Hands-On Activity: Applying SQL

TOTAL POINTS 2

1.

Activity overview

In previous lessons, you learned how to apply formulas in spreadsheets. In this activity, we will practice using formulas with SQL queries.

By the time you complete this activity, you will be able to use SQL to write queries for datasets. This will enable you to explore public datasets in BigQuery, which is important for writing queries in your career as a data analyst.

Set up your data

1. Log in to BigQuery Sandbox. If you have a free trial version of BigQuery, you can use that instead. On the BigQuery page, click the **Go to BigQuery** button.

- **Note:** BigQuery Sandbox frequently updates its user interface. The latest changes may not be reflected in the screenshots presented in this activity, but the principles remain the same. Adapting to changes in software updates is an essential skill for data analysts, and it's helpful for you to practice troubleshooting. You can also reach out to your community of learners on the discussion forum for help.

2. If you have never created a BigQuery project before, click **CREATE PROJECT** on the right side of the screen. If you have created a project before, you can use an existing one or create a new one by clicking the project dropdown in the blue header bar and selecting **NEW PROJECT**.

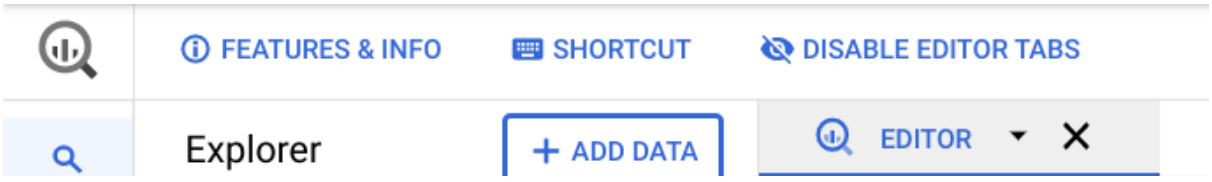
3. Name your project something that will help you identify it later. You can give it a unique project ID or use an auto-generated one. Don't worry about selecting an organization if you don't know what to put.

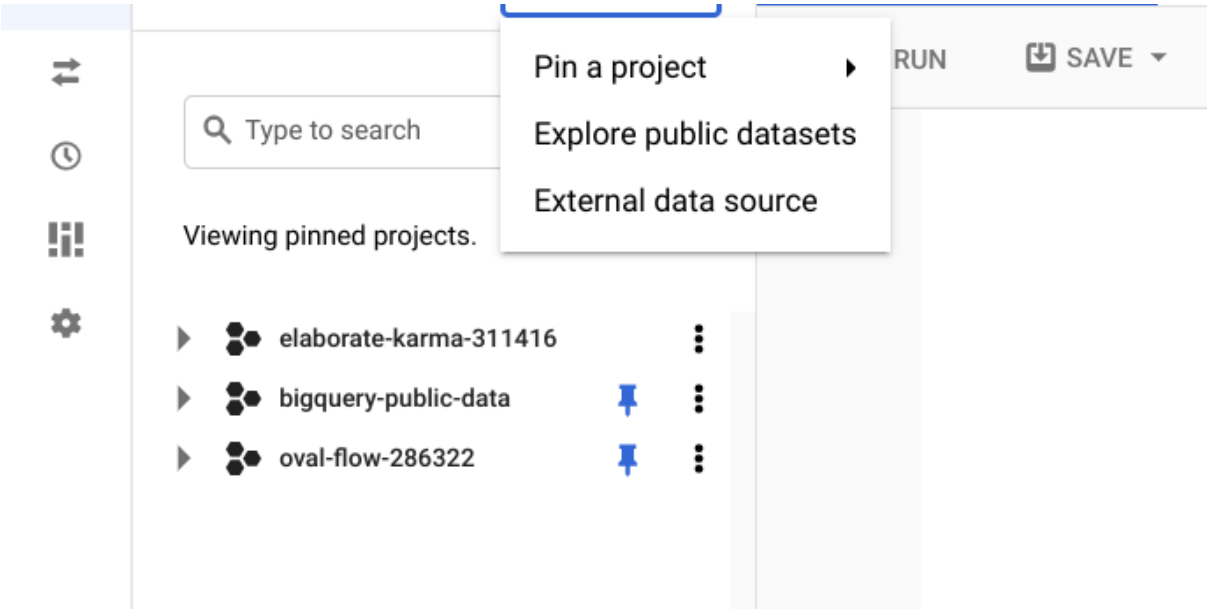
4. Now, you'll see the **Editor** interface. In the middle of the screen is a window where you can type code, and to the left is the **Explorer** menu where you can search for datasets.

