

## Capítulo 7/8

### Projeto do Curso

Parabéns! Você concluiu o curso de SQL. É hora de aplicar o conhecimento e as habilidades que você adquiriu em um projeto: um estudo de caso analítico da vida real que você concluirá por conta própria.

Quando você terminar o projeto, envie seu trabalho para o revisor do projeto para avaliação. Ele te dará feedback dentro de 48 horas. Use o feedback para fazer alterações e, em seguida, envie a nova versão de volta ao revisor do projeto.

Você pode obter mais feedback sobre a nova versão. Isso é completamente normal. Não é incomum passar por vários ciclos de feedback e revisão.

Seu projeto será considerado concluído assim que o revisor do projeto o aprovar.

### Descrição do Projeto

Você está trabalhando como analista da Zuber, uma nova empresa de compartilhamento de caronas que está sendo lançada em Chicago. Sua tarefa é encontrar padrões nas informações disponíveis. Você quer entender as preferências dos passageiros e o impacto de fatores externos nas corridas.

Trabalhando com um banco de dados, você analisará dados de concorrentes e testará uma hipótese sobre o impacto do clima na frequência das viagens.

### Descrição dos dados

Um banco de dados com informações sobre corridas de táxi em Chicago:

tabela `neighborhoods` : dados sobre os bairros da cidade

- `name` : nome do bairro
- `neighborhood_id` : código do bairro

tabela `cabs` : dados sobre os táxis

- `cab_id` : código do veículo
- `vehicle_id` : a identificação técnica do veículo
- `company_name` : a empresa proprietária do veículo

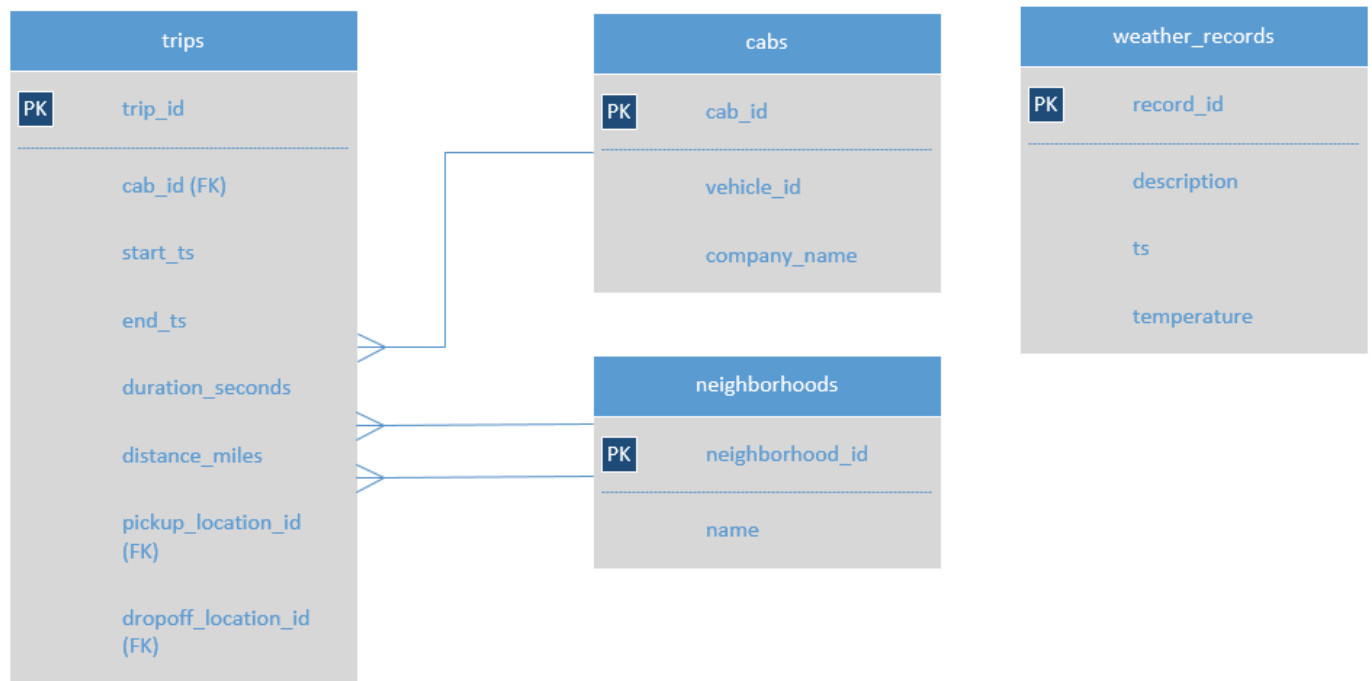
tabela `trips` : dados sobre corridas

- `trip_id` : código da corrida
- `cab_id` : código do veículo que opera a corrida
- `start_ts` : data e hora do início da corrida (tempo arredondado para a hora)
- `end_ts` : data e hora do final da corrida (tempo arredondado para a hora)
- `duration_seconds` : duração da corrida em segundos
- `distance_miles` : distância percorrida em milhas
- `pickup_location_id` : código do bairro de retirada
- `dropoff_location_id` : código do bairro de entrega

tabela `weather_records` : dados sobre o clima

- `record_id` : código de registro meteorológico
- `ts` : grava data e hora (tempo arredondado para a hora)
- `temperature` : temperatura quando o registro foi feito
- `description` : breve descrição das condições meteorológicas, ex. "chuva leve" ou "nuvens esparsas"

## Esquema de tabela



Nota: não há uma conexão direta entre as tabelas `trips` e `weather_records` no banco de dados. Mas você ainda pode usar JOIN e juntá-las usando a hora em que o passeio começou ( `trips.start_ts` ) e a hora em que o registro do tempo foi feito ( `weather_records.ts` ).

