## **Hands-On Activity: Calculations in SQL**

**TOTAL POINTS 2**

1.

Question 1



## Activity overview



In previous lessons, you were introduced to SQL. In this activity, you will practice it by learning how to write basic calculations in an SQL query.

By the time you complete this activity, you will be able to use operators in SQL queries. This will enable you to make calculations using SQL and combine multiple arithmetic operations in a single query. This will help you to work faster and more efficiently in SQL as you quickly discover significant patterns in your data—a key skill for data analysts.

## Operations in SQL



### In SQL, the symbols for the four basic arithmetic operations are:

* + addition
* - subtraction
* \* multiplication
* / division

For this activity, imagine you’ve been asked to analyze subway ridership data to help improve the quality of the city’s public transportation. You can use basic calculations in SQL to help you make sense of this data.

## Use a calculation with a single operator



Log in to [BigQuery Sandbox](https://cloud.google.com/bigquery/docs/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.

In this activity, you’ll use data that describes the average weekly subway ridership in New York City from 2013-2018. The columns include: subway stations, subway routes, weekly ridership for each year (2013-2018), the change in ridership from 2017 to 2018 (both in raw numbers and percent), and the ridership rank of each station in 2018. Each row corresponds to a different station in the subway system.

The change\_2018\_raw column describes the change in weekly ridership from 2017 to 2018 in raw numbers. Suppose you want to find data on the change in weekly ridership from 2013 to 2014. You can use SQL to subtract the number of riders in 2013 from the number of riders in 2014.

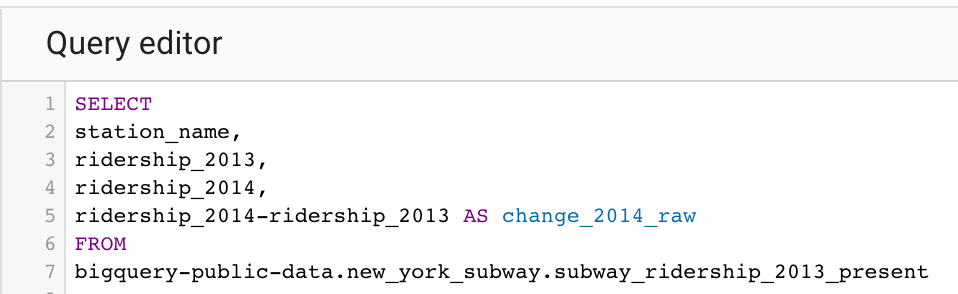
To do that, follow these steps:

1. In the Query editor, type SELECT. Add the names of the columns you want to use in your calculations. You’re selecting several columns, so press Enter (Windows) or Return (Mac) after SELECT. Add a comma after each column name. Include the following columns: station\_name, ridership\_2013, and ridership\_2014.

2. Add the calculation to the query. Enter the names of the two columns with a minus sign between them: ridership\_2014 - ridership\_2013.

3. List the result in a new column. To do this, type AS followed by the new column’s name. Name it change\_2014\_raw since it represents the change in ridership from 2013 to 2014 in raw numbers.

4. End your query with the FROM command and the name of the dataset and subset you’re pulling data from. After FROM, press Enter or Return and type bigquery-public-data.new\_york\_subway.subway\_ridership\_2013\_present.



1. SELECT

2. station\_name,

3. ridership\_2013,

4. ridership\_2014,

5. ridership\_2014-ridership\_2013 AS change\_2014\_raw

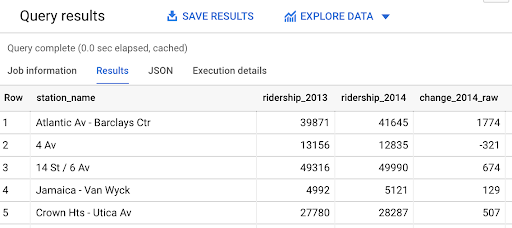
6. FROM

7. bigquery-public-data.new\_york\_subway.subway\_ridership\_2013\_present

5. Now, click Run and get the results.

The results show the change in ridership from 2013 to 2014. For example, the *Atlantic Av - Barclays Ctr* station gained an average of 1,774 riders per week. The *4 Av* station lost 321 riders.

By including a basic calculation in your query, you can get an idea of the change in ridership for each subway station in any given year. Basic calculations help you quickly gain important knowledge about your data.



## Use a calculation with multiple operators



The original dataset describes average weekly ridership for each individual year. Suppose you want to find average weekly ridership for a longer period of time, such as the multi-year period from 2013-2016.

To do this, you can combine multiple arithmetic operations in a query. The average of a set of numbers is the sum of the numbers divided by the total number of values in the set. There are four values in your new set (ridership data for 2013, 2014, 2015, 2016). You can use SQL to sum the numbers for each year and divide that sum by 4.

