Congratulations! You passed!
Grade received 100%To pass 100% or higher

1. Activity overview

So far, you've learned about SQL and used SQL queries to interact with databases. In this activity, you'll practice sorting data by using SQL queries with ORDER BY and WHERE clauses.
By the time you complete this activity, you will be able to write queries that sort data depending on your needs. This will enable you to organize and use data more efficiently in your career as a data analyst.

Sorting with SQL

To practice sorting data with SQL, you'll query the CDC Births Data Summary public dataset. The queries you write will help you obtain some answers about which counties in the United States have the most and least births in the years 2016-2018.

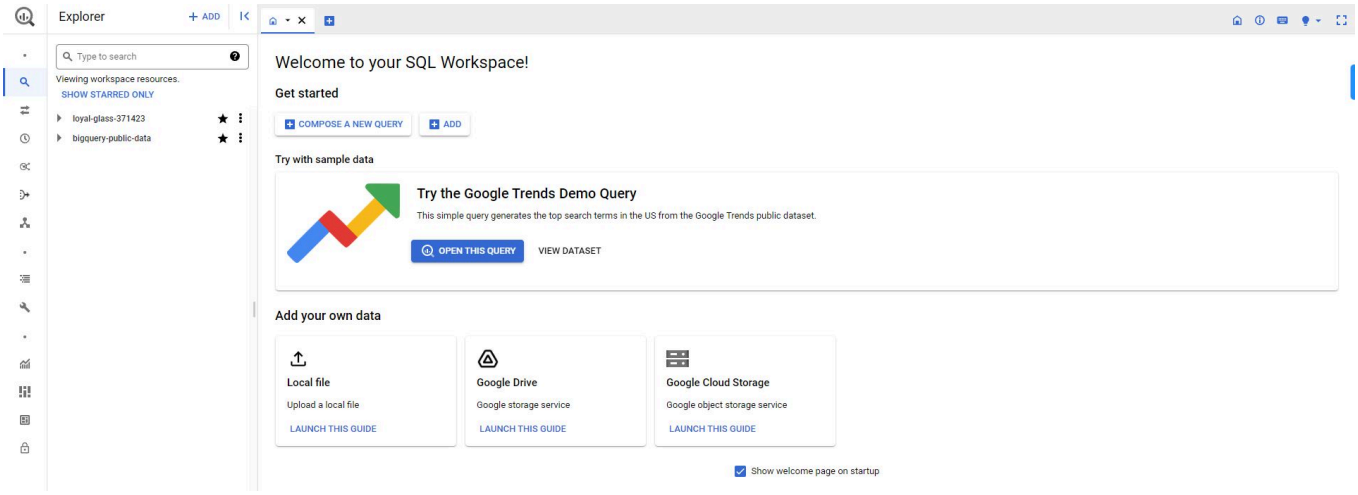
Load the dataset

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.

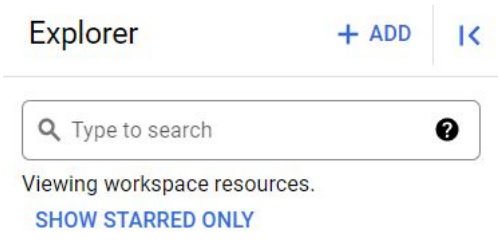
- Note: BigQuery Sandbox frequently updates its user interface. The latest changes may not be reflected in the screenshots presented in this activity, but the principles remain the same. Adapting to changes in software updates is an essential skill for data analysts, and it's helpful for you to practice 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 drop-down in the data box to the right of the *Google Cloud* icon in the top left corner and selecting NEW PROJECT.

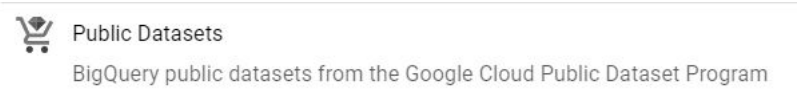
3. Name your project something that will help you identify it later. You can give it a unique project ID or use the auto-generated name that is created for you in the data box. Within the *Location* section, keep the default option set to No organization.
4. Now, you'll see the Editor interface. In the 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.



