



## Internal Audit Analytics Exercise

These exercises are intended for Uber and/or Uber employees only. They are confidential and not to be circulated, shared or redistributed in any format, in part or in their entirety. This includes both the exercises and your answers, analysis and code.

---

Understanding the needs of key stakeholders and performing analysis/prototyping solutions to help them better perform in their role is a key undertaking of the Internal Audit Core Analytics & Science team. In this exercise, you will be using your data analysis skills to analyze EMEA rideshare data and use it to draw a conclusion for the 2020 strategy for the Riyadh market.

### Part 1: SQL

In our day-to-day setting, one of the initial steps in approaching a problem is locating, cleaning, and compiling data from Uber's datastore. Using the CSV sample of trips in Riyadh as your end result and the table schema on the next page, write a single SQL query that would return the provided information. Feel free to adopt any SQL dialect.

### Part 2: Analysis & Presentation

- Please perform an analysis using the dataset provided in Part 1. The goal is to suggest action(s) or change(s) to your stakeholder through your findings. Use R or Python to generate an analysis that could be shared and reviewed by other analysts.
- Prepare a presentation that clearly explains your findings, conclusions and recommendations based on this analysis.

### Part 3: Tooling

- Use the data from Part 1 as an input to build a visualisation tool in R or Python that would help your stakeholders understand and monitor the suggestion that comes out of your analysis over time.
- You should be comfortable explaining how the tool would be utilised by your stakeholders. updated on an on-going basis.
- We will award bonus points to candidates who create the tool in a scaleable manner such that it can be updated on an on-going basis, will be able to handle interchangeable and full population datasets.

Note: If asked to come to an on-site interview, you will have the opportunity to present the tool and your process to the interview panel which will be a cross functional group of stakeholders at Uber.

---

## Table Schema for Part 1

Table 1: trips

| Column Name         | Datatype  |
|---------------------|-----------|
| city_id             | int       |
| completed_trip      | boolean   |
| distance_to_pickup  | int       |
| driver_id           | str       |
| dropoff_geo         | str       |
| dropoff_local_time  | timestamp |
| dropoff_utc_time    | timestamp |
| entered_destination | boolean   |
| esttime_to_pickup   | int       |
| pickup_geo          | str       |
| pickup_local_time   | timestamp |
| pickup_utc_time     | timestamp |
| request_geo         | str       |
| request_local_time  | timestamp |
| request_type        | str       |
| request_utc_time    | timestamp |
| rider_id            | str       |
| surged_trip         | boolean   |
| time_to_pickup      | int       |
| trip_id             | str       |
| trip_status         | str       |
| vehicle_id          | int       |

Table 2: cities

| Column Name   | Datatype |
|---------------|----------|
| city_id       | int      |
| city_name     | str      |
| country_id    | int      |
| country_name  | str      |
| distance_unit | str      |
| lat           | float    |
| lng           | float    |

|                |     |
|----------------|-----|
| local_currency | str |
|----------------|-----|

Table 3: riders

| Column Name        | Datatype  |
|--------------------|-----------|
| active_city_id     | int       |
| first_trip_id      | str       |
| preferred_language | str       |
| rider_app          | str       |
| rider_device       | str       |
| rider_email        | str       |
| rider_id           | str       |
| rider_trip_count   | int       |
| signup_date        | timestamp |

Table 4: drivers

| Column Name        | Datatype  |
|--------------------|-----------|
| active_city_id     | int       |
| driver_app         | str       |
| driver_device      | str       |
| driver_email       | str       |
| driver_id          | str       |
| driver_trip_count  | int       |
| first_trip_id      | str       |
| preferred_language | str       |
| signup_date        | timestamp |
| vehicle_ids        | array     |

Table 5: bills

| Column Name      | Datatype |
|------------------|----------|
| bill_id          | str      |
| cancel_fee_local | float    |
| cancel_fee_usd   | float    |

|                       |         |
|-----------------------|---------|
| completed_trip        | boolean |
| driver_id             | str     |
| entered_destination   | boolean |
| exchange_rate         | float   |
| local_currency        | str     |
| paid_cash             | boolean |
| partner_id            | str     |
| payment_type          | str     |
| product_category      | str     |
| request_type          | str     |
| rider_id              | str     |
| surged_trip           | boolean |
| trip_distance_miles   | float   |
| trip_duration_seconds | int     |
| trip_fare_local       | float   |
| trip_fare_usd         | float   |
| trip_id               | str     |

Table 6: Vehicles

| Column Name        | Datatype |
|--------------------|----------|
| seat_count         | int      |
| vehicle_color      | str      |
| vehicle_id         | int      |
| vehicle_trip_count | int      |
| vehicle_type       | str      |