



# Os PANTufas BANCO PAN

## Controle do Documento

### Histórico de revisões

Data	Autor	Versão	Resumo da atividade
<02/08/2022>	<Vitor Zeferino>	<1.1>	<Nome grupo/ empresa e introdução>
09/08/2022	<Gustavo Ferreira>	<1.2>	<Entregáveis artefato 1>
<11/08/2022>	<Vitor Zeferino>	<1.3>	<Atualização dos Entregáveis artefato 1>
<11/08/2022>	<Gustavo Ferreira>	<1.4>	<Entregáveis artefato 2>
<12/08/2022>	<Gustavo Ferreira>	<1.5>	<Atualização dos entregáveis artefato 2>
<17/08/2022>	<Vitor Zeferino>	<2.1>	<Revisão (Persona e Jornada)>

# Sumário

<b>1. Introdução</b>	<b>4</b>
<b>2. Objetivos e Justificativa</b>	<b>5</b>
2.1. Objetivos	5
2.2. Proposta de Solução	5
2.3. Justificativa	5
<b>3. Metodologia</b>	<b>6</b>
3.1. CRISP-DM	6
3.2. Ferramentas	6
3.3. Principais técnicas empregadas	6
<b>4. Desenvolvimento e Resultados</b>	<b>7</b>
4.1. Compreensão do Problema	7
4.1.1. Contexto da indústria	7
4.1.2. Análise SWOT	7
4.1.3. Planejamento Geral da Solução	7
4.1.4. Value Proposition Canvas	7
4.1.5. Matriz de Riscos	7
4.1.6. Personas	8
4.1.7. Jornadas do Usuário	8
4.2. Compreensão dos Dados	9
4.3. Preparação dos Dados	10
4.4. Modelagem	11
4.5. Avaliação	12
<b>5. Conclusões e Recomendações</b>	<b>13</b>
<b>6. Referências</b>	<b>14</b>
<b>Anexos</b>	<b>15</b>

# 1. Introdução

Nosso parceiro de negócio neste terceiro módulo é o banco Pan, controlado apenas pelo banco BTG Pactual, que atua no mercado com seu foco em clientes de classe C, D e E. Uma empresa que oferece diversos produtos aos seus clientes, como investimento, créditos, saúde, entre outros. Seu objetivo é dar um novo olhar para os desafios de todos que querem vencer os obstáculos da vida, liderando a inclusão financeira digital das famílias brasileiras.

O problema é que, atualmente, o banco Pan não possui um atendimento personalizado para os possíveis propósitos de seus clientes. Se ele está buscando um problema de relacionamento com o banco, contratar novos produtos ou ser um novo cliente.

## 2. Objetivos e Justificativa

### 2.1. Objetivos

Realizar um atendimento personalizado com os possíveis propósitos dos clientes para assim aprimorar o atendimento e trazer maior satisfação no relacionamento.

### 2.2. Proposta de Solução

Descreva resumidamente sua proposta de modelo preditivo e como esse modelo pretende resolver o problema, atendendo os objetivos

### 2.3. Justificativa

Faça uma breve defesa de sua proposta de solução, escreva sobre seus potenciais, seus benefícios e como ela se diferencia.

## 3. Metodologia

Descreva as etapas metodológicas que foram utilizadas para o desenvolvimento, citando o referencial teórico. Você deve apenas enunciar os métodos, sem dizer ainda como ele foi aplicado e quais resultados obtidos.

### 3.1. CRISP-DM

Descreva brevemente a metodologia CRISP-DM e suas etapas de processo

### 3.2. Ferramentas

Descreva brevemente as ferramentas utilizadas e seus papéis (Google Colaboratory)

### 3.3. Principais técnicas empregadas

Descreva brevemente as principais técnicas empregadas, algoritmos e seus benefícios

## 4. Desenvolvimento e Resultados

De maneira geral, você deve descrever nesta seção a aplicação dos métodos aprendidos e os resultados obtidos por seu grupo em seu projeto

### 4.1. Compreensão do Problema

#### 4.1.1. Contexto da indústria

##### **Principais players:**

Dentre os principais players do mercado, atualmente temos o Itaú - Itaú, Caixa econômica, Bradesco e Santander, que mais se assemelham com os objetivos do Banco Pan.

##### **Modelo de negócio:**

Atualmente o Banco Pan possui os modelos de negócio B2B & B2C, com foco nos clientes C, D e E, através da inclusão financeira digital das famílias brasileiras, além de auxiliar na geração de renda, fornecimento e financiamento de produtos.

