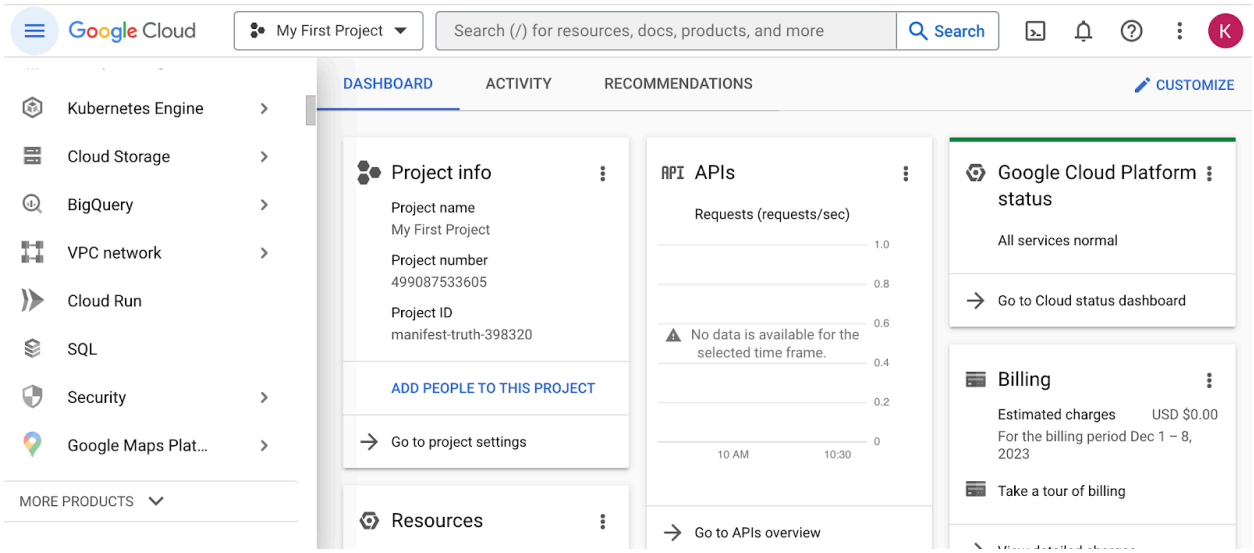
Get started with BigQuery

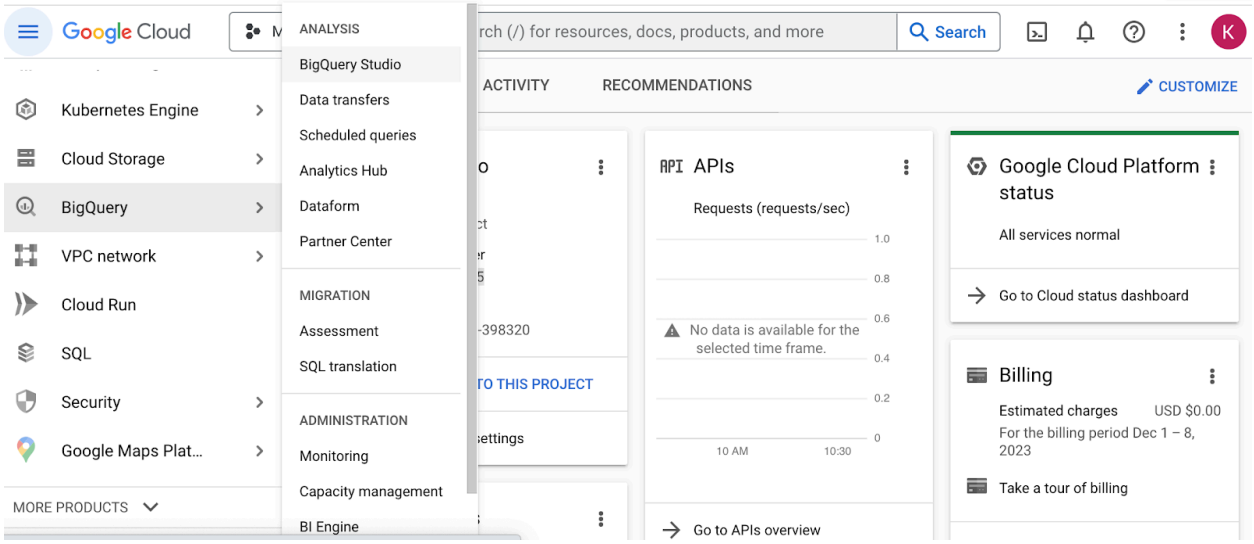
BigQuery is a data warehouse on the Google Cloud Platform used to query and filter large datasets, aggregate results, and perform complex operations. Throughout this program, you're going to use BigQuery to practice your SQL skills and collect, prepare, and analyze data. At this point, you have set up your own account. Now, explore some of the important elements of the SQL workspace. This will prepare you for the upcoming activities in which you will use BigQuery. Note that BigQuery updates its interface frequently, so your console might be slightly different from what is described in this reading. That's okay; use your troubleshooting skills to find what you need!

Log in to BigQuery

When you log in to BigQuery using the landing page, you will automatically open your project space. This is a high-level overview of your project, including the project information and the current resources being used. From here, you can check your recent activity.

