

# Seleção - Estagiário de 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!**

A FortBrasil é 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 **09/08/2020 as 23:59** para os seguintes e-mail's:

- lucas.carneiro@fortbrasil.com.br
- irislene.santos@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

Nos últimos meses, o mundo está vivendo uma situação completamente atípica. A pandemia do Corona Virus mudou a vida de todos e muitas pessoas já foram afetadas por essa doença. Buscando entender o número atuais do Brasil, principalmente os do Ceará, gostaríamos de ter um acompanhamento dos seguintes indicadores.

- Números de casos acumulado;
- Número de mortes acumulado;
- % de mortes sobre infectados;
- % da população já infectada;

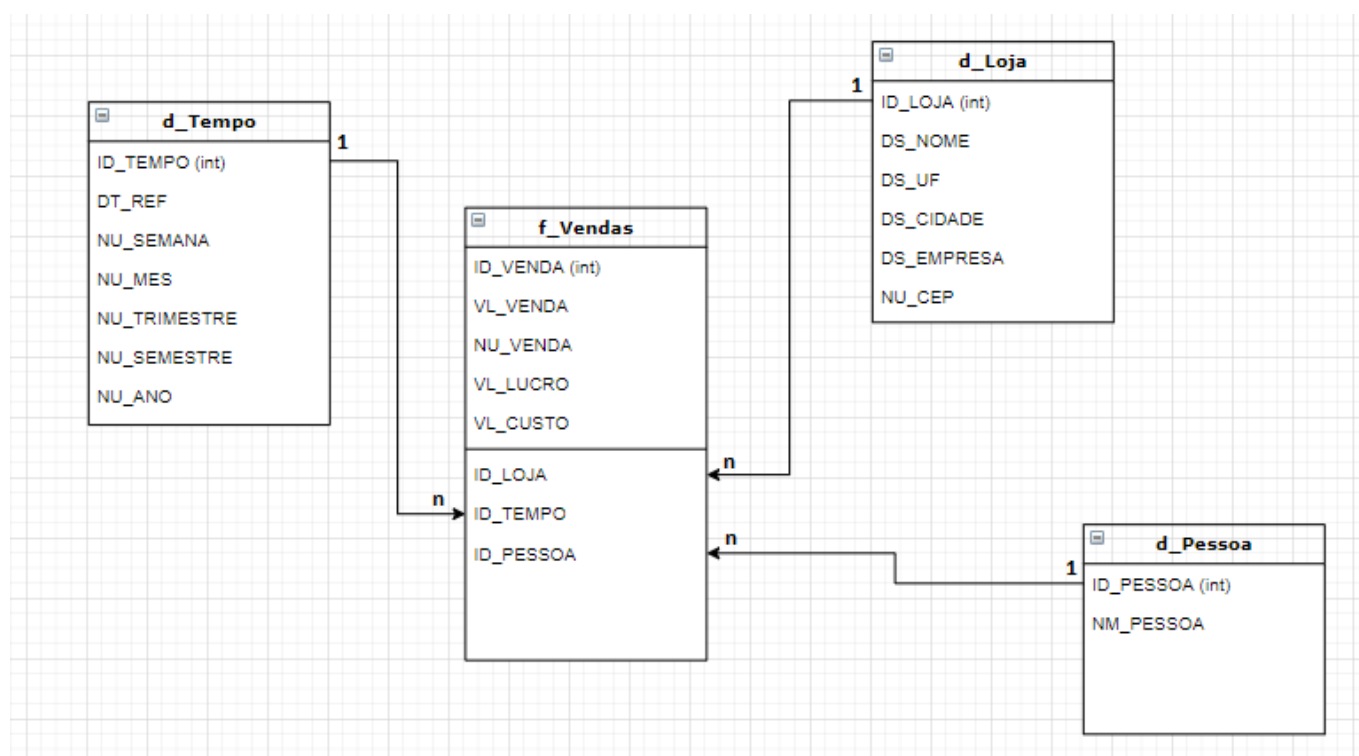
Além de acompanhar esses indicadores de forma geral, gostaríamos de poder acompanhar utilizando como filtro a data e os estados brasileiros.