##### **Tendências:**

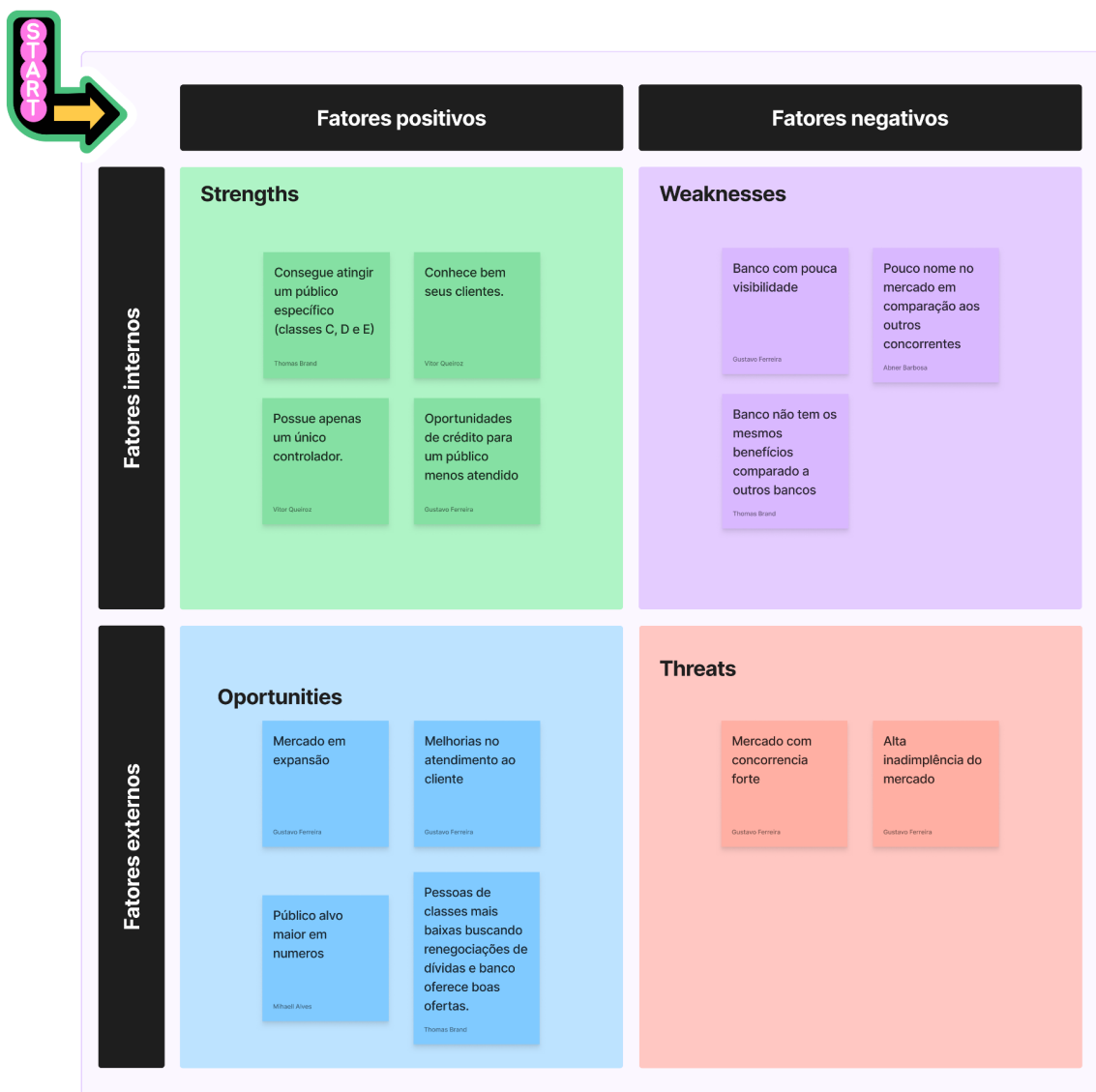
Dentre as tendências de mercado identificadas para o Banco Pan, existe o aumento da tecnologia a ser utilizada pelo banco, atendimento personalizado para cada cliente, engajamento dos clientes e melhora no relacionamento com o cliente.

## 4.1.2. Análise SWOT

Posicione aqui sua análise SWOT

Figma:

<https://www.figma.com/file/49ZRCpjaWUtK9oVL4pK9U/Analise-SWOT?node-id=0%3A1>



## 4.1.3. Planejamento Geral da Solução

a) quais os dados disponíveis

Como dados possuímos algumas informações da área de atendimento e informações de conta, disponibilizadas e explicadas pelo parceiro durante o primeiro encontro, através deles podemos obter algumas informações que nos possibilitam ter uma prévia de como os clientes do parceiros atuam dentro do banco. Para tanto foi nos disponibilizado uma base de dados com as seguintes informações: cpf, quantidade de produtos que o cliente possui contratado, tempo de relacionamento com cliente desde o



primeiro produto, quantidade de atendimentos realizados, quantidade de reclamações abertas, Indica quantidade de atendimentos que o cliente obteve fora do prazo, soma do saldo de todos os produtos de crédito contratados como cliente, classificação do grau de risco de crédito para clientes, soma do saldo de todos os produtos de crédito contratados em outras instituições financeiras, quantidade de produtos que o cliente possui contratado em outras instituições financeiras, Indica se o cliente possui algum registro de negativação no mercado, classificação do grau de risco de crédito no mercado, valor de renda presumido do cliente, Indicador de cliente atritado, indicador de cliente engajado e indicador de potencial novo cliente, através dessas informações espera-se que seja possível analisar previamente a demanda que o usuário possui ao entrar em contato com um atendente que irá atender da melhor forma possível.

