

COMPETIÇÃO DE DATA SCIENCE - FASE 2**SQL E DATA STUDIO****RECOMENDAÇÕES**

Nessa parte da competição você vai precisar de um editor de SQL. Existem outras opções, como o DB Browser (SQLite), mas recomendamos que você utilize o VS Code. É o editor que nós usamos, pois suporta vários tipos de arquivo e possui uma conexão fácil entre eles. Caso ainda não tenha instalado, segue o [link de instalação do VS Code](#). Após a instalação, você vai precisar baixar as extensões para SQL:

**RECEBÍVEIS**

Banco de dados: [olist.db](#)

Exemplo de entregável: [query.txt](#)

Layout dash da IME Jr: [dash](#) (faça uma cópia)

ENTREGÁVEIS: EM ATÉ 19 DIAS APÓS RECEBIMENTO

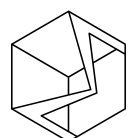
- txt com as queries em statements (divisão por ";" - ponto e vírgula) - você pode escrever queries diferentes, mas na hora de entregar, junte todas num mesmo arquivo.
- Dashboard no Data Studio.

INTRODUÇÃO

Um banco de dados é uma junção de tabelas que possuem conexões através de colunas específicas. O tipo de arquivo é um `.db` e nada mais é do que a junção de vários CSV.

Por exemplo, se você tiver uma tabela com os NUSP e os nomes de todos os alunos e alunas da USP e outra tabela com os NUSP e o curso de cada um(a), você consegue retornar uma terceira tabela com nome, NUSP e curso de cada aluno(a) da USP fazendo o que chamamos de JOIN com o NUSP como chave.

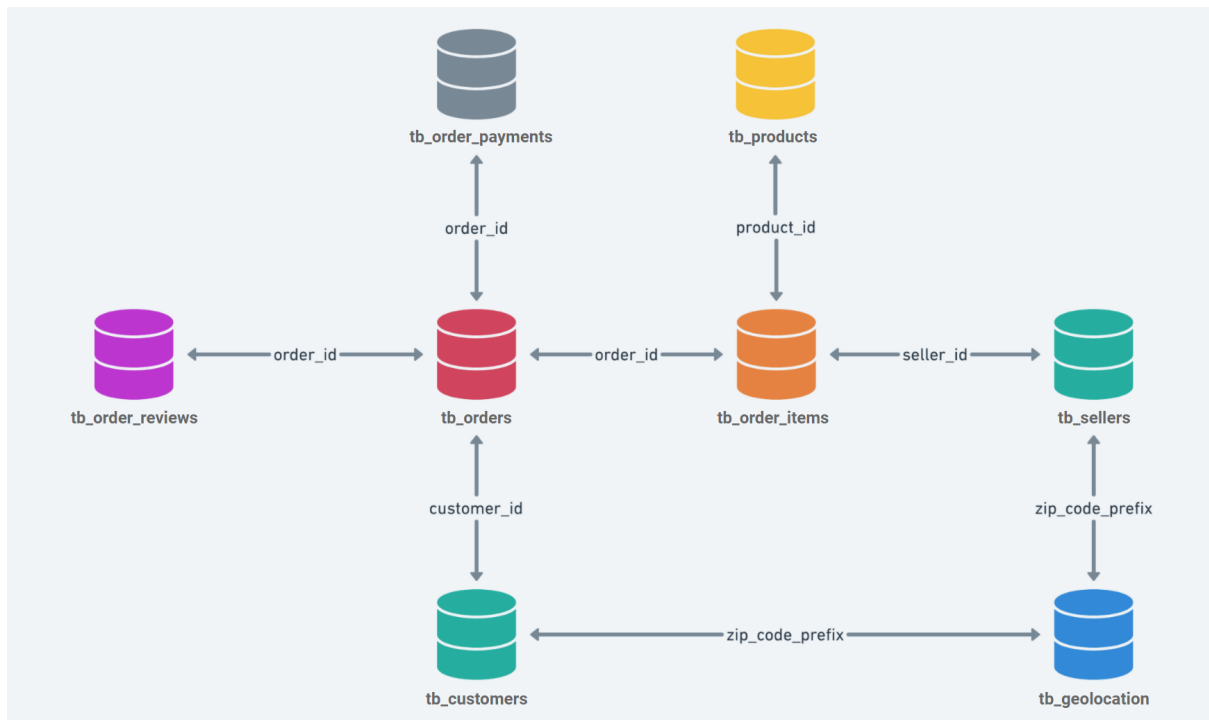
Neste desafio, você está recebendo um banco de dados da Olist, uma plataforma de ecommerce. Suas análises não serão utilizadas para tomada de decisão, mas se envolva nesse desafio como se fossem. Esse tipo de análise é utilizada quase todo dia na rotina de quem tá no mercado de trabalho e começar a carreira já tendo alguma experiência vai ser um baita diferencial.



