

Insights from Data with BigQuery: Challenge Lab

GSP787



Google Cloud Self-Paced Labs

Overview

You must complete a series of tasks within the allocated time period. Instead of following step-by-step instructions, you'll be given a scenario and a set of tasks - you figure out how to complete it on your own! An automated scoring system (shown on this page) will provide feedback on whether you have completed your tasks correctly.

To score 100% you must complete all tasks within the time period!

When you take a Challenge Lab, you will not be taught Google Cloud concepts. To build the solution to the challenge presented, use skills learned from the labs in the quest this challenge lab is part of. You will be expected to extend your learned skills; you will be expected to change broken queries.

This lab is recommended for students who have enrolled in the [Insights from Data with BigQuery](#) quest. Are you ready for the challenge?

Scenario

You're part of a public health organization which is tasked with identifying answers to queries related to the Covid-19 pandemic. Obtaining the right answers will help the organization in planning and focusing healthcare efforts and awareness programs appropriately.

The dataset and table that will be used for this analysis will be : `bigquery-public-data.covid19_open_data.covid19_open_data`. This repository contains country-level datasets of daily time-series data related to COVID-19 globally. It includes data relating to demographics, economy, epidemiology, geography, health, hospitalizations, mobility, government response, and weather.

Setup and Requirements

Before you click the Start Lab button

Read these instructions. Labs are timed and you cannot pause them. The timer, which starts when you click **Start Lab**, shows how long Google Cloud resources will be made available to you.

This Qwiklabs hands-on lab lets you do the lab activities yourself in a real cloud environment, not in a simulation or demo environment. It does so by giving you new, temporary credentials that you use to sign in and access Google Cloud for the duration of the lab.

What you need

To complete this lab, you need:

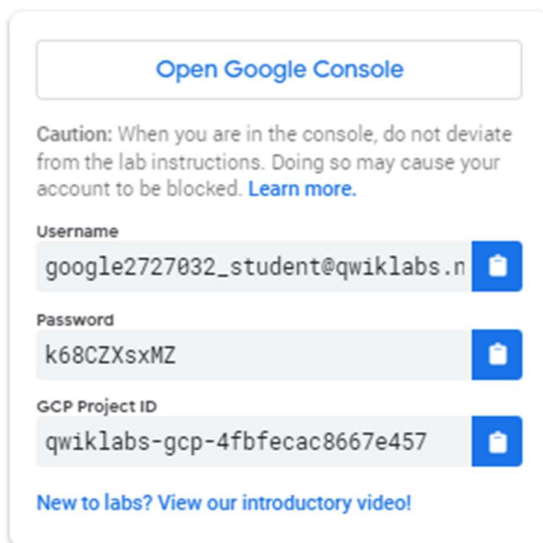
- Access to a standard internet browser (Chrome browser recommended).
- Time to complete the lab.

Note: If you already have your own personal Google Cloud account or project, do not use it for this lab.

Note: If you are using a Pixelbook, open an Incognito window to run this lab.

How to start your lab and sign in to the Google Cloud Console

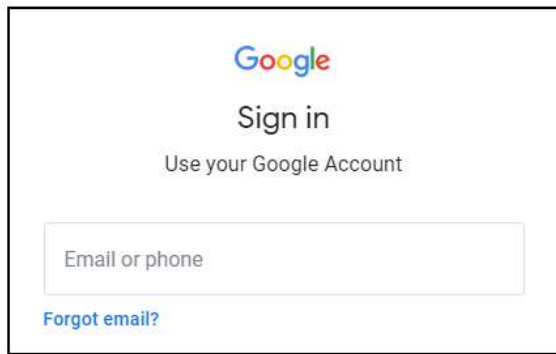
1. Click the **Start Lab** button. If you need to pay for the lab, a pop-up opens for you to select your payment method. On the left is a panel populated with the temporary credentials that you must use for this lab.