b) qual a solução proposta

Criar uma inteligência artificial para otimizar o atendimento para o cliente. A solução forneceria mais informações sobre os clientes e assim o funcionário que está em contato com o cliente teria uma base muito maior para satisfazê-lo, sendo na sugestão de um produto ou até para tirar dúvidas.

c) qual o tipo de tarefa (regressão ou classificação)

O tipo de tarefa é classificação, pois queremos prever em qual categoria o cliente se encaixa com base em dados não observados.

d) como a solução proposta deverá ser utilizada

No momento que o cliente realizar uma ligação ou outra forma de contatar o banco, a IA irá fazer a análise de acordo com os dados que estiverem ligados ao cliente, ajudando em um atendimento de forma mais plausível para ele.

e) quais os benefícios trazidos pela solução proposta

Os benefícios trazidos pela proposta serão a melhor satisfação do cliente e a motivação para, possivelmente, adquirir novos produtos do Banco Pan, além dos já utilizados.

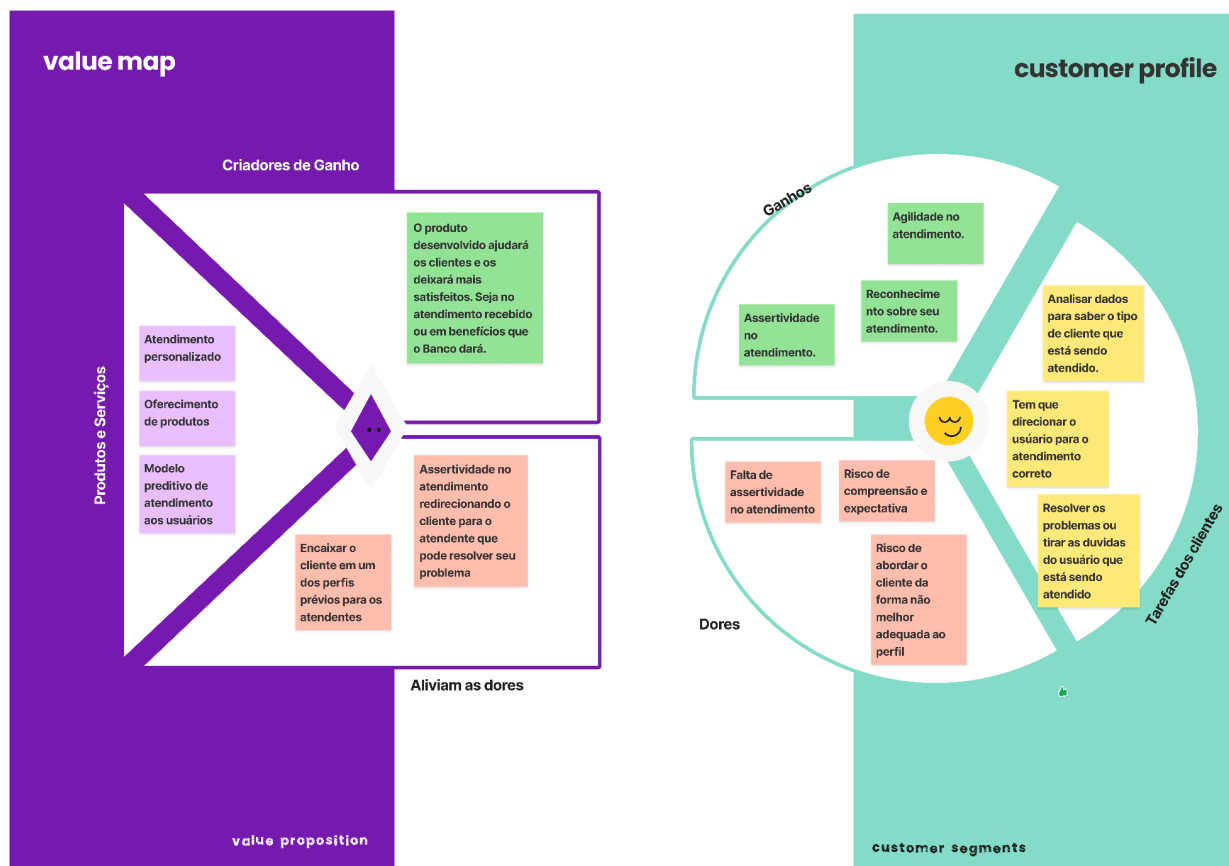
f) qual será o critério de sucesso e qual medida será utilizada para o avaliar

O critério de sucesso é se realmente o algoritmo está funcionando. Quando o cliente ligar e for direcionado para uma das classificações; atritado ou em busca de novos produtos e realmente for isso, significa que o projeto foi um sucesso.

