Seleção Analista Data Science

# Introdução

Essa atividade é a segunda etapa do processo seletivo para a vaga de analista do setor de Data Science da Fortbrasil. O principal objetivo dessa etapa é conseguir entender o nível técnico de cada participante e principalmente as soluções de negócio que o participante trará nas questões abaixo.

# A Empresa

Prazer, somos a FortBrasil S/A, uma fintech genuinamente cearense que há mais de 15 anos atua no segmento financeiro, na concessão de crédito rápido e consciente para parceiros e para as classes emergentes. Atualmente, nossa maior operação está focada na região nordeste, mas estamos em expansão, a todo vapor, para todo o restante do Brasil. Nos últimos meses, já começamos operações em São Paulo, Brasília, Goiânia, dentre outros. Ao longo da nossa história, já conquistamos quase 2 milhões de clientes e mais de 300 clientes varejistas (privates). Em 2019, fechamos o ano com, aproximadamente, um bilhão em vendas. Uma grande conquista ocorrida no ano de 2019 foi o início da operação do cartão Fortbrasil Mastercard, onde consolidamos nossa expansão nacional.

# Pontos Importantes

O case será composto por 3 questões e deve ser enviado até o dia combinado em e-mail, para o seguinte e-mail:

[elaynne.nascimento@fortbrasil.com.br](mailto:elaynne.nascimento@fortbrasil.com.br)

Para solucionar o case, o participante poderá utilizar **qualquer** **ferramenta gratuita** (Python, R, Power BI, Julia, Pentaho, mysql...) ou **ferramentas pagas que temos licenças** (Microsoft Office). Todas as respostas do case e sua respectiva apresentação deverá ser publicada no **GitHub ou Gitlab** e disponibilizado o link para que possa ser avaliado.

# Case

## Questão 1

O negócio da Fortbrasil consiste na oferta de crédito rápido e consciente para parceiros e clientes de classes emergentes, sendo nosso principal produto o cartão de crédito. Para a primeira questão, disponibilizamos algumas variáveis relacionadas à fatura dos clientes:

**ID\_CONTA:** Identificador da conta;

**DT\_VENCIMENTO:** Data vencimento da fatura;

**DS\_ROLAGEM:** Indicador que define se o cliente pagou a fatura do Mês anterior ou não. Se o valor for **FX0** quer dizer que o cliente pagou a fatura do mês anterior e se **FX1** quer dizer que o cliente não pagou a fatura do mês anterior;

**VL\_FATURA:** Valor da Fatura atual.

Baseado nessas informações, responda:

**1.1 –** Qual o percentual de faturas emitidas por mês no qual os clientes não pagaram a fatura anterior?

**1.2 –** Tendo como referência todos os clientes que tiveram fatura emitida no mês de setembro, gere uma base para esses clientes com os seguintes calculados:

• **Total de faturas emitidas** nos últimos 6 meses (sem contar com a fatura de setembro);

(Ou seja, de março a agosto.)

- Quero todos os clientes no qual têm data\_vencimento\_fatura em setembro (4ª coluna na tabela abaixo)

• **O valor médio de fatura** nos últimos 6 meses (sem contar com a fatura de setembro);

• **Quantidade de vezes que ele ficou sem pagar a fatura anterior** nos últimos 6 meses (sem contar com a fatura de setembro).

Segue exemplo:



**1.3 –** Utilizando como referência a base calculada na questão anterior, identifique qual das 3 variáveis calculadas tem o maior potencial de preditivo em relação a variável DS\_ROLAGEM do mês de setembro.