The screenshot shows a sign-in panel with the following elements:

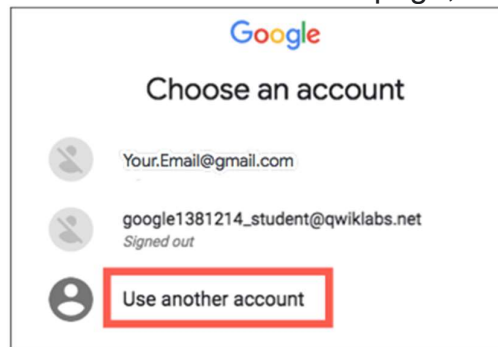
- A button at the top labeled "Open Google Console".
- A caution message: "Caution: When you are in the console, do not deviate from the lab instructions. Doing so may cause your account to be blocked. [Learn more.](#)"
- Three input fields, each with a copy icon to its right:
 - Username:** google2727032_student@qwiklabs.n
 - Password:** k68CZXsxMZ
 - GCP Project ID:** qwiklabs-gcp-4fbfecac8667e457
- A link at the bottom: "New to labs? [View our introductory video!](#)"

2. Copy the username, and then click **Open Google Console**. The lab spins up resources, and then opens another tab that shows the **Sign in** page.



Tip: Open the tabs in separate windows, side-by-side.

If you see the **Choose an account** page, click **Use Another**



Account.

3. In the **Sign in** page, paste the username that you copied from the Connection Details panel. Then copy and paste the password.

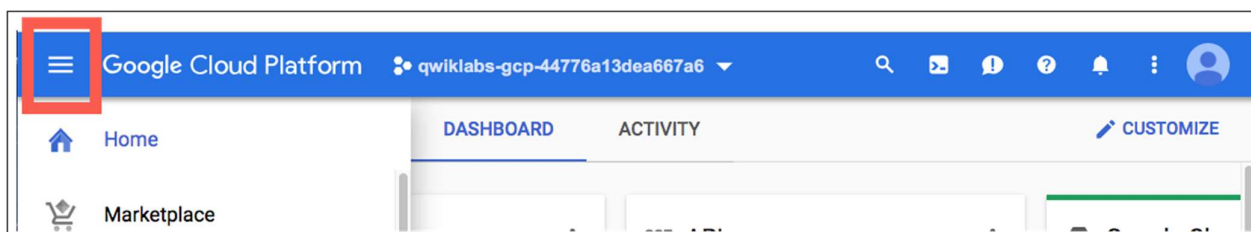
Important: You must use the credentials from the Connection Details panel. Do not use your Qwiklabs credentials. If you have your own Google Cloud account, do not use it for this lab (avoids incurring charges).

4. Click through the subsequent pages:

- Accept the terms and conditions.
- Do not add recovery options or two-factor authentication (because this is a temporary account).
- Do not sign up for free trials.

After a few moments, the Cloud Console opens in this tab.

Note: You can view the menu with a list of Google Cloud Products and Services by clicking the **Navigation menu** at the top-left.



Query 1: Total Confirmed Cases

Build a query that will answer "What was the total count of confirmed cases on Apr 15, 2020?" The query needs to return a single row containing the sum of confirmed cases across all countries. The name of the column should be **total_cases_worldwide**.

Click *Check my progress* to verify the objective.

Query 2: Worst Affected Areas

Build a query for answering "How many states in the US had more than 100 deaths on Apr 10, 2020?" The query needs to list the output in the field **count_of_states**. **Hint:** Don't include NULL values.

Click *Check my progress* to verify the objective.

Query 3: Identifying Hotspots

Build a query that will answer "List all the states in the United States of America that had more than 1000 confirmed cases on Apr 10, 2020?" The query needs to return the State Name and the corresponding confirmed cases arranged in descending order. Name of the fields to return **state** and **total_confirmed_cases**.

Click *Check my progress* to verify the objective.

Query 4: Fatality Ratio

Build a query that will answer "What was the case-fatality ratio in Italy for the month of April 2020?" Case-fatality ratio here is defined as $(\text{total deaths} / \text{total confirmed cases}) * 100$. Write a query to return the ratio for the month of April 2020 and containing the following fields in the output: **total_confirmed_cases**, **total_deaths**, **case_fatality_ratio**.

Click *Check my progress* to verify the objective.

Query 5: Identifying specific day

Build a query that will answer: "On what day did the total number of deaths cross 10000 in Italy?" The query should return the date in the format **yyyy-mm-dd**.

Click *Check my progress* to verify the objective.