## 4.1.4. Value Proposition Canvas

Figma:

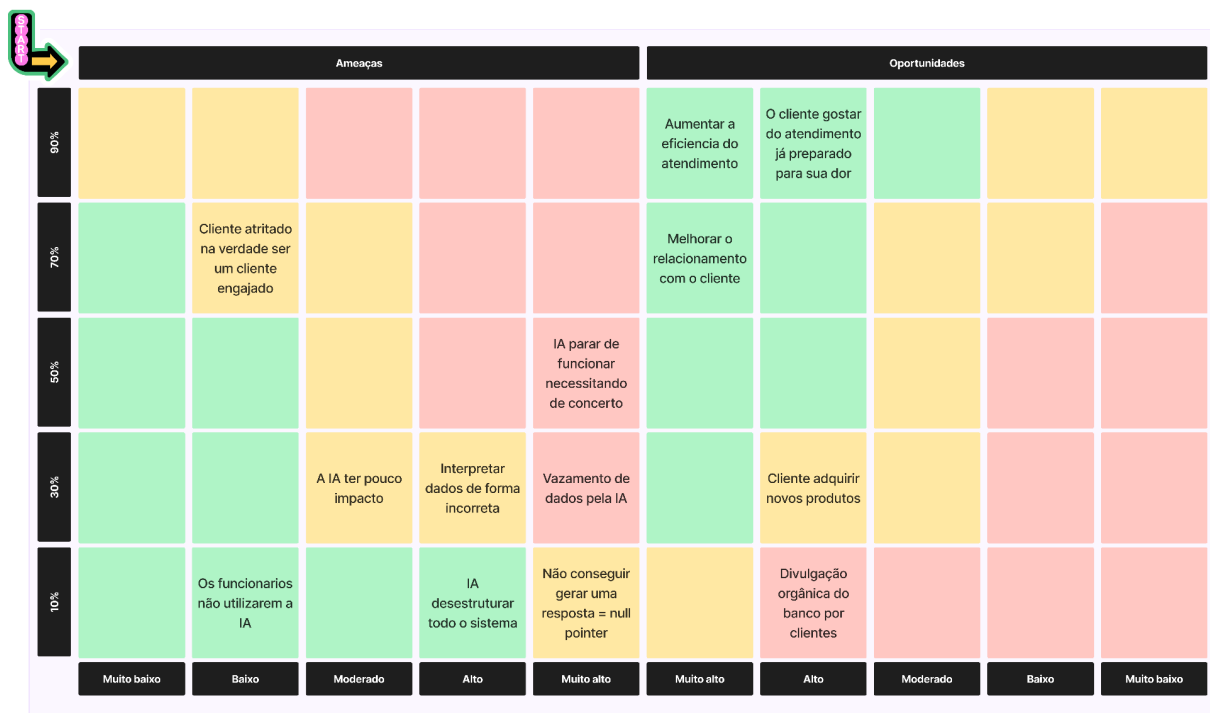
<https://www.figma.com/file/YuQoIP6Rlo3VGWTu8W3CSL/Value-Proposition-Canvas?node-id=0%3A1>



## 4.1.5. Matriz de Riscos

Figma:

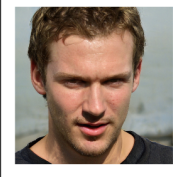
<https://www.figma.com/file/c4mNtPuWBJ3TXZ1RVuysio/Matriz-de-Risco?node-id=0%3A1>



## 4.1.6. Personas

Posicione aqui suas Personas (as que utilizam o modelo e as que são afetadas pelo modelo)

Figma: <https://www.figma.com/file/X77AccOOj4Fw5dAKwSwLux/Personas?node-id=0%3A1>



**João Ferreira**

**Age:** 33  
**Occupation:** Estudante de Ciências da Computação  
**Location:** São Paulo, SP  
**Education:** Cursando a graduação  
**Status:** Solteiro

**Bio**

"Não faz muito tempo que entrei na faculdade e já tive a chance de realizar um estágio no banco Pan. A equipe de lá me acolheu muito bem e pretendo sempre dar o melhor de mim no meu trabalho, mesmo sendo muito cansativo.

Eu trabalho na parte de atendimento e as vezes é muito complicado atender por estar despreparado ao tipo de cliente que irei responder. As vezes ele já chega irritado, é um cliente conflitado. Seria melhor se eu soubesse desse possível comportamento antes mesmo de mandar mensagem ou falar no telefone."

