



Theory

The Zuber Database

You're working as an analyst for Zuber, a new ride-sharing company that's launching in Chicago. Your task is to find patterns in the available information. You want to understand passenger preferences and the impact of external factors on rides.

You'll study a database, analyze data from competitors, and investigate the impact of weather on ride frequency.

Description of the data

A database with info on taxi rides in Chicago:

neighborhoods table: data on city neighborhoods

- *name*: name of the neighborhood
- *neighborhood_id*: neighborhood code

cabs table: data on taxis

- *cab_id*: vehicle code
- *vehicle_id*: the vehicle's technical ID
- *company_name*: the company that owns the vehicle

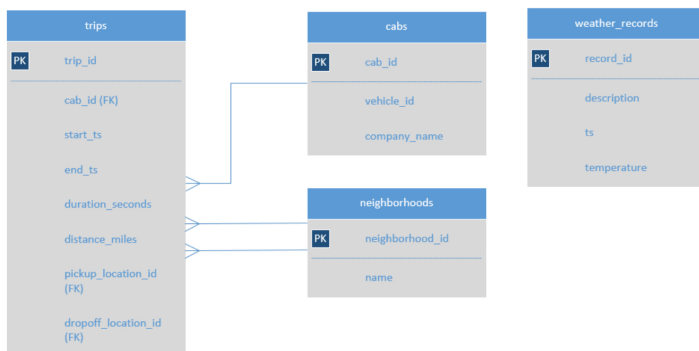
trips table: data on rides

- *trip_id*: ride code
- *cab_id*: code of the vehicle operating the ride
- *start_ts*: date and time of the beginning of the ride (time rounded to the hour)
- *end_ts*: date and time of the end of the ride (time rounded to the hour)
- *duration_seconds*: ride duration in seconds
- *distance_miles*: ride distance in miles
- *pickup_location_id*: pickup neighborhood code
- *dropoff_location_id*: dropoff neighborhood code

weather_records table: data on weather

- *record_id*: weather record code
- *ts*: record date and time (time rounded to the hour)
- *temperature*: temperature when the record was taken
- *description*: brief description of weather conditions, e.g. "light rain" or "scattered clouds"

Table scheme



Note: there isn't a direct connection between the tables *trips* and *weather_records* in the database. But you can still use JOIN and link them using the time the ride started (*trips.start_ts*) and the time the weather record was taken (*weather_records.ts*).

You've already done the first part of the project: you wrote a code to parse the

```
2     CASE
3     WHEN c.company_name IN ('Flash Cab', 'Taxi Affiliation
Services') THEN c.company_name
4     ELSE 'Other'
5     END AS company,
6     COUNT(t.trip_id) AS trips_amount
7 FROM
8     trips t
9 JOIN
10    cabs c ON t.cab_id = c.cab_id
11 WHERE
12    t.start_ts BETWEEN '2017-11-01 00:00:00' AND '2017-11-07 23:59:59'
13 GROUP BY
14    company
15 ORDER BY
```

Result

weather data from a website. Now you'll do the second and third parts:

Tasks 1-4: Exploratory data analysis

Tasks 5-7: Investigate whether the duration of rides from the the Loop to O'Hare International Airport changes on rainy Saturdays

Task 3 / 6



Passed
You can continue

Run

Submit



Rate



Task 2

Task 4

