


## Challenge SPRINT 3 – DATA GOVERNANCE – RODRIGO DIAS.

### 1. IDENTIFICAÇÃO

 Time Scrum

*Papel*                      *Integrantes*

Product Owner   **Paulo Vinicius Medeiros RM:559768**

Scrum Master     **Lucas Guedes Gianini RM:560454**

Devs/Analistas   **Priscila Bueno RM:561142**  
**Mabel Mayer RM:561145**  
**Kauan Guimaraes RM: 559472**

<b>Nome do Projeto</b>	Projeto DeepChurn – Inteligência Preditiva para Retenção
<b>Nome do Grupo</b>	Axistech
<b>Categoria da Demanda</b>	Desenvolvimento MVP Inicial
<b>Data</b>	16/04/2025

## 2. PEDIDO INICIAL e MUDANÇA DE RETAIL MEDIA PARA CHURN.

### Mudança de Escopo

Durante as etapas iniciais de planejamento e levantamento de dados, a equipe identificou desafios relevantes na obtenção de datasets públicos consistentes e na viabilidade de entrega de resultados relevantes dentro do tempo e recursos disponíveis. Diante disso, optou-se por uma mudança de escopo estratégica, direcionando o foco do projeto para a predição de churn — uma temática que mantém a aplicabilidade prática, permite a utilização de dados públicos confiáveis e se alinha fortemente com demandas reais de mercado.

### Novo Foco: Predição de Churn

- O projeto passou então a concentrar esforços no desenvolvimento de um MVP voltado à identificação preditiva de clientes com alta probabilidade de cancelamento (churn), utilizando algoritmos de aprendizado de máquina sobre dados reais. A nova proposta permitiu uma abordagem mais robusta em termos analíticos, com entregas viáveis e resultados tangíveis.

### Objetivo do MVP

- Desenvolver um modelo preditivo com boa acurácia para identificação de churn.
- Construir um dashboard simples com visualização dos resultados.
- Demonstrar aplicabilidade do modelo para áreas de negócio como marketing e relacionamento.


### Justificativa

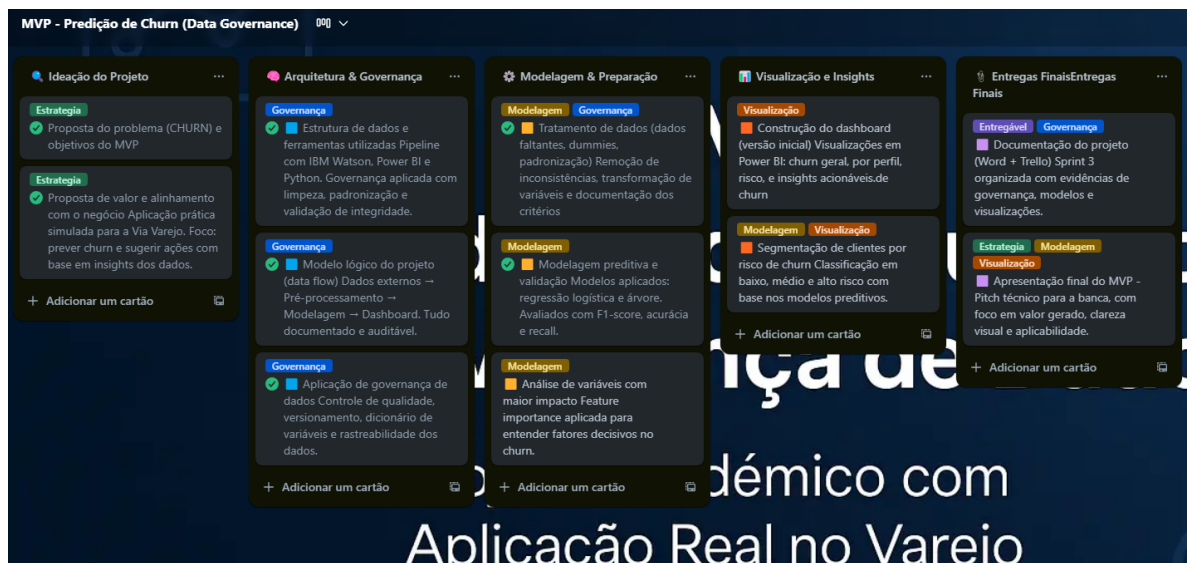
O churn representa uma das maiores preocupações de empresas que operam com base em recorrência. Antecipar cancelamentos pode **aumentar a receita, reduzir o CAC** (custo de aquisição de cliente) e melhorar o relacionamento com o cliente. A escolha deste tema se deu pela **viabilidade técnica, acesso a bases públicas e relevância prática para o mercado**.