**Personality**

Introvert	Extrovert
Analytical	Creative
Busy	Time rich
Messy	Organized
Independent	Team player

## 2. More about this persona

**Interests**

Meu maior interesse é meu desenvolvimento físico. Malho a mais de dois anos mas não tenho tido resultados tão grandes. Isso deve mudar daqui para frente.

Vitor Queiroz

**Influences**

Sou influenciado pelos influenciadores digitais fitness. Eles sempre possuem ótimas dicas de treino e também de dieta. Me inspiro neles para crescer cada dia mais.

Vitor Queiroz

**Goals**

Minha maior meta é independencia financeira. Estou começando agora no mercado coorporativo e amo chegar no final do mês e ver o dinheiro cair na conta.

Vitor Queiroz

**Needs & Expectations**

Eu preciso de ajuda para atender melhor os clientes. Entender se eles possuem um problema, vontade de ter algo a mais no banco ou se é novo na empresa. Com isso, eu acho que facilitaria muito meu trabalho.

Vitor Queiroz

**Motivations**


Minha motivação é já saber o provável motivo dele estar ligando para ter praticidade na hora de atender o telefone. Assim ele poderia sentir que o serviço é bom e que vale a pena continuar conosco.

Vitor Queiroz

**Pain Points / Frustrations**

Meu maior medo é que essa tecnologia não dê certo. Eu atenda o cliente errado, oferecendo um produto quando ele ainda vai criar uma conta no banco e nem nos conhece direito. Isso poderia ser ruim para gerar confiança.

Vitor Queiroz



**Mauro Augusto**

**Age:** 53  
**Occupation:** Engenheiro  
**Location:** Piracicaba  
**Education:** Superior completo  
**Status:** Casado

**Bio**

Mauro Augusto, nascido e residente de Piracicaba, é um homem casado, com 2 filhos e que é apaixonado em engenharia. Adora ir ao parque com sua família e aos domingos costuma assistir aos jogos do seu time do coração. Sua condição financeira não é das melhores. Atuando na área de engenharia, Mauro não tem muita visibilidade no mercado visto que a cidade em que mora não tem uma grande atuação nessa área.

**Personality**

Introvert ————— Extrovert

Analytical ————— Creative

Busy ————— Time rich

Messy ————— Organized

Independent ————— Team player

## 2. More about this persona

**Interests:**

- Engenharia
- Futebol
- Tecnologia
- Matemática

Gustavo Ferreira

**Influences:**

- Família
- Sucesso
- Dinheiro

Gustavo Ferreira

**Goals:**

- Promoções no trabalho
- Estabilidade financeira

Thomas Brand

**Needs & Expectations**

- Outras fontes de renda
- Ajuda de bancos
- Renegociação de dívidas

Gustavo Ferreira

**Motivations**


- Construir um patrimônio fixo, para a minha aposentadoria

Gustavo Ferreira

**Pain Points / Frustrations**

- Atendimento demorado quando tenta entrar em contato ao banco e que quase sempre não resolve os seu problemas

Gustavo Ferreira



**Maria das Neves**

**Age:** 41  
**Occupation:** Enfermeira  
**Location:** Belo Horizonte  
**Education:** Superior  
**Status:** Casada

**Bio**

Maria das Neves, nascida e residente de Belo Horizonte, é casada com José Antonio. Adora cozinhar nos fins de semana quando está de folga de seus plantões e viajar com seu esposo durante as férias, além de gostar de se exercitar. Sua condição financeira não é das melhores, mas mesmo assim a vida é muito movimentada por trabalhar em dois hospitais. Sendo assim, possui o desejo de ter uma vida mais calma quando se aposentar.

**Personality**

Introvert	Extrovert
Analytical	Creative
Busy	Time rich
Messy	Organized
Independent	Team player

## 2. More about this persona

**Interests**

- Cozinhar
- Assistir reality shows
- Fazer academia
- Ensinar e aprender
- Viajar

Gustavo Ferreira

**Influences**

- Seu gato "Algodão", por ser peludo e branco
- A vontade de possuir uma vida calma

Gustavo Ferreira

**Goals**

- Participar de campanhas para assistência social

Gustavo Ferreira

**Needs & Expectations**

- Aprender mais sobre o mercado financeiro
- Ter uma renda boa com base no que investiu

Gustavo Ferreira

**Motivations**

- Ajudar quem precisa
- Construir uma renda sólida, gerenciar seus gastos para investir mais e mais

Gustavo Ferreira

**Pain Points / Frustrations**

- Pouco conhecimento sobre produtos do mercado financeiro
- Dificuldade de contratar produtos diretamente com o banco

Gustavo Ferreira

### 4.1.7. Jornadas do Usuário

Posicione aqui seus mapas de jornadas do usuário que utiliza o modelo

Figma:

[https://www.figma.com/file/wSMUkm2yUPpjDgwR6ai694/User-Journey-Map-Template-\(Community\)?node-id=2%3A16](https://www.figma.com/file/wSMUkm2yUPpjDgwR6ai694/User-Journey-Map-Template-(Community)?node-id=2%3A16)



#### Mauro Augusto

Age: 53

Occupation: Engenheiro

Location: Piracicaba

#### Cenário

Mauro está sofrendo dificuldades com o uso de algumas opções no aplicativo do Banco Pan e gostaria de uma ajuda do suporte do Banco Pan pelas redes sociais.

