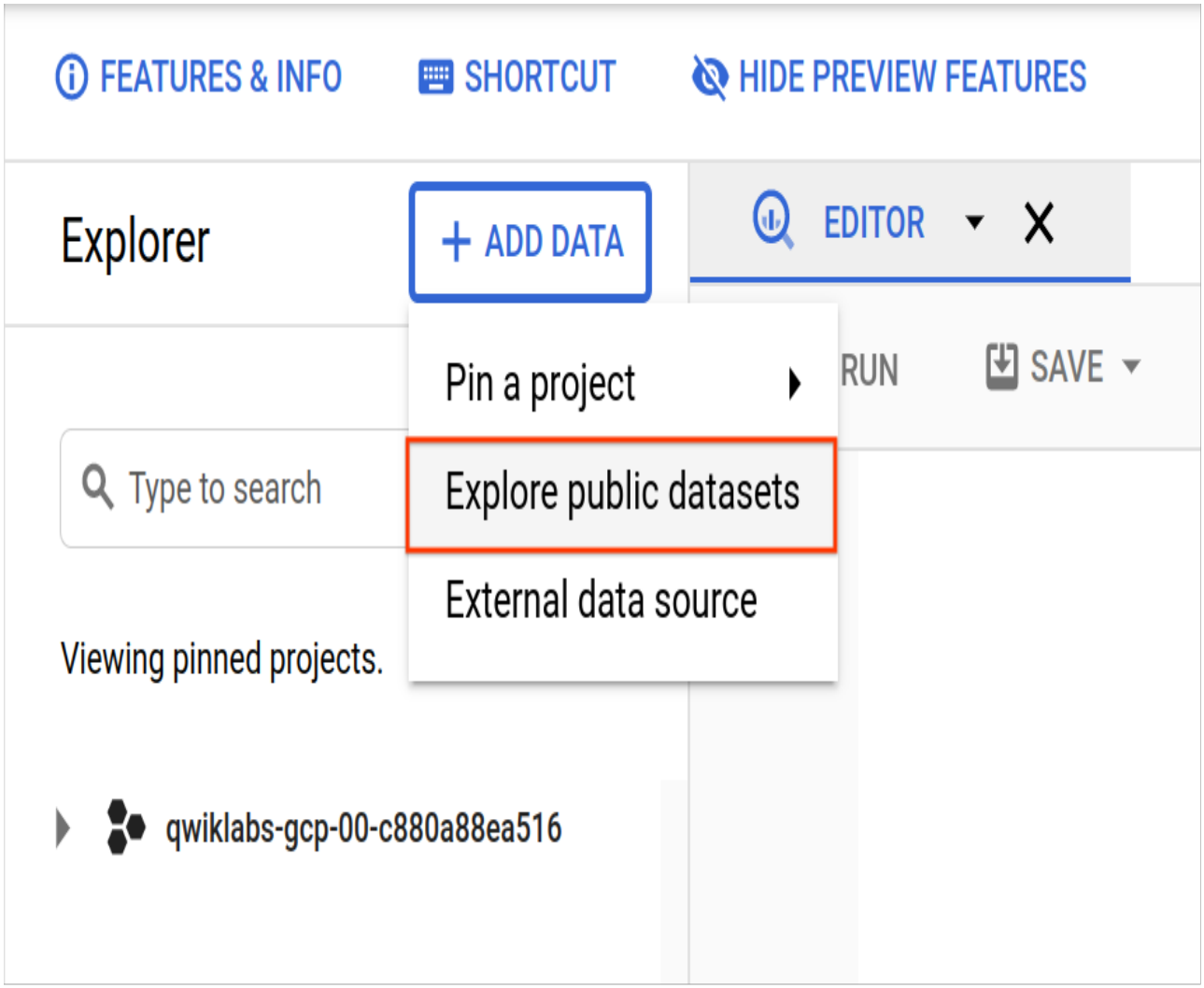
Query a public dataset

In this section, you load a public dataset, USA Names, into BigQuery, then query the dataset to determine the most common names in the US between 1910 and 2013.

Load USA Name dataset

1.In the left pane, click **ADD DATA** > **Explore public datasets**.