### 3. PLANO DE PROJETO SCRUM

#### Ferramenta de Gestão

O projeto foi conduzido com base em metodologia ágil, utilizando o Trello como ferramenta de planejamento e controle. Todas as sprints foram organizadas em listas, com descrições detalhadas, checklists e entregas por etapa.

-  **Link do board no Trello:**  
<https://trello.com/invite/b/67eb50eaae3adc1b6fc77f67/ATTIfafa2365c8446eeaae6b158d3c11b6c71BF0B8A8/mvp-predicao-de-churn-data-governance>



- O Trello evidencia não apenas o fluxo de trabalho, mas também a aplicação dos conceitos de Governança de Dados aprendidos ao longo da disciplina. É possível visualizar a conexão entre arquitetura, modelos, métricas, limpeza e visualização de dados, em total conformidade com as boas práticas de projeto.

#### Comunicação da Equipe

- Diálogos e decisões realizadas por grupo no **WhatsApp**
- Reuniões pontuais de alinhamento via Discord e Teams
- Orientações dos professores incorporadas a cada sprint

## Sprints Definidas

Sprint	Entregas Principais
<b>Sprint 1</b>	Definição do escopo, identificação da dor, montagem de apresentação PowerPoint, apresentação à IBM
<b>Sprint 2</b>	Definição da arquitetura de dados, escolha dos modelos e ferramentas
<b>Sprint 3</b>	Coleta dos dados, tratamento, avaliação de acurácia, demonstrações parciais por disciplina
<b>Sprint 4</b>	Treinamento final do modelo, construção do dashboard, documentação e apresentação do MVP

## 4. BACKLOG DO PRODUTO

### 4.1 Épicos


<b>1. Definição e Preparação do Problema</b> Compreensão do conceito de churn, pesquisa de datasets, definição da dor do negócio e preparação do PowerPoint para apresentação.
<b>2. Arquitetura e Ferramentas de Projeto</b> Escolha das ferramentas de machine learning, definição do fluxo de dados, arquitetura do modelo e preparação da base.
<b>3. Modelagem e Avaliação de Dados</b> Coleta e tratamento dos dados, aplicação do modelo preditivo, avaliação de desempenho e entrega parcial para disciplinas específicas.
<b>4. Construção e Finalização do MVP</b> Treinamento final do modelo, construção do dashboard com insights acionáveis, documentação e apresentação final.

### 4.2 Funcionalidades / Features

- Selecionar e validar dataset público com registros de churn
- Realizar tratamento completo dos dados (limpeza, transformação, balanceamento)
- Implementar modelo de machine learning supervisionado (árvore de decisão)

- Avaliar o desempenho com métricas (F1-score, accuracy, recall)
- Identificar variáveis que mais influenciam o churn
- Desenvolver dashboard para visualização de clientes em risco
- Gerar documentação técnica e de apresentação

