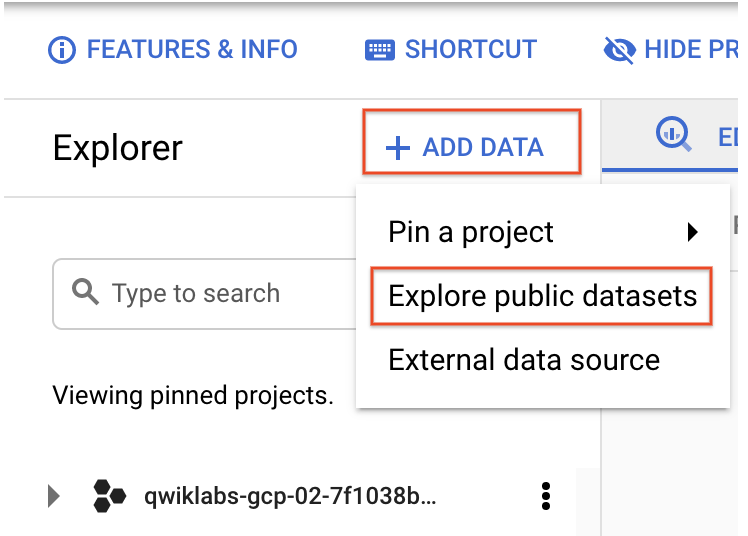
**Task 1. Query a public dataset**

In this task, 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.