Pick a dataset

Follow these steps to find and pick a dataset for this activity:

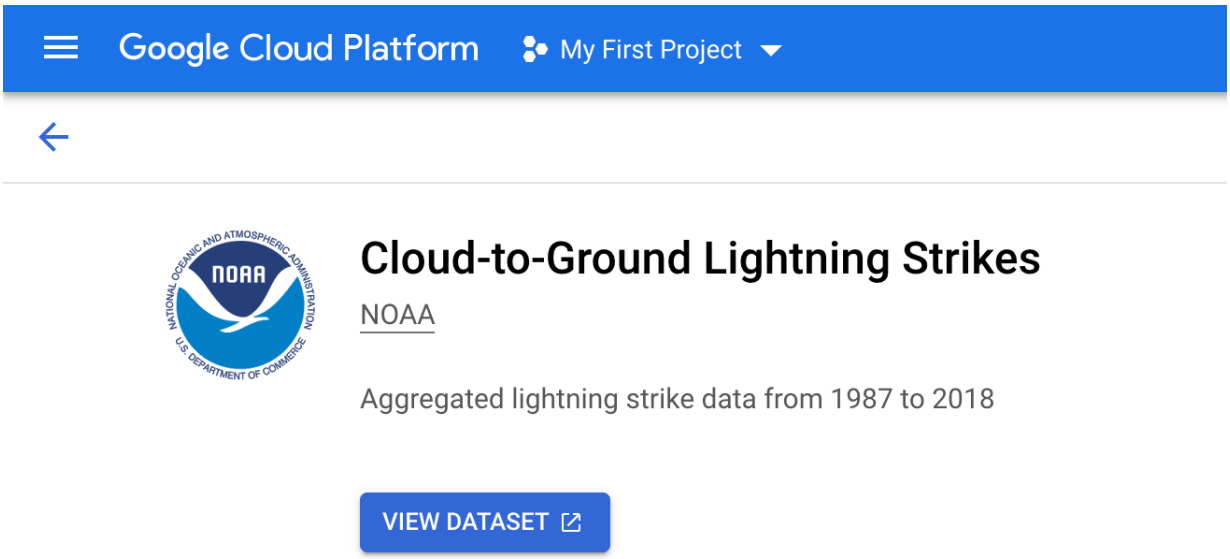
1. Locate the **Explorer menu** on the left side of your screen. Click on **+ ADD DATA** and then **Explore public datasets**.