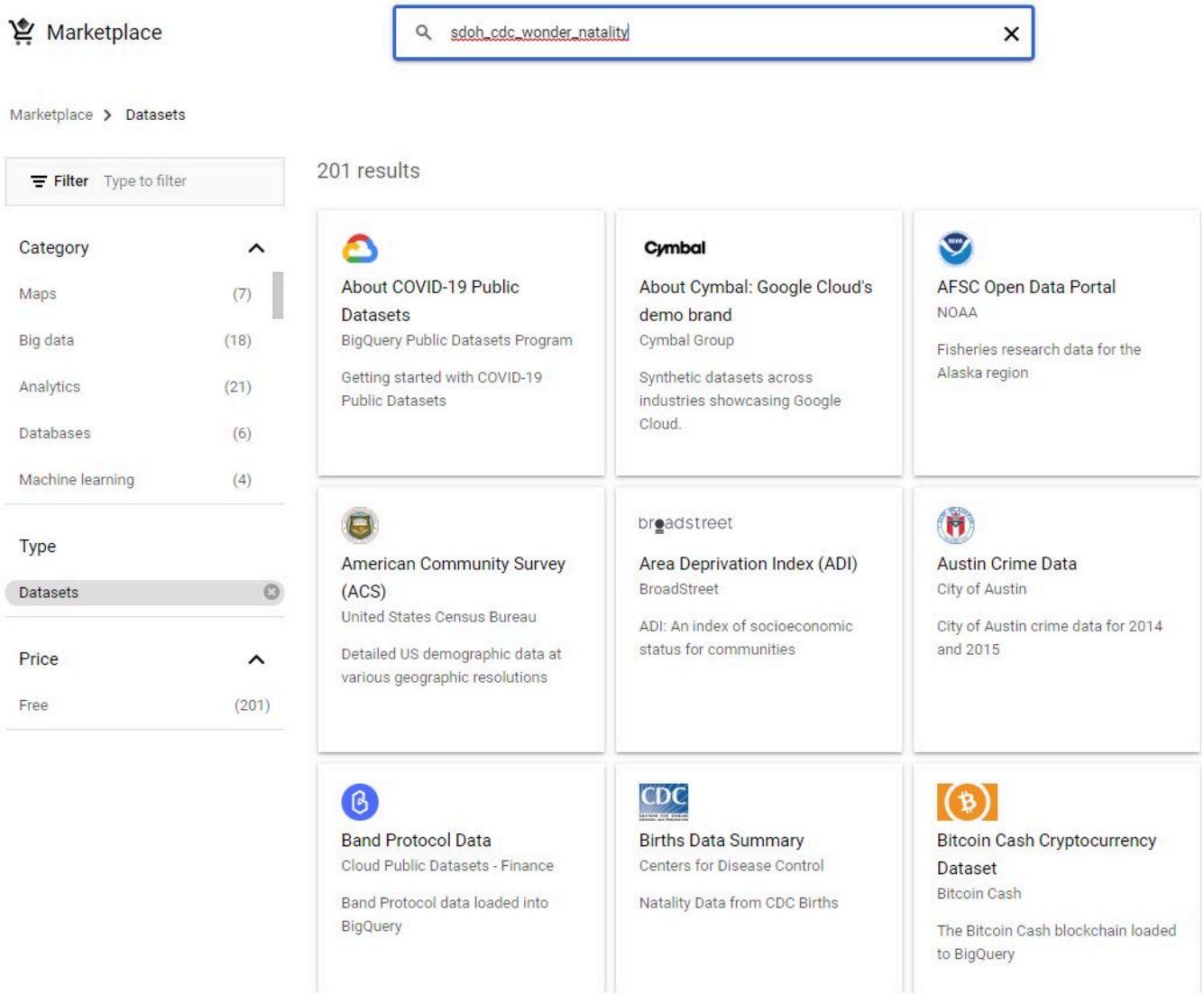
5. Click + ADD at the top of the Explorer menu pane, which will call up the Add menu of various types of public data repositories.



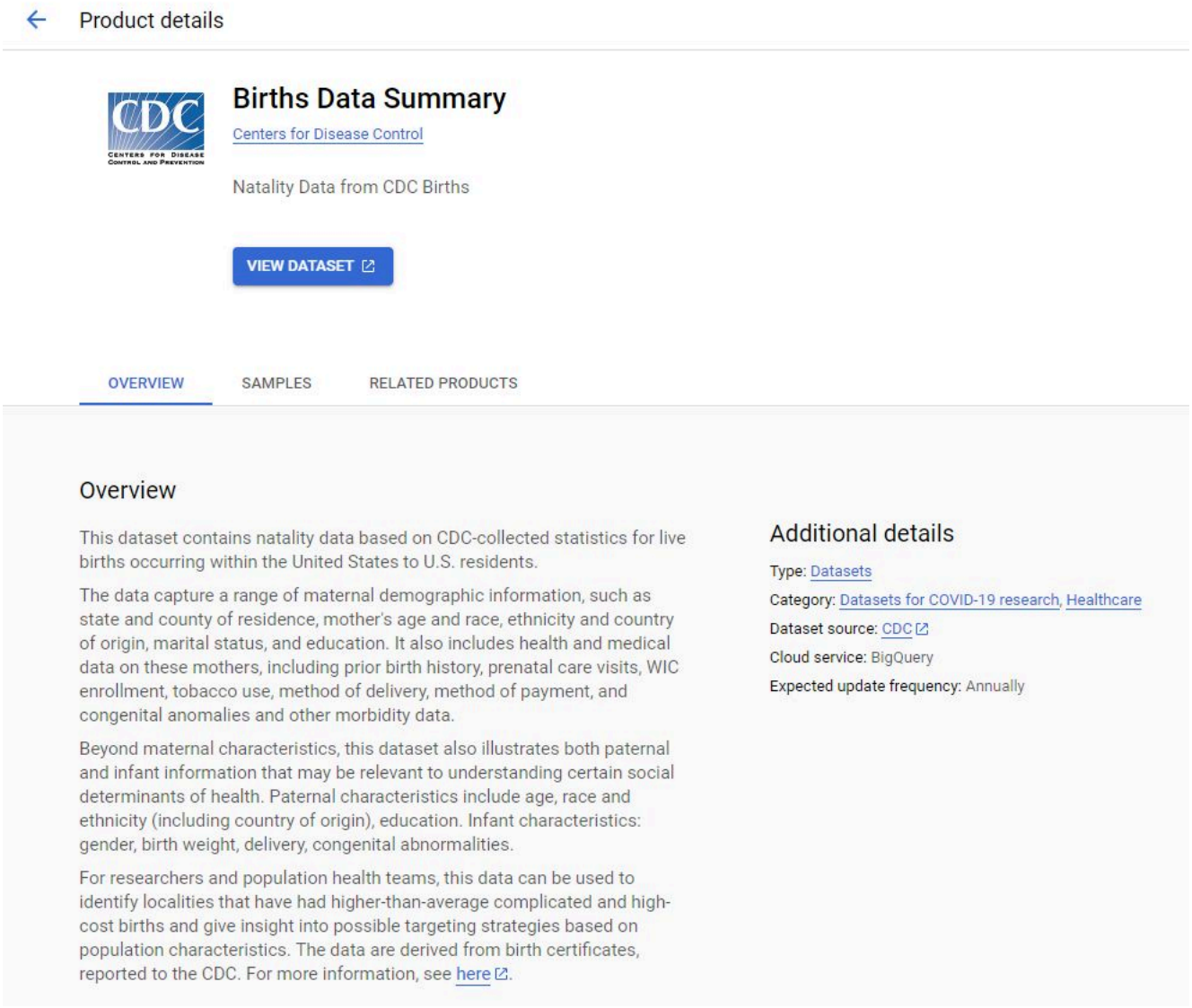
6. Navigate down the list in the Add window to select the Public Datasets option.