Query 6: Finding days with zero net new cases

The following query is written to identify the number of days in India between 21 Feb 2020 and 15 March 2020 when there were zero increases in the number of confirmed cases. However it is not executing properly. You need to update the query to complete it and obtain the result:

```
WITH india_cases_by_date AS (  
  SELECT  
    date,  
    SUM(cumulative_confirmed) AS cases  
  FROM  
    `bigquery-public-data.covid19_open_data.covid19_open_data`  
  WHERE  
    country_name="India"  
    AND date between '2020-02-21' and '2020-03-15'  
  GROUP BY  
    date  
  ORDER BY  
    date ASC  
)  
  
, india_previous_day_comparison AS  
(SELECT  
  date,  
  cases,  
  LAG(cases) OVER(ORDER BY date) AS previous_day,  
  cases - LAG(cases) OVER(ORDER BY date) AS net_new_cases  
FROM india_cases_by_date  
)
```

Click *Check my progress* to verify the objective.

Query 7: Doubling rate

Using the previous query as a template, write a query to find out the dates on which the confirmed cases increased by more than 10% compared to the previous day (indicating doubling rate of ~ 7 days) in the US between the dates March 22, 2020 and April 20, 2020. The query needs to return the list of dates, the confirmed cases on that day, the confirmed cases the previous day, and the percentage increase in cases between the days. Use the following names for the returned

fields: **Date**, **Confirmed_Cases_On_Day**, **Confirmed_Cases_Previous_Day** and **Percentage_Increase_In_Cases**.

Click *Check my progress* to verify the objective.

Query 8: Recovery rate

Build a query to list the recovery rates of countries arranged in descending order (limit to 10) on the date May 10, 2020. Restrict the query to only those countries having more than 50K confirmed cases. The query needs to return the following fields: **country**, **recovered_cases**, **confirmed_cases**, **recovery_rate**.

Click *Check my progress* to verify the objective.

Query 9: CDGR - Cumulative Daily Growth Rate

The following query is trying to calculate the CDGR on May 10, 2020 (Cumulative Daily Growth Rate) for France since the day the first case was reported. The first case was reported on Jan 24, 2020. The CDGR is calculated as:

$$((\text{last_day_cases} / \text{first_day_cases})^{1/\text{days_diff}} - 1)$$

Where :

- **last_day_cases** is the number of confirmed cases on May 10, 2020
- **first_day_cases** is the number of confirmed cases on Feb 02, 2020
- **days_diff** is the number of days between Feb 02 - May 10, 2020

The query isn't executing properly. Can you fix the error to make the query execute successfully?

```
WITH
  france_cases AS (
    SELECT
      date,
      SUM(cumulative_confirmed) AS total_cases
    FROM
      `bigquery-public-data.covid19_open_data.covid19_open_data`
    WHERE
      country_name="France"
      AND date IN ('2020-01-24',
        '2020-05-10')
    GROUP BY
      date
    ORDER BY
      date)
, summary as (
SELECT
  total_cases AS first_day_cases,
  LEAD(total_cases) AS last_day_cases,
  DATE_DIFF(LEAD(date) OVER(ORDER BY date), date, day) AS days_diff
```



```

FROM
  france_cases
LIMIT 1
)

select first_day_cases, last_day_cases, days_diff,
SQRT((last_day_cases/first_day_cases),(1/days_diff))-1 as cdgr
from summary

```

Note: Refer to the following [page](#) to learn more about the SQL function referenced `LEAD()`. Click *Check my progress* to verify the objective.

