



DATA CHALLENGE



BUSCAMOS ANALISTAS DE DADOS PARA
NOS AJUDAR NESSA MISSÃO



DATA CHALLENGE 2022

Uma iniciativa Stone

"You must gain control over your money or the lack of it will forever control you"

Dave Ramsey

Sumário

[Objetivo deste documento](#)

[Introdução](#)

[Descrição do case](#)

[Qual é a curva ideal de vezes que acionamos um cliente?](#)

[A régua de acionamento](#)

[Os dados](#)

[O que esperamos de você?](#)

[Processo de avaliação](#)

[Processo de envio do case](#)

[Anexos](#)

[Ponto de apoio](#)

1. Objetivo deste documento

Se você chegou até aqui, **parabéns!!** Você avançou para a próxima etapa do nosso desafio Stone Data Challenge!

O objetivo principal deste documento é apresentar o problema abordado nesta edição, além de conceituarmos tudo que você precisa saber para resolvê-lo.

Todos os documentos a serem utilizados na resolução do case estão disponíveis na pasta de arquivos enviada à você e também serão descritos detalhadamente neste material.

Separe alguns minutinhos do seu dia, pegue um café e **leia este documento com calma**, ele será um ótimo aliado para você durante o hackathon!

Boa sorte! :)

Atenção! Aviso legal e conformidade com a LGPD!!

Todos os fluxos, processos, dados utilizados e disponibilizados neste desafio foram gerados de forma aleatória e exclusivamente para este hackathon e não representam dados de nossos clientes ou processos da StoneCo.

2. Processo de avaliação

O Stone Data Challenge conta com 5 etapas principais conforme ilustrado no cronograma abaixo. Caso você esteja entre os melhores colocados, você terá oportunidade de refinar sua apresentação entre a avaliação técnica e a avaliação final a partir dos feedbacks da banca de jurados técnicos.



Todos os participantes da 2ª etapa terão uma mentoria para discussão do case e dos dados disponíveis no dia 27/04 às 18h30 através da ferramenta zoom (link enviado ao seu e-mail).

3. Introdução

O Data Challenge da Stone é um desafio de dados com o objetivo de **impulsionar** o desenvolvimento de **pessoas brilhantes** através de uma **experiência mão na massa**.

O evento quer conectar profissionais e entusiastas do universo de **dados e analytics**, através da abordagem de problemas reais que você encontrará no meio corporativo, sendo também uma oportunidade de **treinar** e **desafiar seus conhecimentos**. E tudo isso **concorrendo a prêmios incríveis!**

Preparamos um desafio baseado nas experiências dos nossos times de dados aqui da Stone e, neste documento, **vamos contar tudo que você precisa saber** para construir este case, todos os dados disponíveis para resolução e demais materiais de suporte que você precisará durante sua jornada.

4. Descrição do case

A Stone tem como objetivo ajudar o empreendedor brasileiro a **vender, gerir e crescer** seu negócio através dos nossos produtos e serviços que vão desde a maquininha de cartão de crédito à sistemas ERP para gestão.

Um dos produtos que oferecemos aos nossos clientes é o **crédito**, voltado para facilitar a vida do lojista e impulsionar o crescimento do seu negócio. A Stone trabalha desde 2019 para construir um produto sólido e maduro. Mesmo no momento mais intenso da pandemia provocada pela **COVID-19**, em 3 anos, já liberamos mais de 190 mil linhas de crédito, que somam mais de 3,4 bilhões de reais desembolsados para apoiar nossos clientes a crescer e manter seus negócios.

Nosso produto é diferenciado: **os clientes pagam seus empréstimos através de um percentual de retenção aplicado sobre as transações realizadas pela maquininha de cartão Stone**. Desta forma, conseguimos promover um pagamento sustentável do empréstimo que acompanha as variações de fluxo de caixa do lojista, que não precisa se comprometer com um valor fixo que possa pesar em um mês de baixo movimento em seu negócio. **Um cliente pode ter mais de um contrato de empréstimo simultâneo**, desde que o contrato anterior esteja em dia.