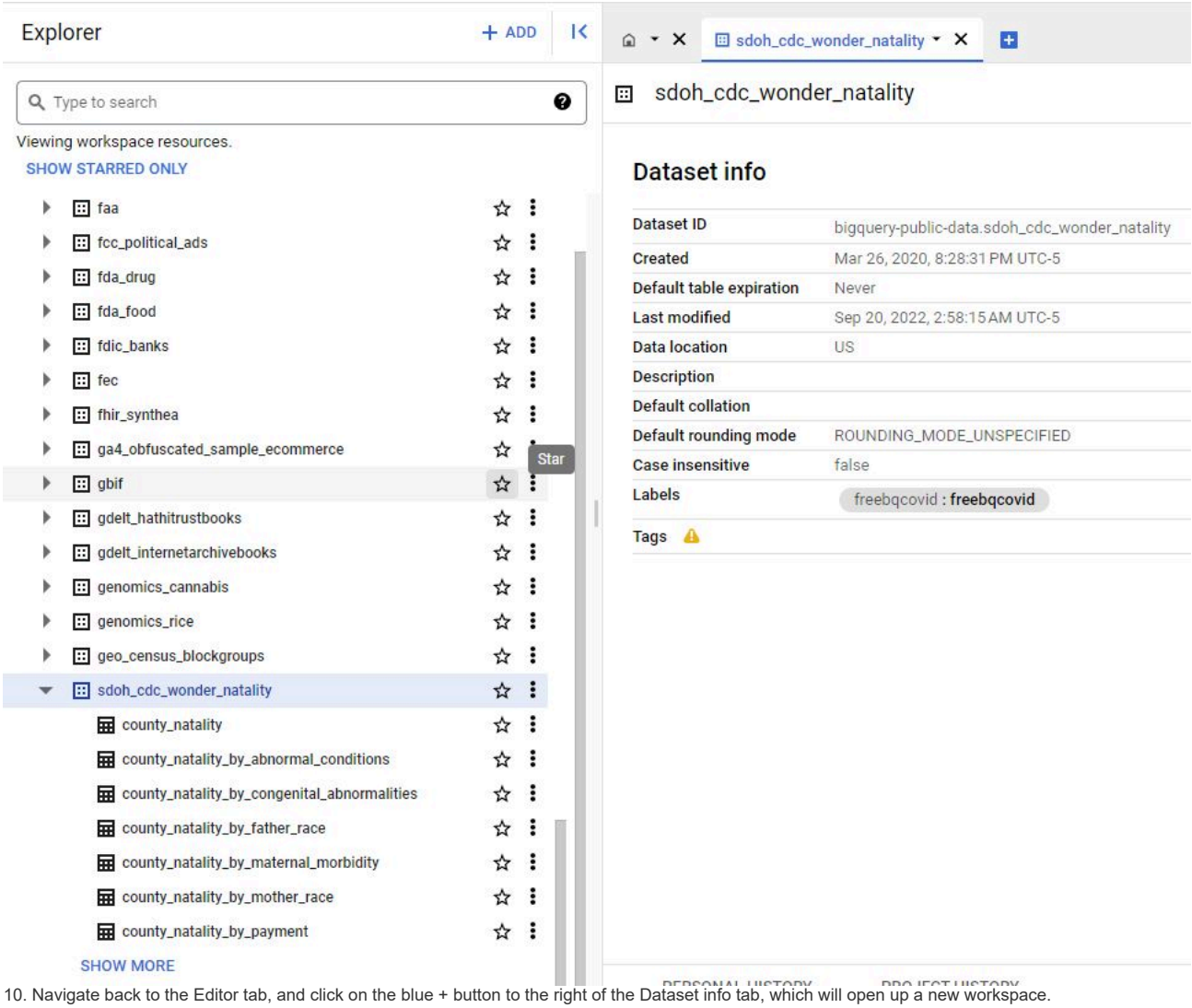
7. In the Marketplace search bar, type sdoh_cdc_wonder_natality.



