerável. Já características como limite de crédito e número de dependentes tiveram menor impacto relativo no modelo.

Essa análise de *feature importance* pelos modelos de árvore confirmou nossos achados de negócio: indicadores de **engajamento** são determinantes – clientes com mais transações e maior utilização do cartão tendem a permanecer, enquanto baixa atividade (refletida implicitamente por poucas transações ou muita inatividade) aumenta o risco de saída <sup>19</sup>. Da mesma forma, outros modelos destacaram o **número de contatos de SAC e meses inativos** entre os principais preditores (variáveis altamente classificatórias para churn vs. não churn) <sup>19</sup> <sup>33</sup>. Em suma, a modelagem forneceu tanto um **ferramental preditivo robusto** (AUC ~96%) quanto **insights interpretáveis** sobre os fatores de churn – combinando o melhor dos dois mundos.

## Segmentação de Clientes (Clustering K-Means)

Além da predição individual de churn, o projeto incorporou uma análise não supervisionada para **segmentar a base de clientes** e entender perfis distintos. Optamos pelo algoritmo **K-Means** pela simplicidade e fácil interpretabilidade inicial. Testamos diferentes valores de *K* (número de clusters) utilizando métodos de avaliação como a curva de cotovelo e silhouette, e definimos **K = 3 clusters** como o ponto ótimo de segmentação para os clientes ativos (não churn) <sup>34</sup> <sup>35</sup>. *Nota:* Faz mais sentido segmentar os clientes **ativos**, já que são eles o alvo de estratégias de retenção; clientes já perdidos (churnados) servem como referência, mas a clusterização foi feita sobre os ~84% que permanecem, considerando suas características atuais.

Os **3 clusters** identificados representam perfis bem distintos na base. Caracterizamos cada segmento analisando as médias das variáveis em cada grupo e visualizando gráficos comparativos (ex.: radar charts de perfil, gráficos de barras) para destacar diferenças. Os segmentos encontrados foram descritos assim <sup>36</sup> <sup>37</sup>:

- **Cluster 1 – "Clientes VIP Engajados"** (~40% dos clientes): Apresentam as **maiores médias** de limite de crédito, gasto anual e número de transações. São clientes de altíssimo valor, que usam bastante o cartão e possuem relacionamento amplo – o maior número médio de produtos no banco <sup>38</sup>. Também possuem renda e escolaridade superiores à média. Em contrapartida, têm baixíssimos índices de inatividade (praticamente não ficam meses sem usar o cartão) e quase não contatam o SAC, indicando alta satisfação. **Churn histórico praticamente nulo** nesse cluster – dificilmente esses clientes cancelam. *Estratégia:* manter a excelência no serviço, recompensá-los (programas de fidelidade exclusivos, upgrades VIP) e aproveitar para cross-sell, pois são defensores da marca.
- **Cluster 2 – "Clientes Moderados e Estáveis"** (~35% dos clientes): Perfil intermediário. Uso moderado do cartão (gastos e transações próximos da média geral), muitos com 2-3 produtos no banco <sup>39</sup>. Idade mediana em torno de 45-50 anos. Tendem a ser conservadores financeiramente: utilizam pouco crédito rotativo (pagam fatura em dia) e não abusam do limite. Também não contatam muito o banco (poucas reclamações). **Churn observado baixo** – relativamente fiéis, embora não tão rentáveis quanto o cluster VIP. *Estratégia:* incentivar maior engajamento para aumentar o valor (oferecer aumento de limite, novos produtos ou benefícios para uso mais frequente do cartão), migrando parte deles para o perfil VIP.
- **Cluster 3 – "Clientes de Risco Desengajados"** (~25% dos clientes): Este cluster concentrou a maioria dos **churners identificados**. Caracteriza-se por **baixo uso** – menor número de transações e gasto anual mais baixo entre os grupos <sup>40</sup>. Muitos desses clientes estão quase inativos, com o **maior número médio de meses inativos** (ex.: >3 meses/ano sem usar) e também o **maior número de contatos no SAC** (sinal de reclamações ou problemas) <sup>41</sup> <sup>42</sup>. Em geral possuem apenas 1 produto (só o cartão) e limites de crédito baixos. Há uma tendência demográfica de serem um pouco mais jovens ou de renda mais baixa, mas o traço marcante é comportamental mesmo (desengajamento). A **taxa de churn dentro deste cluster é elevada** – muitos já saíram ou estão propensos a sair em breve. *Estratégia:* este é o segmento crítico para

ações de **retenção**. Intervenções sugeridas incluem: campanhas de “reengajamento” (estímulo ao uso do cartão via promoções, upgrade de benefícios básicos) e contato proativo para resolver eventuais queixas (melhorar atendimento, oferta de isenção de anuidade, etc.) <sup>43</sup>. Entretanto, dado que vários são clientes de baixo valor, deve-se **priorizar esforços nos subgrupos de maior valor** dentro deste cluster – ou seja, focar retenção onde há **LTV significativo** a proteger, evitando gastar demasiados recursos com quem pouco contribui à receita <sup>44</sup>.

