

Case Study: Analyzing Rides Data in SQLite

Objective:

The goal of this case study is to analyze data from three related tables: `customers`, `rides`, and `drivers`, using SQLite. These tables represent the interactions between customers who book rides and the drivers who complete them. You will explore the dataset to answer specific business questions, such as identifying the month with the highest number of rides and determining the best-performing driver(s).

Provided Data:

You have been provided with 3 tables formatted as 3 CSV files:

1. `customers.csv`

- Represents individuals who use the ride-hailing service. Each customer has a unique ID and basic personal information.
- **Fields:**
 - `id`: Unique identifier for each customer (Primary Key).
 - `first_name`: The customer's first name.
 - `last_name`: The customer's last name.

2. `drivers.csv`

- Represents drivers working for the ride-hailing service. Each driver has a unique ID and personal information such as their name and country code.
- **Fields:**
 - `id`: Unique identifier for each driver (Primary Key).
 - `full_name`: The driver's full name.
 - `country_code`: The driver's country of operation.

3. `rides.csv`

- Contains the details of each ride taken by customers, including the customer who booked the ride and the driver who completed it.
- **Fields:**

- **id**: Unique identifier for each ride (Primary Key).
- **customer_id**: Refers to the customer who booked the ride (Foreign Key referencing **customers.id**).
- **driver_id**: Refers to the driver who completed the ride (Foreign Key referencing **drivers.id**).
- **rating**: The rating provided by the customer for the ride (e.g., 1-5 scale).
- **is_driver_late**: A boolean flag indicating whether the driver was late for the ride (**true** if the driver was late, **false** if the driver was not late).
- **date**: The date and time the ride took place.

How They Relate:

- **Customers** book **rides** through the platform, and their IDs are stored in the **rides** table.
- **Drivers** fulfill these **rides**, and their IDs are also stored in the **rides** table.
- The **rides** table acts as a connection between the **customers** and **drivers**, linking each ride to the corresponding customer and driver.

You will import these CSV files into SQLiteOnline to perform the analysis.

Instructions for Importing CSV Data into SQLiteOnline:

Step 1: Download the CSV Files

1. Download the provided CSV files:
 - customers.csv
 - rides.csv
 - drivers.csv

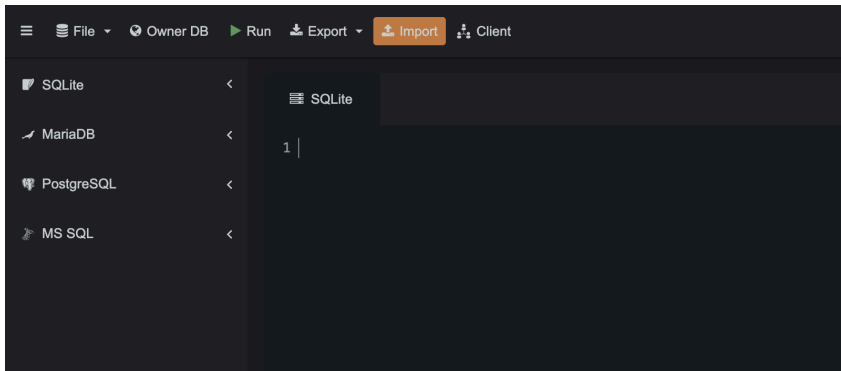
Step 2: Open SQLite Online

1. Go to <https://sqliteonline.com/>.

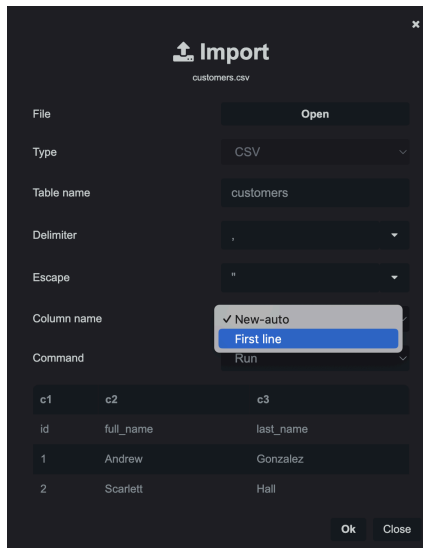
Step 3: Import the CSVs as tables

For each of the 3 files:

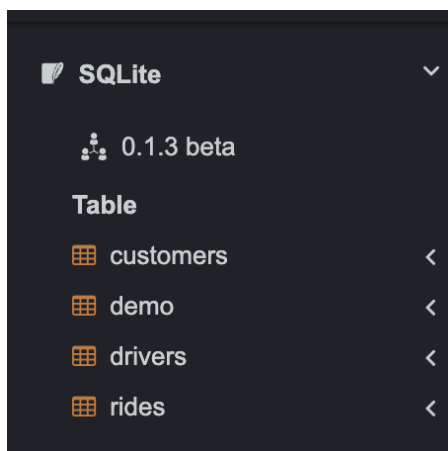
1. Click **Import > Open > file.csv**



2. For Column name, replace **New-auto** with **First line**, then click **Ok**. No need to change anything else.

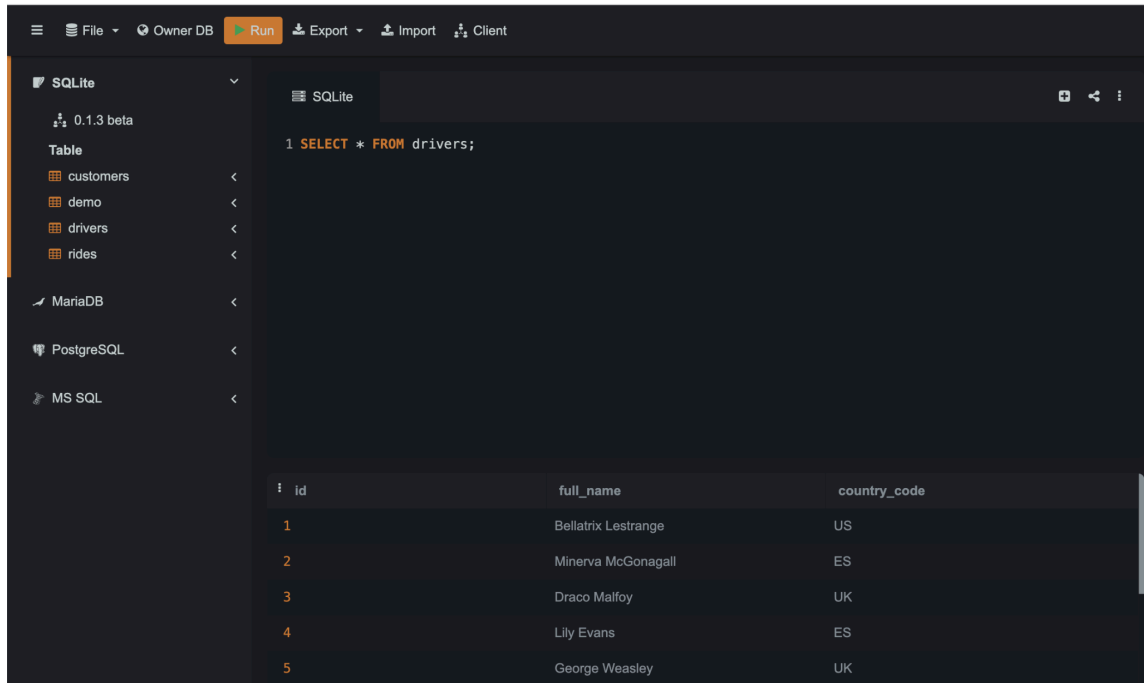


3. The tables should appear on the left column (ignore the **demo** table)



Step 4: Run a test query

1. In the console, write `SELECT * FROM drivers;` and click **Run**. It should return the following:



The screenshot shows a database client interface with a dark theme. On the left, a sidebar lists database types: SQLite (selected), MariaDB, PostgreSQL, and MS SQL. Under SQLite, there's a version '0.1.3 beta' and a 'Table' section listing 'customers', 'demo', 'drivers', and 'rides'. The main area displays the SQL query `1 SELECT * FROM drivers;` in a text editor. Below the editor, the results of the query are shown in a table with three columns: 'id', 'full_name', and 'country_code'. The table contains five rows of data.

id	full_name	country_code
1	Bellatrix Lestrange	US
2	Minerva McGonagall	ES
3	Draco Malfoy	UK
4	Lily Evans	ES
5	George Weasley	UK

Tasks:

Task 1: Determine the Month with the Highest Number of Rides and the Month with the Lowest Number of Rides

- Write a query to find which month had the highest number of rides. Do the same for the month with the lowest number of rides.
- **Hint:** *In SQLite, we use the function `strftime` to extract a part of a date. For instance `strftime('%Y-%m-%d', date)` will return the day of the date time string*

Task 2: Who do you Consider to be the Best Performing Driver(s), and Why?

- Define what makes a "best performing driver" based on the data available.
 - Justify your criteria(s) for the best performing driver(s) and explain how you explored the data to reach your conclusion.
-

Submission

- Submit the SQL queries, exploration summary, and conclusions in a **Google Doc**, **Google Slides** or **PDF** format. If you use a Google Doc or Google Slides, make sure to make it accessible to anyone with the link and share the link.
- Include **screenshots** of the query results from SQLiteOnline to support your findings where applicable.