

*SQL Para Data Science*

# SQL Para Data Science

## Exercício Cenário de Análise de Dados 4



Este exercício é praticamente um projeto e poderá ser um pouco desafiador, pois, vamos unir todo o conhecimento abordado no curso até aqui. Siga as instruções abaixo com bastante atenção.

### 1- Download dos dados

Vamos trabalhar com dados reais disponíveis publicamente. Acesse o link abaixo para fazer o download dos arquivos (será necessário criar sua conta gratuita no Kaggle). Baixe todos os arquivos csv.

<https://www.kaggle.com/nosbielcs/brazilian-delivery-center>

### 2- Leia a descrição dos dados

O link acima, além dos dados para download, contém a descrição dos dados e detalhes sobre o cenário em que os dados foram extraídos. Leia com atenção para compreender o que cada dataset representa. O texto está em português.

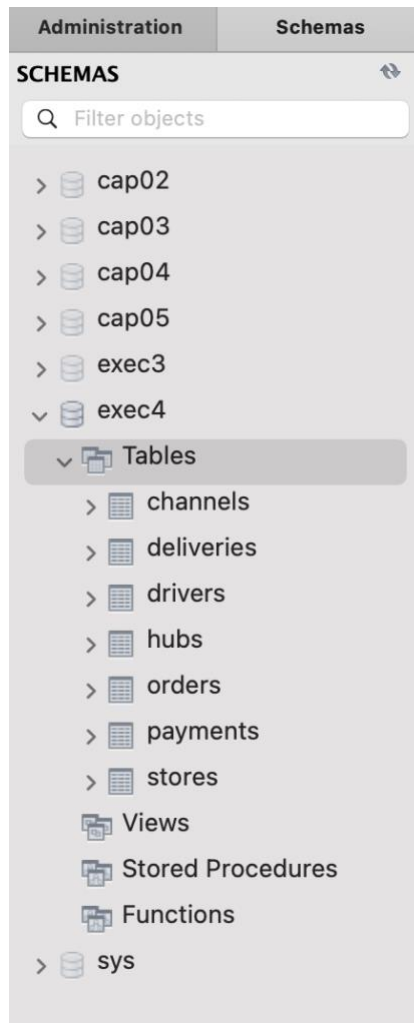
### 3- Carregue os dados no banco de dados

Você pode usar o **Table Import Wizard** (que foi mostrado na solução do exercício 3) e importar os dados no MySQL. Crie uma tabela para cada arquivo (pode usar o nome da tabela como o mesmo nome de cada arquivo). A carga de dados deve funcionar sem problemas.

**Mas há alguns detalhes:**

- Alguns arquivos têm milhões de registros e o tempo de carga pode ser bem alto dependendo do seu computador. Seja paciente e aguarde cada arquivo ser carregado. A lentidão é causada pelo MySQL Workbench. Caso queira acelerar o processo faça a carga via linha de comando conforme mostrado na solução do exercício 3.
- Um dos arquivos, o arquivo **hubs.csv**, tem texto com caracteres especiais e você terá problemas pelo MySQL Workbench. Adotamos como solução simples, editar o arquivo com um editor de texto (usamos o Sublime Text) e alteramos todas as linhas de SÃO PAULO para SAO PAULO (sem acento). Isso resolve o problema e os dados podem ser carregados sem dificuldade.
- O arquivo **deliveries.csv** pode apresentar erro de carga em um dos IDs de driver. Não há problema e é só ignorar. Mas se quiser corrigir, abra o arquivo em um editor de texto, busque por duas vírgulas juntas e substitua por isso: ,NULL,

Após carregar todos os arquivos seu ambiente deve ser parecer com a imagem abaixo.



Trabalhar com dados reais não é fácil, mas o resultado é um aprendizado valioso.

Agora você estará pronto para criar suas queries SQL que respondam às perguntas abaixo. A solução será apresentada no começo do próximo capítulo. Bom trabalho.

- # 1- Qual o número de hubs por cidade?
- # 2- Qual o número de pedidos (orders) por status?
- # 3- Qual o número de lojas (stores) por cidade dos hubs?
- # 4- Qual o maior e o menor valor de pagamento (payment\_amount) registrado?
- # 5- Qual tipo de driver (driver\_type) fez o maior número de entregas?
- # 6- Qual a distância média das entregas por tipo de driver (driver\_modal)?
- # 7- Qual a média de valor de pedido (order\_amount) por loja, em ordem decrescente?
- # 8- Existem pedidos que não estão associados a lojas? Se caso positivo, quantos?
- # 9- Qual o valor total de pedido (order\_amount) no channel 'FOOD PLACE'?
- # 10- Quantos pagamentos foram cancelados (chargeback)?
- # 11- Qual foi o valor médio dos pagamentos cancelados (chargeback)?
- # 12- Qual a média do valor de pagamento por método de pagamento (payment\_method) em ordem decrescente?
- # 13- Quais métodos de pagamento tiveram valor médio superior a 100?
- # 14- Qual a média de valor de pedido (order\_amount) por estado do hub (hub\_state), segmento da loja (store\_segment) e tipo de canal (channel\_type)?
- # 15- Qual estado do hub (hub\_state), segmento da loja (store\_segment) e tipo de canal (channel\_type) teve média de valor de pedido (order\_amount) maior que 450?
- # 16- Qual o valor total de pedido (order\_amount) por estado do hub (hub\_state), segmento da loja (store\_segment) e tipo de canal (channel\_type)? Demonstre os totais intermediários e formate o resultado.
- # 17- Quando o pedido era do Hub do Rio de Janeiro (hub\_state), segmento de loja 'FOOD', tipo de canal Marketplace e foi cancelado, qual foi a média de valor do pedido (order\_amount)?
- # 18- Quando o pedido era do segmento de loja 'GOOD', tipo de canal Marketplace e foi cancelado, algum hub\_state teve total de valor do pedido superior a 100.000?

# 19- Em que data houve a maior média de valor do pedido (order\_amount)? Dica: Pesquise e use a função SUBSTRING().

# 20- Em quais datas o valor do pedido foi igual a zero (ou seja, não houve venda)? Dica: Use a função SUBSTRING().

Divirta-se. 😊