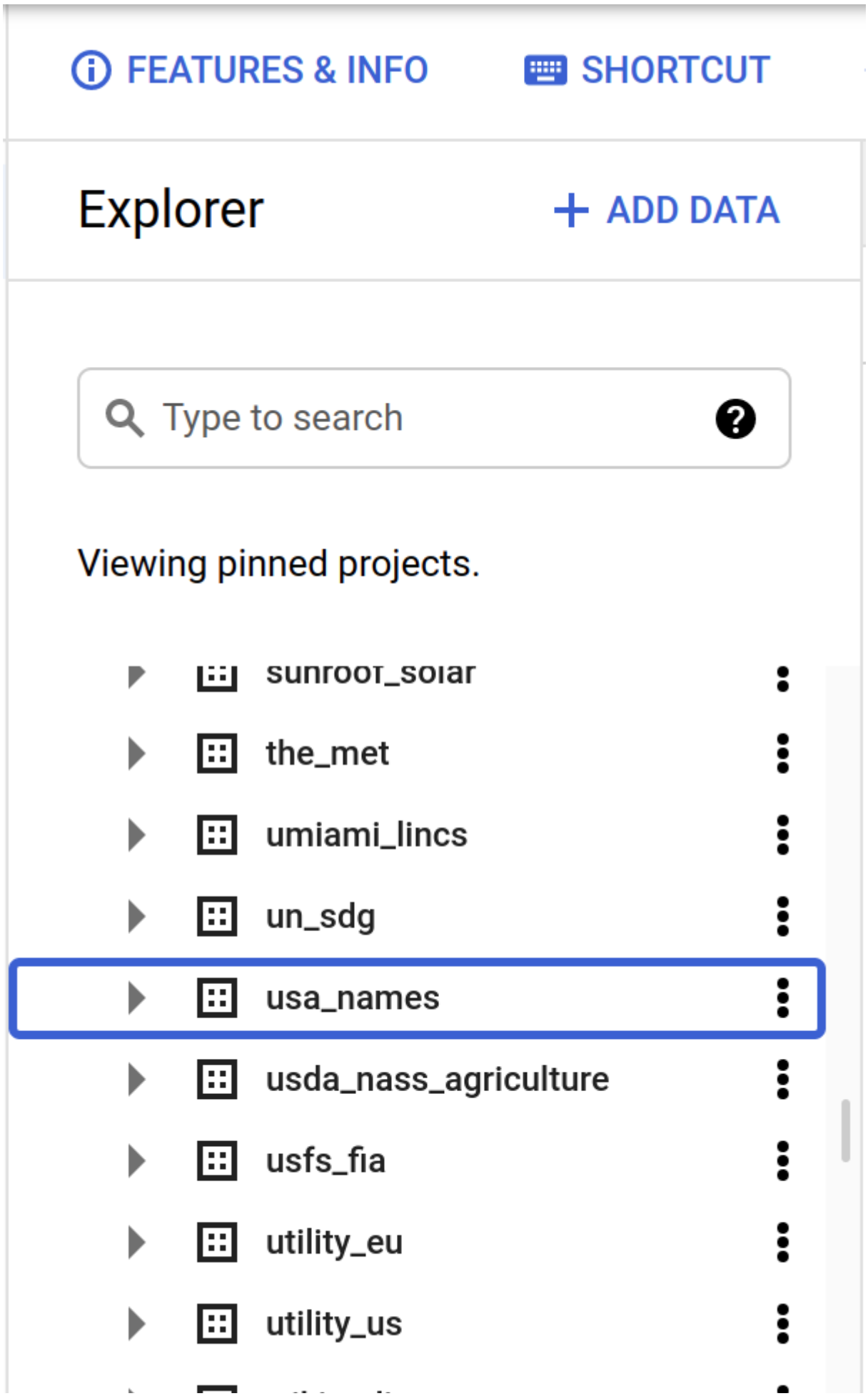
The Datasets window opens.

2.In the searchbox, type "USA Names" then enter.

3.Click on the **USA Names** tile you see in the search results.

4.Click **VIEW DATASET**.

BigQuery opens in a new browser tab. The project bigquery-public-data is added to your resources and you see the dataset usa_names listed in the left pane in your **Explorer** section.


