

Teste técnico para o cargo de Analista de Dados Delivery Much

Descrição e proposta

Este teste está subdividido em questões teóricas e práticas, cada qual com seu devido valor, totalizando 10 pontos.

Todos os conceitos abordados neste teste fazem parte dos requisitos mínimos ou desejáveis para desempenhar o cargo de analista de dados na Delivery Much Brasil.

Para as **questões práticas**:

- Ao final deste arquivo, nos **Anexos**, se encontram as diretrizes para download do dataset necessário para a realização das atividades;
- é necessário escrever o código na linguagem Python ou Node.js, e disponibilizá-lo em algum repositório de código acessível para correção (Github, Gitlab, Bitbucket, etc).
- Questões relacionadas a SQL não exigem escrita em uma sintaxe específica (MySQL, PostgreSQL, Redshift), por isso podem ser respondidas de acordo com a escolha do candidato. Não é necessário enviar nenhum arquivo ou listagem do resultado obtido na consulta. Basta enviar a consulta em si.

Boa sorte!

Questão 1. Importando o dataset AB_NYC_2019.csv para um SGBD a sua escolha (MySQL, PostgreSQL, Redshift), **crie uma consulta SQL** para cada item listado a seguir.

- a) Lista de hosts distintos (coluna name), agrupados pela coluna room_type. (0,5 ponto)
- b) Total de reviews e média de reviews por mês para hosts localizados no Brooklyn (neighbourhood_group). (1 ponto)
- c) Trazer a quantidade de dias disponíveis, valores mínimo, máximo e média da coluna price para cada host que contenha a palavra room na coluna name e estiveram disponíveis ao menos um dia para locação (considere availability_365 como o número de dias disponíveis para locação). O resultado deve estar agrupado por quantidade de dias disponíveis. (1,5 pontos)

Questão 2. Crie um script em Python ou Node.js que carrega os dados do dataset e, como saída, crie um arquivo CSV somente as seguintes colunas do dataset original:

id, host_id, price, number_of_reviews, last_review (0,5 ponto).

Questão 3. Visando analisar a variação do preço de locação dos hosts, descreva de forma clara e organizada, o tratamento de dados você utilizaria sobre o dataset em questão, englobando pré-processamento, análise textual, algoritmos de mineração de dados, ferramentas de data discovery ou qualquer outra forma/técnica que seja de seu conhecimento. Para esta questão será considerado o processo aplicado sobre o conjunto de dados, visando extrair informações relevantes sobre os dados que possam influenciar no preço de locação do host. (2 pontos)

Questão 4. Imagine que você é um analista de dados no Airbnb (empresa fornecedora dos dados) e lhe foi solicitada a criação de visualizações para entender melhor o que vem acontecendo com os aluguéis na cidade de Nova Iorque. Utilizando alguma ferramenta de visualização (PowerBI, Tableau, Excel, Metabase, Data Studio, etc) para realizar a ingestão do dataset, crie visualizações conforme as questões abaixo solicitadas. **Justifique cada escolha de visualização** (tipo de gráfico e cores utilizadas), descreva qual ferramenta utilizou e **disponibilize o respectivo arquivo exportado da própria ferramenta**, para correção.

- a) Quantidade de reviews agrupados pela coluna neighbourhood. (0,5 ponto). Justifique suas escolhas (0,5 ponto);
- b) Média da coluna minimum_nights agrupado pela coluna room_type. (0,5 ponto). Justifique suas escolhas (0,5 ponto);
- c) Utilizando as colunas latitude e longitude, crie uma visualização geográfica (mapa) dos hosts presentes no dataset (0,5 ponto). Enriqueça o mapa com a utilização de cores e/ou tamanho de ponto diferenciado para outras características como preço, número mínimo de noites, número de reviews, tipo de quarto, por exemplo (0,5 ponto). Justifique cada escolha (0,5 ponto);
- d) Suponhamos que foi solicitada uma visualização que mostre o número de reviews recebidos por mês em um determinado imóvel. Estes dados fictícios foram disponibilizados no **item 3. dos anexos**. Crie uma visualização que mostre o número de reviews recebidos por mês (0,5 ponto). Justifique suas escolhas (0,5 ponto).

Anexos

1. Dataset e explicação do conjunto de dados **AB_NYC_2019.csv** disponível em:
<https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data/>
2. Link direto para download do arquivo **AB_NYC_2019.csv**:
<https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data/download>
3. Dados fictícios para construção do gráfico da **Questão 4.d**:
https://drive.google.com/file/d/1GoKp1WfVSKvNIJhWI7hzFwraC_j9rXon/view?usp=sharing