Em resumo, a clusterização permitiu **classificar a base em segmentos acionáveis** com estratégias diferenciadas. Temos um **segmento Top (VIP)** de alto valor e baixo risco – foco em fidelização e aumento de valor; um **segmento Intermediário** de valor médio e baixo risco – foco em aumentar engajamento e upsell; e um **segmento Crítico (Risco)** de baixo valor (em média) e alto risco – foco em mitigar churn dos que valem a pena e eventualmente aceitar a saída dos menos lucrativos <sup>45</sup>. Essa visão segmentada enriquece a análise de churn, pois entendemos **quem são os churners**: majoritariamente parte de um segmento específico (desengajados), ao invés de dispersos aleatoriamente pela base <sup>46</sup>. Isso orienta ações muito mais eficazes – por exemplo, em vez de uma campanha genérica para todos os clientes, o banco pode **direcionar esforços ao Cluster 3**, ou até mais especificamente aos clientes de alto valor dentro do Cluster 3.

## Cálculo de Lifetime Value (LTV) Proxy

Outro pilar importante do projeto foi incorporar a dimensão de **valor do cliente** nas análises. Saber *quem* pode churnar é essencial, mas também é crucial saber **quanto cada cliente vale**, para priorizar retenções. Implementamos então uma estimativa de **Lifetime Value (LTV) proxy** – uma métrica aproximada do valor financeiro de longo prazo de cada cliente para o banco <sup>47</sup>. O conceito de LTV representa o valor presente dos lucros futuros esperados de um cliente enquanto ele permanecer ativo. Sem acesso a dados financeiros detalhados (lucro por cliente, etc.), construímos um **proxy razoável baseado em dados disponíveis** <sup>48</sup>, combinando principalmente o gasto no cartão, o uso de crédito e o relacionamento em produtos, conforme abaixo:

- **Gasto Anual no Cartão:** Usamos a variável `Total_Trans_Amt` (soma anual das transações) como principal proxy de valor. Embora nem todo gasto no cartão seja lucro direto (apenas uma fração via **interchange fee** ou juros do rotativo), assumimos que o gasto reflete engajamento e gera receita proporcional <sup>49</sup>.
- **Uso de Crédito Rotativo:** Clientes com alta utilização de limite (`Avg_Utilization_Ratio`) tendem a pagar juros e tarifas – adicionamos um peso extra para capturar valor proveniente de crédito rotativo/finance charges <sup>50</sup>.
- **Número de Produtos no Banco:** Clientes com múltiplos produtos (ex.: cartão + empréstimo consignado + seguros) geram múltiplas fontes de receita (juros de empréstimo, anuidades, tarifas diversas). Portanto, incluímos um componente de valor relativo ao `Total_Relationship_Count` para diferenciar quem tem mais produtos <sup>51</sup>.

Em termos práticos, definimos uma fórmula linear simples do tipo:

$$** LTV\_proxy = (Total\_Trans\_Amt \times \alpha) + (Saldo\_Rotativo \times \beta) + (Total\_Relationship\_Count \times \gamma) **,$$

onde calibramos os pesos  $\alpha$ ,  $\beta$ ,  $\gamma$  com base em suposições de negócio <sup>52</sup> <sup>53</sup>. Escolhemos, por exemplo,  $\alpha = 1.0$  (cada  $\$1$  gasto anual conta como  $\$1$  de valor base),  $\beta = 0.2$  (dando um crédito adicional de  $\$0.20$  por dólar de saldo rotativo médio, assumindo margem de juros) e  $\gamma = 500$  (cada produto adicional agrega um valor fixo equivalente a  $\$500$  em valor presente). Esses valores foram ajustados heurísticamente para gerar uma escala coerente – **não são absolutos**, mas servem para ranquear os clientes por valor. Após o cálculo contínuo, normalizamos o

LTV proxy numa escala comparativa (por exemplo, pontuação 0 a 100 ou categorização **Baixo/Médio/Alto LTV**) <sup>54</sup> .