8. In the resulting window, click the CDC Births Data Summary. After reviewing the detailed dataset info, click on the VIEW DATASET button.



9. This will bring you back to the BigQuery Sandbox interface in a new tab. This will also open up the bigquery-public-data drop-down list in the Explorer menu. You can use this to browse datasets and tables.

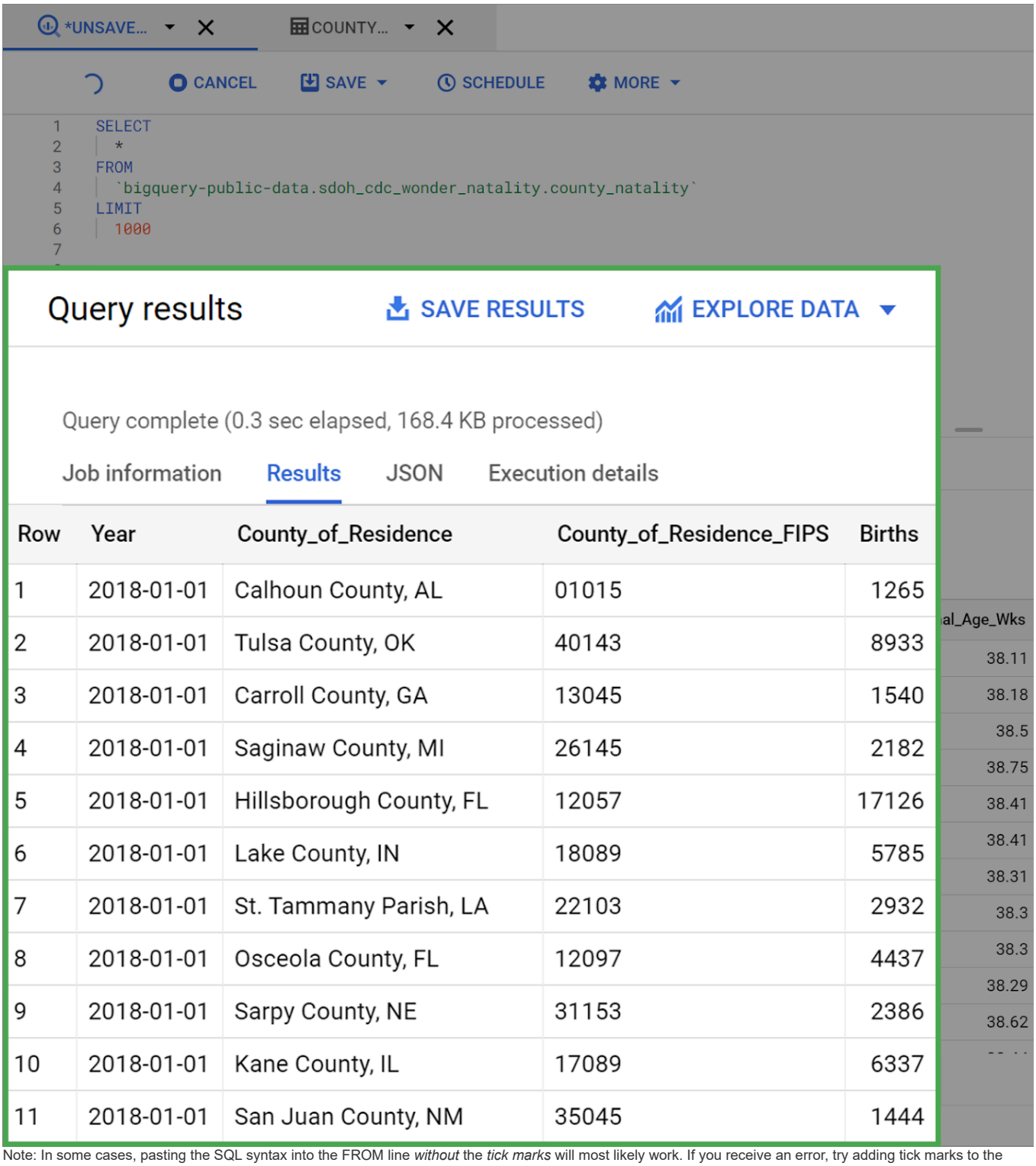


10. Navigate back to the Editor tab, and click on the blue + button to the right of the Dataset info tab, which will open up a new workspace.