Durante a pandemia, naturalmente tivemos um aumento de clientes inadimplentes, ou seja, com problemas no pagamento de seus contratos e, consequentemente, sentimos a necessidade de **evoluir** nossas **ações de comunicação e acionamento** para **recuperação dos saldos devedores** que alguns clientes passaram a deixar em aberto.

Dentro desta evolução, uma das ações tomadas internamente foi refinar nossas estratégias de comunicação com os clientes, **visando recuperar clientes que apresentavam problemas no ritmo de pagamento**, e claro, sempre usando dados para melhor guiar nossas decisões!

Nesta edição do Stone Data Challenge, queremos dividir com você um dos problemas reais que nosso negócio enfrentou:

Qual é a curva ideal de vezes que acionamos um cliente?

Quando um cliente começa a apresentar dificuldade na liquidação de seus contratos, ele passa a ser elegível para ser acionado por nossa **régua de comunicação** com intuito de estimular a retomada do ritmo saudável de pagamento de acordo com o perfil e o momento de cada cliente.

As **comunicações são disparadas por contrato**, sendo assim, um cliente pode receber diversas comunicações com diferentes conteúdos, um para cada momento da jornada de seu respectivo contrato. Desde o envio da comunicação até o recebimento, são gerados flags que indicam o engajamento do cliente com as comunicações da seguinte forma:

1. Envio do comunicado > 2. Comunicado recebido > 3. Comunicado lido

Dada a sequência da comunicação acima, ocasionalmente são observados **erros no processo de envio do comunicado**, gerando um percentual (%) de falhas por diversos motivos (servidor indisponível, número incorreto etc.). Após o envio bem-sucedido, observamos também que nem todos os comunicados são lidos.

Além do engajamento do cliente com as comunicações, temos outra questão importante: a efetividade na recuperação dos contratos ativos. Uma **comunicação eficaz é aquela que estimula e converte um cliente com problemas na liquidação do seu contrato a retomar o pagamento saudável**, seja por meio de quitação ou do retorno ao ritmo de transação esperado.

Para mensurar a efetividade de uma ação, temos dois ótimos termômetros: os dias sem pagamento (dsp) e os dias sem pagamento do principal de um cliente (dspp).

Os **dias sem pagamento (dsp)** representam o total de dias corridos que um contrato apresenta sem realizar nenhum pagamento. Já os **dias sem pagamento do principal (dspp)**, representam o total de dias corridos que um contrato apresenta sem reduzir o valor do saldo principal. Neste último conceito, mesmo que o contrato apresente algum pagamento, se este

montante não for suficiente para cobrir juros + impostos, **valores deduzidos prioritariamente**, o **saldo principal do contrato permanecerá sem pagamento**.

Dado o problema apresentado, queremos que você nos conte qual a sua abordagem para analisar os dados disponibilizados a fim de **entender e explorar a curva de engajamento x quantidade de acionamentos** e nos traga **insights** sobre a efetividade destas ações em termos de pagamento.

Consideramos um **diferencial** a exploração de variáveis complementares (ex. ramo do lojista, dados geográficos, sazonalidade etc.) para entender se existem **diferentes curvas para diferentes perfis**, visando customizar as ações de comunicação de acordo com o perfil de cada cliente. Estas análises extras também podem (e devem) se estender a performance de pagamento e efetividade das ações de cobrança.

A régua de acionamento

Para este hackathon, criamos uma **régua de comunicação simplificada** para determinar quais ações de comunicação cada cliente receberá de acordo com seu momento de inadimplência a partir dos dias sem pagamento (dsp) e dias sem pagamento de principal (dspp).

