Predição de Churn

Edmauro, Eveline, Isabel, Jefferson, Tales, Vitória Rabelo Santos

PARTE I: Estrutura do projeto

# Contexto

O cenário encontrado para o case foi baseado em um dataset do kaggle que nos trazia a seguinte informação: uma preocupação da instituição financeira em relação ao número de consumidores deixando os seus serviços de cartão de crédito.

Em um mercado cada vez mais repleto de opções, os clientes se tornam muito mais propensos à mudança de suas contas para instituições que têm mais a ver com o seu perfil e que oferecem maiores benefícios. Um grande desafio que elas enfrentam é saber qual o comportamento dos clientes que saem de suas instituições e para descobrir tudo que está envolvido nesse processo é necessário um olhar técnico e analítico para todas as informações relativas a esse perfil de cliente.

Em nosso projeto iremos analisar variáveis que estejam relacionadas com a saída do cliente e explorar tendências como seu perfil, renda, serviços que possui, frequência de atividade e inatividade na instituição e as características de quem sai ou fica na instituição para entender em quais fatores a instituição deve trabalhar para uma melhora na retenção de seus clientes.

Nesta etapa é interessante explicar um pouco de contexto em que o seu projeto está inserido. Pode falar sobre a situação do mercado, tendências que são importantes observar e os principais desafios que existem. Caso estejam fazendo trabalho para alguma empresa específica, podem dar um breve panorama sobre a empresa também.

# Problema de negócio

Optamos por esse case pelo fato dele expressar um dilema muito comum em muitos negócios: a saída dos clientes. Muitas vezes o negócio consegue alcançar muitos novos clientes mas a retenção desses clientes deixa muito a desejar. Entendendo mais quais fatores explorar para entender os pontos que geram insatisfação ou satisfação ao cliente, essa retenção se torna uma tarefa muito mais assertiva para os gestores da operação.

Se o nosso problema está em torno da retenção desses clientes, algumas perguntas são chaves para que possamos entender esse cenário e gerar insights para lidarmos com esse desafio. Um exemplo é: qual o perfil do cliente que deixa a empresa? Qual o perfil do cliente que está a mais tempo na instituição? Tem alguma relação entre o cliente que saiu mais com a opção de serviço que estamos ofertando pra ele? O produto que esse cliente possui é o que mais se adequa ao seu perfil?

Esta é provavelmente a etapa mais importante da etapa de elaboração do projeto, pois vai definir muito os próximos passos.

Pode ser feito em dois parágrafos: o primeiro no qual você delimita bem o escopo do trabalho e justifica essa escolha e um segundo que consistirá do problema em si.

Explique de maneira objetiva qual o problema de negócio que o projeto está propondo resolver. Formular como uma pergunta pode trazer mais clareza, exemplo:

* Com os dados disponíveis, quais clientes têm maior propensão à compra?
* Quais os perfis de motoristas que procuram a minha marca?

# Impacto

O intuito dessa análise será trazer para as áreas de negócio ferramentas para a construção de estratégias que possam sanar os problemas que fazem as pessoas deixarem o serviço contratado e trazer alternativas de análises que contribuam para a melhoria desse serviço e de sua reputação no mercado em que se encontra.

Assim, a empresa poderá não só conquistar novos clientes, mas trazer longevidade para esse relacionamento tão essencial para a manutenção de suas operações.

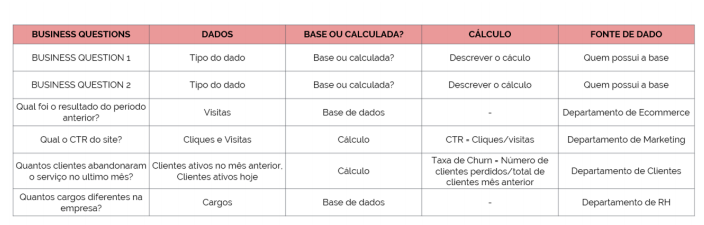
Descreva qual impacto você espera que o seu projeto tenha no negócio caso seja implementado.

Por que é importante que este trabalho seja feito? Quais vantagens competitivas uma pessoa, empresa ou instituição pode ter com este projeto finalizado?

PARTE II: Dados e solução

# Dados

## Data Tracking Sheet



## Fonte

Liste a fonte dos dados que vão ser utilizados.

Podem colocar os principais desafios para consegui-los também: Scrapping, solicitação para entidades públicas, negociação com área da empresa, etc...