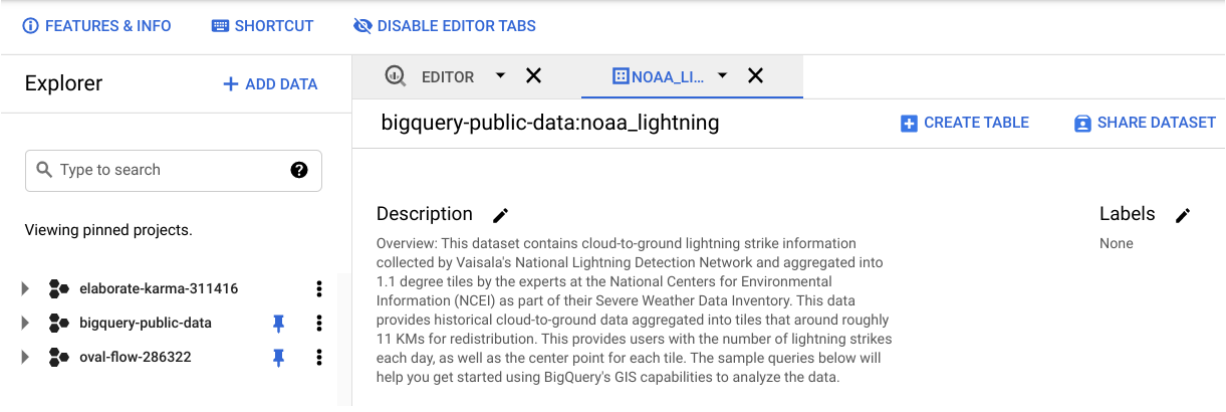


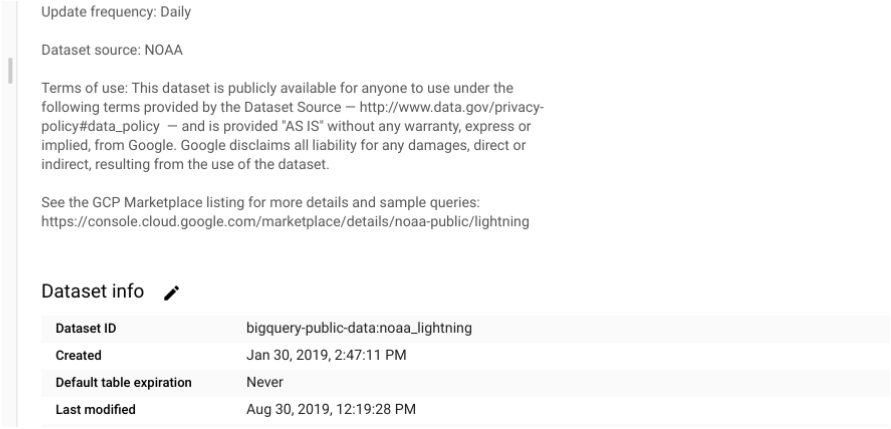
2. In the **Search Marketplace** bar, search for “Cloud-to-Ground Lightning Strikes,” a NOAA dataset. Click the result, then click **View Dataset**. This will bring you back to the BigQuery Sandbox interface in a new tab.

- **Note:** This may pin the **bigquery-public-data** dropdown to the **Explorer** menu. You can use this to browse datasets and tables.



3. In BigQuery, you’ll find information on the dataset you selected. Review the description of the dataset.





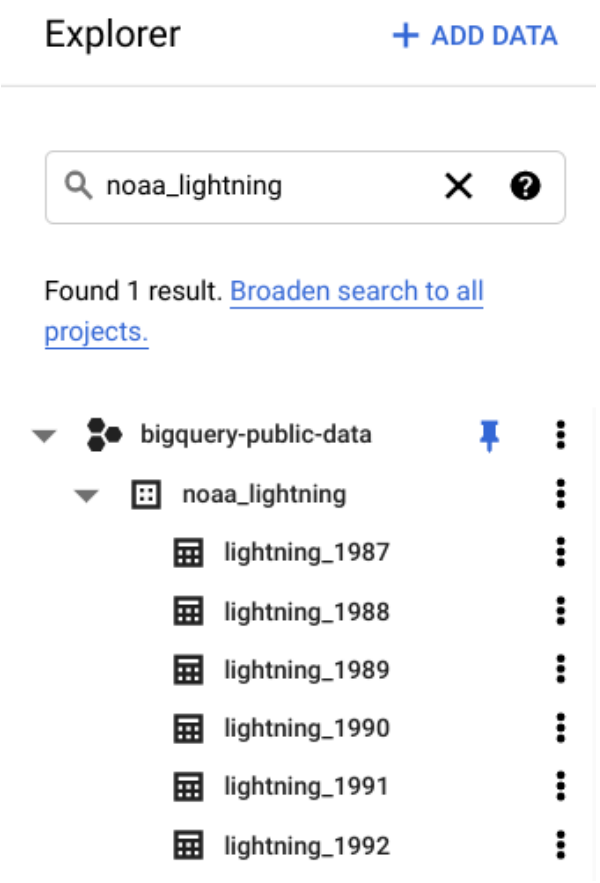
For instance, you can locate the Dataset ID. You will need this in order to write an SQL query, so that you can tell what database, dataset, and table you're targeting. In this case, the database connection is "bigquery-public-data" and the Dataset ID is "noaa_lightning". You'll still need to identify what table you want to query, so begin with a close review of the dataset.