FEATURES & INFO
SHORTCUT
Explorer
+ ADD DATA
Type to search
Viewing pinned projects.
sunroot_solar
the_met
umiami_lincs
un_sdg
usa_names
usda_nass_agriculture
usfs_fia
utility_eu
utility_us

Query the USA Name dataset

Query bigquery-public-data.usa_names.usa_1910_2013 for the name and gender of the babies in this dataset, and then list the top 10 names in descending order.

1.Copy and paste the following query into the query **EDITOR** text area:

```
SELECT
  name, gender,
  SUM(number) AS total
FROM
  `bigquery-public-data.usa_names.usa_1910_2013`
GROUP BY
  name, gender
ORDER BY
  total DESC
LIMIT
  10content_copy
```

2.In the upper right of the window, view the query validator.

BigQuery displays a green check mark icon if the query is valid. If the query is invalid, a red exclamation point icon is displayed. When the query is valid, the validator also shows the amount of data the query processes when you run it. This helps to determine the cost of running the query.

3.Click **Run**.

The query results opens below the Query editor. At the top of the Query results section, BigQuery displays the time elapsed and the data processed by the query. Below the time is the table that displays the query results. The header row contains the name of the column as specified in GROUP BY in the query.

Query results SAVE RESULTS EXPLORE DATA

Query complete (1.5 sec elapsed, 99.9 MB processed)

Job information | Results | JSON | Execution details

Row	name	gender	total
1	James	M	4924235
2	John	M	4818746
3	Robert	M	4703680
4	Michael	M	4280040
5	William	M	3811998
6	Mary	F	3728041
7	David	M	3541625
8	Richard	M	2526927
9	Joseph	M	2467298
10	Charles	M	2237170

Click **Check my progress** to verify the objective.

Create a custom table

In this section, you create a custom table, load data into it, and then run a query against the table.

Download the data to your local computer

The file you're downloading contains approximately 7 MB of data about popular baby names, and it is provided by the US Social Security Administration.

1.Download the baby names zip file to your local computer.

2.Unzip the file onto your computer.

3.The zip file contains a NationalReadMe.pdf file that describes the dataset. Learn more about the dataset.

4.Open the file named yob2014.txt to see what the data looks like. The file is a comma-separated value (CSV) file with the following three columns: name, sex (M or F), and number of children with that name. The file has no header row.

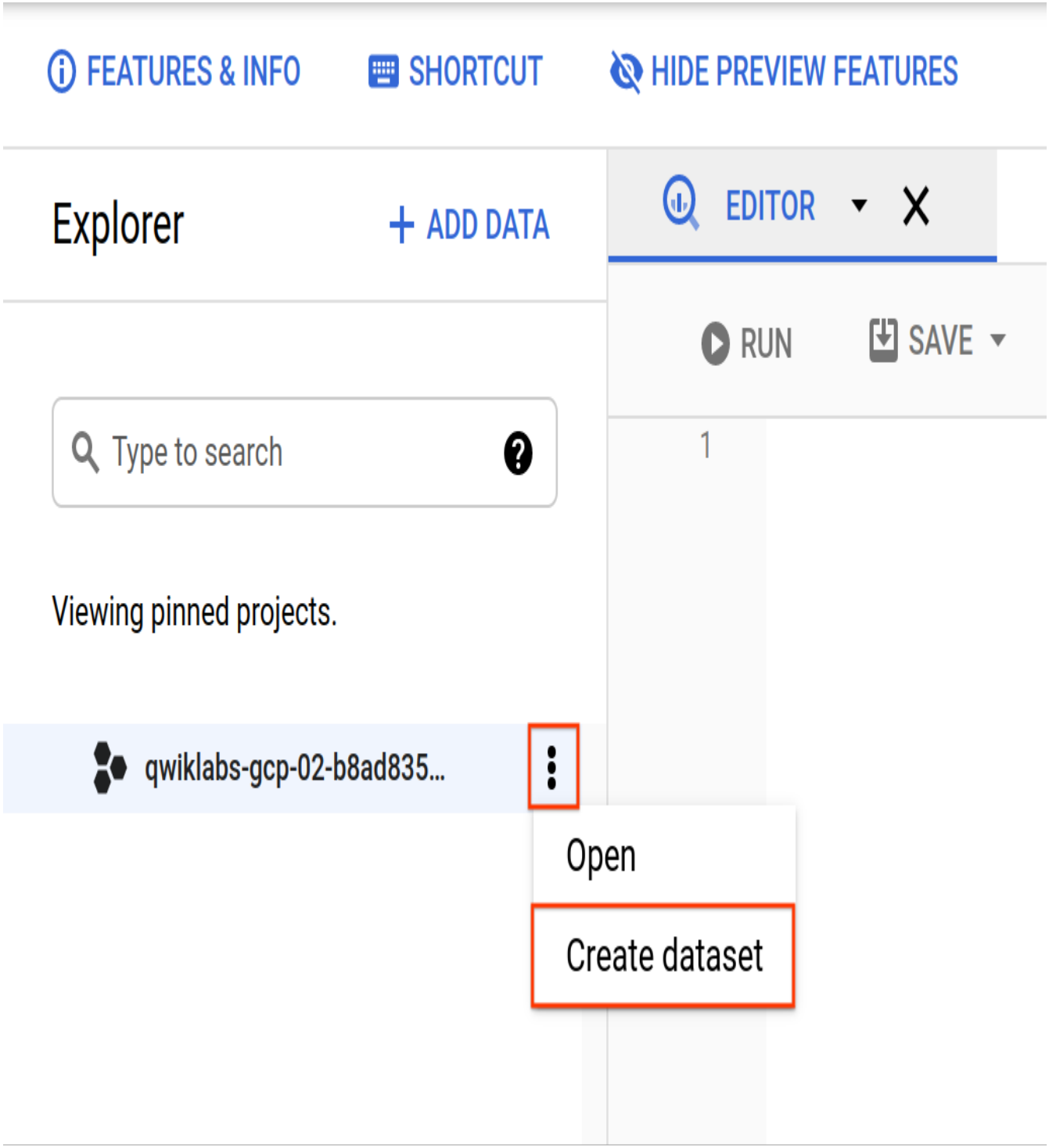
5.Note the location of the yob2014.txt file so that you can find it later.

