**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

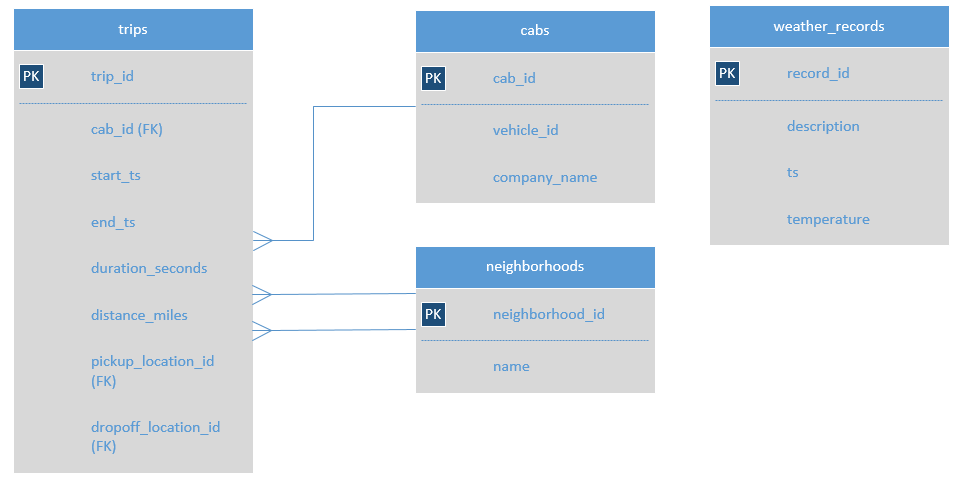
tabela trips: dados sobre corridas

* 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>

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 todos 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:

* dropoff\_location\_name: bairros de Chicago onde as corridas terminaram
* average\_trips: o número médio de viagens que terminaram em cada bairro em novembro de 2017.

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