### 4.3 Itens do Backlog do Produto

 *Itens bem detalhados, com foco na execução real do projeto de churn.*

ID	Item	Descrição	Épico Relacionado	Prioridade
01	Selecionar dataset público de churn	Buscar e validar uma base de dados pública, com registros de clientes e variável “churn” (binária).	2. Arquitetura e Ferramentas	Alta
02	Definir critérios de uso da base e avaliar qualidade	Avaliar se os dados possuem volume, consistência e colunas relevantes para predição.	2. Arquitetura e Ferramentas	Alta
03	Criar apresentação executiva do projeto para a IBM	Consolidar escopo, problema e solução proposta no PowerPoint institucional.	1. Definição e Preparação	Média
04	Definir arquitetura do projeto de dados	Montar o fluxo técnico da solução: entrada, pré-processamento, modelagem, saída e visualização.	2. Arquitetura e Ferramentas	Alta
05	Escolher e configurar ferramentas para o desenvolvimento	Escolher IDEs, bibliotecas, ambientes e ferramentas (Python, pandas, scikit-learn, etc.).	2. Arquitetura e Ferramentas	Média

<b>06</b>	Realizar tratamento de dados	Executar limpeza (missing), transformação (dummies), normalização e balanceamento de classes.	2. Arquitetura e Ferramentas	Alta
<b>07</b>	Aplicar modelo supervisionado de churn	Treinar modelo preditivo (regressão logística, árvore de decisão) com dados tratados.	3. Modelagem e Avaliação	Alta
<b>08</b>	Avaliar desempenho do modelo	Calcular métricas: F1-score, accuracy, recall, precision. Identificar necessidade de ajustes.	3. Modelagem e Avaliação	Média
<b>09</b>	Identificar variáveis mais relevantes para o churn	Usar feature importance, SHAP ou análise estatística para destacar fatores com maior peso na predição.	3. Modelagem e Avaliação	Média
<b>10</b>	Criar dashboard com clientes em risco	Visualizar resultados: churn previsto, categorias de risco e variáveis mais impactantes.	4. Construção e Finalização	Alta
<b>11</b>	Preparar documentação final do projeto	Relatar metodologia, decisões, código-fonte e resultados em PDF para entrega final.	4. Construção e Finalização	Média
<b>12</b>	Apresentar MVP com storytelling e visão executiva	Apresentar o modelo, resultados e impacto do projeto para professores e stakeholders.	4. Construção e Finalização	Alta

## 5. HISTÓRIAS DE USUÁRIO E CRITÉRIOS DE ACEITE

---

### ◆ Req-0010 – Seleção de Dados

**História:**

Acessar uma base pública com registros de clientes e churn, para alimentar o modelo preditivo.

**Critérios de Aceite:**

- A base deve conter ao menos 200 registros reais
  - A variável alvo “churn” deve estar claramente definida (0 = ativo, 1 = cancelado)
  - A estrutura dos dados deve ser compatível com aplicação de machine learning
- 

◆ **Req-0020 – Preparação da Base de Dados**

**História:**

preparar os dados com limpeza, normalização e codificação, para que o modelo seja treinado com dados de qualidade.

**Critérios de Aceite:**

- Dados nulos tratados ou eliminados
  - Variáveis categóricas transformadas em variáveis numéricas
  - Base final balanceada para reduzir viés do modelo
- 

◆ **Req-0030 – Modelagem Preditiva de Churn**

**História:**

Treinar o modelo supervisionado para prever churn, para que eu possa identificar clientes em risco com base no histórico.

**Critérios de Aceite:**

- Modelo treinado com separação treino/teste (ex: 70/30)
  - Algoritmo implementado
  - Resultados devem ser consistentes e replicáveis
-

#### ◆ Req-0040 – Avaliação do Modelo

##### **História:**

Avaliar a performance da predição, para garantir que os resultados sejam confiáveis.

##### **Critérios de Aceite:**

- Métricas calculadas: F1-score, accuracy, recall, precision
  - F1-score mínimo aceitável: 70%
  - Gráfico de matriz de confusão gerado e analisado
- 

#### ◆ Req-0050 – Interpretação das Variáveis Relevantes

##### **História:**

Saber quais características influenciam mais o churn, para direcionar campanhas de retenção.