Create a dataset

In this section, you create a dataset to hold your table, add data to your project, then make the data table you'll query against.

Datasets help you control access to tables and views in a project. This lab uses only one table, but you still need a dataset to hold the table.

1.Back in the console, in the **Explorer** section, click on the **View actions** icon next to your project ID and select **Create dataset**.

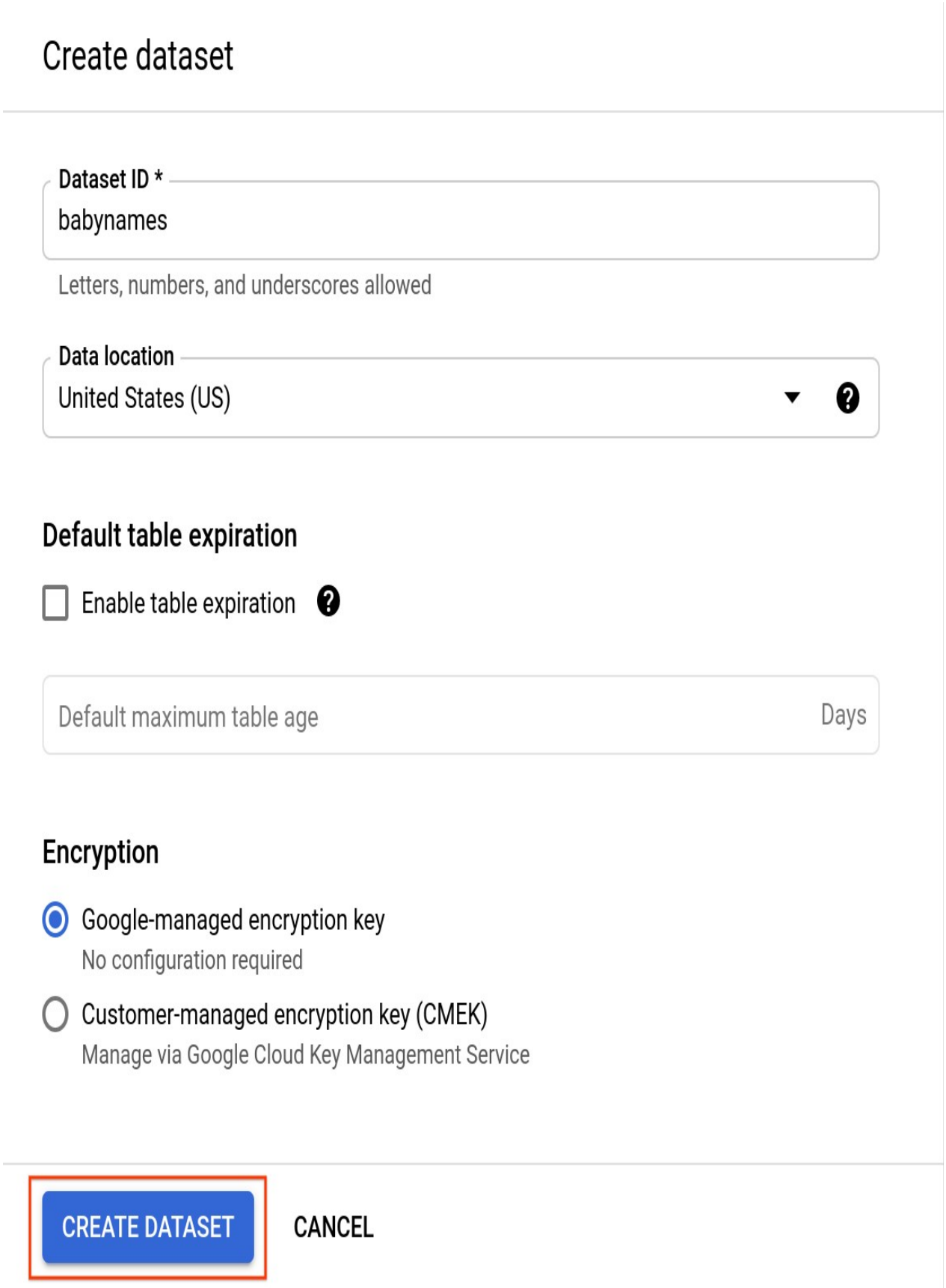


2.On the **Create dataset** page:

•For **Dataset ID**, enter babynames.
•For **Data location**, choose **United States (US)**.

•For **Default table expiration**, leave the default value.

Currently, the public datasets are stored in the US multi-region location. For simplicity, place your dataset in the same location.



3.Click **Create dataset** at the bottom of the panel.

Click **Check my progress** to verify the objective.

Load the data into a new table

In this section, you load data into the table you made.

1.Create a table by clicking on the **View actions** icon next to your **babynames** dataset in the **Explorer** section. Select **Open**, then click **Create table**.
Use the default values for all settings unless otherwise indicated.

2.On the **Create table** page:

•For **Source**, choose **Upload** from the dropdown menu.

•For **Select file**, click **Browse**, navigate to the yob2014.txt file and click **Open**.

•For **File format**, choose **CSV** from the dropdown menu.

•For **Table name**, enter names_2014.

•In the **Schema** section, click the **Edit as text** toggle and paste the following schema definition in the text box.