11. This is where you will generate a new query from the public data list. Copy, paste, and run the following query to display the first 1,000 rows of the county_natality table:

```
SELECT
  *
FROM
  bigquery-public-data.sdoh_cdc_wonder_natality.county_natality
LIMIT
  1000
```

After the query has run, your results should appear like this:



Note: In some cases, pasting the SQL syntax into the FROM line *without* the *tick marks* will most likely work. If you receive an error, try adding tick marks to the ends of the Table ID syntax, and this may fix the error. Follow this method for the rest of the *copy + paste* code chunks for the remainder of the activity.

Use the ORDER BY clause

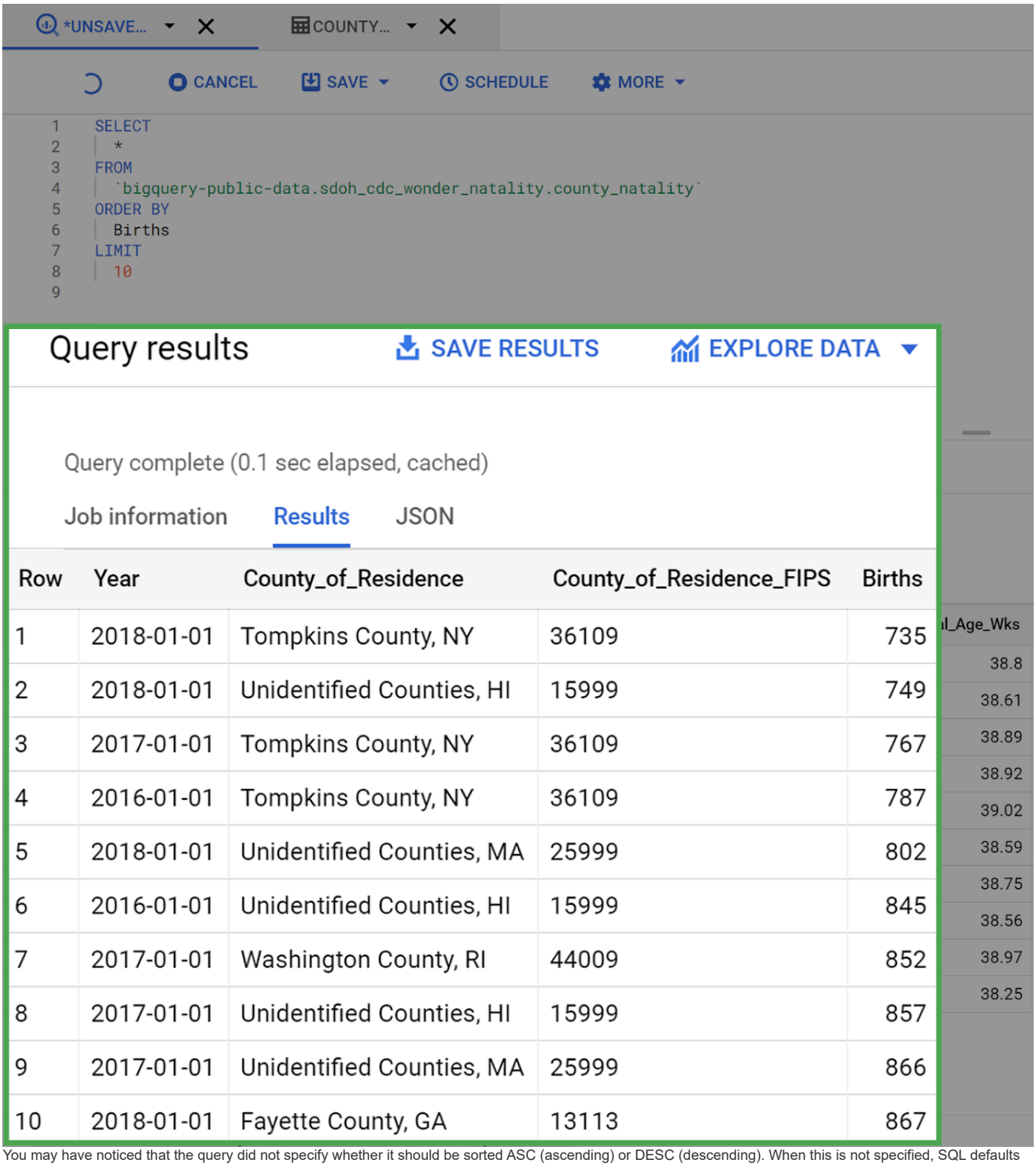
Examine the dataset you just loaded. Take a moment to familiarize yourself with the columns and get a feel for what each can tell you.

Now, imagine you were asked by your manager to figure out which 10 counties had the lowest birth count for 2016-2018. You could accomplish this by modifying your query to use the ORDER BY clause.