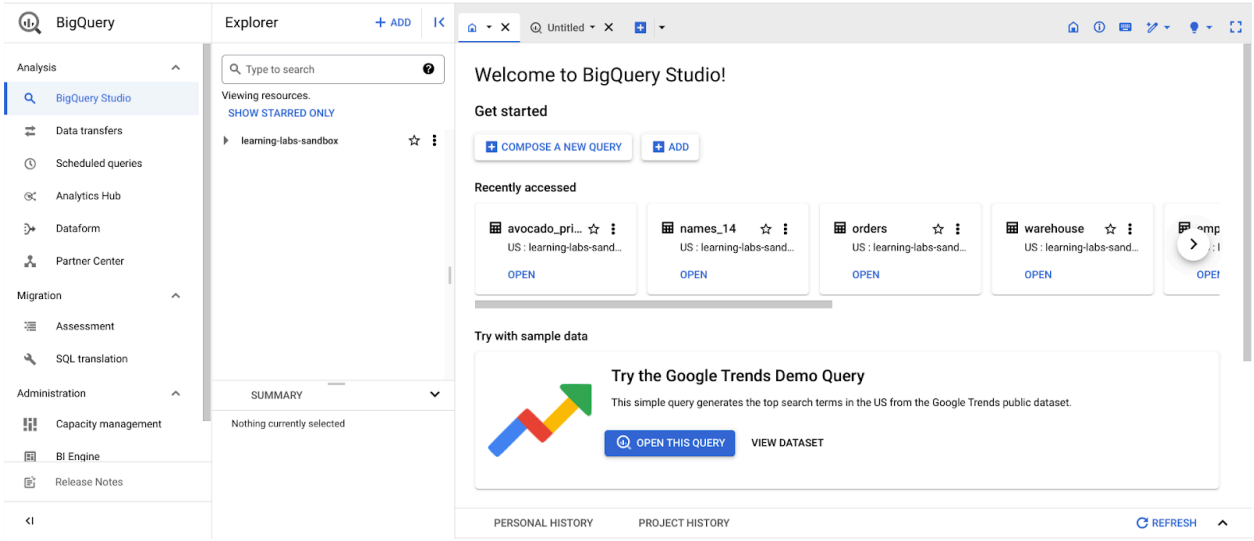


Navigate to your project's BigQuery Studio by selecting BigQuery from the navigation menu and BigQuery Studio from the dropdown menu.



BiqQuery Studio components

Once you have navigated to BigQuery from the project space, most of the major components of the BigQuery console will be present: the **Navigation** pane, the **Explorer** pane, and the **SQL Workspace**.