## Datasets

Attrition\_Flag (situação da conta) object

Customer\_Age (idade do cliente) int64

Gender (gênero) object Education\_Level (nível educacional) object

Marital\_Status (status civil) object

Income\_Category (renda) object

Card\_Category (categoria de cartão) object

Months\_on\_book (período de relacionamento com o banco) int64

Total\_Relationship\_Count (números de produtos que o cliente possui) int64

Months\_Inactive\_12\_mon (meses de inatividade) int64

Contacts\_Count\_12\_mon (número de contatos com o banco-ultimos 12 meses) int64

Credit\_Limit (limite de crédito) float64

Avg\_Open\_To\_Buy (linhas de crédito disponíveis) float64

Total\_Trans\_Amt (valor total de transações) int64

Total\_Trans\_Ct (quantidade total de transações) int64

Avg\_Utilization\_Ratio (taxa de utilização do cartão) float64

Naive\_Bayes\_ float64

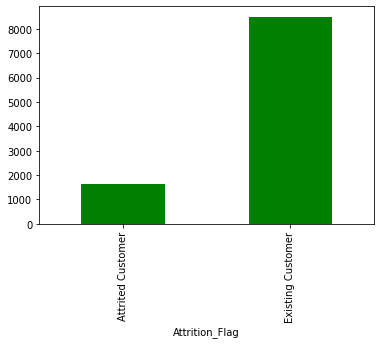
Naive\_Bayes\_Classifier\_Attrition\_Flag\_float64

dtypes: float64(5), int64(7), object(6)

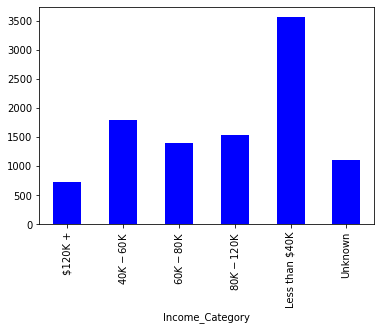
Descreva o dataset e suas variáveis. Definição, tipo de dado (string, int, float), tipo de medida (R$, metro, litro, velocidade, etc)

# EDA (Análise Exploratória de dados)

Através das análises dos dados disponíveis verificamos o panorama geral do momento em que o banco estava em relação à saída de clientes. Do total de clientes cadastrados (10.000 cadastros) tínhamos uma média de 1627 que haviam deixado os serviços do banco e 8500 que ainda mantinham suas contas ativas nesse período, nos levando à seguinte distribuição:



A distribuição de renda de todos os clientes e ex clientes se distribuía de uma forma mais recorrente em valores menores (isso também diria muito sobre o tipo de cartão que essas pessoas de menor renda iriam usar). Como podemos ver na imagem a seguir, as rendas se concentravam em torno de 40 mil dólares anuais.



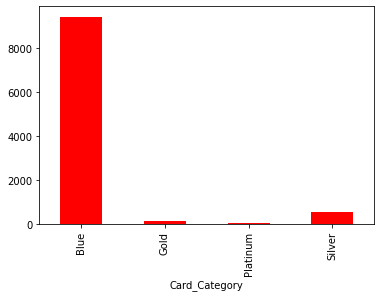
E, em relação aos produtos mais consumidos tivemos a seguinte distribuição:

Blue 9436

Gold 116

Platinum 20

Silver 555



Como pudemos perceber, a renda dos clientes parecia ter influência direta com o tipo de cartão de crédito destinado a esses clientes, nesse caso, o cartão de entrada da instituição, o Blue.

Nesta etapa é quando ocorre a análise dos dados. Descreva novamente os datasets, porém desta vez de maneira analítica. Fique à vontade para colocar gráficos, tabelas, medidas resumo e tudo que julgar importante para um melhor entendimento dos dados e que gerem *insights* para a resolução do problema.

# Abordagem

Descreva qual foi o racional por trás do desenvolvimento da solução:

* Como os dados foram tratados
* Quais modelos foram utilizados
* Métricas do modelo

# Desenho da solução

Quais técnicas vocês preveem que serão utilizadas para resolver este problema? Por qual caminho vocês seguiram?

PARTE III: Colocando a solução à prova

# Interface

No começo do projeto podem colocar aqui um rascunho de como imaginam que será a interface final do projeto. Terá gráficos? Receberá input do usuário? Fará uma previsão?

[Colocar imagem exemplo]

Quando o projeto estiver finalizado podem acrescentar um *print* da tela real do modelo em produção.

Implicações e próximos passos

Pensando no impacto gerado, quais são as externalidades (positivas ou negativas) da solução?

A solução está sujeita a alguma distorção, viés ou interferência de usuários? Quais?

Quais cuidados devem ser tomados a médio e longo prazo? (mudanças comportamentais em usuários e no contexto de negócio, possível descalibragem do modelo)

Quais são os próximos passos do projeto?

# Conclusão

Para finalizar escrevam quais foram os principais aprendizados sobre o problema e sobre o projeto em si, principais desafios, aprendizados e próximos passos.