Mesmo sendo um proxy simples, essa métrica nos permitiu **ordenar os clientes por valor** e observar diferenças claras: o Cluster 1 (VIP) teve LTV proxy **alto**, o Cluster 3 (Desengajados) teve LTV muito **baixo** em média, e o Cluster 2 intermediário apresentou variabilidade <sup>55</sup> <sup>56</sup> . Isso era esperado e validou o proxy. Mais importante, **combinamos o LTV proxy com a probabilidade de churn** estimada pelo modelo para direcionar ações de retenção <sup>57</sup> . Em outras palavras, criamos uma matriz **Risco vs. Valor** segmentando os clientes em categorias como:

- **Alto Risco & Alto Valor:** clientes críticos (“diamantes prestes a sair”). Acharmos, por exemplo, clientes no **top 20% de LTV** com previsão de churn > 50% – estes são prioridade máxima, pois perder um deles gera grande impacto em lucro <sup>58</sup> . Devem receber intervenção imediata (ofertas personalizadas, contato do gerente, etc.).
- **Alto Risco & Baixo Valor:** muitos destes estão no Cluster 3. Se recursos de retenção são limitados, esses clientes podem **não justificar esforços caros** – pode-se tentar retenção via meios automatizados de baixo custo (email marketing, ofertas genéricas), mas, caso insistam em sair, o impacto financeiro é pequeno <sup>59</sup> <sup>60</sup> .
- **Baixo Risco & Alto Valor:** tipicamente clientes VIP leais – não há urgência de ação corretiva, apenas continuar programas de fidelização e monitorar <sup>61</sup> .
- **Baixo Risco & Baixo Valor:** clientes estáveis de pouco valor – pouca preocupação de churn e baixo impacto. Estratégia aqui pode ser trabalhar **upsell** para aumentar o LTV, já que eles provavelmente ficarão, mas geram pouco lucro.

Essa priorização Risco-Valor foi central nas recomendações: garantiu que esforços de retenção fossem **economicamente racionais**, investindo mais onde há maior retorno esperado. Em suma, incorporar LTV permitiu responder “**quem vale a pena reter?**”, complementando a previsão de churn que responde “quem provavelmente vai sair?”. Como enfatiza a literatura de Data Science aplicada a negócios, maximizar retenção passa por **maximizar LTV** – não basta segurar clientes indiscriminadamente, mas sim focar naqueles que trazem mais valor <sup>62</sup> .

## Simulações de Cenários de Negócio

Com o modelo de churn treinado, a segmentação e os valores (LTV) estimados, realizamos simulações de **cenários hipotéticos** para quantificar impactos e embasar decisões estratégicas do banco. Construímos quatro cenários principais, comparando a situação atual vs. possíveis ações ou mudanças externas <sup>63</sup> <sup>64</sup> :

1. **Cenário 1 – Status Quo (Sem Ação de Retenção):** Usando as previsões do modelo, estimamos **quantos clientes churnariam em 12 meses se nenhuma ação for tomada**. Aproximadamente 16% dos ~10.000 clientes (ou seja, ~1.600 clientes) tenderiam a sair no próximo ano <sup>65</sup> . Podemos estimar o custo disso somando o LTV proxy desses churners previstos – essa seria a **receita futura perdida** caso nada seja feito, servindo de base para comparação nos cenários seguintes (perda potencial de \$\$ milhões, proporcional ao valor médio dos churners).
2. **Cenário 2 – Foco em Retenção de Alto Risco/Alto Valor:** Neste cenário, simulamos uma **campanha de retenção dirigida aos clientes mais valiosos em risco**. Por exemplo, focamos nos **top 200 clientes de maior LTV que foram preditos como churn**. Supondo que, através de incentivos (ex.: ofertas exclusivas, upgrade de serviço), o banco consiga reter **50% desses clientes que churnariam**, removemos metade deles da previsão de churn. Isso reduziria o churn total previsto de ~1.600 para ~1.400 clientes/ano <sup>66</sup> . Calculamos então **quanto de valor**

**(LTV) é salvo:** se inicialmente os 1.600 churners somavam \X em LTV perdido, ao salvar 100 deles (50% de 200), digamos que recuperamos \Y daquele valor. Houve custo? Sim, consideramos o custo da campanha – por exemplo, supondo \\$100 de custo médio para reter um cliente (brindes, descontos, etc.), para 200 clientes seria \\$20k investidos. Ainda assim, se salvamos \\$100k em valor futuro, o **ROI** é positivo (cada \\$1 investido rendeu \\$5 em valor retido) <sup>67</sup>. Esse cenário demonstrou que **focar em poucos clientes “High Value, High Risk” traz alto retorno** – estratégias personalizadas para esse grupo são altamente efetivas.