Create a Datastudio report

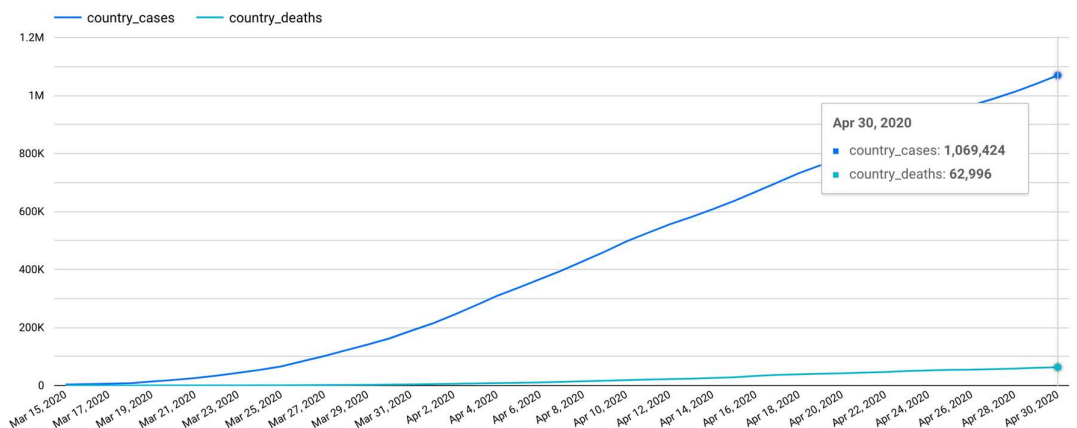
Create a Datastudio report that plots the following for the United States:

- Number of Confirmed Cases
 - Number of Deaths
 - Date range : 2020-03-15 to 2020-04-30
- Click *Check my progress* to verify the objective.

Create a Datastudio report

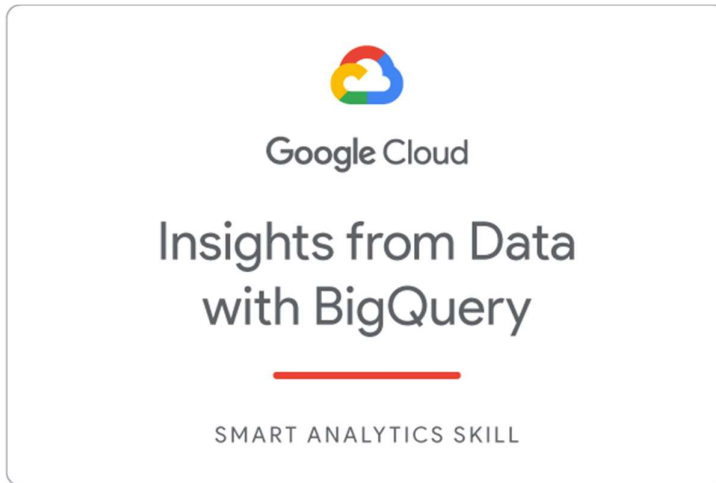
Check my progress

Hint: Use the below image as a reference while building the report and make sure your report matches it.



Congratulations!

In this lab you used BigQuery to query the public Covid dataset (`bigquery-public-data.covid19_open_data.covid19_open_data`) to answer important questions related to the Covid-19 pandemic. You also created a Datastudio report for one of the query outputs to visualize the cases and deaths trend.



Finish Your Quest

This self-paced lab is part of the Qwiklabs [Insights from Data with BigQuery](#) Quest. A Quest is a series of related labs that form a learning path. Completing this Quest earns you the badge above, to recognize your achievement. You can make your badge public and link to them in your online resume or social media account. [Enroll in this Quest](#) and get immediate completion credit if you've taken this lab. [See other available Qwiklabs Quests](#). This skill badge quest is part of Google Cloud's [Data Analyst](#) learning path. Continue your learning journey by enrolling in the [Create ML Models with BigQuery ML](#) quest.

Next Steps / Learn More

This lab is also part of a series of labs called Challenge Labs. These labs are designed test your Google Cloud knowledge and skill. Search for "Challenge Lab" in the [lab catalog](#) and challenge yourself!

Google Cloud Training & Certification

...helps you make the most of Google Cloud technologies. [Our classes](#) include technical skills and best practices to help you get up to speed quickly and continue your learning

journey. We offer fundamental to advanced level training, with on-demand, live, and virtual options to suit your busy schedule. [Certifications](#) help you validate and prove your skill and expertise in Google Cloud technologies.

Manual Last Updated January 7, 2021

Lab Last Tested September 29, 2020

Copyright 2021 Google LLC All rights reserved. Google and the Google logo are trademarks of Google LLC. All other company and product names may be trademarks of the respective companies with which they are associated.

Solutions:

<https://www.youtube.com/watch?v=9cEzh3Jis6Y>

<https://onlineintercollege.blogspot.com/2021/03/insights-from-data-with-bigquery.html>