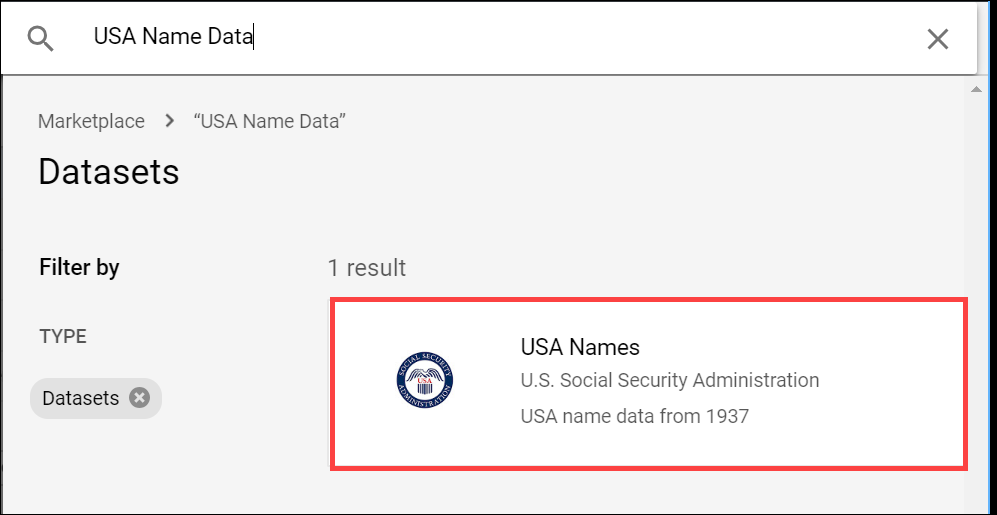
Load the USA Names dataset

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



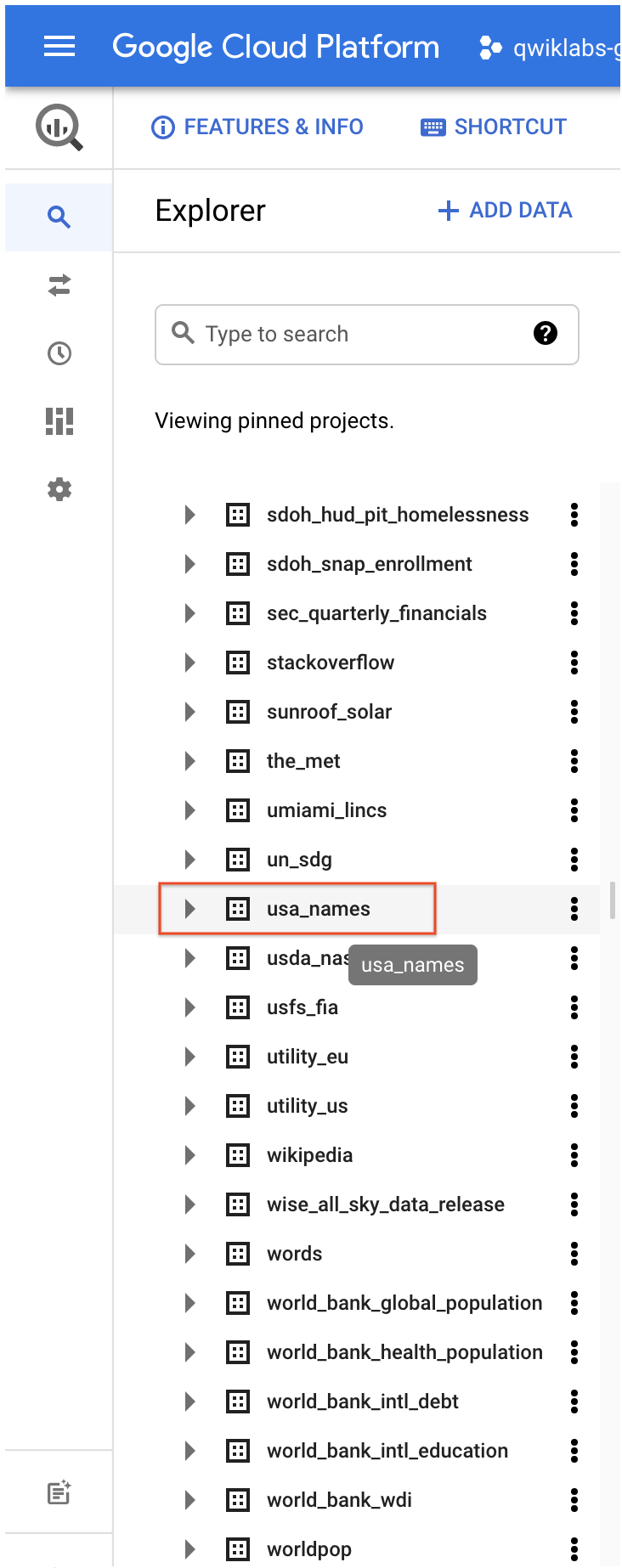
The Datasets window opens.

1. In the searchbox, type USA Names then press ENTER.
2. Click on the **USA Names** tile you see in the search results.



1. 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 Resources tree.