## Questão 2

O ciclo de crédito do cartão Fortbrasil consiste em 3 etapas: Aquisição, Manutenção e Cobrança. O cliente entra em cobrança com 5 dias de atraso, e quando atinge a marca de 65 dias tem o cartão cancelado e então para realizar o pagamento da dívida precisa realizar um acordo. A FortBrasil irá fazer uma nova campanha para aumentar a adesão aos acordos e gostaria de traçar uma estratégia de cobrança utilizando as informações históricas e técnicas de modelagem preditivas. Segue abaixo as definições das variáveis:

**ID\_CONTA:** Identificador da conta;

**DT\_ACORDO:** Data do acordo;

**NU\_DIAS\_ATRASO:** Número de dias após o vencimento que esse cliente está sem pagar;

**VALOR\_CRELIQ:** Valor da dívida do cliente quando ele ultrapassa os 65 dias de atraso;

**DIVIDA\_ATUAL:** Dívida atualizada com juros;

**RESPOSTA:** Se for 1, significa que o cliente aderiu aquele acordo. Ou seja, efetuou o pagamento da entrada. Se for 0, caso contrário.

**QTD\_PARCELAMENTO\_XM**: Quantidade de parcelamentos de fatura aderidos X meses antes da fatura que ele deixou de pagar.

**LIMITE:** Limite total do cliente;

**QTD\_EXTRATOS:** Quantidade de Faturas emitidas;

**QTD\_FX0:** Quantidade de faturas emitidas em dia;

**QTD\_FX1:** Quantidade de faturas emitidas no qual o cliente não pagou a fatura do mês anterior;

**QTD\_FX2:** Quantidade de faturas emitidas no qual o cliente não pagou as duas ultimas faturas anteriores;

**QTD\_ACIONAMENTO:** Quantidade de vezes que ligamos para os clientes para cobrar a sua dívida;

**QTD\_CP:** Quantidade de vezes que a ligação foi atendida, mas não houve contato com o cliente;

**QTD\_CPC:** Quantidade de vezes que o próprio cliente atendeu a ligação de cobrança.

**DEFINIÇÃO**: Acordo é uma proposta de negociação da dívida por parte da Fortbrasil para clientes que estão com mais de 180 dias de atraso.

Baseado nessas informações, responda:

**2.1** – Qual o percentual de adesão mensal por faixa de atraso (Histórico)?

**2.2** – Qual modelo você utilizaria para traçar uma estratégia objetivando o aumento da adesão dos acordos? (Descreva a técnica utilizada)