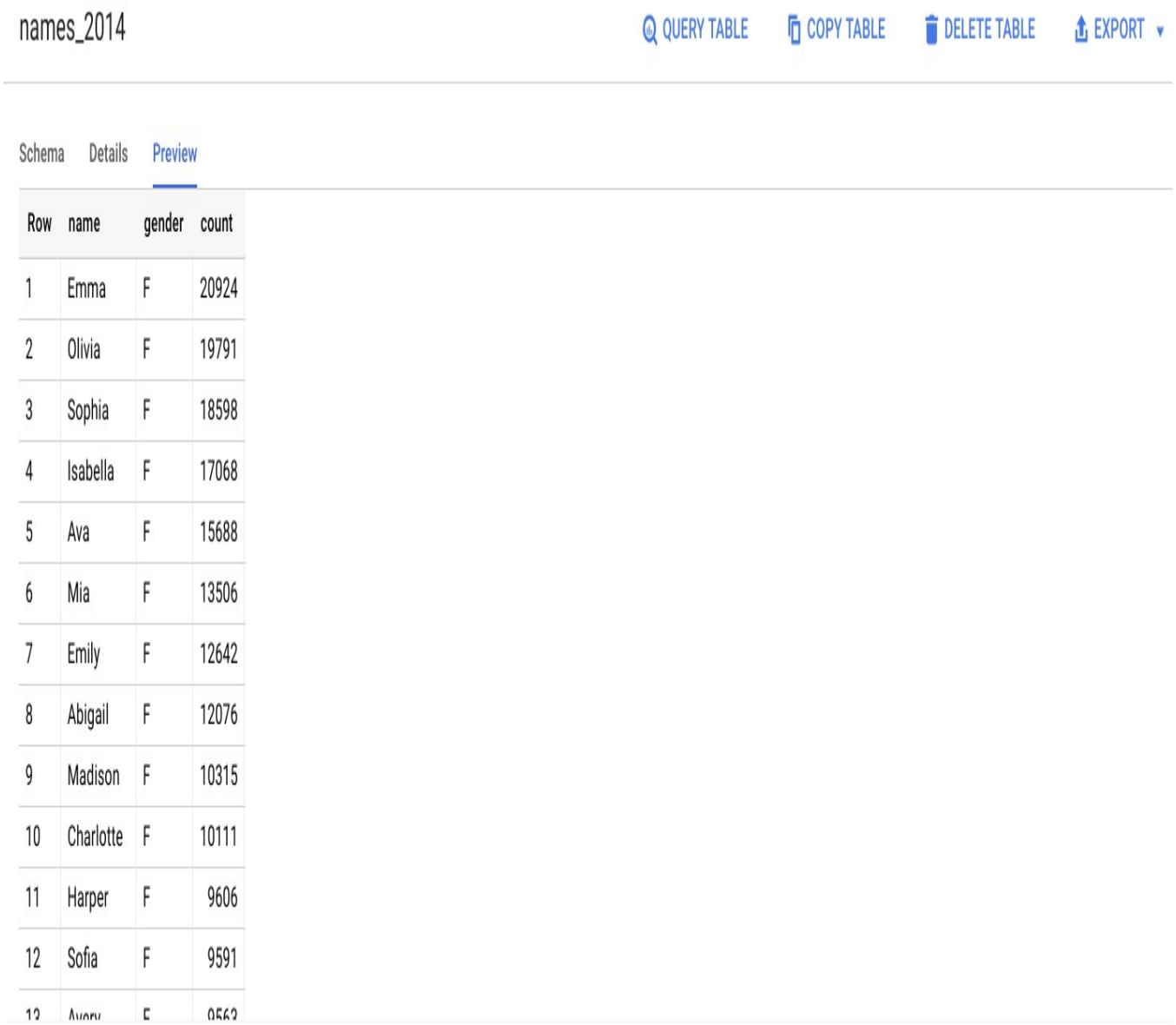
```
name:string,gender:string,count:integer
```
content_copy

3.Click **Create table** (at the bottom of the window).

4.Wait for BigQuery to create the table and load the data. While BigQuery loads the data, a **(1 running)** string displays beside the **Job history** in the left pane. The string disappears after the data is loaded.

Preview the table

1. In the left pane, select **babynames** > **names_2014** in the navigation panel.
2. In the details panel, click the **Preview** tab.



Row	name	gender	count
1	Emma	F	20924
2	Olivia	F	19791
3	Sophia	F	18598
4	Isabella	F	17068
5	Ava	F	15688
6	Mia	F	13506
7	Emily	F	12642
8	Abigail	F	12076
9	Madison	F	10315
10	Charlotte	F	10111
11	Harper	F	9606
12	Sofia	F	9591
13	Avery	F	9563