Conforme demonstrado abaixo, de acordo com os dias de inadimplência de um contrato, ele receberá um tipo de campanha diferente com intuito de **estimular o retorno saudável de pagamento ou realizar a quitação do saldo devedor**:

DSP		
Régua DSP		
Campanha	Regra	Canal acionamento
Campanha de Observação	Enviada com 5 dias sem pagamento (5 DSP)	E-mail / Whatsapp
Campanha Parcelamento	Enviada com 10 dias sem pagamento (10 DSP)	E-mail / Whatsapp
Campanha Boleto quitado	Enviada com 15 dias sem pagamento (15 DSP)	E-mail / Whatsapp
Campanha Pré negativação	Enviada com 30 dias sem pagamento (30 DSP)	E-mail / Whatsapp
Campanha Negativação	Enviada com 60 dias sem pagamento (60 DSP)	E-mail / Whatsapp
Campanha Boleto quitado	Enviada com 90 dias sem pagamento (90 DSP)	E-mail / Whatsapp
Régua DSPP		
Campanha	Regra	Canal acionamento
Campanha de Observação	Enviada com 15 dias sem pgto principal (15DSPP)	E-mail / Whatsapp
Campanha Parcelamento	Enviada com 30 dias sem pgto principal (30DSPP)	E-mail / Whatsapp
Campanha Boleto quitado	Enviada com 45 dias sem pgto principal (45DSPP)	E-mail / Whatsapp

5. Os dados

Abaixo você encontrará a descrição de todos os arquivos de dados que separamos para que você possa **construir a resolução do problema proposto**. Você também pode **(e deve)** sugerir **análises complementares** para **defender a sua abordagem** do tema sugerido. Todo e qualquer diferencial em suas análises garantirá **pontos extras** que podem ser determinantes para **garantir seu lugar na final!**

Queremos que você demonstre suas habilidades para extrair valor dos dados, mostrando todo o potencial de um verdadeiro profissional **data driven**.

Para este case, disponibilizamos uma **amostra de 14 mil** contratos gerados aleatoriamente tendo como inspiração o cenário de dados de uma empresa de verdade, onde nem sempre temos todos os dados prontos para consumo. A preparação dos dados disponibilizados **também será avaliada**, pois demonstrará sua habilidade de lidar com problemas reais do mundo corporativo.

Também disponibilizamos um **dicionário de dados** contendo mais detalhes sobre todos os campos presentes nas tabelas descritas e o **modelo de entidade-relacionamento (MER)** para te apoiar no processo de junção e preparação dos dados:

Base PORTFOLIO_GERAL.csv

Esta tabela contém uma amostra aleatória do histórico de alguns contratos de crédito. Ou seja, cada contrato contém N registros, um para cada dia do histórico do contrato desde o momento de sua origem.

Base PORTFOLIO_TPV.csv

Esta base de dados possui informações dos registros de transações relacionadas à maquininha Stone do cliente, onde é possível amortizar o empréstimo através das transações e pode ser utilizada para entender a performance de pagamento de forma geral.

Base PORTFOLIO_COMUNICADOS.csv

Aqui você encontrará todas as informações relacionadas às ações de comunicação enviadas aos clientes bem como seu respectivo status e demais informações.

Base PORTFOLIO_CLIENTES.csv

E nesta tabela você encontrará todas as informações relacionadas aos dados cadastrais dos clientes que contrataram empréstimos, bem como suas respectivas informações geográficas e de segmentos de negócios.

6. O que esperamos de você?

Dado o problema apresentado e os dados disponíveis, esperamos que você seja capaz de chegar a uma resposta sobre **“Qual é a curva ideal de vezes que devemos acionar um cliente?”**.