#### Expectativas

Que seu problema no aplicativo seja resolvido da forma mais rápida e direta possível, com um ótimo atendimento e com um suporte bem qualificado.

Fases	Fase 1 Contato pelas redes de suporte	Fase 2 IA analisa o perfil do usuário e redireciona para a sua categoria	Fase 3 Encaminhado para o setor que conseguirá resolver o seu problema específico
Tarefas	<ul style="list-style-type: none"> <li>Escolher uma plataforma que possua contato com o suporte do Banco Pan.</li> <li>Mandar mensagem para o suporte (IA)</li> </ul>	<ul style="list-style-type: none"> <li>Esperar pela resposta da IA</li> <li>Se não estiver logado, mandar as informações que a IA necessitar.</li> <li>Enviar de forma descritiva o seu problema</li> </ul>	<ul style="list-style-type: none"> <li>Esclarecer as informações para o funcionário do Banco Pan.</li> <li>Seguir todos os passos que ele pedir.</li> </ul>
Pensamentos	"Em qual eu vou ser respondido mais rápido e de forma simples?"	"Qual a melhor forma de eu escrever o meu problema e minhas informações se necessário, e posso confiar na IA?"	"Esse funcionário está resolvendo o meu problema da forma que eu pensei que fosse, algo prático e não muito demorado."
Emoções			

## 4.2. Compreensão dos Dados

1. Descreva os dados a serem utilizados (disponibilizados pelo cliente e outros se tiverem sido incluídos), detalhando a fonte, o formato (CSV, XLSX, banco de dados, etc.), o conteúdo e o tamanho.

Nos foi enviado pelo Banco Pan, usando a plataforma do Slack, um arquivo CSV com 1,36GB de dados. Dentre eles há informações como; CPF, quantidade de produtos do cliente, tempo de relacionamento com o cliente, quantidade de atendimentos realizados, reclamações abertas, quantidade de atendimentos que o cliente obteve fora do prazo, soma do saldo de todos os produtos de créditos contratados para o cliente, classificação do grau de risco de crédito para clientes, soma do saldo de todos os produtos de créditos contratados em outra instituição financeira, dentre outras informações. Também nos foi passado recomendações de segurança; não enviar o documento para ninguém, não deixar o link público, ou descriptografar o hash do cpf. Além disso, existem muitos campos vazios, resultando em, talvez, uma avaliação

não tão precisa e suposições equivocadas. Devido ao tamanho do conjunto de dados, usaremos apenas o primeiro 1 milhão de linhas.