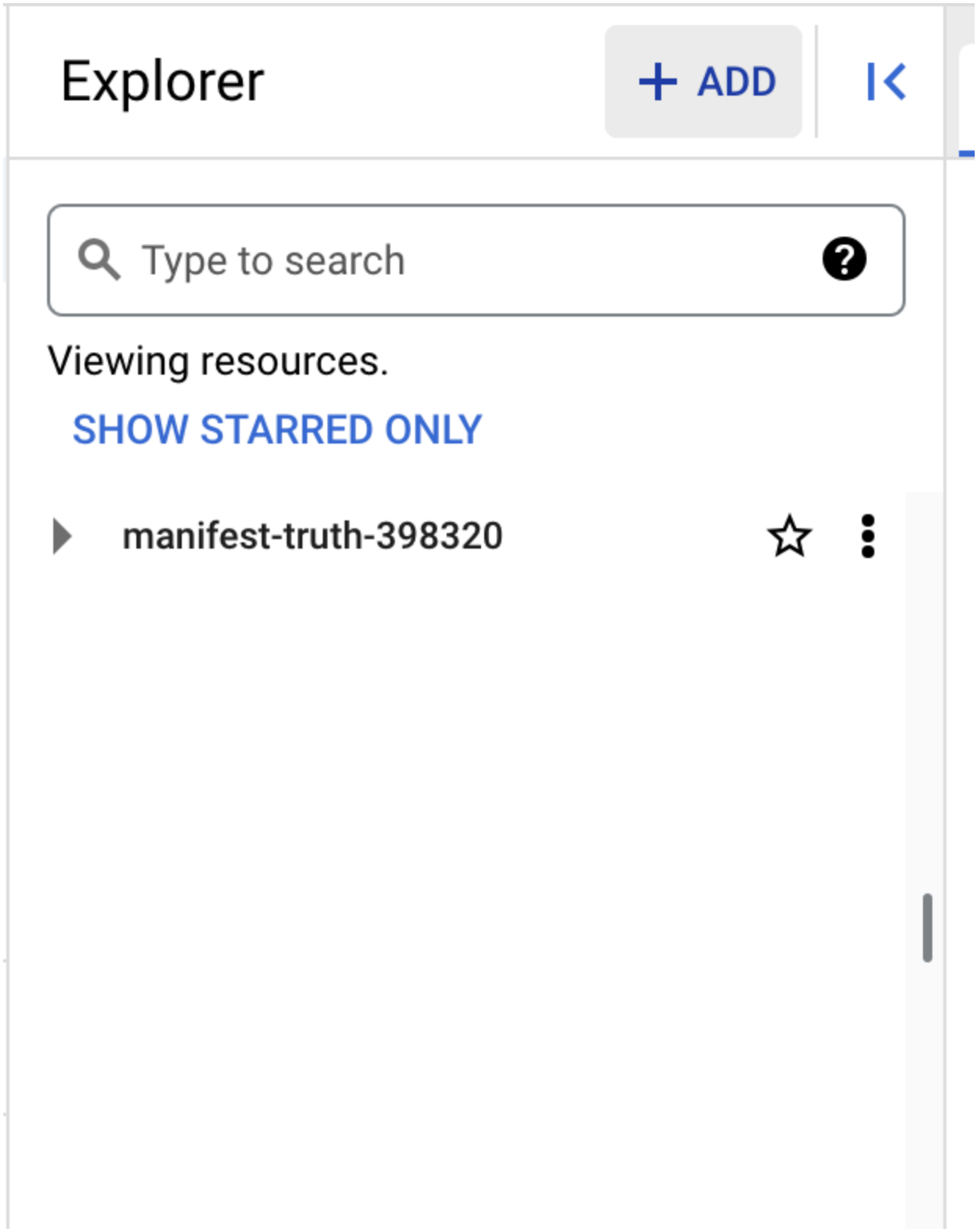


The Navigation pane

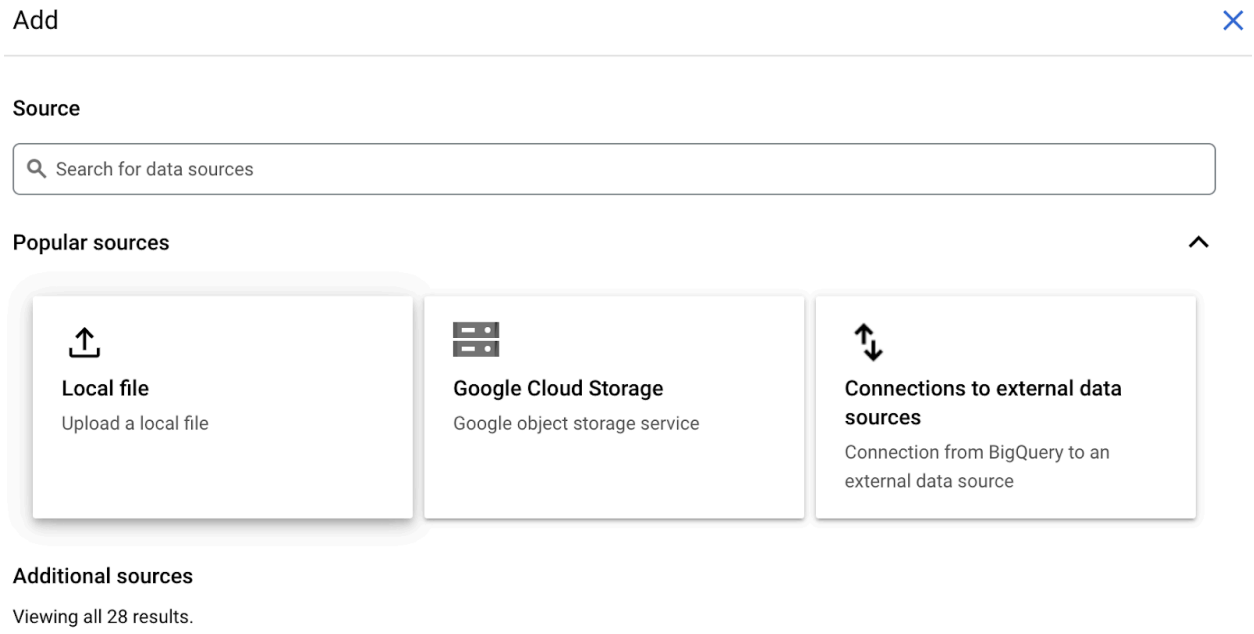
On the console page, find the **Navigation** pane. This is how you navigate from the project space to the BigQuery tool. This menu also contains a list of other Google Cloud Project (GCP) data tools. During this program, you will focus on BigQuery, but it's useful to understand that the GCP has a collection of connected tools data professionals use every day.

The Explorer pane

The **Explorer** pane lists your current projects and any starred projects you have added to your console. It's also where you'll find the **+ ADD** button, which you can use to add datasets.

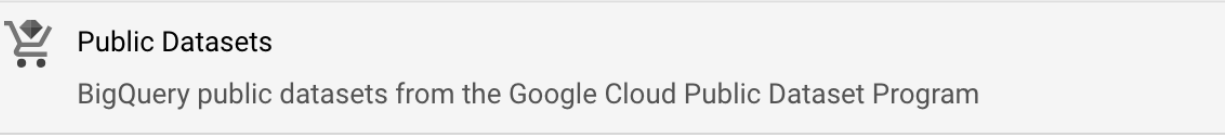


This button opens the **Add** dialog that allows you to open or import a variety of datasets.

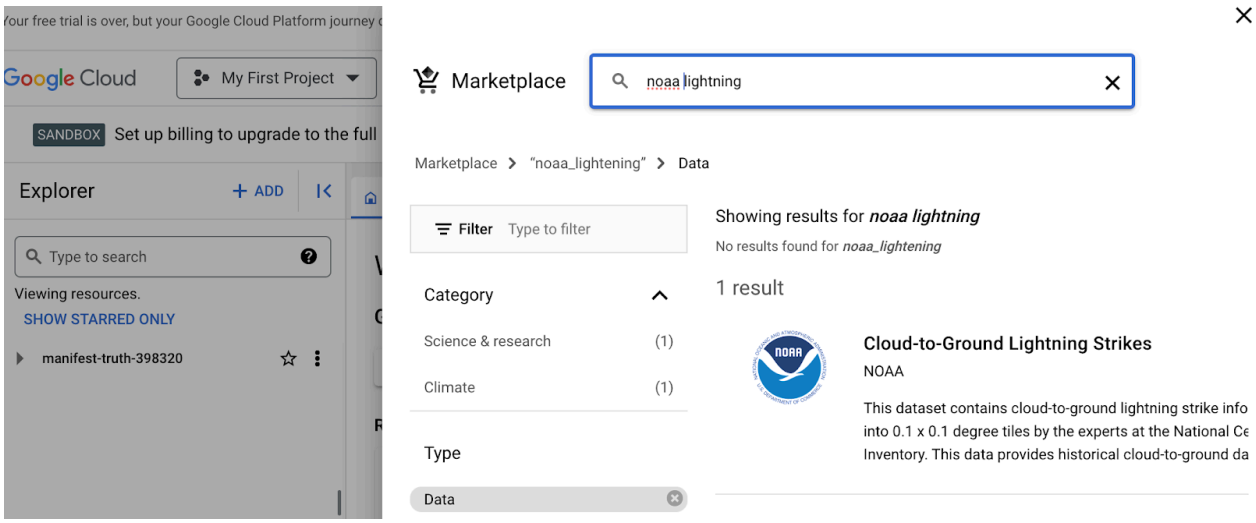


Add Public Datasets

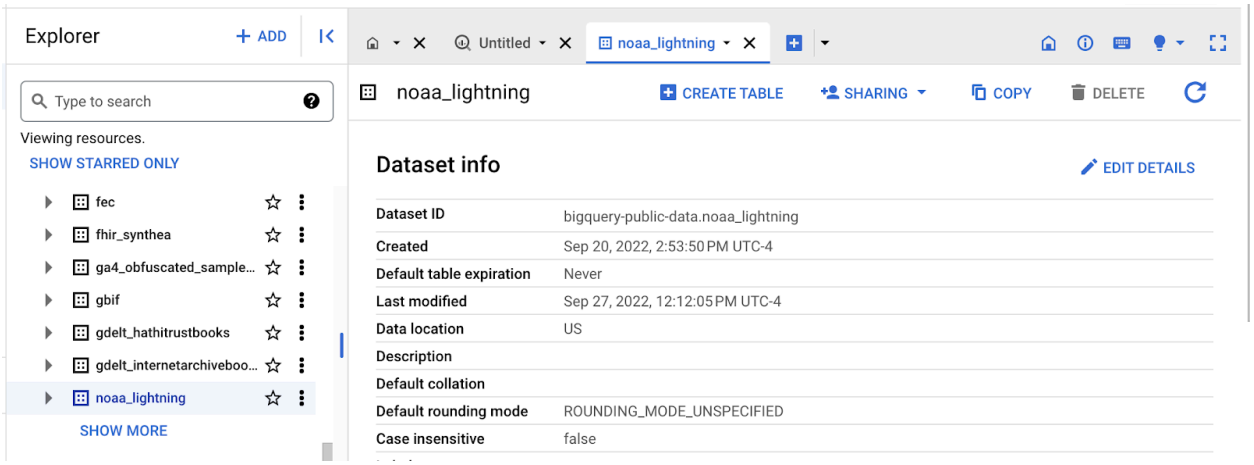
BigQuery offers a variety of public datasets from the Google Cloud Public Dataset Program. Scroll down the **Add** dialog to the **Public Datasets** option.