Copy, paste, and run the following query:

```
SELECT
  *
FROM
  bigquery-public-data.sdoh_cdc_wonder_natality.county_natality
ORDER BY
  Births
LIMIT
  10
```

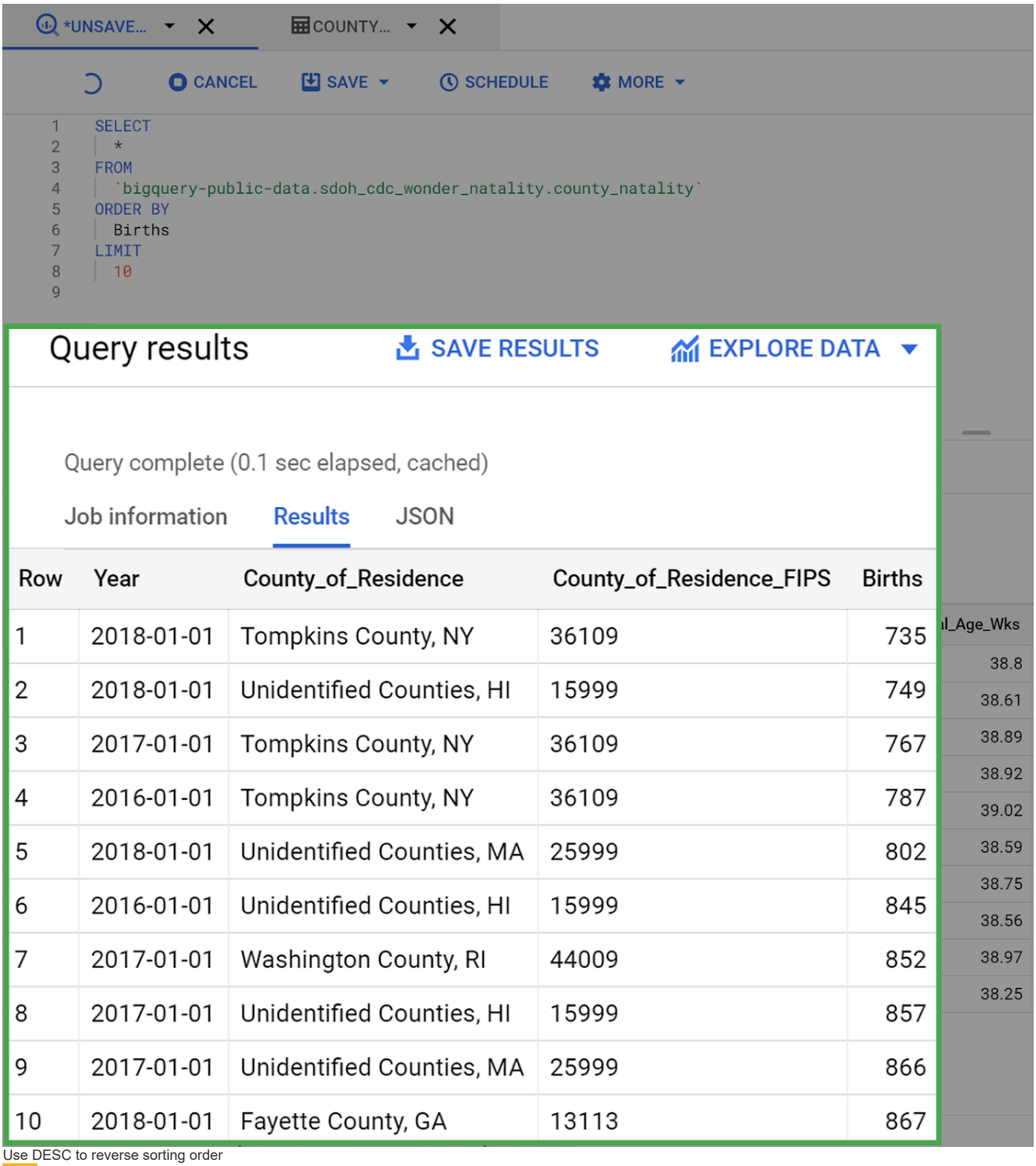
The results of your query should appear like this:



You may have noticed that the query did not specify whether it should be sorted ASC (ascending) or DESC (descending). When this is not specified, SQL defaults to sorting by ascending order. You can run another query to confirm this.
Copy, paste, and run the following query that includes ASC:

```
SELECT
  *
FROM
  bigquery-public-data.sdoh_cdc_wonder_natality.county_natality
ORDER BY
  Births
ASC
LIMIT
  10
```

You'll find that the results did not change. Notice that Tompkins County, NY, had just 735 births in 2018—the lowest birth count of any county in the US between 2016-2018.