~~True~~

False

Query the table

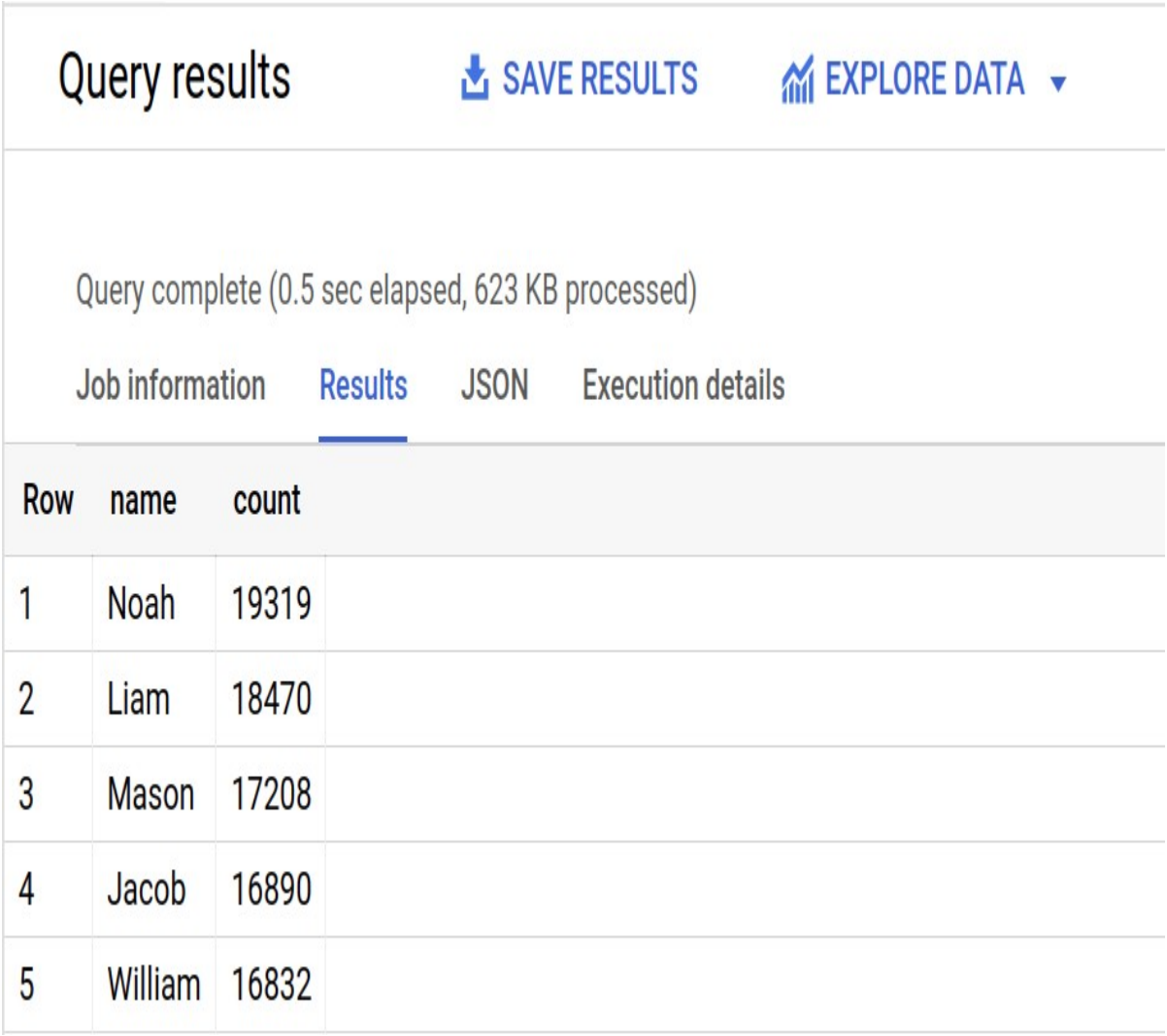
Now that you've loaded data into your table, you can run queries against it. The process is identical to the previous example, except that this time, you're querying your table instead of a public table.

1.In the Query editor, click **Compose new query**.

2.Copy and paste the following query into the query **EDITOR**. This query retrieves the top 5 baby names for US males in 2014.

```
SELECT
  name, count
FROM
  `babynames.names_2014`
WHERE
  gender = 'M'
ORDER BY count DESC LIMIT 5content_copy
```

3.Click **Run**. The results are displayed below the query window.



Row	name	count
1	Noah	19319
2	Liam	18470
3	Mason	17208
4	Jacob	16890
5	William	16832

Click **Check my progress** to verify the objective.

Congratulations!

You queried a public dataset, then created a custom table, loaded data into it, and then ran a query against that table.

Finish Your Quest

This self-paced lab is part of the Qwiklabs NCAA® March Madness®: Bracketology with Google Cloud, and Data Catalog Fundamentals Quests. 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 (or badges) public and link to them in your online resume or social media account. Enroll in a Quest and get immediate completion credit if you've taken this lab. See other available Qwiklabs Quests.