Query the USA Names 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

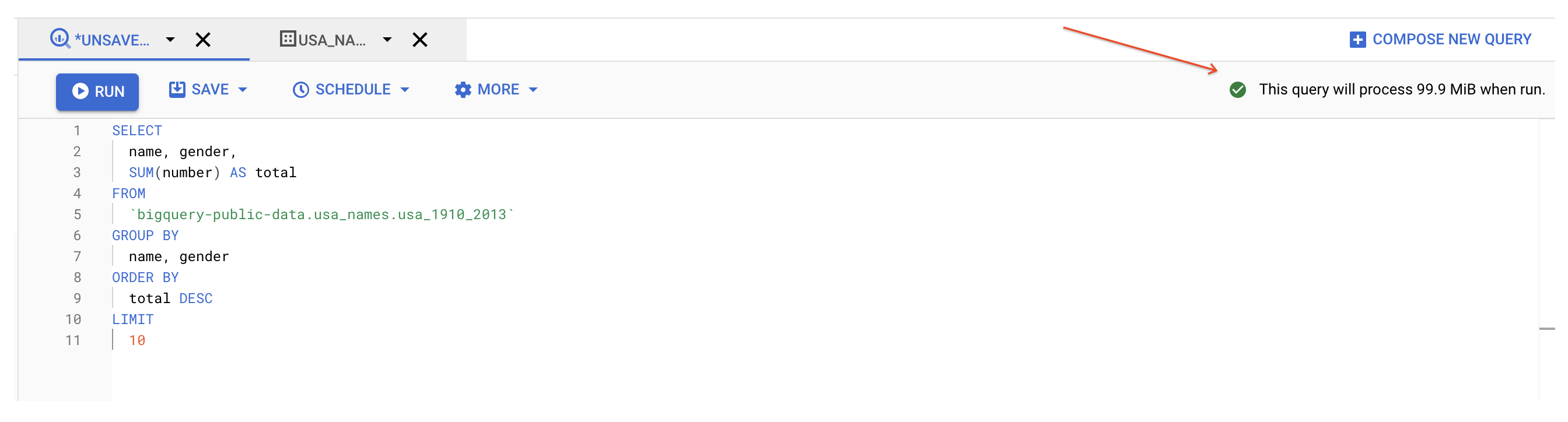
LIMIT

10

Copied!

content\_copy

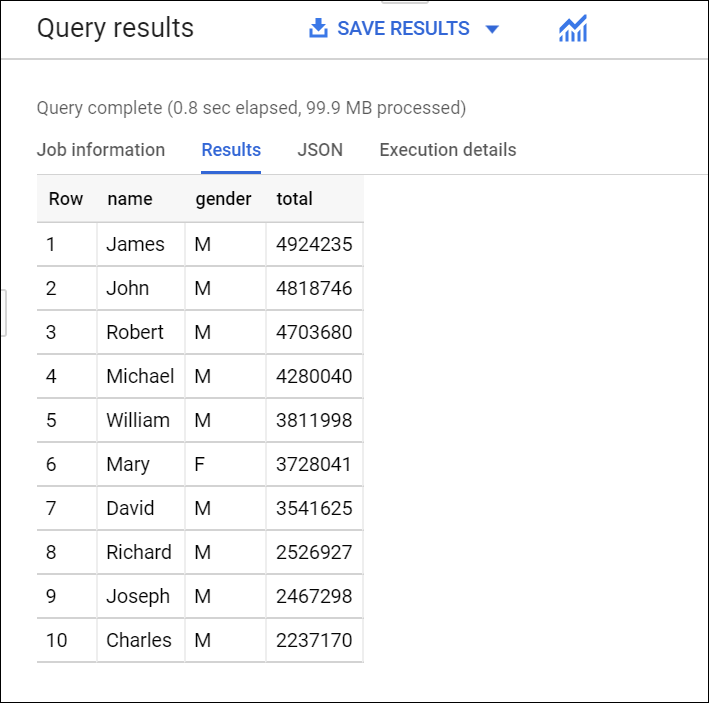
1. In the upper right of the window, view the query validator.



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.

1. 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.



**Task 2. Create a custom table**

In this task, 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](https://www.ssa.gov/OACT/babynames/names.zip) to your local computer.

**Note:** If this download link fails please copy the baby names zip file from the student resources on the left pane of the instruction guide.