##### **Critérios de Aceite:**

- Top 5 variáveis mais relevantes identificadas
  - Resultados apresentados com base em SHAP, coeficientes ou feature importance.
  - As variáveis devem ser compreensíveis para áreas não técnicas
- 

#### ◆ Req-0060 – Visualização de Resultados

##### **História:**

Visualizar os clientes com maior risco de churn e os motivos, para agir de forma antecipada.

##### **Critérios de Aceite:**

- Dashboard apresenta classificação de risco por cliente (alto, médio, baixo)
- Gráfico com principais variáveis exibido
- Interface simples e navegável



---

### ◆ Req-0070 – Entrega Final e Apresentação

#### História:

Entregar o MVP documentado e apresentá-lo com clareza, para mostrar o impacto e viabilidade da solução.

#### Critérios de Aceite:

- Documento em PDF com resumo técnico e de negócio entregue
- Apresentação oral realizada com o grupo
- MVP funcional e visual apresentado para banca

## 6. PONTUAÇÃO DO ESFORÇO (STORY POINTS)

Código	Item	Story Points
Req-0010	Selecionar dataset público de churn	2
Req-0020	Preparar base de dados (limpeza, dummies, balanceamento)	5
Req-0030	Treinar modelo de churn	5
Req-0040	Avaliar desempenho do modelo	3
Req-0050	Interpretar variáveis mais relevantes	3
Req-0060	Criar dashboard de visualização	5
Req-0070	Documentar o projeto e apresentar o MVP	2
◆ Total estimado de esforço: 25 Story Points		

## ⚙️ Requisitos Não Funcionais

<b>Categoria</b>	<b>Requisito</b>
Usabilidade	O dashboard deve ser acessível em tela cheia e com navegação intuitiva.
Desempenho	O tempo de execução do modelo deve ser inferior a 10 segundos.
Manutenibilidade	O código e as etapas do projeto devem ser organizados e documentados.
Portabilidade	O projeto deve funcionar em ambiente local (Jupyter, VSCode).
Segurança	Nenhum dado sensível será utilizado; todas as bases serão públicas.

#### **Contingência/Exceções**

- **Se o modelo não atingir a acurácia mínima de 70%, será feita reavaliação da preparação dos dados e ajuste de hiperparâmetros.**
- **Caso algum membro da equipe fique indisponível, as tarefas serão redistribuídas pelo grupo para garantir entrega no prazo.**
- **Problemas técnicos (ex: instabilidade de internet, falha no notebook) serão minimizados com backups locais regulares dos arquivos e planilhas.**

## **7. PLANO DE RELEASE**

	Nome	Descrição	Justificativa
<b>01</b>	Integração com APIs externas	Conectar com sistemas de CRM reais, plataformas de e-mail, etc.	Complexidade técnica alta e risco elevado
<b>02</b>	Autenticação de usuários	Criar sistema de login ou autenticação no MVP	Não é foco do projeto (modelo preditivo de churn)
<b>03</b>	Interface Web interativa	Desenvolver plataforma web funcional com filtros e navegação	Exige stack adicional que foge ao objetivo do MVP

Sprint	Entregas Principais	Data
<b>Sprint 1</b>	Definir escopo do projeto, identificar a dor do negócio, montar PowerPoint e apresentar proposta à IBM	30/10/2024
<b>Sprint 2</b>	Definir arquitetura de dados, escolher ferramentas e modelos, validar o fluxo técnico com professores	25/11/2024
<b>Sprint 3</b>	Coletar dados, realizar tratamento, aplicar modelo, calcular métricas e apresentar progresso por disciplina	16/04/2025
<b>Sprint 4</b>	Treinar versão final do modelo, criar dashboard visual, finalizar documentação e realizar apresentação do MVP	(a definir)