3. **Cenário 3 – Ação Preventiva Geral (Amplia Retenção na Base Toda):** E se o banco **fizesse uma ação abrangente para reduzir churn em geral?** Simulamos, por exemplo, que melhorias no serviço e programas de engajamento reduzam a taxa de churn do **Cluster 3 (desengajados)** em 5 pontos percentuais (de ~??% para ~??% naquele grupo). Isso poderia diminuir o churn global de ~16% para ~14% <sup>68</sup>. Estimamos o ganho de valor por evitar essas saídas extras – mesmo clientes de baixo valor, em volume, somam valor considerável retido. Porém, comparado ao cenário 2, o **ROI tende a ser menor**, já que inclui muitos clientes de baixo valor e envolve ações massivas (custos maiores para cobrir toda a base) <sup>69</sup> <sup>70</sup>. Em contrapartida, esse cenário **melhora a satisfação geral** e pode ter efeitos de longo prazo (marca, indicações). É um balanço entre retenção ampla vs. focada.
4. **Cenário 4 – Mudança de Mercado (Piora Externa):** Consideramos também cenários adversos fora do controle do banco, por exemplo uma **elevação na propensão de churn devido a concorrência ou crise econômica**. Suponhamos que a taxa básica de churn anual subiria para **20%** (mais clientes insatisfeitos ou buscando concorrentes) <sup>71</sup> <sup>72</sup>. Projetamos o impacto: ~2.000 saíam em vez de 1.600, aumentando a perda de LTV significativamente. Essa simulação gerou **argumentos para investir em fidelização** mesmo em tempos difíceis – por exemplo, mostrou que se nada for feito num cenário de mercado adverso, as perdas anuais adicionais seriam enormes, enquanto um investimento preventivo em retenção poderia atenuar esse efeito <sup>73</sup> <sup>74</sup>. Em essência, o cenário 4 reforçou a importância de ter **planos de contingência** de retenção quando sinais de aumento de churn aparecem.

Comparando os cenários, obtivemos insights valiosos: o **Cenário 2 (ação focada em clientes de alto valor)** mostrou o melhor retorno por real investido, salvando receita futura substancial com custo relativamente baixo <sup>64</sup>. O **Cenário 3 (ação ampla)**, embora menos eficiente financeiramente, apresentou benefício em volume total de clientes retidos e pode ser justificado para melhorar a experiência de toda a base. Já o **Status Quo** quantificou o “custo da inação” – uma perda anual considerável, que provavelmente supera os custos de investir em retenção. E o cenário de **piora externa** serviu como alerta: se indicadores de churn subirem, deve-se reagir rápido com campanhas adicionais para evitar perda ainda maior de clientes. Todos os cenários e premissas foram documentados (notebooks e reports técnicos), e os números, embora baseados em proxies, ajudam a **convencer a diretoria do banco sobre onde investir o budget de retenção** de forma racional <sup>75</sup> <sup>76</sup>.

## Principais Recomendações Estratégicas

Integrando todos os resultados – modelo preditivo, análise de drivers, clusters, LTV e cenários – chegamos a um conjunto de **recomendações práticas** para o Banco Montes Claros maximizar retenção e valor dos clientes:

- **Ações Pró-Ativas nos Drivers de Churn:** Endereçar diretamente os fatores críticos identificados. Por exemplo, implementar um **programa de reativação** para clientes inativos: monitorar clientes com 2+ meses seguidos sem uso do cartão e oferecer incentivos (milhas bônus, descontos na fatura, isenção de anuidade) para trazê-los de volta antes que atinjam 3+ meses



inativos <sup>77</sup>. Em paralelo, criar um fluxo especial para clientes com **muitos contatos de SAC** (reclamações recorrentes): tratá-los de forma prioritária (oferecer um atendimento dedicado, solucionar problemas rapidamente) antes que percam a paciência e abandonem <sup>19</sup> <sup>78</sup>. Essas ações preventivas atacam as causas raiz do churn (insatisfação e desengajamento) nos momentos certos.