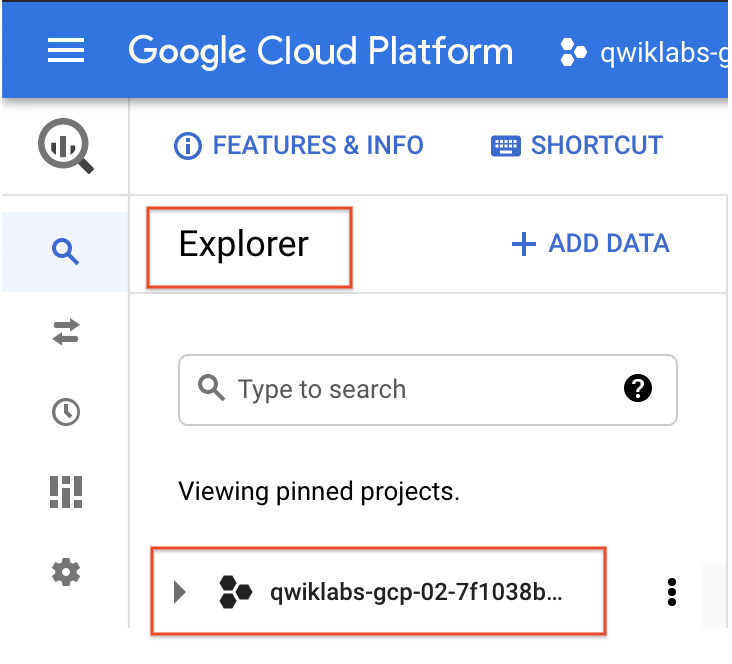
1. Unzip the file onto your computer.
2. The zip file contains a NationalReadMe.pdf file that describes the dataset. [Learn more about the dataset](https://www.ssa.gov/OACT/babynames/background.html).
3. Open the file named yob2014.txt to see what the data looks like. The file is a comma-separated value (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.
4. Note the location of the yob2014.txt file so that you can find it later.

**Task 3. Create a dataset**

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

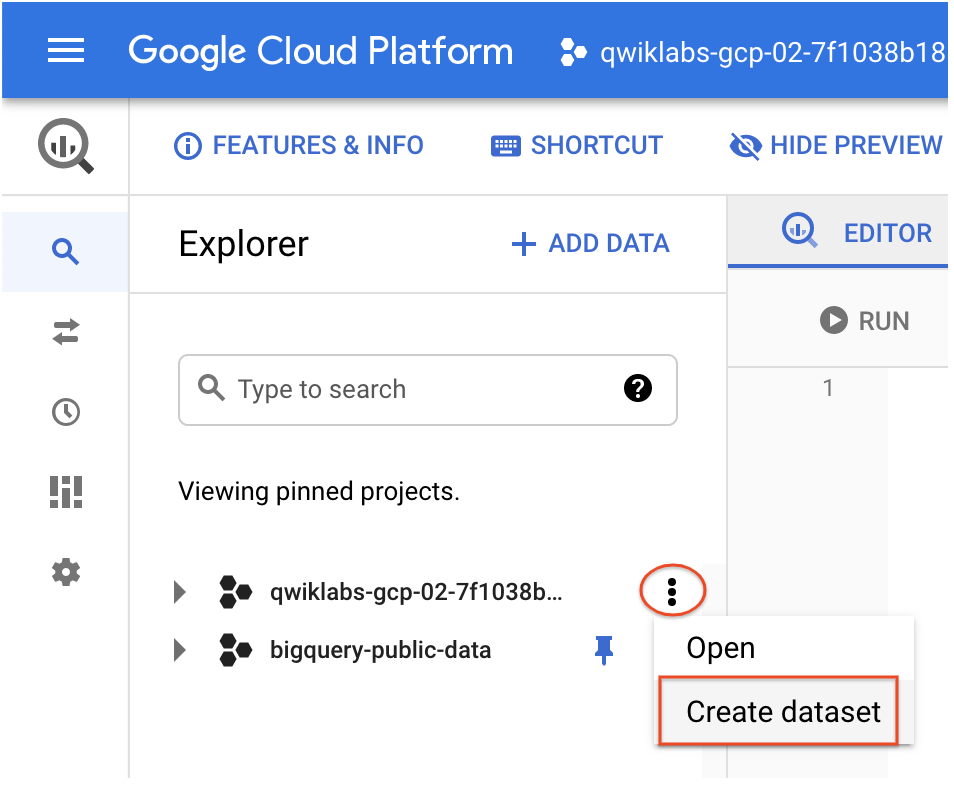
Datasets help you control access to tables and views in a project. This lab uses only one table, but you still need a dataset to hold the table.

1. Back in the Cloud Console, in the left pane, in the **Explorer** section, click your Project ID (it will start with qwiklabs).



Your project opens under the Query editor.

