

PROCESSO SELETIVO (QI)³ - DESAFIO C

Uma grande loja de varejo, a BigStore, deseja utilizar os seus dados para realizar análises inteligentes com o objetivo de entender melhor a situação do negócio e alavancar suas vendas. Ela possui uma grande base de dados com informações sobre as vendas dos últimos anos.

Parte 1:

Para pôr seu plano em prática, BigStore decidiu contratar uma empresa de consultoria em ciência de dados para analisar e tratar os dados disponíveis. Imagine que você é o analista que ficará responsável por tratar esses dados. Utilize seus conhecimentos em **Python**, R ou outras ferramentas de **ETL** para desenvolver um código para o **tratamento da base de dados da BigStore** ([clique aqui](#)). Normalmente os dados chegam “**sujos**” com **linhas duplicadas**, **dados faltantes**, **datas em formato numérico**, etc, e precisam ser limpos antes de serem utilizados. O tratamento dos dados será realizado com o objetivo de responder às **questões da Parte 2 e 3**.

Parte 2:

Caso tenha alguma familiaridade com **ferramentas de visualização**, como **Power BI**, **Tableau**, **Excel** (ou qualquer outra ferramenta), utilize os dados tratados para criar visualizações que respondam a essas perguntas do cliente. Seja **criativo e tente conquistar** esse cliente para nossa empresa! **Fique à vontade para criar outras análises que achar pertinente com os dados disponibilizados**.

Lembre-se que o objetivo é que o cliente consiga responder às questões abaixo:

1. Quais são os três estados com maior valor de vendas para o ano de 2018?
2. Retorne os 3 primeiros dígitos para cada país e crie uma relação de país e estados. Exemplo: “BRA-Acre”.
3. Quanto o produto ‘Carregador’ contribuiu percentualmente para cada país alcançar sua meta no ano de 2015?
4. Existem pedidos sem itens registrados? Se sim, para quantos pedidos e qual o valor total desses pedidos sem itens?
5. Quantos clientes compraram mais de 2 vezes, durante 2015 e 2018, e gastaram no total mais de 1 mil reais?

Parte 3:

Escreva o código de um **algoritmo de Machine Learning em Python** ou R para **segmentar os clientes em categorias a partir do valor total gasto**. Dessa forma é possível identificar os grupos de clientes que mais gastaram na empresa. Apresente esse resultado de maneira que o cliente consiga facilmente identificar os grupos de clientes e tomar decisões com base nesses resultados. Use sua criatividade para incluir outras visões com base nos dados existentes na base de dados da Big Store, que possam auxiliar na tomada de decisão. (Dica: **utilize os dados da tabela “Pedidos”**).

Pontos de atenção:

- Se houver alguma dúvida, *mande-a para este e-mail até o dia 10/03*. Avaliaremos as dúvidas e, caso necessário, responderemos a sua dúvida a todos para que o entendimento seja nivelado entre os candidatos.
- **Lembre-se de documentar sua linha de raciocínio, tudo que foi desenvolvido (códigos, análises, arquivos, etc), mesmo que você não chegue ao resultado final completo.** Consolide em um arquivo .zip com o seu nome seguindo o formato (Nome_Sobrenome.zip) para este e-mail, **até o dia 14/03, 23:59h**. Os candidatos que não entregarem até essa data, consideraremos como desclassificados.
- Assim que recebermos o material, entraremos em contato agendando uma apresentação para entendermos melhor a sua linha de raciocínio diante o problema.

Um abraço e boa sorte!