- **Retenção Personalizada por Segmento:** Aproveitar a clusterização para **customizar estratégias de retenção**. Clientes **VIP Engajados** (Cluster 1) têm churn baixíssimo, mas altíssimo valor – manter eles felizes é fundamental. Recomenda-se **programas premium**: upgrades de cartão (Platinum+), convites a eventos exclusivos, aumento proativo de limites e ofertas de investimentos, tornando-os ainda mais leais <sup>79</sup> <sup>80</sup>. Clientes **Moderados** (Cluster 2) apresentam oportunidade de crescimento – **incentivar maior uso** com promoções de cashback direcionadas aos que usam pouco, combos de produtos (oferecer seguro, empréstimo, etc.) para aumentar envolvimento <sup>81</sup> <sup>82</sup>. Já o segmento **Desengajado de Risco** (Cluster 3) requer retenção básica urgente: aqui sugerimos uma campanha de “**Reconquista**” focada – contato proativo oferecendo, por exemplo, **isenção de anuidade ou redução de juros do empréstimo** para aqueles que voltarem a usar ativamente <sup>83</sup>. Além disso, realizar pesquisas de satisfação direcionadas a esse grupo para entender queixas específicas. **Dentro do cluster 3, priorize subgrupos de maior valor** (conforme LTV proxy) para investimento mais intenso, pois alguns clientes não compensarão o custo de retenção – é melhor deixá-los sair do que gastar sem retorno <sup>44</sup>.
- **Focar no Binômio Risco vs. Valor:** Nem todos os clientes têm o mesmo valor – logo, as ações de retenção devem ser priorizadas conforme a matriz **Risco/Valor**. Recomendamos estabelecer uma lista VIP de clientes **High Value, High Risk** (LTV alto + prob. churn alta) a ser acompanhada de perto pelo time de sucesso do cliente <sup>84</sup>. Esses clientes “diamante” devem receber atenção humana imediata ao menor sinal de insatisfação (ex.: ligação do gerente, oferta personalizada), pois perdê-los impacta muito o lucro. Em contraste, clientes **Low Value, High Risk** podem receber ações automatizadas mais simples (e-mail marketing com ofertas padrão) e, se ainda assim decidirem sair, o banco pode aceitar a saída com pouco investimento de retenção <sup>85</sup>. Essa gestão segmentada garante melhor alocação de recursos de marketing.
- **Ajustes em Produto e Operações:** Alguns **insights colaterais** sugerem melhorias nos serviços do banco. Observamos que clientes de **maior renda/escolaridade têm churn relativamente maior** <sup>18</sup>, indicando que talvez o banco não esteja atendendo plenamente esse público exigente. Recomenda-se avaliar o portfólio de cartões e serviços **premium** – por exemplo, criar um **cartão Platinum+** ou benefícios exclusivos para alta renda, evitando que esses clientes migrem para concorrentes internacionais <sup>86</sup>. Adicionalmente, o fato de clientes com muitos contatos no SAC tenderem a churnar indica possíveis falhas no atendimento ou processos burocráticos. O banco deve investigar as reclamações frequentes e melhorar a **experiência de atendimento**: reduzir tempos de espera, treinar atendentes para resolver problemas na primeira ligação, e oferecer canais mais ágeis (chat, gerente dedicado) para clientes insatisfeitos <sup>87</sup>. Tais melhorias operacionais têm efeito direto na retenção.
- **Implantação de Monitoramento Contínuo:** Sugerimos integrar o modelo de churn aos sistemas internos para **geração de alertas preventivos**. Por exemplo, mensalmente rodar o modelo nos clientes atuais e sinalizar aqueles com probabilidade de churn acima de um certo threshold (digamos >50%) <sup>88</sup>. Esses clientes entrariam numa lista de ações prioritárias naquele mês. Esse processo pode ser automatizado via o **web app** desenvolvido (ver adiante) ou integrado ao CRM do banco. Assim, a equipe de relacionamento passa a agir de forma **proativa em vez de reativa**, contatando o cliente antes que ele feche a conta <sup>89</sup>. Recomendamos

também acompanhar continuamente métricas de sucesso da retenção: churn mensal, LTV médio da base, e taxas de sucesso das ofertas de retenção (quantos clientes abordados aceitaram ficar) <sup>90</sup>. Por fim, a diretoria deve **alocar orçamento fixo para campanhas de retenção** e refinar o cálculo de LTV no futuro incorporando dados financeiros reais (receita por cliente, custo de aquisição), melhorando a precisão do ROI das iniciativas <sup>91</sup> <sup>92</sup>.

Em conclusão, o projeto *Previsão de Churn Bancário* forneceu ao Banco Montes Claros uma **solução completa e integrada** para gerir a retenção de clientes: desde um modelo preditivo acurado que identifica quem tem risco de sair, passando por insights dos motivos e perfis desses clientes, até ferramentas para estimar o valor financeiro de cada um e simular impactos de ações de negócio. Essas análises convergiram em recomendações concretas de marketing, atendimento e produto que, se implementadas, devem **reduzir a taxa de churn e aumentar a rentabilidade** da base de clientes, alinhando-se às metas estratégicas de fidelização. O próximo passo é operacionalizar essas soluções – integrar o modelo nos sistemas do banco e executar as estratégias sugeridas – criando assim um ciclo virtuoso de melhoria contínua, onde os **insights de Data Science são traduzidos em ações reais de negócio** e resultados mensuráveis.