Select **Public Datasets**. This takes you to the **Public Datasets Marketplace**, where you can search for and select public datasets to add to your BigQuery console. For example, search for the "noaa lightning" dataset in the Marketplace search bar. When you search for this dataset, you will find NOAA's Cloud-to-Ground Lightning Strikes data.



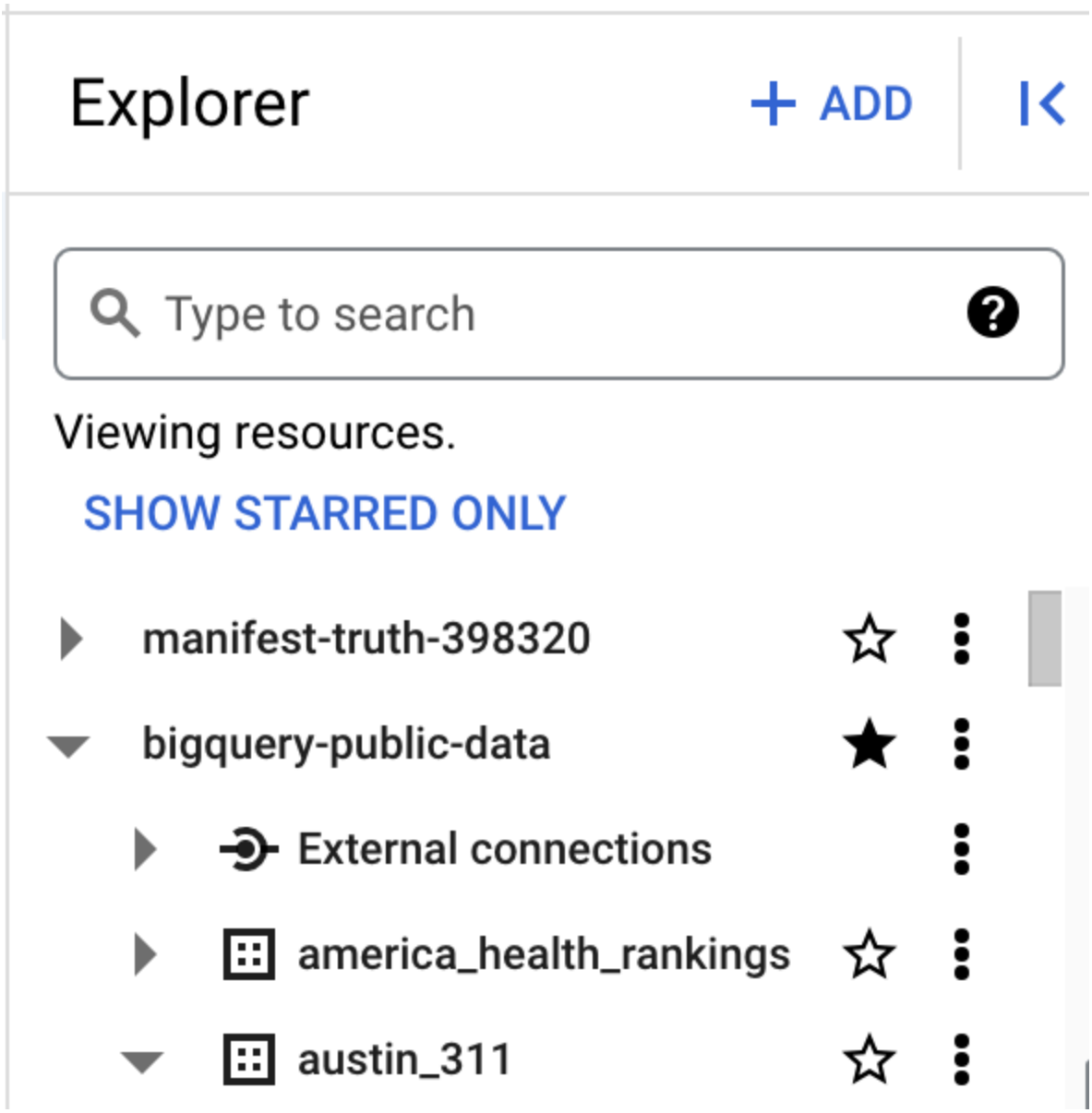
Select the dataset to read its description. Select **View dataset** to create a tab of the dataset's information within the SQL workspace.