Também esperamos que você seja capaz de **criar análises adicionais** relacionadas ao tema apresentado, **explorando todos os dados disponibilizados**, usando bases de dados públicas e inovando! Esse é o nosso espírito Stone! \o/

Ao final da resolução do case, esperamos que você nos entregue os itens descritos abaixo:

- 1. Arquivo de dados:** Seu primeiro entregável será um arquivo de texto (**extensão .csv**) contendo a(s) tabela(s) final(is) de dados com todas as preparações e análises que você desenvolveu para chegar na resposta final.
O que iremos avaliar? O principal objetivo desta entrega é validar se os joins, cálculos e operações realizadas estão corretas, em relação ao problema e aos dados disponibilizados.
- 2. Códigos e scripts:** Também precisamos que você nos envie seus códigos compilando todas as etapas de processamento e análise de dados. Podem ser enviados scripts em SQL e/ou Python.
O que iremos avaliar? O principal objetivo desta entrega é validar tanto as operações de preparação e transformação dos dados bem como suas análises e insights.
- 3. Documento de resolução do case:** O terceiro item a ser entregue é o documento (.doc) **‘Framework de avaliação’** que você preencherá conforme o template disponibilizado, o qual nos ajudará a compreender todas as etapas que você executou durante a resolução do problema proposto.
O que iremos avaliar? Vamos avaliar todas as etapas que você desenvolveu durante a resolução do case, tais como **preparação, limpeza, análise e exploração dos dados** disponíveis até, finalmente, os **insights** que você encontrou ao final do processo. **Capriche nas descrições!**
- 4. Dashboard:** Como este é um case de análise de dados, **não poderia faltar um bom dashboard!** Sinta-se à vontade para criá-lo no software de sua preferência

desde que ele possa ser **exportado para formato .pdf** pois durante a avaliação dos cases, é importante que todos os jurados da banca avaliadora consigam abrir seu dashboard. Softwares pagos podem precisar de licenças para o uso e o arquivo pdf garante que isso não acontecerá!

O que iremos avaliar? Nesta etapa serão avaliados critérios como ergonomia de interface (design claro, não poluído/objetivo), storytelling e coerência das análises apresentadas.

- 5. Apresentação final (bônus):** Como diferencial, valendo nota extra, você poderá construir uma **apresentação** em um arquivo (.ppt, prezi ou .mp4) que contenha a defesa do seu case e os insights encontrados. Caso você esteja entre nossos finalistas, você terá a oportunidade de apresentar suas descobertas para a banca de jurados, para nossos convidados e times dados!

Na apresentação sugerimos que dê destaque para a sua **abordagem de resolução** do problema, insights obtidos e soluções propostas. Você também pode incluir de forma macro, as etapas que você seguiu durante o processo de preparação e análise dos dados.

7. Entrega do case

Ao final da resolução do case, você deverá enviar todos os entregáveis descritos no capítulo anterior para a etapa de validação e avaliação. Este processo será **feito via Github** e precisamos que você deixe as permissões do seu projeto, contendo todos os itens de entrega, **no modo público**.

O link do seu projeto deverá ser enviado para o email tech@stone.com.br até o dia 04/05 às 23h59 com o assunto no seguinte padrão:

STONE DATA CHALLENGE 2022 - NOME PARTICIPANTE

Checklist de entregas

Para te ajudar no processo de entrega final, separamos este checklist para garantir que você não deixe nada para trás:

- Editar a permissão de seu projeto Github para **público**;
- Upload dos códigos e scripts no Github;

- Upload do arquivo de dados final no Github;
- Upload do documento 'Framework de avaliação' no Github;
- Upload do dashboard no Github;
- Upload da apresentação final (opcional);
- Enviar o link do seu projeto Github para o tech@stone.com.br até o dia 04/05 às 23h59

Você será avaliado nos seguintes critérios:

- a - Avaliação do arquivo de dados
- b - Queries e scripts
- c - Documento 'Framework de avaliação'
- d - Dashboard
- e - Nota extra: apresentação

Cada item será avaliado de 0 a 10 e a nota final será o resultado da seguinte fórmula.

$$Nota\ final = ((a + b + c + d) \div 4) + e$$

8. Ponto de apoio

Se você não conseguir abrir algum dos anexos ou tiver algum problema que não esteja relacionado ao desenvolvimento do case ou a data de entrega, fale conosco através do tech@stone.com.br.

Nós do time Stone agradecemos a sua participação!

Boa sorte!