## Instruções para completar o projeto

**Passo 1.** Escreva um código para analisar os dados sobre o clima em Chicago em novembro de 2017 no site:

[https://practicum-content.s3.us-west-1.amazonaws.com/data-analyst-eng/moved\\_chicago\\_weather\\_2017.html](https://practicum-content.s3.us-west-1.amazonaws.com/data-analyst-eng/moved_chicago_weather_2017.html)

## Passo 2. Análise Exploratória de Dados

1. Encontre o número de corridas de táxi para cada empresa de táxi de 15 a 16 de novembro de 2017. Nomeie o campo resultante como `trips_amount` e imprima-o junto com o campo `company_name` . Ordene os resultados pelo campo `trips_amount` em ordem decrescente.
2. Encontre o número de corridas para cada empresa de táxi cujo nome contém as palavras "Yellow" ou "Blue" ("Amarelo" ou "Azul", respectivamente) de 1º a 7 de novembro de 2017. Nomeie a variável resultante como `trips_amount` . Agrupe os resultados pelo campo `company_name` .
3. Em novembro de 2017, as empresas de táxi mais populares eram Flash Cab e Taxi Affiliation Services. Encontre o número de corridas para essas duas

empresas e nomeie a variável resultante como `trips_amount`. Junte as corridas para todas as outras empresas no grupo "Other". Agrupe os dados por nomes de empresas de táxi. Nomeie o campo com os nomes das empresas de táxi como `company`. Classifique o resultado em ordem decrescente por `trips_amount`.

### Passo 3. Teste a hipótese de que a duração das corridas do Loop até ao Aeroporto Internacional O'Hare muda em sábados chuvosos.

1. Recupere os identificadores dos bairros O'Hare e Loop da tabela `neighborhoods`.
2. Para cada hora, recupere os registros de condições meteorológicas da tabela `weather_records`. Usando o operador `CASE`, divida todas as horas em dois grupos: "Bad" se o campo `description` contiver as palavras "rain" (chuva) ou "storm" (tempestade) e "Good" para outros. Nomeie o campo resultante como `weather_conditions`. A tabela final deve incluir dois campos: data e hora (`ts`) e `weather_conditions`.
3. Recupere da tabela `trips` todas as corridas que começaram no Loop (`neighborhood_id : 50`) e terminaram em O'Hare (`neighborhood_id : 63`) em um sábado. Obtenha as condições meteorológicas para cada corrida. Use o método que você aplicou na tarefa anterior. Também recupere a duração de cada corrida. Ignore corridas para as quais os dados sobre as condições meteorológicas não estão disponíveis.

### Passo 4. Análise exploratória de dados (Python)

Além dos dados recuperados nas tarefas anteriores, você recebeu um segundo arquivo. Agora você tem estes dois CSVs:

`project_sql_result_01.csv`. Ele contém os seguintes dados:

- `company_name` : nome da empresa de táxi
- `trips_amount` : o número de corridas para cada empresa de táxi de 15 a 16 de novembro de 2017.

`project_sql_result_04.csv`. Ele contém os seguintes dados:



Para esses dois conjuntos de dados, agora você precisa:

- importar os arquivos
- estudar os dados que eles contêm
- verifique se os tipos de dados estão corretos
- identificar os 10 principais bairros em termos de destinos
- fazer gráficos: empresas de táxi e número de corridas, 10 principais bairros por número de corridas em que esse - bairro é destino
- tirar conclusões com base em cada gráfico e explicar os resultados

## Passo 5. Testando hipóteses (Python)

`project_sql_result_07.csv` —o resultado da última consulta. Ele contém dados sobre viagens do Loop para o Aeroporto Internacional O'Hare. Lembre-se, estes são os valores dos campos da tabela:

- `start_ts` — data e hora do começo da corrida
- `weather_conditions` — condições meteorológicas no momento em que a corrida começou
- `duration_seconds` — duração da viagem em segundos

Teste a hipótese:

"A duração média dos passeios do Loop para o Aeroporto Internacional O'Hare muda nos sábados chuvosos."

Defina o valor do nível de significância (alfa) por conta própria.

Explique:


- como você formou as hipóteses nula e alternativa
- qual critério você usou para testar a hipótese e porque

## Como o meu projeto será avaliado?

Aqui estão os critérios de avaliação do projeto. Leia-os cuidadosamente antes de começar a trabalhar.

Aqui está o que o revisor do projeto procurará ao avaliar seu projeto:

- como você recupera dados do site

- como você faz fatias de dados
- como você agrupa os dados  Avaliar
- se você usa os métodos corretos para unir tabelas
- como você formula hipóteses
- quais critérios você usa para testar as hipóteses e por quê
- a que conclusões você chega
- se você deixa comentários a cada passo

As folhas de conclusões e resumos das lições anteriores têm tudo o que você precisa para completar o projeto.

Boa sorte!