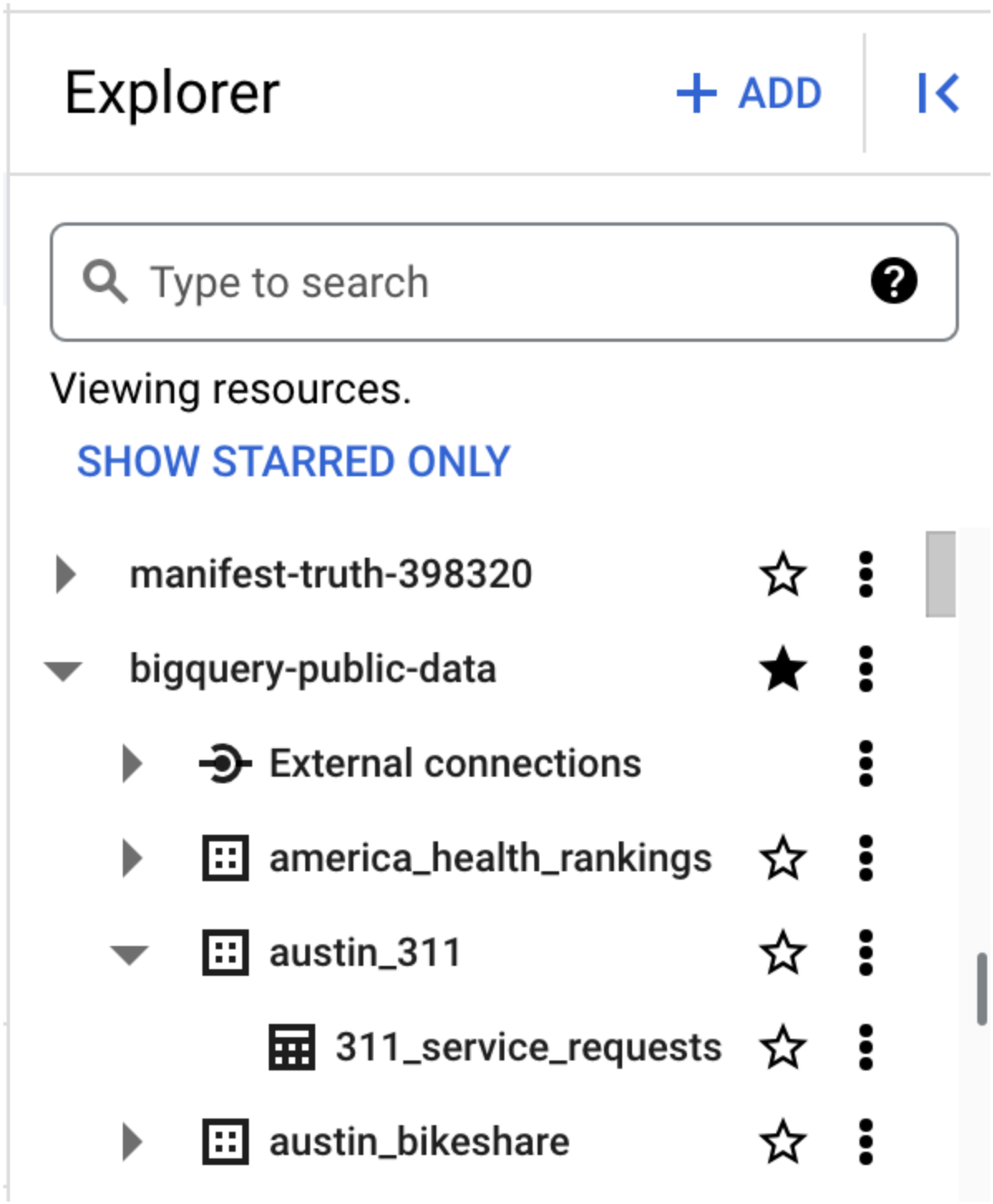
The Explorer Pane lists the noaa_lightning and other public datasets.

Star and examine Public Datasets

You added the public noaa_lightning dataset to your BigQuery Workspace, so the **Explorer** pane displays the noaa_lightning dataset, along with the list of other public datasets. These datasets are nested under bigquery-public-data. Star bigquery-public-data by navigating to the top of the **Explorer** pane and selecting the star next to bigquery-public-data.



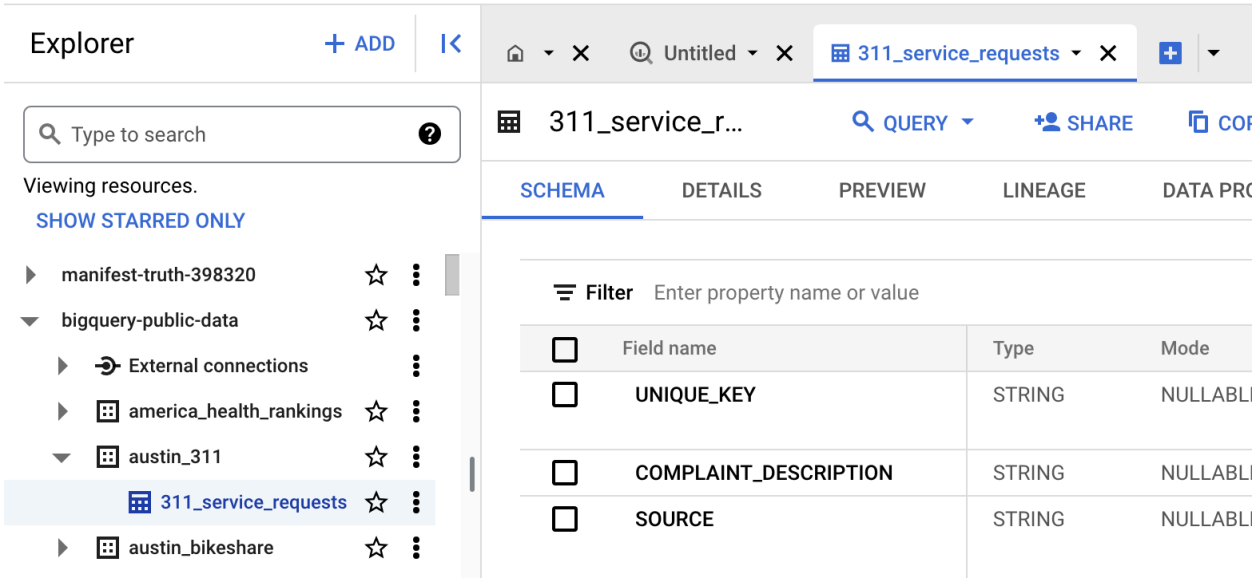
Starring bigquery-public-data will enable you to search for and add public datasets by scrolling in the **Explorer** pane or by searching for them in the **Explorer** search bar.
For example, you might want to select a different public dataset. If you select the second dataset, "austin_311," it will expand to list the table stored in it, “311_service_requests.”