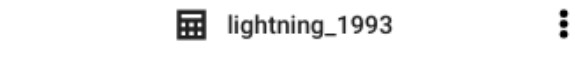
Choose a table

1. Enter the Dataset ID, "noaa_lightning," in the search bar of the **Explorer menu**.

If this doesn't pull anything up, you can find it manually by deleting the text from the search bar, clicking on the arrow next to **bigquery-public-data**, and scrolling until you find the right dataset.

2. Once you've found the "noaa_lightning" dataset, click on the arrow next to it to expand the dataset to examine the tables it contains.

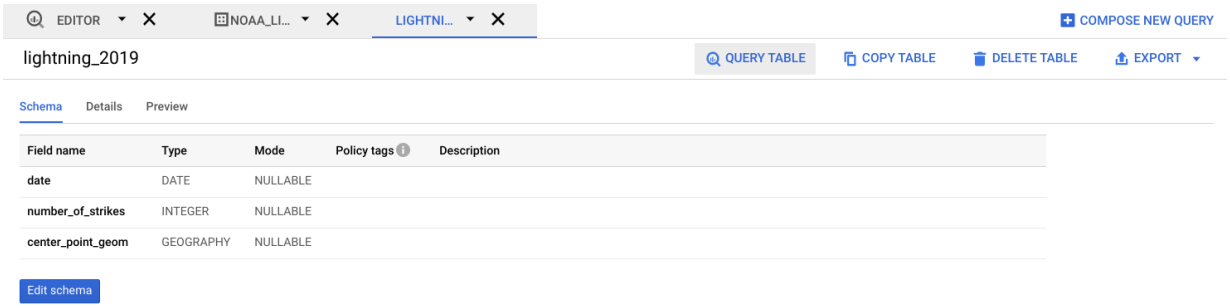




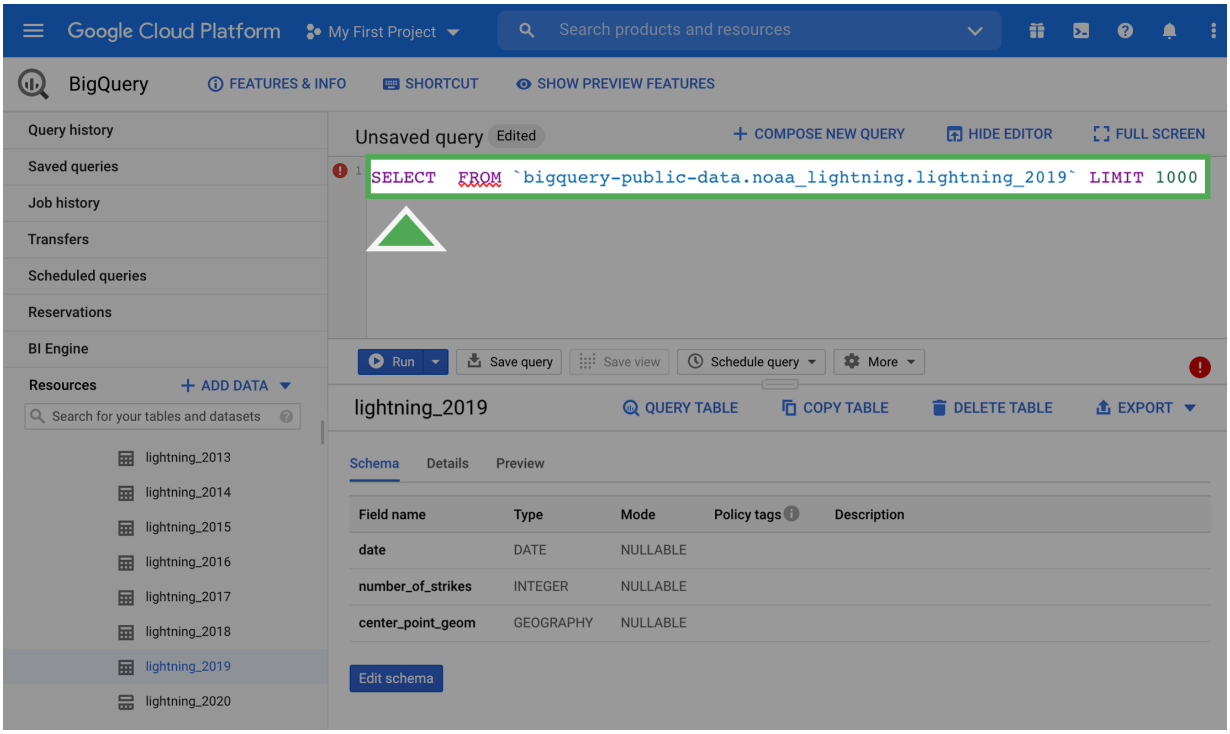
These are all tables contained in the dataset. You can check out the data for 2019, which is contained in the table “lightning_2019”.