## 8. ITENS FORA DE ESCOPO

## 9. GLOSSÁRIO DE TERMOS

Termo	Definição
-------	-----------

<b>Churn</b>	Indicador de cancelamento de clientes. Alta taxa de churn impacta diretamente receita e LTV.
<b>Retenção de Clientes</b>	Estratégia para manter clientes ativos, reduzindo custos de aquisição e aumentando lucratividade.
<b>MVP (Produto Mínimo Viável)</b>	Primeira versão funcional de uma solução, com foco em testar viabilidade e entregar valor inicial.
<b>Modelo Preditivo</b>	Ferramenta analítica que antecipa eventos futuros com base em dados históricos.
<b>Insight Acionável</b>	Informação clara e objetiva que pode gerar ação imediata no negócio.
<b>Base Preditiva de Decisão</b>	Fundamento técnico que dá suporte a decisões estratégicas com dados e estatísticas.
<b>Acurácia de Modelo</b>	Grau de assertividade da previsão em relação à realidade. Essencial para gerar confiança nos dados.
<b>Segmentação de Risco</b>	Classificação de clientes por potencial de cancelamento (alto, médio, baixo), para ações específicas.
<b>Dashboard Executivo</b>	Painel visual que resume os dados mais relevantes de forma acessível para gestores.

## 10. Resumo Consolidado dos Feedbacks da Sprint 1

### Pontos Positivos

- A ideia do projeto foi elogiada por ser interessante, objetiva e sensacional.
- O objetivo do projeto está claro, com uma proposta bem alinhada ao escopo.
- O grupo demonstrou boa pesquisa, dedicação e organização.
- A entrega foi destacada como excelente por vários professores.
- A abordagem prática e os temas escolhidos (como fidelização) foram vistos como relevantes para o mercado.

---

### Pontos de Melhoria

#### **Escopo e viabilidade:**

- O projeto, apesar de promissor, pode estar ambicioso para o 1º ano.
- É essencial manter os objetivos realistas e possíveis de entregar até a Sprint 4.
- Evitar incluir elementos que não são factíveis, como integrações complexas que exigem conhecimento em desenvolvimento web (a menos que a equipe domine isso).

**Apresentação visual:**

- Cuidado com a cor da fonte (evitar texto esmaecido).
- Consistência visual entre os slides (negrito, tamanhos, estilos).
- Slides com muito texto e fontes pequenas dificultam a leitura — mais objetividade visual é recomendada.
- Erros de ortografia e repetições devem ser corrigidos.

**Organização da apresentação:**

- Certos slides estavam fora de contexto (ex: "Investimento em Retail Media...").
- As fases do projeto precisam seguir as sprints definidas no challenge para refletir corretamente o cronograma.

**Fundamentação e fontes:**

- Todos os dados e estudos devem conter fontes claras, especialmente os relacionados a ROI, investimento e impacto.
- A análise de concorrentes deve ir além dos nomes e incluir funcionalidades comparativas.

**Detalhamento técnico:**

- O treinamento do modelo KNN precisa estar bem definido.
- As recomendações feitas pelo sistema devem ser tangíveis e bem explicadas — como elas serão apresentadas?
- Se o grupo quiser seguir com alguma interface web, estudar Flask (microframework Python) pode ser um bom caminho, conforme sugerido.

**Recomendações Específicas**

- Informar quem é o responsável pelo projeto dentro do grupo.
- Focar em métricas de fidelização em vez de ticket médio, que é mais difícil de influenciar.
- Avaliar a viabilidade econômica do projeto: custo de desenvolvimento, manutenção, precificação, retorno esperado.
- Considerar o uso de ferramentas como Flask, caso optem por uma aplicação web simples.