The Explorer pane with the “bigquery-public data” and “austin_311” datasets expanded, revealing the “311_service_requests” table

When you select a table, its information is displayed in the SQL Workspace. Select the 311_service_requests table to examine several tabs that describe it, including:

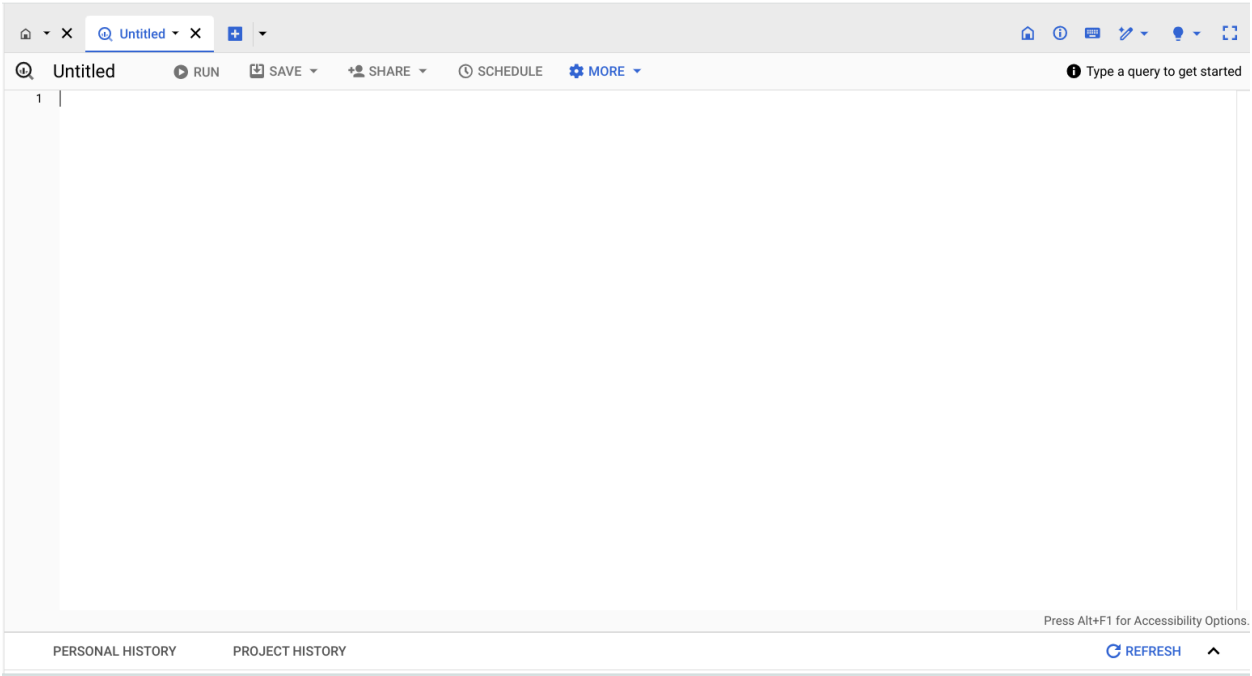
- **Schema**, which displays the column names in the dataset
- **Details**, which contains additional metadata, such as the creation date of the dataset
- **Preview**, which shows the first rows from the dataset



Additionally, you can select the **Query** button from the menu bar in the SQL Workspace to query this table.

The SQL Workspace

The final menu pane in your console is the SQL Workspace. This is where you will actually write and execute queries in BigQuery.



The SQL Workspace also gives you access to your personal and project history, which stores a record of the queries you've run. This can be useful if you want to return to a query to run it again or use part of it in another query.

Upload your data

In addition to offering access to public datasets, BigQuery also gives you the ability to upload your own data directly into your workspace. Access this feature by opening the **+ ADD** menu again or by clicking the three vertical dots next to your project's name in the Explorer pane. This will give you the option to create your own dataset and upload your own tables. You will have the opportunity to upload your own data in an upcoming activity to practice using this feature!

Key takeaways

BigQuery's SQL workspace allows you to search for public datasets, run SQL queries, and even upload your own data for analysis. Whether you're working with public datasets, running SQL queries, or uploading your own data, BigQuery's SQL workspace offers a range of features to support all kinds of data analysis tasks. Throughout this program, you will be using BigQuery to practice your SQL skills, so being familiar with the major components of your BigQuery console will help you navigate it effectively in the future!

Step-by-Step: BigQuery in action

This reading provides you with the steps the instructor performs in the following video, BigQuery in action. The video focuses on how to create a query to view a small section of data from a large dataset.

Keep this guide open as you watch the video. It can serve as a helpful reference if you need additional context or clarification while following the video steps. This is not a graded activity, but you can complete these steps to practice the skills demonstrated in the video.