Para responder essa questão, disponibilizamos a base de dados "**questao1\_coronavirus.csv**". Esse dataset tem as seguintes colunas:

Coluna	Tipo	Título	Descrição
date	string (max_length=10)	Data	Data de coleta dos dados no formato YYYY-MM-DD.
state	string (max_length=2)	UF	Sigla da unidade federativa, exemplo: SP.
city	string (max_length=64)	Município	Nome do município
new_confirmed	integer	Confirmações no dia	Número de novos casos confirmados desde o último dia.
new_deaths	integer	Óbitos no dia	Número de novos óbitos desde o último dia
estimated_population_2019	integer	População estimada 2019	População estimada para esse município/estado em 2019, segundo o IBGE.

## Questão 2

O time de engenharia de dados da Fortbrasil criou uma base de dados aonde os analistas da fortbrasil conseguiram analisar as informações de vendas no cartão de crédito utilizando algumas aberturas.



Baseado nesse modelo relacional, gostaria que fosse disponibilizado 4 queries .sql para que seja possível obter as seguintes informações.

- Todas as compras realizadas no mês de janeiro de 2020 em lojas do estado do Ceará (CE).
  - ID da pessoa
  - Nome da pessoa
  - Data Referencial da Venda
  - Valor da Venda

- Quantidade de compras por cliente no mês de março de 2020.
  - ID da pessoa
  - Quantidade de compras
- Todos os clientes que não fizeram compras no mês de março de 2020.
- Data da última compra por cliente.

### Questão 3

Uma empresa de crédito taiwanês está querendo prever o pagamento próxima fatura dos seus clientes. Tendo esse objetivo em mente, os cientistas de dados dessa empresa coletaram informações cadastrais dos clientes e informações sobre as suas faturas e seus pagamentos.

Esses dados seriam utilizado para criação de estratégias de comunicação para clientes que tivessem alta probabilidade de não pagar as suas proximas faturas, objetivando a diminuição da inadimplencia.

OBS: Já é comprovado por essa empresa que enviar comunicação faz com que os clientes lembrem mais de pagar suas faturas.

Utilizando a base de dados "**questao3\_creditcard.csv**" e o seu conhecimento técnicos e de negócio, responda as seguintes perguntas:

2.1 - Que tipo de problema estamos enfrentando e qual técnica você utilizaria para resolver esse problema?

2.2 - Se você só pudesse enviar comunicação para 10% dos clientes devido ao alto custo, para quais clientes abaixo você enviaria? Explique sua resposta. Utilizar os clientes da base "**questao33\_creditcard\_clientes.csv**"

#### -> Atributos das bases enviadas

- Y: Pagamento da Próxima Fatura (1: Sim, 0: Não)
- X1: Amount of the given credit (NT dollar): it includes both the individual consumer credit and his/her family (supplementary) credit.
- X2: Gender (1 = male; 2 = female).
- X3: Educação (1 = graduate school; 2 = university; 3 = high school; 4 = others).
- X4: Marital status (1 = married; 2 = single; 3 = others).
- X5: Age (year).
- X6: X11: History of past payment. We tracked the past monthly payment records (from April to September, 2019) as follows: X6 = the repayment status in September, 2019; X7 = the repayment status in August, 2019; ... ; -2: No consumption; 0: The use of revolving credit
- X11: the repayment status in April, 2019. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; ... ; 8 = payment delay for eight months; 9 = payment delay for nine months and above.
- X12-X17: Amount of bill statement (NT dollar). X12 = amount of bill statement in September, 2019; X13 = amount of bill statement in August, 2019; ... ; X17 = amount of bill statement in April, 2019.
- X18-X23: Amount of previous payment (NT dollar). X18 = amount paid in September, 2019; X19 = amount paid in August, 2019; ... ; X23 = amount paid in April, 2019.

Qualquer dúvida sobre as perguntas do case, pode entrar em contato via whatsapp (85999356908) ou pelo e-mail ([lucas.carneiro@fortbrasil.com.br](mailto:lucas.carneiro@fortbrasil.com.br)).