---

## Apresentação para a Banca (Roteiro de Slides)

### Slide 1: Introdução ao Problema

- **Contexto:** Banco Montes Claros (fictício) atua com crédito consignado e cartões. Sofre com **churn** de clientes (cancelamento de cartões/contas).
- **Motivação:** Reter clientes é vital – custa *5 a 25x menos* do que adquirir novos <sup>1</sup>. Além disso, +5% retenção pode elevar lucros em até 95% <sup>2</sup>. Churn alto = perda de receita e aumento de custos (marketing de aquisição).
- **Problema de Negócio:** Identificar *antes* quem vai sair, para podermos agir e evitar a evasão. Atualmente ~16% dos clientes saem ao ano (índice preocupante). Objetivo é baixar isso via inteligência de dados.

### Slide 2: Objetivos do Projeto

- **Objetivo Geral:** Desenvolver um modelo preditivo de churn que antecipe quais clientes estão propensos a abandonar o banco <sup>5</sup>, permitindo ações proativas de retenção.
- **Objetivos Específicos:**
  - Analisar o **perfil dos clientes** e descobrir **fatores-chave** que influenciam o churn (quem sai e por quê).
  - Construir e comparar **modelos de Machine Learning** (Logística, Árvores, Random Forest, XGBoost, etc.) para prever churn com acurácia.
  - **Segmentar** os clientes em grupos homogêneos (via *clustering*) para estratégias de retenção personalizadas.
  - Estimar um **Lifetime Value (LTV)** proxy para cada cliente, integrando risco de churn + valor financeiro, focando retenção nos clientes mais valiosos.
  - Simular **cenários de negócio**: qual impacto se nada for feito vs. se realizarmos ações de retenção específicas? Gerar recomendações estratégicas a partir disso.

### Slide 3: Dados e Exploração Inicial

- **Dataset:** ~10 mil clientes de cartão de crédito, 16% churn (BankChurners case adaptado) <sup>11</sup>. Inclui dados demográficos (idade ~46 anos média, 54% casados), perfil financeiro (renda predominante < \ \$40k anuais) e comportamental (gastos, transações, contatos SAC etc.).
- **Qualidade:** Dados completos (sem missing significativos). Variáveis já adequadas; apenas traduzidas e

recodificadas para análise (e.g., churn mapeado para 0/1).

- **Churn Rate:** ~1 em cada 6 clientes sai por ano – representa possível **perda de receita considerável** se forem clientes valiosos.

- **Padrões Observados:** Indicativos de churn mais alto em clientes **pouco engajados**: por ex., grupos com poucas transações e vários meses inativos tiveram taxas de churn maiores. Clientes que contataram muito o SAC (possíveis insatisfeitos) também saem mais. Por outro lado, clientes com uso frequente do cartão e relacionamento amplo (mais produtos) tendem a permanecer. *Insight:* Satisfação e engajamento parecem ser os fatores críticos por trás do churn.

#### Slide 4: Modelagem Preditiva (ML)

- **Abordagem:** Testamos modelos supervisionados de classificação: **Regressão Logística** (baseline interpretável), **Árvore de Decisão**, **Random Forest**, **XGBoost** e **LightGBM**. Avaliamos por AUC, Recall, Precision, F1, dado o desbalanceamento (84% não-churn).

- **Desempenho:** Os modelos de ensemble (árvores) superaram os simples. *XGBoost* e *LightGBM* obtiveram os melhores resultados (AUC ~0.96). O modelo final escolhido foi o **XGBoost**, que atingiu ~92% acurácia com **Recall ~82%** (captura 4 em 5 churners) e **Precision ~71%** <sup>25</sup> – um bom equilíbrio entre detectar churners e evitar falsos alarmes.

- **Importante:** A Regressão Logística, apesar de AUC menor (~0.84), confirmou teorias – ex.: coeficientes mostraram que **meses inativos e muitos contatos aumentam a chance de churn**, enquanto maior número de produtos reduz. Essa interpretabilidade ajudou a explicar os drivers para stakeholders, complementando os modelos de alta performance.

- **Principais Variáveis:** Os modelos identificaram fatores de **engajamento** como os mais relevantes. No XGBoost, as variáveis de **uso do cartão** lideraram a importância (ex.: quantidade de transações, volume gasto, porcentagem de limite usado). Isso reflete que **menos uso = maior risco**. Variáveis de **inatividade** e **atendimento (SAC)** também figuram entre as top influências no churn, alinhadas com a intuição de negócio.