1. Click on the three dots next to your project ID and then click **Create dataset**.



1. On the **Create dataset** page:

* For **Dataset ID**, enter babynames.
* For **Data location**, choose **us (multiple regions in United States)**.
* For **Default table expiration**, leave the default value.
* For **Encryption**, leave the default value.

1. Click **Create dataset** at the bottom of the pane.

**Task 4. Load the data into a new table**

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

1. Click **babynames** found in the left pane in the **Explorer** section, and then click **Create table**.

Use the default values for all settings unless otherwise indicated.

1. On the **Create table** page:

* For **Source**, choose **Upload** from the Create table from: 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

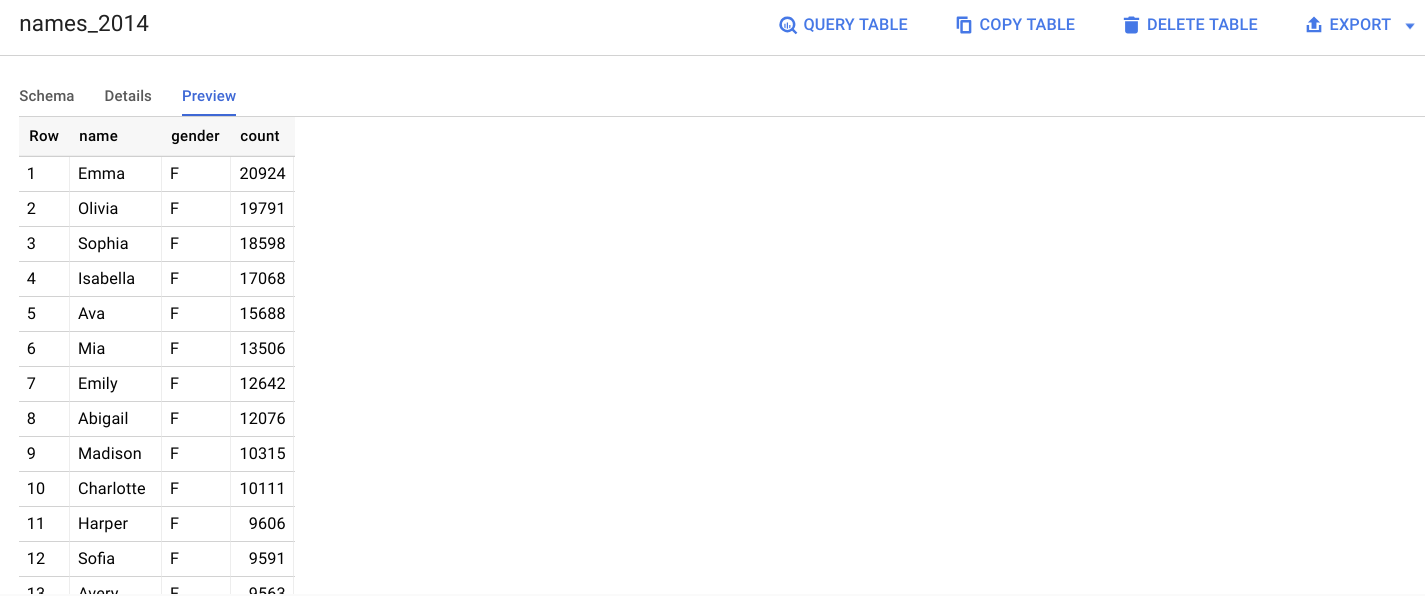
Copied!

content\_copy

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

Preview the table

1. In the left pane, select **babynames** > **names\_2014** in the navigation pane.
2. In the details pane, click the **Preview** tab.



**Task 5. 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.

Note: Inside '' it does distinguish upper vs. lower case, therefore make sure to align exactly the names of the dataset and the table you created.

SELECT

name, count

FROM

`babynames.names\_2014`

WHERE

gender = 'M'

ORDER BY count DESC LIMIT 5

Copied!

content\_copy

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