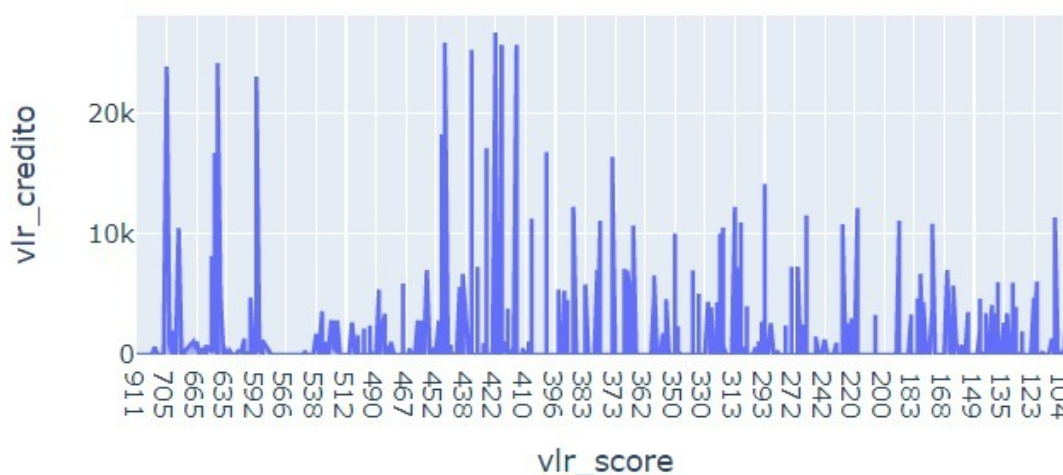
2. Descrição estatística básica dos dados, principalmente dos atributos de interesse, com inclusão de visualizações gráficas e como essas análises embasam suas hipóteses.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12505293 entries, 0 to 12505292
Data columns (total 13 columns):
#   Column      Dtype
---  -
0   anomes      int64
1   vlr_credito float64
2   vlr_saldo   float64
3   num_atend_atrs float64
4   vlr_score   float64
5   num_produtos float64
6   num_atend   float64
7   num_cpf     object
8   qtd_oper    float64
9   qtd_reclm   float64
10  qtd_restr    float64
11  vlr_renda   object
12  cod_rating  object
dtypes: float64(9), int64(1), object(3)
memory usage: 1.2+ GB
```

Figura 1: Tabela de informações.

Index	anomes	vlr_credito	vlr_saldo	num_atend_atrs	vlr_score	num_produtos	num_atend	qtd_oper	qtd_reclm	qtd_restr
count	12505293.0	7032474.0	6900003.0	6604.0	8738902.0	9888796.0	26545.0	7032474.0	1364.0	8330173.0
mean	202135.33767701403	32065.026640424185	5864.500675229832	1.11765596081163	466.78216920157706	1.0193715580502082	1.3735166698059897	11.980313044882925	1.000733137829912	2.857061431977463
std	42.85810319160321	65672.94291319748	28558.148290187528	0.38907955984007266	207.45916856812772	0.9849523710378423	0.6974935271539803	10.274287798464782	0.027076518053694085	3.5611560558544166
min	202104.0	0.0	0.01	1.0	0.0	1.0	1.0	0.0	1.0	1.0
25%	202107.0	2974.3225	994.69	1.0	329.0	1.0	1.0	5.0	1.0	1.0
50%	202110.0	14245.005	2358.2399999999999	1.0	429.0	1.0	1.0	10.0	1.0	2.0
75%	202201.0	33959.935	6748.9499999999995	1.0	580.0	2.0	2.0	16.0	1.0	3.0
max	202204.0	10348109.079999998	32102768.81	7.0	1000.0	15.0	17.0	306.0	2.0	413.0

Figura 2: Tabela de descrição.





**Figura 6:** gráfico de maiores valores de números de ligações por categoria.

3. Descrição da predição desejada (“target”), identificando sua natureza (binária, contínua, etc.)

Para as análises de casos de atendimentos dos clientes do banco Pan iremos segmentar como target os clientes atritados em 3, sendo eles: ‘pouco atritados’, ‘meio atritados’ e ‘muito atritados’, sua natureza é discreta, com possibilidades de respostas fechadas.

## 4.3. Preparação dos Dados

### Início

Nossa equipe está utilizando as bibliotecas do pandas e sklearn. Elas são extremamente importantes para reduzir o uso de códigos no programa e nos ajudar na análise e manipulação de dados, além da prática de machine learning. Também, fizemos a importação do plotly.express para a criação de gráficos que nos auxiliam a entender os dados e idealizar nosso padrão. Além de ter sido necessário o uso das bibliotecas pydrive.auth e pydrive.drive para conseguirmos acessar o CSV, dentre outras para a realização de algumas tarefas.

Por nossos dados serem confidenciais, somente quem possui o “@inteli” pode ter o acesso. Devido a isso, o link para acessar o CSV não pode ser público, então é necessário realizar uma autenticação para ter a permissão e enfim realizar o download dos dados no colab.

O download dos dados foi feito diretamente do armazenamento em nuvem disponibilizado pelo parceiro. A biblioteca PyDrive é utilizada para baixar esses dados, que estão comprimidos, no armazenamento do Google Collaboratory. Em seguida utilizamos a biblioteca Shutil do python para descompactar os arquivos, extraindo o arquivo CSV diretamente no Colaboratory. Tudo isso aliado às funções de autenticação do PyDrive, que garante que apenas alunos do Inteli consigam acessar as pastas em nuvem para baixar os dados do drive do parceiro. Assim, temos o arquivo já pronto para ser trabalhado.

### Tratamento

A importação do banco de dados foi feita para o próprio tratamento das informações que nos foram proporcionadas. Usando o Google Colab e a biblioteca pydrive, conseguimos acessar as informações de uma maneira eficiente. Além disso, estamos colocando esses dados em uma variável chamada df(DataFrame) e, dentro dela, é possível manipular e assim, tratá-los de forma precisa.