#### Slide 5: Segmentação de Clientes (Clustering)

- **Por que segmentar?** Além de prever churn individualmente, é útil entender **grupos de clientes** para ações coletivas. Usamos **K-Means** (k=3) nos clientes ativos para encontrar clusters naturais de comportamento.

- **3 Clusters Identificados:**

- **Cluster 1: VIP Engajados** (~40% da base). Clientes de **alto valor** – maiores limites, gastos e transações. Muito ativos (quase sem inatividade) e satisfeitos (raros contatos SAC). **Churn praticamente zero** aqui. *Estratégia:* manter e premiar (programas VIP, upgrades).

- **Cluster 2: Moderados Estáveis** (~35%). Uso e engajamento médios, perfil conservador (pagam em dia, pouco rotativo), alguns produtos. Churn baixo. *Estratégia:* upsell e incentivos para aumentar uso e valor (migrar alguns para VIP).

- **Cluster 3: Desengajados de Risco** (~25%). **Baixo uso** do cartão (poucas transações, gastos baixos), muitos **meses inativos**, porém **muitos contatos SAC** (insatisfação). Maioria dos churners vêm daqui – churn alto no grupo. *Estratégia:* ações imediatas de retenção (campanhas de reengajamento, melhoria de serviço). Focar principalmente nos que ainda têm bom potencial de valor e recuperar confiança desses clientes.

- **Aplicabilidade:** A segmentação revelou **quem são os churners** (concentrados no cluster 3) e permitiu planejar **ações direcionadas** em vez de genéricas. Ex.: Em vez de abordar 10k clientes, foque nos ~2.500 do cluster de risco com comunicações específicas. Para VIPs, estratégia diferente (fidelização); moderados, estímulo de uso. Isso eleva a efetividade das iniciativas de marketing e relacionamento.

#### Slide 6: Cenários Simulados

- **Objetivo:** Quantificar impacto financeiro de agir ou não agir sobre churn, guiando decisões. Elaboramos cenários comparativos:

- **Status Quo (Sem ação):** ~16% churn anual = ~1.600 clientes perdidos ≈ grande perda de receita futura (LTV somado desses clientes). Este é o custo de não fazer nada.
- **Ação Focada (High Value/Risk):** Focar retenção nos **Top 200 clientes em risco** (maior LTV). Se retivermos 50% deles, churn cai p/ ~14% (~1.400 saídas). Salvamos, p.ex., \ \$X00 mil em valor futuro a um custo de campanha baixo (~\ \$20 mil), **ROI altamente positivo** (5x+) <sup>67</sup>. *Insight:* Priorizar poucos clientes valiosos traz ótimo retorno.
- **Ação Ampla (Toda Base):** Investir em **melhorias gerais** (serviço, benefícios) que reduzam churn em toda a base – ex.: baixar churn do cluster de risco em 5 p.p. Isso também leva churn global ~14%. Retém valor significativo, porém custos maiores e ROI mais diluído. Benefício: melhora satisfação de muitos clientes, não só os tops.
- **Cenário Adverso:** E se churn aumentar para **20%** (crise, concorrência forte)? Teríamos ~2.000 saídas/ano – +400 vs atual. Simulação mostra perdas adicionais de \ \$X mi. Destaca que, em tempos difíceis, **investir em retenção é ainda mais crítico** para segurar a base.
- **Conclusão dos Cenários:** A comparação evidenciou que **agir compensa**. O cenário focado em alto valor mostrou o melhor **custo-benefício** – deve ser primeira medida (segurar os “big spenders” em risco). A ação ampla, se houver orçamento, complementa, melhorando a saúde da base no geral. E os cenários ajudaram a sensibilizar executivos: não agir custa caro, e num cenário pior seria desastroso. Assim, os números reforçam as recomendações de investimento em programas de retenção.