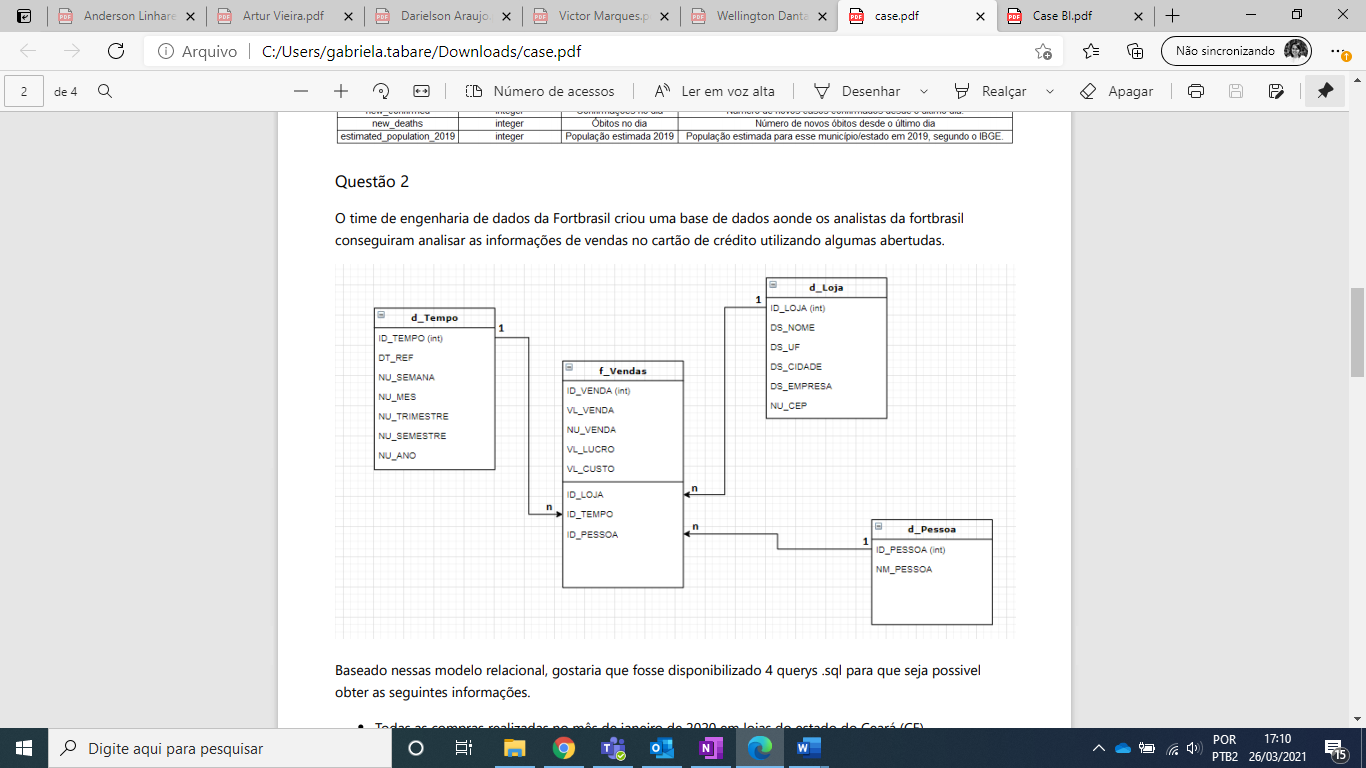
**2.3** – Quais indicadores e ferramentas você utilizaria para avaliar a performance/aderência desse modelo? (Descreva os indicadores utilizados)

**2.4** – Apresente o modelo desenvolvido utilizando a técnica do item (2.2) e as técnicas de avaliação descritas no item (2.3).

**2.5 –** Crie um relatório analítico no Power BI para acompanhar a adesão dos acordos e criar insights de como melhorar essa adesão

## Questão 3

O time de engenharia de dados da Fortbrasil criou uma base de dados onde os analistas conseguiram avaliar as informações de vendas no cartão de crédito utilizando algumas aberturas:



Baseado nesse modelo relacional, disponibilize 4 querys .sql para que seja possível obter as seguintes informações:

**3.1** - Todas as compras realizadas no mês de janeiro de 2020 em lojas do estado do Ceará (CE)

* ID da pessoa
* Nome da pessoa
* Data Referência da Venda
* Valor da Venda

SELECT P.ID\_PESSOA, P.NM\_PESSOA, T.NU\_MES, T.NU\_ANO, F.VL\_VENDA FROM d\_pessoa P , f\_vendas F, d\_Tempo T, d\_loja L WHERE T.NU\_ANO = 2020 AND T.NU\_MES = 1 AND D.DS\_UF = 'CE' AND P.ID\_PESSOA = F.ID\_PESSOA AND F.ID\_LOJA = L.ID\_LOJA AND F.ID\_TEMPO = T.ID\_TEMPO;

**3.2** - Quantidade de compras por cliente no mês de março de 2020

* ID da pessoa
* Quantidade de compras

SELECT P.ID\_PESSOA, COUNT(\*) AS QTD\_DE\_COMPRAS FROM d\_pessoa P, f\_vendas V, d\_tempo T WHERE P.ID\_PESSOA = V.ID\_PESSOA AND V.ID\_TEMPO = T.ID\_TEMPO AND T.NU\_ANO = 2020 AND T.NU\_MES = 3;

**3.3** – Todos os clientes que não fizeram compras no mês de março de 2020

**3.4** – Data da última compra por cliente