3. Click on the **lightning_2019** table. This will bring up information for the table.

4. On the right side of the information window, click **Query Table**.



This will populate the query window with a query. Notice that the query doesn’t contain anything in between “SELECT” and “FROM”.



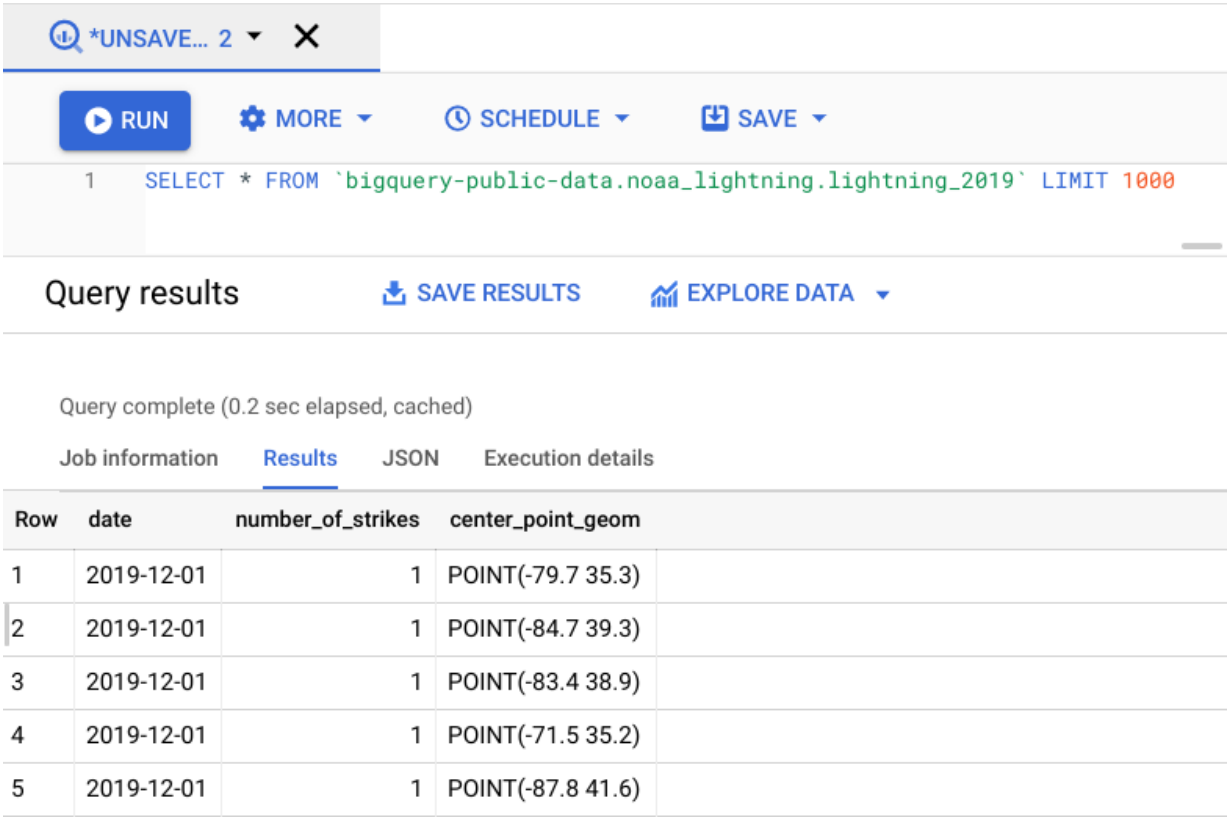
Write a query

Query the data

You’ll still need to complete the query by adding what you want.

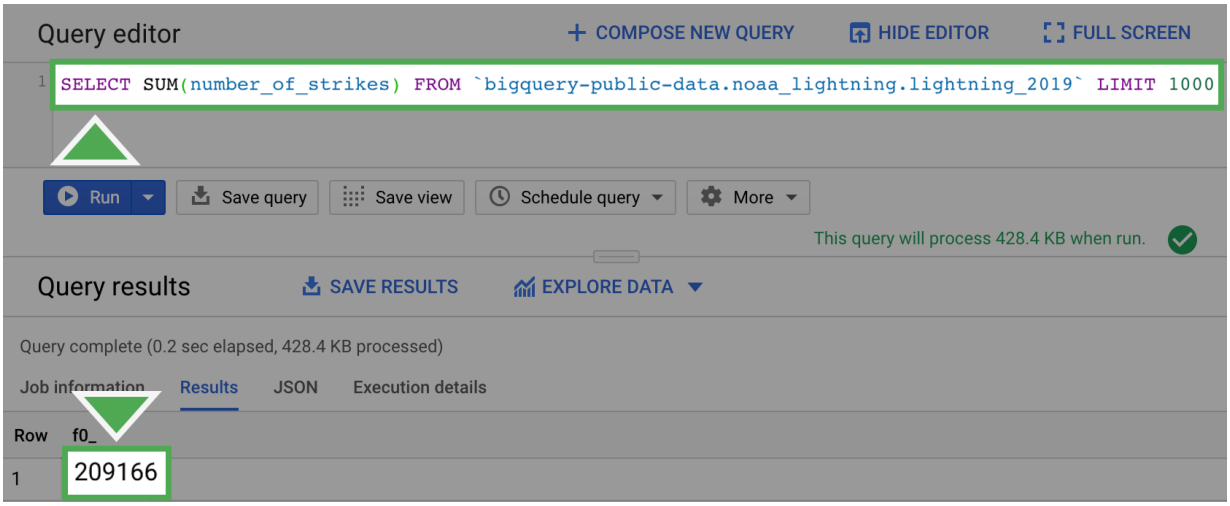
1. Insert an asterisk * after the select, so that your query reads **SELECT * FROM** followed by your table location.

2. Run the query. In the example provided, your result should be something like this:



This query returns all columns for the first 1,000 rows from the table.

3. Write a query to see how many total lightning strikes happened in 2019. Instead of an asterisk, type **SUM(number_of_strikes)**.



This returns your answer, **209,166**.

Write your own queries

Now, come up with some questions and answer them with your own SQL queries. For instance, with the example dataset, try finding out how many lightning strikes happened in a different year.

You are also free to choose another publicly available dataset in BigQuery and write your own queries for extra practice—there are a lot of interesting choices!