DOCUMENTAÇÃO DA TABELA

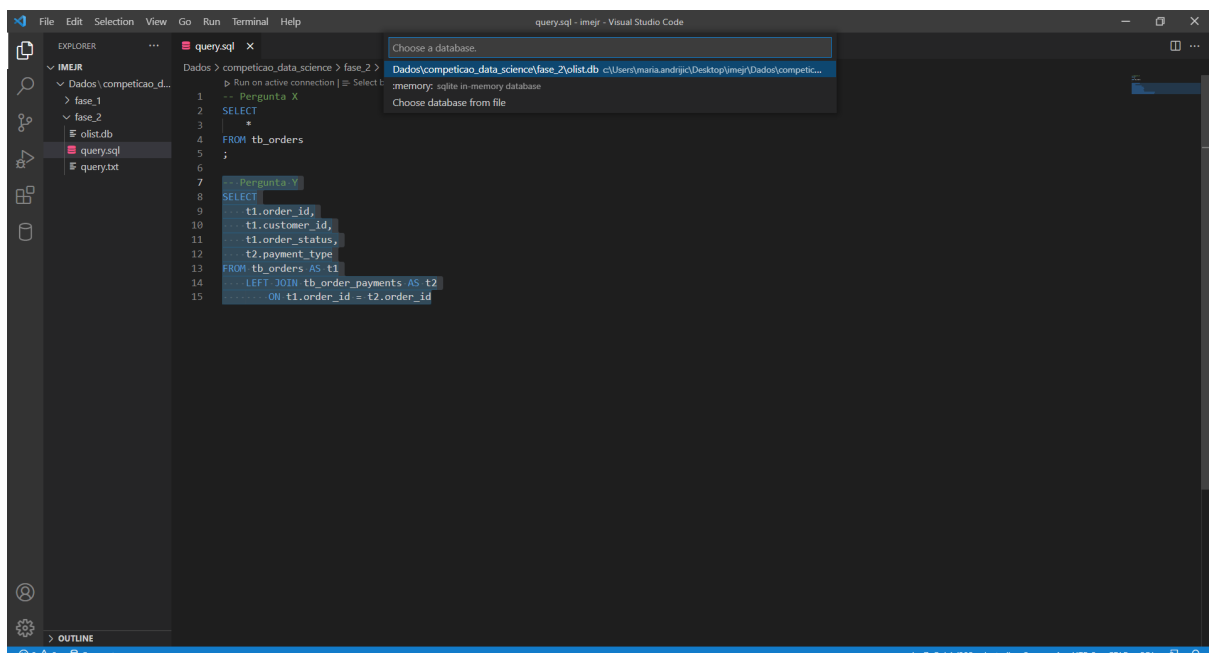
As tabelas foram retiradas do Kaggle e convertidas em um .db.

Toda a [documentação das tabelas se encontra no Kaggle](#), incluindo a descrição de cada coluna em cada tabela. As tabelas no banco de dados foram renomeadas. Segue o esquema igual ao da documentação do Kaggle com os nomes das tabelas atualizados:



INSTRUÇÕES

Após baixar o banco de dados, escreva sua query em um arquivo .sql. Para rodar, selecione a query e, com o botão direito, clique em “Run Query”. Selecione o olist.db e veja a mágica acontecer!



PERGUNTAS

Utilize os dados recebidos para responder às perguntas abaixo.

Cada pergunta possui uma pontuação específica - não divulgada. Busque responder à maior quantidade de questões que conseguir para alcançar mais pontos.

Procure escrever códigos limpos e comente explicações da resposta.

1. Quantas compras foram feitas no total?
2. Quantos vendedores temos? E quantos compradores?
3. Qual o produto vendido mais pesado?
4. Quantas compras foram entregues?
5. Qual o ticket médio de cada método de pagamento?
6. Qual a média de itens por compra?
7. Qual estado possui mais compras? E qual possui menos?
8. Qual estado possui mais vendas? E qual possui menos?
9. Qual foi o valor total das compras realizadas no primeiro semestre de 2018?
10. Há pedidos reembolsados? Se sim, quantos?
11. Ordene os estados pelo tempo de demora da chegada do pedido do maior pro menor
12. Ordene as categorias com mais vendas da mais vendida pra menos vendida
13. Quais os 10 compradores que compraram mais itens?
14. Quais são os 5 vendedores com maior review score?
15. Qual a taxa de conversão de cada método de pagamento?
16. Qual é o mês do ano com maior média de vendas?
17. Qual é o método de pagamento mais usado em cada estado? Qual é o share (porcentagem do total) desse método de pagamento?
18. Qual a porcentagem de clientes que já usou mais de um método de pagamento diferente?
19. Liste os 10 melhores vendedores e seus respectivos estados
20. Escolha o output de alguma pergunta ou crie uma tabela própria e construa um dashboard no Google Data Studio a partir dela - o dash pode ter mais de uma tabela

VISUALIZAÇÃO DE DADOS

INTRODUÇÃO

Nesta etapa, você irá desenvolver suas habilidades de comunicação, habilidades artísticas, habilidades técnicas para criação de gráficos e - claro -, sua criatividade.

A ideia é que você corra atrás de dados - ou crie - para criação de visualizações.

RECEBÍVEIS

Recomendações: [planilha Google](#)

Exemplos de códigos: [repositório no GitHub](#)

Sugestões de bases de dados:

[Preço da Netflix em diferentes Países](#)

[Dataset completo sobre Covid-19](#)

[Consumo de energia no Mundo](#)

[Abordagens da polícia no tráfico de Rhode Island](#)

ENTREGÁVEIS: EM ATÉ 19 DIAS APÓS RECEBIMENTO

Construa a máxima quantidade de visualizações que você puder. Quanto mais visualizações boas fizer, mais pontos e experiência ganhará. Das visualizações que você fizer, armazene o código e o gráfico. Você pode criar slides para treinar um pouco o storytelling, um notebook com os gráficos nele, ou apenas os arquivos de imagem.

Crie seus próprios datasets, utilize os que sugerimos, ou busque outros para criar bons gráficos em cima. Você pode extrair datasets da etapa de SQL dessa fase 2 ou até mesmo utilizar o dataset da fase 1!

Exercite sua criatividade e habilidades artísticas!

Você deverá entregar todos os arquivos - os código e os gráficos - via e-mail.

Crie.