## Slide 7: Recomendações Finais & Impacto Estratégico

- **Programa de Retenção Proativo:** Implantar medidas preventivas atacando os fatores de churn. Ex.: **Monitorar uso** – quando cliente ficar 2 meses sem usar, acionar oferta de reativação (bônus, descontos) antes que ele se desligue <sup>77</sup>. **Monitorar reclamações** – clientes com muitas interações de SAC devem receber contato de um gerente com solução personalizada (mostrando que o banco se importa) <sup>20</sup>. Antecipar-se ao churn melhora muito a retenção.
- **Estrategiar por Segmento:** *One size fits all* não funciona. Implementar **planos de ação diferenciados**:
  - **VIPs:** manter e engajar ainda mais (benefícios exclusivos, tratamento premium) – são clientes fiéis e altamente lucrativos <sup>79</sup>.
  - **Moderados:** oportunidade de crescimento – oferecer incentivos (upgrade de limite, combos de produtos, cashback) para aumentar uso e torná-los VIPs <sup>82</sup> <sup>81</sup>.
  - **Desengajados:** foco total em **mitigar churn** – campanha de reconquista, melhoria no atendimento, ofertas para trazê-los de volta a atividade <sup>83</sup>. Porém, **focar nos que valem a pena** (analisar LTV): alguns de baixíssimo valor talvez não justifiquem esforço intenso <sup>44</sup>.
- **Foco nos “High Value, High Risk”:** Criar uma lista prioritária de clientes “VIP em risco” a ser acompanhada de perto mensalmente <sup>84</sup>. Esses clientes devem ser tratados quase individualmente – ex.: ligação do gerente, oferta customizada – pois perder um deles tem impacto grande. Por outro lado, para clientes de baixo valor que estão em risco, usar soluções de baixo custo (e-mail, SMS) e aceitar que alguns churns aqui não prejudicam tanto.
- **Melhorias de Produto/Serviço:** Resolver pontos cegos revelados pelos dados: clientes premium (alta renda) com churn acima do esperado indicam necessidade de **melhorar oferta para segmento alta renda** <sup>18</sup> – considerar novos produtos (cartão Black/Platinum+ com benefícios melhores) ou condições diferenciadas para não perdê-los para concorrência. Também, churn alto associado a muitos contatos SAC indica **problemas no atendimento** – investir em treinamento, reduzir burocracia, agilizar resoluções vai diretamente diminuir churn por insatisfação <sup>87</sup>.
- **Integração e Monitoramento:** Colocar o modelo de churn para rodar regularmente no ambiente do banco. Cada mês, gerar um **alerta dos clientes com maior probabilidade de churn** <sup>88</sup> e compartilhar com equipes de relacionamento/marketing. Acompanhar métricas de retenção (churn mensal, LTV médio, sucesso das ofertas de retenção) e **fechar o ciclo**: usar o feedback para melhorar continuamente tanto o modelo quanto as estratégias. Recomendamos destinar um **budget fixo para retenção** e avaliar ROI das campanhas periodicamente <sup>90</sup> – assim, a direção enxerga claramente o benefício financeiro das ações tomadas (p. ex., “salvamos \ \$X retendo clientes, com custo \ \$Y”). Com o

modelo e sistema implementados, o banco passa a agir de forma **mais eficaz e informada** para reter seus clientes valiosos, transformando inteligência de dados em resultados de negócio.

#### Slide 8: Próximos Passos & Demo (App)

- **Implementação Técnica:** O código do modelo, engenharia de dados e notebooks de análise estão organizados no repositório (pipeline reproduzível). Próximo passo é integrar ao **CRM do banco** ou sistemas internos para automatizar o uso nas bases de clientes reais.

- **Aplicação Web (Protótipo):** Desenvolvemos um web app simples em **Streamlit** para demonstrar a solução. Nele, é possível inserir os dados de um cliente e obter em tempo real a **probabilidade de churn** prevista pelo modelo, além de indicar em que cluster e faixa de LTV ele se enquadra. Esse *dashboard* serve como **prova de conceito** de como a área de negócios poderia consumir o modelo (por exemplo, por um gerente verificar o risco de churn de um cliente e já acionar uma oferta de retenção).

- **Uso do App:** A aplicação é local/privada (não publicada online por confidencialidade). Para executá-la internamente, basta rodar: `streamlit run webapp/app.py` (ambiente Python configurado) <sup>93</sup>. A interface permite testar vários cenários de clientes e visualizar as predições.

- **Encerramento:** Com essa entrega, o projeto alcança seu objetivo de oferecer ao Banco Montes Claros uma ferramenta poderosa de **Inteligência Artificial aplicada ao negócio**, com impacto direto em **redução de churn e aumento de valor**. Esperamos que as recomendações e sistemas desenvolvidos sejam úteis para orientar a estratégia de retenção do banco nos próximos anos. Estamos à disposição para perguntas e para discutir os próximos passos de implementação!

---

1 2 Carolina Fernandes: Cliente bom é cliente que volta | Exame

<https://exame.com/bussola/carolina-fernandes-cliente-bom-e-cliente-que-volta/>

3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 22 23 24 25 26 27 28 29 30 31 32  
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61  
62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90

91 92 Projeto de Previsão de Churn – Banco Montes Claros.pdf

<file:///file-9LBGFDj8GB7STSPHYGgEw>

21 93 README.md

<file:///file-EbM4MrqWPKkSvrQ6gTQAts>