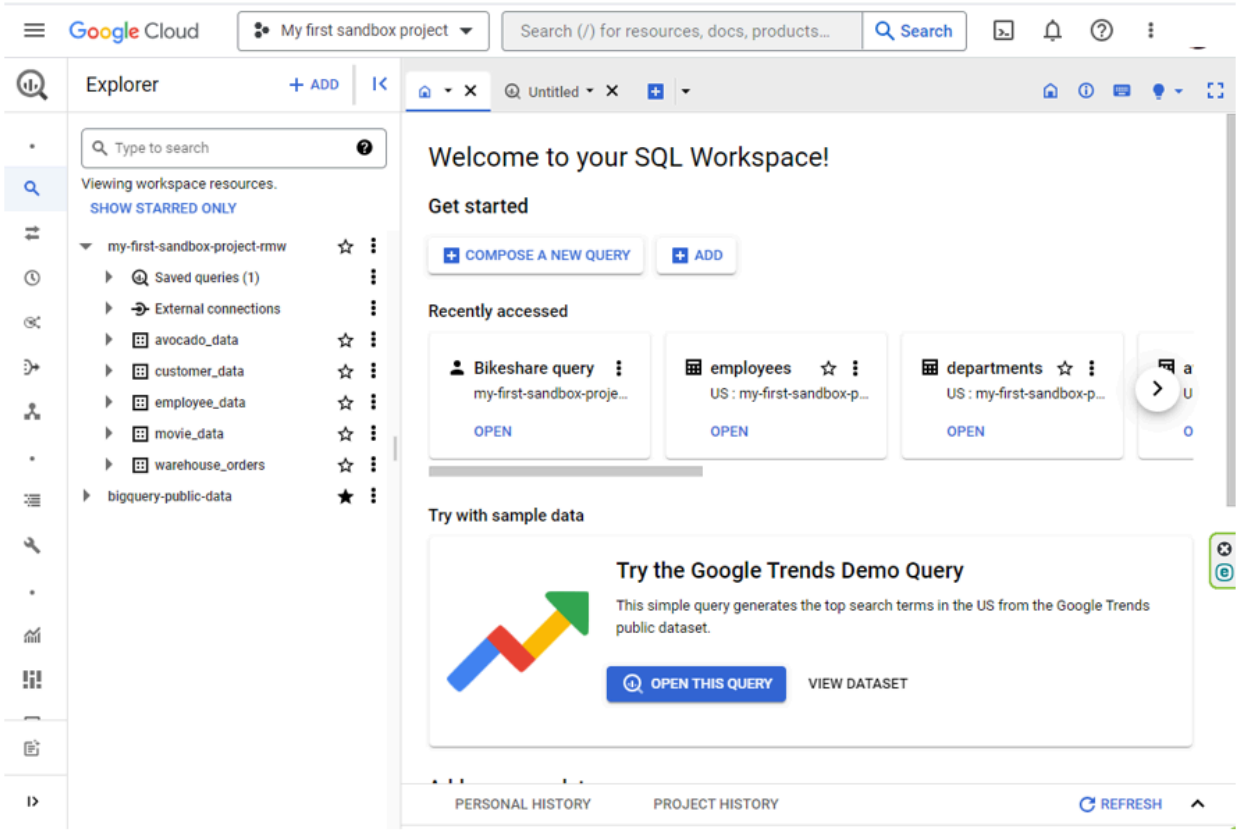
What you'll need

To follow along with the examples in this video, log in to your BigQuery account and follow the instructions to star `bigquery-public-data` in **The Explorer pane** section of the previous reading, Get Started with BigQuery.

Example 1: Preview a section from a table viewer

A database is a collection of data stored in a computer system. Query languages such as SQL enable communication between databases and data analysts. You discovered earlier that a relational database is made up of several tables that may be joined together to create relationships. Primary and foreign keys serve as representations of these relationships. To extract data from these tables, data analysts use queries. To learn more about that, explore BigQuery in action:

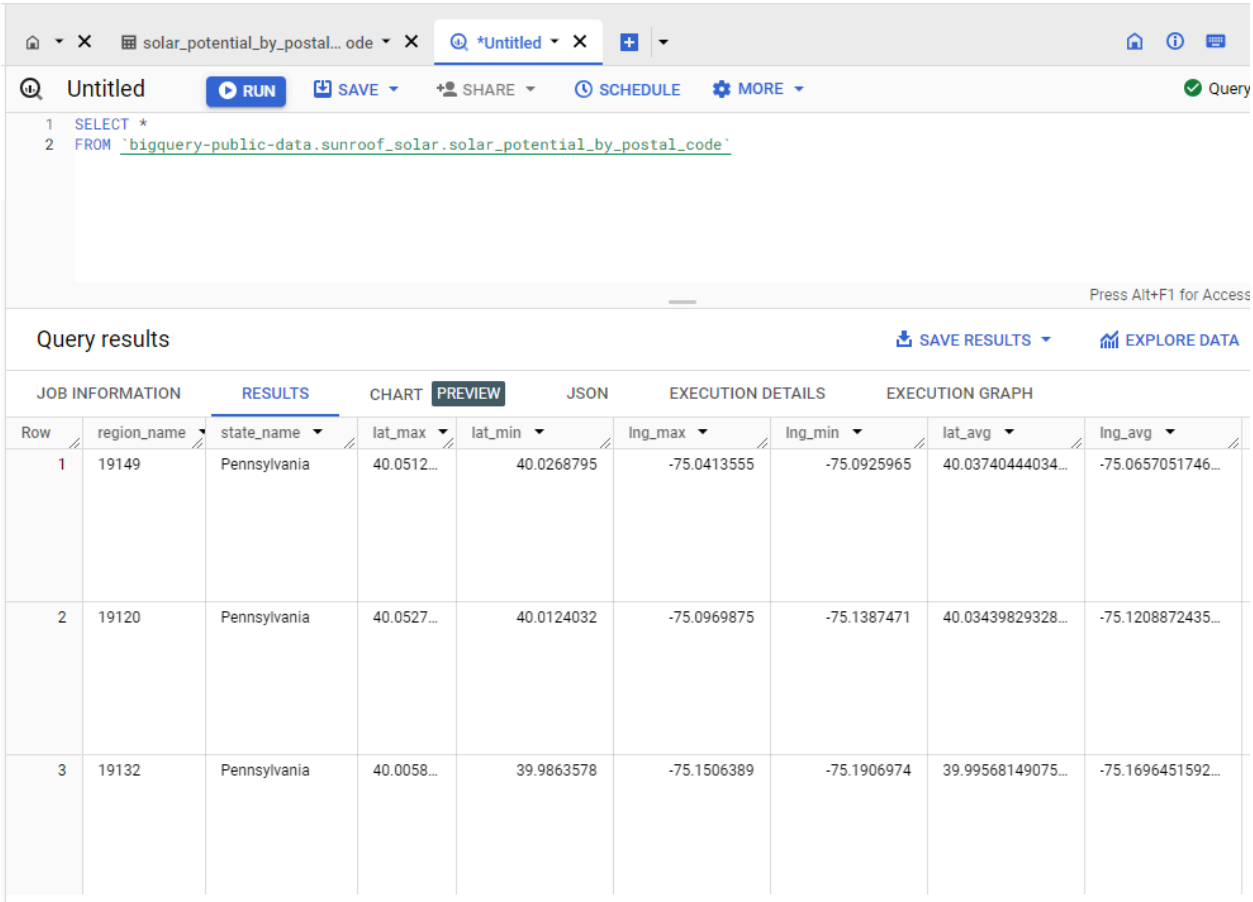
1. Log in to BigQuery and go to your console. You should find the **Welcome to your SQL Workspace!** landing page open. Select **COMPOSE A NEW QUERY** In the Bigquery console. Make sure that no tabs are open so that the entire workspace is displayed, including the **Explorer** pane.
2. Enter **sunroof** in the search bar. In the search results, expand `sunroof_solar` and then select the `solar_potential_by_postal_code` dataset.
3. Observe the **Schema tab** of the **Explorer** pane to explore the table fields.
4. Select the **Preview** tab to view the regions, states, yearly sunlight, and more.



