



Teste Técnico de Engenharia de Dados

Olá, candidato. Parabéns por ter chegado a esta etapa! Estamos ansiosos por ver mais de suas habilidades de analisar, tratar e processar dados. Quaisquer dúvidas que surjam durante o processo, sinta-se livre para entrar em contato com recrutamento@datasprints.com. Vamos lá?

Objetivo

O objetivo deste teste é avaliar sua proficiência nos **requisitos básicos** para nossa vaga:

- Programação básica com SQL
- Programação básica com Python
- Experiência com Cloud Computing
- Experiência com Linux
- Experiência com Data Science/Engineering

E também **habilidades diferenciais**:

- Habilidades em Análise de Grandes Volumes de Dados;
- Habilidades em construção de *pipelines* de dados.

Logo, foque em mostrar que você tem o **domínio dos requisitos**, respondendo às questões dos [quesitos mínimos](#) utilizando um mix entre *Python* e *SQL*. Deixe para cobrir os diferenciais nos pontos levantados em [quesitos bônus](#).

Prazo

Daremos um prazo de **7 dias** para a entrega deste teste a partir da **data de envio deste email**. Consideramos que você tenha, pelo menos, 1 final de semana de esforço para concluí-lo. Se quiser mais tempo, entre em contato para definirmos um prazo que melhor ajuste para nós todos :).

Sobre os dados

Disponibilizamos para você os dados da base **NYC Taxi Trips**. São três conjuntos de arquivos diferentes:

Dataset	Descrição
Trips (1 , 2 , 3 e 4)	Dados sobre corridas de táxi em <i>Nova York</i>
Vendor Lookup	Dados sobre empresas de serviços de táxi
Payment Lookup	Mapa entre prefixos e os tipos reais de pagamento

Quesitos mínimos

Queremos que, a partir dos dados disponibilizados, você responda, de preferência com gráficos às seguintes questões:

1. Qual a distância média percorrida por viagens com no máximo 2 passageiros;
2. Quais os 3 maiores *vendors* em quantidade total de dinheiro arrecadado;
3. Faça um histograma da distribuição mensal, nos 4 anos, de corridas pagas em dinheiro;
4. Faça um gráfico de série temporal contando a quantidade de gorjetas de cada dia, nos últimos 3 meses de 2012.

Quesitos bônus

Estas são coisas que iremos curtir demais se você fizer e, com certeza, irão lhe destacar dentre os outros candidatos:

- Qual o tempo médio das corridas nos dias de sábado e domingo;
- Fazer uma visualização em mapa com latitude e longitude de *pickups and dropoffs* no ano de 2010;
- Simular um [streaming dos dados](#) dos *JSON* e fazer uma visualização acompanhando uma métrica em tempo-real;
- Conseguir provisionar todo seu ambiente em uma cloud pública, de preferência *AWS*.

Como entregar

Você deve entregar para nós 3 pacotes essenciais:

- Arquivo **README.md**: arquivo contendo as instruções para reproduzirmos suas análises;
- Arquivo **Análise.html**: arquivo com as respostas de cada questão da seção [quesitos mínimos](#) e com as análises extras da seção [bônus](#). Utilize, quando possível, [técnicas de storytelling](#) para deixar sua linha de pensamento clara para nós;
- Código-fonte: códigos e *queries* que você utilizou para construir suas análises. (bônus para entrega em repositório do **Github** ou **Gitlab**).