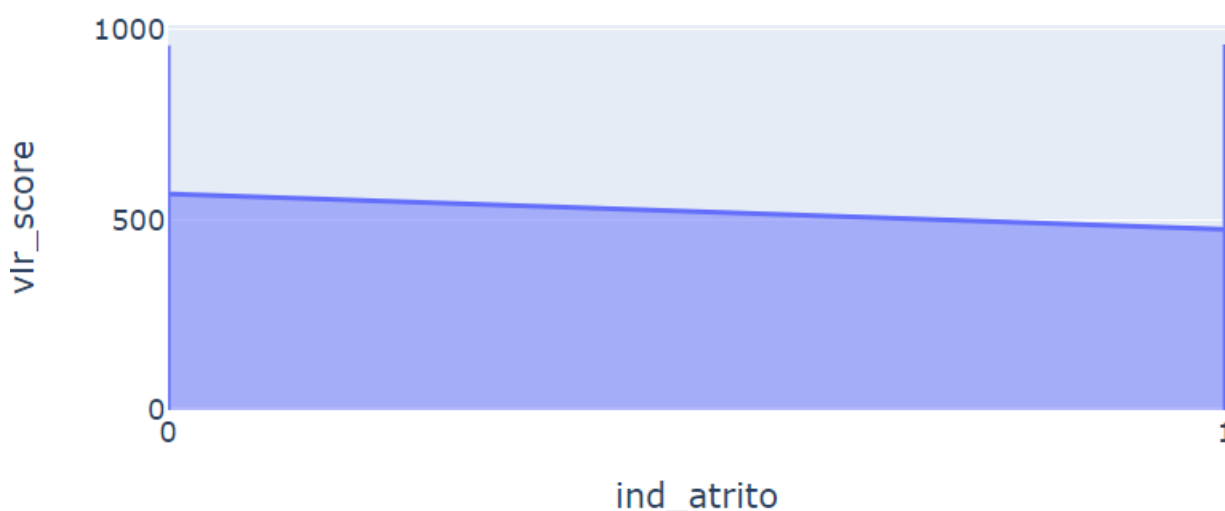
Por meio da seleção dos dados anteriormente, analisamos que existem dados que precisam ser tratados com métodos corretos, como limpeza de dados, integração de dados e transformação de dados, porém o que mais foi usado foi a limpeza de dados onde colocamos 'qtd\_reclm', 'num\_atend\_atrs', 'num\_atend', 'ind\_atrito', 'cod\_rating' e 'vlr\_score' com valores nulos (“vazio”) na tabela para 0 pois são colunas que trabalham com valores float64 logo só representam valores numéricos.

Após a primeira triagem, corrigimos a coluna 'cod\_rating' transformando todos os valores com caracteres (tipo String) para correlatos numéricos, seguindo a seguinte escala: ['A', 'AA', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'HH', NaN] para [10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0]

## Testes e Conclusões

Na fase de seleção de dados, procuramos selecionar algumas colunas de dados sobre os clientes que compõem a base de dados fornecido pelo Banco Pan. Sendo elas as colunas: 'qtd\_reclm' (quantidade de reclamações), 'num\_atend\_atrs' (número de atendimento atrasados), 'num\_atend' (número de atendimento), 'ind\_atrito' (índice de atrito), 'cod\_rating' (código de nota interna) e 'vlr\_score' (valor de score).

Por meio da análise inicial desses dados, concluímos que há algumas correlações entre algumas colunas, uma delas a correlação direta entre as quantidades de reclamações com o índice de atrito e o número de atendimentos atrasados, também a existência de correlação entre o índice de atrito e valor de score (figura 1), pois grande parte dos clientes que possui valor de score maior que 0 possuem algum tipo de atrito com o Banco Pan. Através desses dados, espera-se ser possível montar as bases de treinamento para o modelo preditivo a ser criado.



**Figura 1:** Gráfico de valor de score por índice de atrito.

## 4.4. Modelagem

Para a Sprint 3, você deve descrever aqui os experimentos realizados com os modelos (treinamentos e testes) até o momento. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus experimentos e resultados.

Para a Sprint 4, você deve realizar a descrição final dos experimentos realizados (treinamentos e testes), comparando modelos. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus experimentos e resultados.

## 4.5. Avaliação

Nesta seção, descreva a solução final de modelo preditivo, e justifique a escolha. Alinhe sua justificativa com a seção 4.1, resgatando o entendimento do negócio e explicando de que formas seu modelo atende os requisitos. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus argumentos.

## 5. Conclusões e Recomendações

Escreva, de forma resumida, sobre os principais resultados do seu projeto e faça recomendações formais ao seu parceiro de negócios em relação ao uso desse modelo. Você pode aproveitar este espaço para comentar sobre possíveis materiais extras, como um manual de usuário mais detalhado na seção “Anexos”.

Não se esqueça também das pessoas que serão potencialmente afetadas pelas decisões do modelo preditivo, e elabore recomendações que ajudem seu parceiro a tratá-las de maneira estratégica e ética.

## 6. Referências

Nesta seção você deve incluir as principais referências de seu projeto, para que seu parceiro possa consultar caso ele se interessar em aprofundar.

Utilize a norma ABNT NBR 6023 para regras específicas de referências. Um exemplo de referência de livro:

LUCK, Heloisa. **Liderança em gestão escolar**. 4. ed. Petrópolis: Vozes, 2010.

SOBRENOME, Nome. **Título do livro**: subtítulo do livro. Edição. Cidade de publicação: Nome da editora, Ano de publicação.

## Anexos

Utilize esta seção para anexar materiais como manuais de usuário, documentos complementares que ficaram grandes e não couberam no corpo do texto etc.