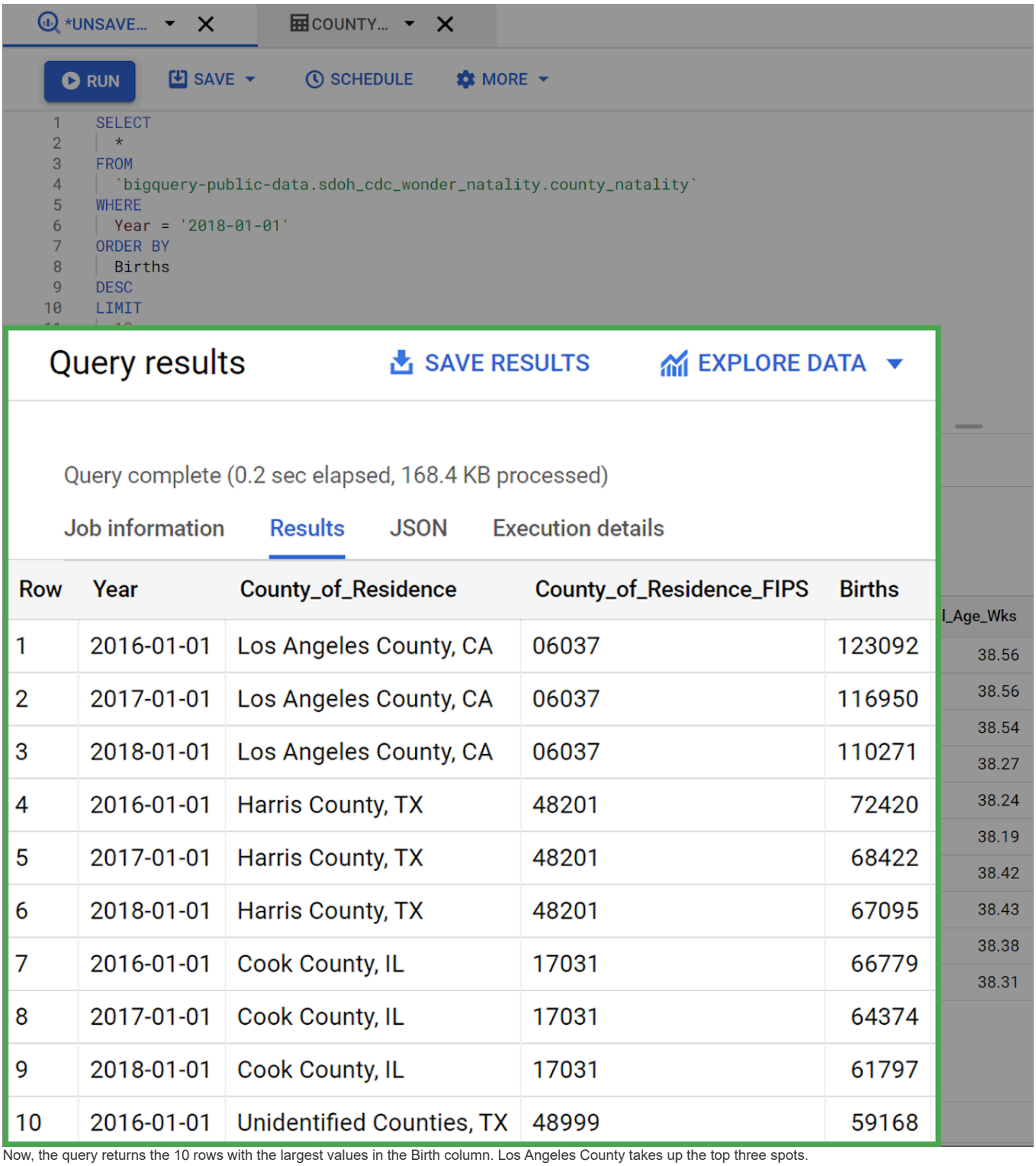
Use DESC to reverse sorting order

Now, modify the query to sort in the other direction, returning the top 10 counties with the highest yearly birth counts between 2016-2018.

Copy, paste, and run the following query:

```
SELECT
  *
FROM
  bigquery-public-data.sdoh_cdc_wonder_natality.county_natality
ORDER BY
  Births
DESC
LIMIT
  10
```

Your results should appear like this:



Now, the query returns the 10 rows with the largest values in the Birth column. Los Angeles County takes up the top three spots.