## ***Resumo Consolidado dos Feedbacks da Sprint 2***

### **Pontos Positivos**

- A ideia do projeto continua sendo elogiada: clara, promissora e bem estruturada.
- Professores destacaram a dedicação da equipe e o cuidado com o protótipo e arquitetura apresentados.
- A proposta de utilizar storytelling na apresentação foi bem recebida.
- O uso de ROI foi considerado válido após explicação, mostrando que o grupo conseguiu sustentar bem os argumentos.
- O protótipo e a arquitetura foram bem avaliados por estarem caminhando na direção certa.

### **Pontos de Melhoria**

#### **1. Arquitetura e Fluxo de Dados**

- A arquitetura precisa de mais detalhamento técnico, explicando claramente:
- O papel de cada ferramenta (ex: MySQL, Kafka, JSON etc.).
- Justificativa das escolhas tecnológicas.
- Fluxo completo dos dados: da ingestão até o consumo final.
- Incluir um diagrama mais robusto, ilustrando como os dados fluem entre os componentes.

#### **2. Prototipação**

- Descrever detalhadamente as funcionalidades de cada parte do protótipo.
- Explicar como cada protótipo se integra com o sistema completo.
- Criar um storyboard da experiência do usuário.
- Sugestão: avaliar uso de ferramentas como Streamlit ou Dash para prototipação.

#### **3. Análise Exploratória de Dados (EDA)**

- Detalhar melhor as métricas utilizadas (ex: média, desvio padrão).
- Focar em insights acionáveis e segmentações relevantes (ex: categorias de vendas).
- Utilizar bibliotecas como Pandas Profiling ou Sweetviz para gerar relatórios visuais automáticos.
- Indicar claramente quais ferramentas foram usadas nos gráficos (Python, Power BI, etc.).

#### **4. Funcionalidades**

- Tornar as funcionalidades mais objetivas e específicas.
- Frases genéricas como "base rica e acionável" devem ser substituídas por descrições práticas de como os dados serão obtidos.
- Separar claramente as etapas de Data Engineering (estruturação) das de Data Science (análise/modelagem).
- Criar dashboards separados para públicos diferentes, como recomendado para visualizações como ROI.

## 5. Concorrência

- A análise de concorrentes precisa ser mais comparativa, destacando:
- Prós e contras de cada um.
- O que diferencia o projeto de vocês em relação à concorrência.
- Evitar apenas descrever — trazer funcionalidades comparadas (ex: em forma de tabela).

## 6. Implementação do Modelo

- Esclarecer se o grupo conseguirá realmente entregar KNN com Scikit-learn.
- Para a apresentação, é melhor pré-treinar o modelo com bases públicas grandes e depois ir refinando com dados reais.
- Mostrar como o modelo será utilizado e reaproveitado ao longo do tempo.

## 7. Engajamento do Usuário

- Escanear o QR Code pode ser um desafio — sugeriu-se incluir alguma recompensa para o cliente, como moedas, descontos ou brindes, para aumentar o engajamento.

## COMENTÁRIOS E OBSERVAÇÕES FINAIS

- O projeto de predição de churn foi construído com foco em **responder a um problema real de negócios**: a evasão de clientes e seus impactos sobre receita e retenção.
- O grupo optou por uma abordagem pragmática, priorizando **dados públicos acessíveis, métodos supervisionados robustos** e entregáveis que gerassem **valor visual e estratégico**.
- Cada sprint foi conduzida com entregas concretas, desde a estruturação da dor até a visualização dos resultados com potencial de aplicação prática em empresas reais.

- O grupo se organizou de forma colaborativa, utilizando ferramentas simples como **Trello, Discord e WhatsApp**, mas mantendo um **nível de responsabilidade elevado sobre prazos, tarefas e entregas**.
- Durante a execução, desafios como o balanceamento de dados e a seleção de variáveis foram superados com estudo e orientação, fortalecendo o aprendizado técnico e de negócios.
- O projeto segue em evolução para a Sprint 4, com foco em **refinamento da solução, storytelling executivo na apresentação final e validação completa do MVP**.