99 Data Analytics Challenge

Dear candidate, we are so excited with your interest in working with us! This challenge is an opportunity for us to know a little bit of the great talent we know you have. It was built to simulate real-case scenarios that you would face while working at 99 and is organized in 2 parts:

1. A technical part of close-ended questions with specific answers that are meant to assess your ability to **analyze large amounts of data with SQL to answer key questions**.

2. An analytical part of open-ended questions to assess your ability to **build data-backed recommendations to support decision making**. Expect further questions and discussions on top of your answers in the next phase of our hiring process.

In order to answer these questions, we built a synthetic dataset and made it available online in a PostgreSQL database. You can find the entity-relationship model and data dictionary, as well as connection instructions below.

There's no time limit, but we estimate this challenge should take around 3-4 hours of work for the mandatory questions. We tried to make it short and enjoyable, but be mindful of your answers. Some tips include: pay attention to the order of calculations that can have an impact on the final answer, and always inspect your data. We will give you a bonus point if you let us know what kind of checks you performed. As of delivery format our preference is for a set of slides, but feel free to use any form you find most clear and concise.

We hope to hear back from you soon.

Good luck!

Database information

Connection parameters

Host	candidates99.cqela5ipg9g5.us-east-1.rds.amazonaws.com
Port	5432
Database	postgres
User	candidate
Password	L3b9ePTN6c
Version	v12.5

Resources:

- [Connection Guide](#) by AWS
- Popular clients: [pandas+sqlalchemy](#), [DBeaver](#), [pgAdmin](#)
- Or, if you prefer to use psql:

      psql --host=candidates99.cqela5ipg9g5.us-east-1.rds.amazonaws.com --port=5432 --username=candidate --password --dbname=postgres

ER Model
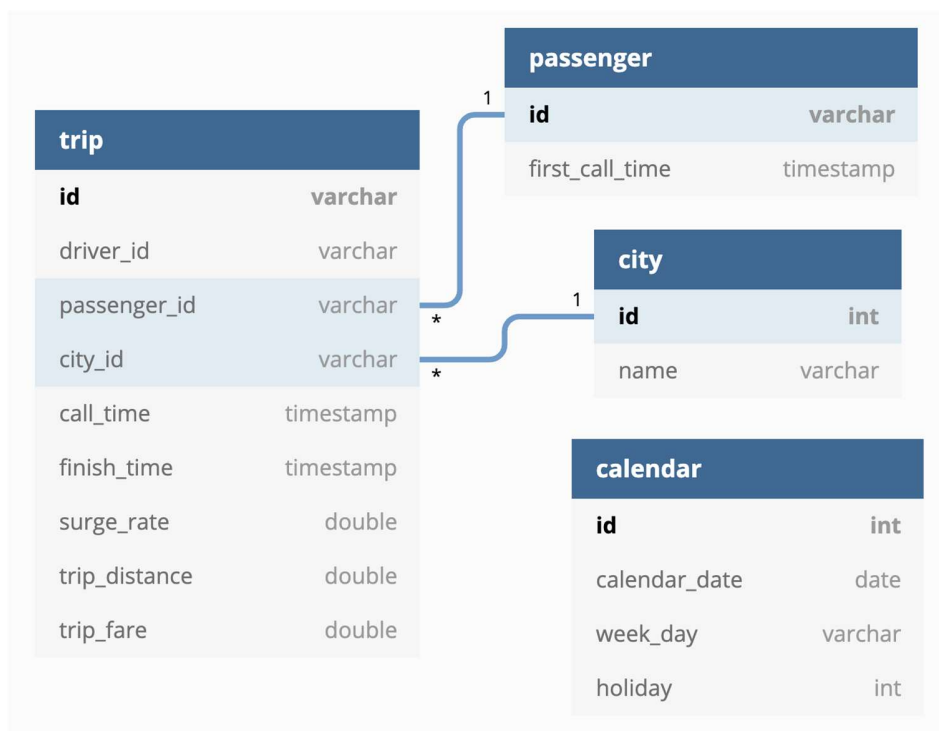
Data dictionary

calendar

Column Name	Column Type	Description
id [pk]	integer	Identification number of calendar date
calendar_date	date	Calendar date
week_day	varchar	The name of the week day
holiday	integer	Binary flag to indicate if the calendar date is a holiday (1) or non-holiday (0)

city

Column Name	Column Type	Description
id [pk]	integer	Unique identifier of city
name	varchar	City name

passenger

Column Name	Column Type	Description
id [pk]	varchar	Identification of passenger
first_call_time	datetime	Date and time of the first call of a passenger

trip

Column Name	Column Type	Description
id [pk]	varchar	Identification of trip
driver_id	varchar	Identification of driver
passenger_id	varchar	Identification of passenger
city_id	varchar	Identification of city
call_time	datetime	Datetime in which the trip was called by the passenger
finish_time	datetime	Datetime in which the trip was finished
surge_rate	double	Additional multiplier of base fare. 0 means no additional surge, 0.3 means 1.3x
trip_distance	double	Trip distance
trip_fare	double	Final trip cost

Questions

Part I - Technical

Provide both the answer and the SQL code used.

1. What is the average trip cost of holidays? How does it compare to non-holidays?

2. Find the average call time (the time in which a trip was requested) of the first time passengers make a trip.

3. Find the average number of trips per driver for every week day.

4. Which day of the week drivers usually drive the most distance on average?

5. What was the growth percentage of rides month over month?

6. *Optional*. List the top 5 drivers per number of trips in the top 5 largest cities.

Part II - Analytical

99 is a marketplace, where drivers are the supply and passengers, the demand. One of our main challenges is to keep this marketplace balanced. If there's too much demand, prices would increase due to surge and passengers would prefer not to run. If there's too much supply, drivers would spend more time idle impacting their revenue.

1. Let's say it's 2019-09-23 and a new Operations manager for *The Shire* was just hired. She has 5 minutes during the Ops weekly meeting to present an overview of the business in the city, and since she's just arrived, she asked your help to do it. What would you prepare for this 5 minutes presentation? Please provide 1-2 slides with your idea.

2. She also mentioned she has a budget to invest in promoting the business. What kind of metrics and performance indicators would you use in order to help her decide if she should invest it into the passenger side or the driver side? Extra point if you provide data-backed recommendations.

3. One month later, she comes back, super grateful for all the helpful insights you have given her. And says she is anticipating a driver supply shortage due to a major concert that is going to take place the next day and also a 3 day city holiday that is coming the next month. What would you do to help her analyze the best course of action to either prevent or minimize the problem in each case?

4. *Optional*. We want to build up a model to predict "Possible Churn Users" (e.g.: no trips in the past 4 weeks). List all features that you can think about and the data mining or machine learning model or other methods you may use for this case.