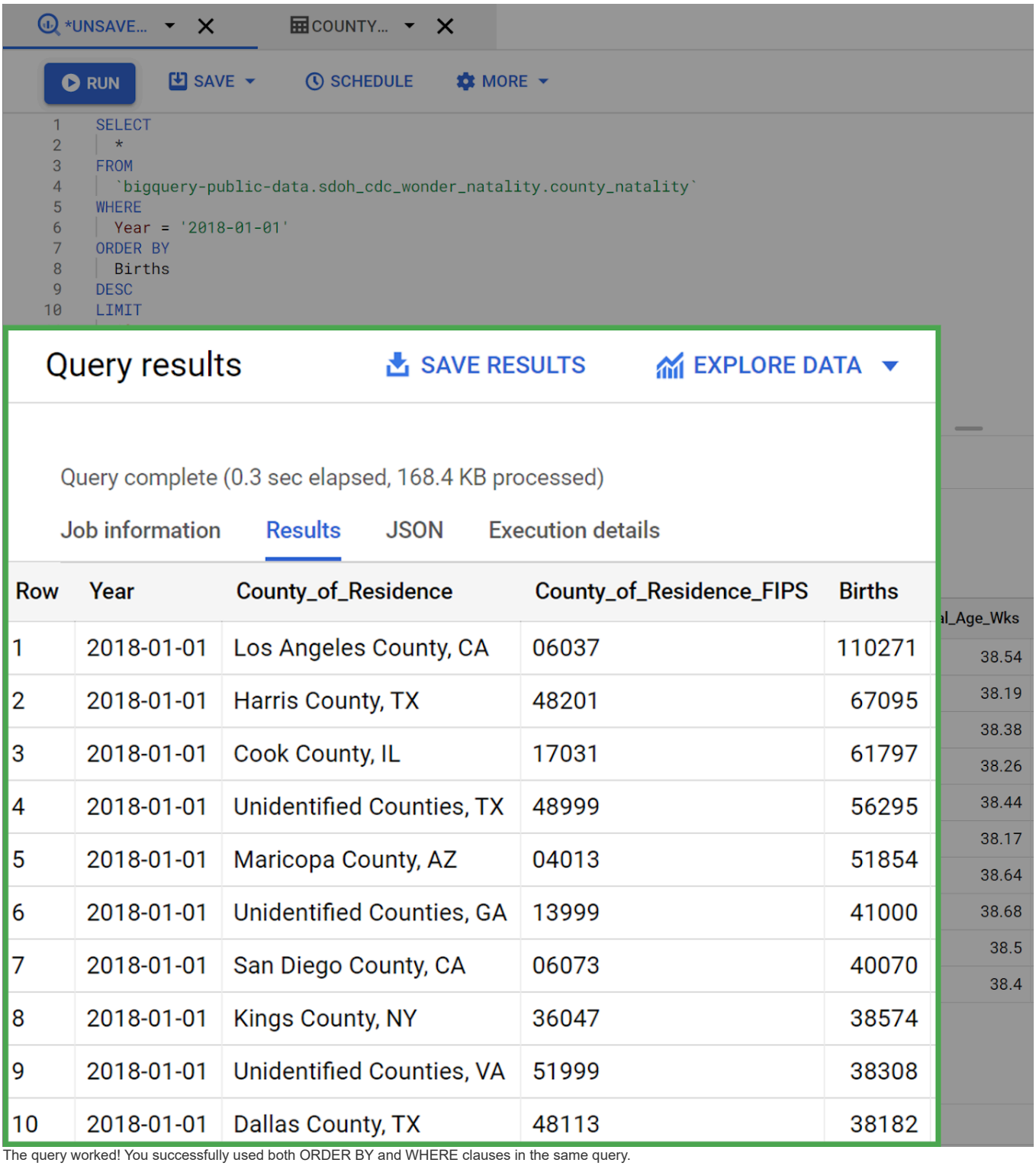
Combine ORDER BY with WHERE clauses

Next, modify the query so that it returns the top 10 counties with the highest birth counts for 2018 only. To do this, add a WHERE clause to the query that specifies only rows that have a Year value equal to 2018-01-01. Note how the ORDER BY clause comes after the WHERE clause.

Copy, paste, and run the following query:

```
SELECT
  *
FROM
  bigquery-public-data.sdoh_cdc_wonder_natality.county_natality
WHERE
  Year = '2018-01-01'
ORDER BY
  Births
DESC
LIMIT
  10
```

Your results should appear something like this:



The query worked! You successfully used both ORDER BY and WHERE clauses in the same query.

Confirmation and reflection

The last query you ran returned the top 10 counties with the highest birth counts for 2018 only. Remove the LIMIT statement and run the query again. What is the county with the 11th highest birth count?

- (●) Orange County, CA
- ( ) Dallas County, TX
- ( ) Unidentified Counties, KY
- ( ) Miami-Dade County, FL

✓ Correct
The county with the 11th highest birth count in 2018 is Orange County, CA. To find this answer, you ran a query with an ORDER BY clause and a WHERE clause. Going forward, you can use this knowledge of SQL to better organize and structure your data.

2. In this activity, you practiced sorting data using SQL queries with ORDER BY and WHERE clauses. In the text box below, write 2-3 sentences (40-60 words) in response to each of the following questions:
   - How can the ORDER BY clause help you organize and structure your data?
   - Why is it helpful to use the ORDER BY and WHERE clauses together when sorting and filtering data?
   - Can you think of a business question that you could answer using this method?

How can the ORDER BY clause help you organize and structure your data?
The ORDER BY clause allows you to sort data by specific columns in ascending or descending order, making it easier to identify patterns, rank entries, and structure information based on numerical or categorical values. This organization is crucial for quick insights, especially when looking for top or bottom values in large datasets.

Why is it helpful to use the ORDER BY and WHERE clauses together when sorting and filtering data?
Using ORDER BY with WHERE enables targeted sorting by first filtering specific criteria, making analysis more focused and efficient. It allows you to retrieve relevant data points within a particular subset, reducing data clutter and making insights easier to interpret by focusing on specific conditions.

Can you think of a business question that you could answer using this method?
A business could ask, "Which products had the highest sales in 2023 within the electronics category?" This can be answered by filtering with WHERE to limit data to 2023 and the electronics category, and using ORDER BY in descending order to list the top-selling products.

Correct
Congratulations on completing this hands-on activity! A good response would include that sorting the data you return in your queries is a crucial tool for analyzing and understanding data.
You can also answer business questions by sorting the dataset according to a given metric. For instance, a store may want to know which products they sell the most or least. Sorting helps you answer business questions that involve phrases such as “how much,” “how many,” “best,” or “worst”—which will be a valuable skill in your career as a data analyst.