Example 2: Writing a query

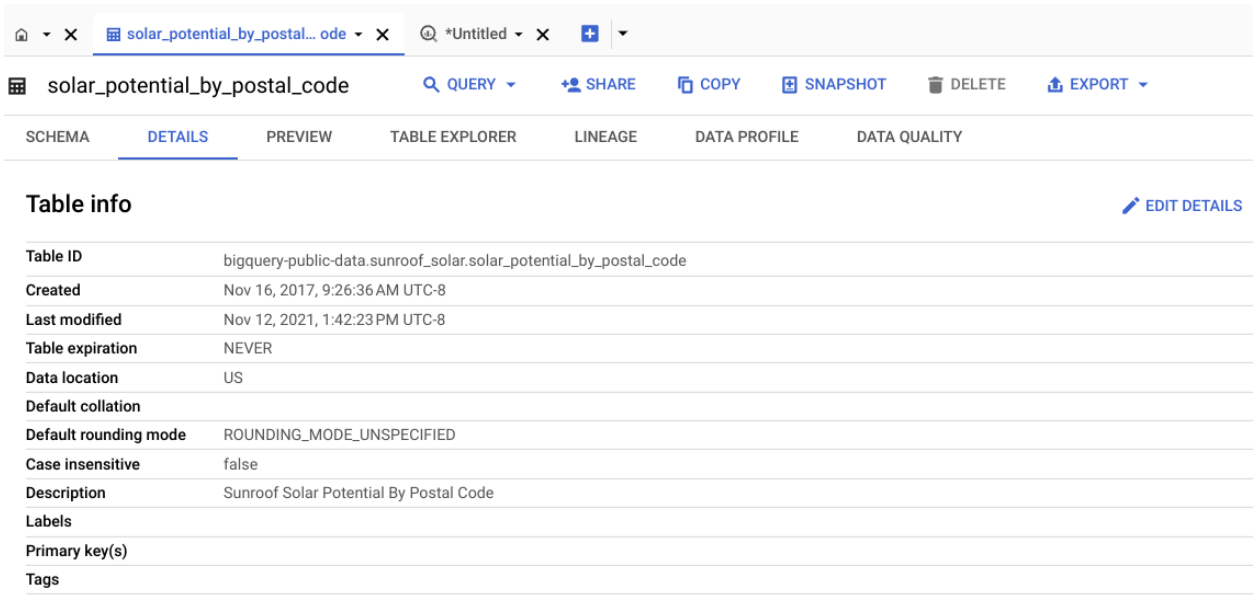
In order to view the entire dataset, you will need to write a query.

1. The first step is finding out the complete, correct path to the table you want to work with. Select the **ellipses** (three vertical dots) by the dataset `solar_potential_by_postal_code`, then select **Query**. A new tab will populate on your screen. Select the tab. The path to the table should be written inside two backticks.
2. Select the full path by highlighting the text including the backticks and copy it. (**Note:** You can also get the full path to the project, database, and table directly by clicking the ellipses next to the table's name in the **Explorer** panel on the left and selecting **Copy ID**.)
3. Now, click on the **plus sign** to create a new query. Notice that BigQuery doesn't automatically generate a `SELECT` statement in this window. Enter `SELECT` and add a space after it.
4. Put an asterisk `*` after `SELECT` to indicate you want to return the entire dataset. The asterisk lets the database know to include all columns. Without this shortcut, you would have to manually enter every column name!
5. Next, press the **Enter/Return** key and enter `FROM` on the second line. `FROM` indicates where the data is coming from. After `FROM`, add another space.
6. Paste in the path to the table that you copied earlier. It will read `` `bigquery-public-data.sunroof_solar.solar_potential_by_postal_code` ``
7. Execute the query by selecting the **RUN** button.



Row	region_name	state_name	lat_max	lat_min	lng_max	lng_min	lat_avg	lng_avg
1	19149	Pennsylvania	40.0512...	40.0268795	-75.0413555	-75.0925965	40.03740444034...	-75.0657051746...
2	19120	Pennsylvania	40.0527...	40.0124032	-75.0969875	-75.1387471	40.03439829328...	-75.1208872435...
3	19132	Pennsylvania	40.0058...	39.9863578	-75.1506389	-75.1906974	39.99568149075...	-75.1696451592...

Important!
Many of the public databases on BigQuery are living records and, as such, are periodically updated with new data. Throughout this course (and others in this certificate program), if your results differ from those you encounter in videos or screenshots, there's a good chance it is due to a data refresh. You can verify when a table has been refreshed by selecting it from the **Explorer** panel and clicking **Details**. You'll find the date the table was created, when it was last modified, as well as other useful information.



Example 3: Use SQL to view a piece of data

If the project doesn't require every field to be completed, you can use SQL to see a particular piece, or pieces, of data. To do this, specify a certain column name in the query.

1. For example, you might only need data from Pennsylvania. You'd begin your query the same way you just did in the previous examples: Click on the **plus sign**, enter `SELECT`, add a space, an asterisk (`*`), and then press **Enter/Return**.
2. Enter `FROM` and then paste `` `bigquery-public-data.sunroof_solar.solar_potential_by_postal_code` ``. Press **Enter/Return**.
3. This time, add `WHERE`. It will be on the same line as the `FROM` statement. Add a space and enter `state_name` with a space before state and a space after name. `state_name` is a column name in the table.
4. Because you only want data from Pennsylvania, add = and `'Pennsylvania'` on the same line as `state_name`. In SQL, single quotes represent the beginning and ending of a string.
5. Execute the query with the **RUN** button.
6. Review the data on solar potential for Pennsylvania. Scroll through the query results.

Keep in mind that SQL queries can be written in a lot of different ways and still return the same results. You might discover other ways to write these queries!

Get started with other databases (if not using BigQuery)

It's easiest to follow along with the course activities if you use BigQuery, but you may use other SQL platforms, if you prefer. If you decide to practice SQL queries on other database platforms, here are some resources to get started:

- Getting Started with MySQL
- Getting Started with Microsoft SQL Server
- Getting Started with PostgreSQL
- Getting Started with SQLite