1. First, click the COMPOSE NEW QUERY button to refresh the query editor.

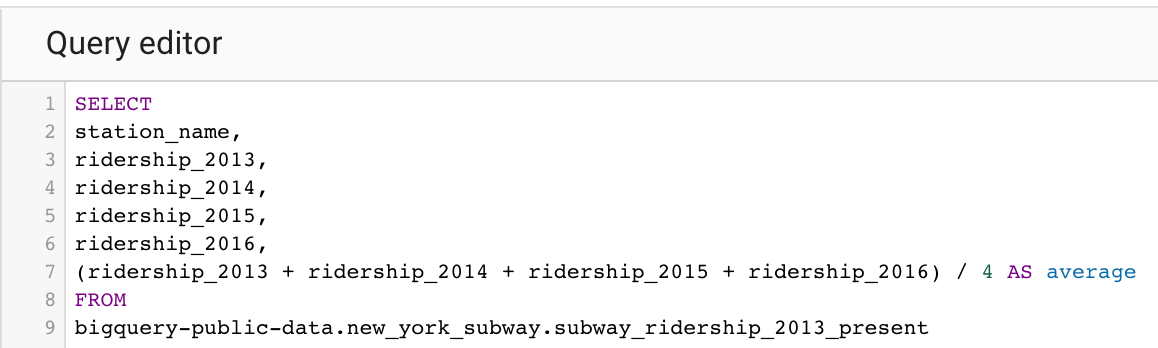
2. Type SELECT to select the columns you want to pull from the table. You’re selecting several columns, so press Enter or Return after SELECT and add a comma after each column name. For this query, enter station\_name, ridership\_2013, ridership\_2014, ridership\_2015, and ridership\_2016.

3. Add the calculation to the query. If you use more than one arithmetic operator in a calculation, you need to use parentheses to control the order of the operations. In this case, you want to sum the years, and then divide the sum by 4. So, put parentheses around the sum of the four column names. Type (ridership\_2013 + ridership\_2014 + ridership\_2015 + ridership\_2016). Then type a division operator / and the number 4.

4. Next, list the result in a new column by typing AS followed by the new column’s name. Call the new column average since it represents average weekly ridership for the period 2013-2016.

5. End your query with the FROM command and the name of the dataset and subset that you’re pulling data from. After FROM, press Enter and type bigquery-public-data.new\_york\_subway.subway\_ridership\_2013\_present.

At this point, your screen should display like this:



1. SELECT

2. station\_name,

3. ridership\_2013,

4. ridership\_2014,

5. ridership\_2015,

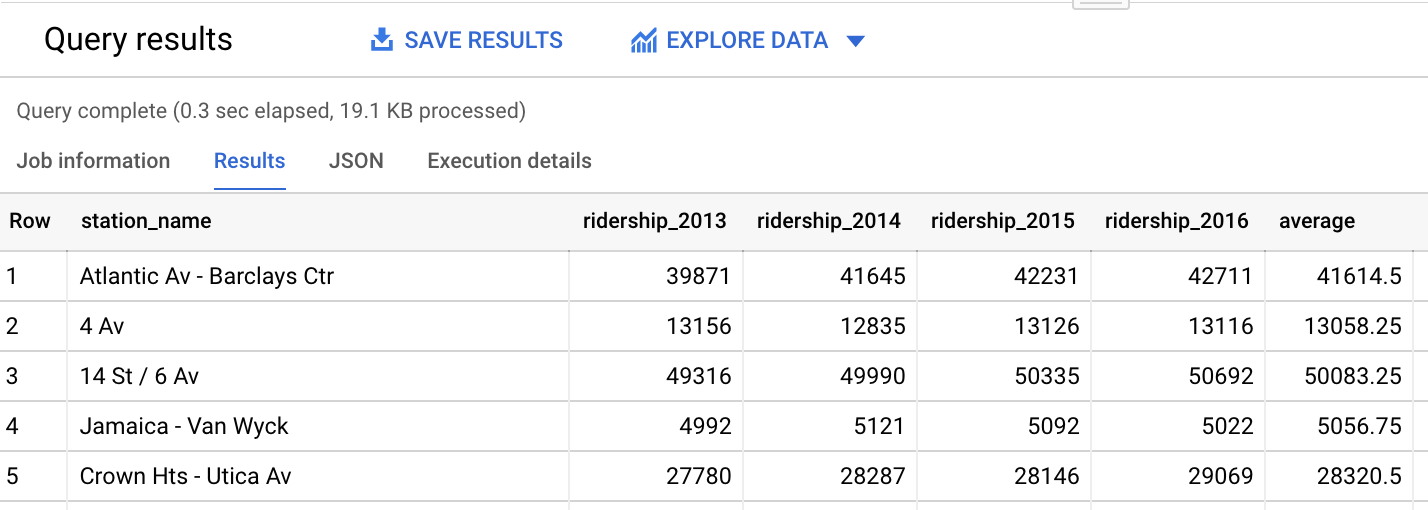
6. ridership\_2016,

7. (ridership\_2013 + ridership\_2014 + ridership\_2015 + ridership\_2016) / 4 AS average

8. FROM

7. bigquery-public-data.new\_york\_subway.subway\_ridership\_2013\_present

6. Click the Run button to get the results. The results clearly show the trend in ridership at each station from 2013 to 2016. For example, weekly ridership at the *Atlantic Av - Barclays Ctr* station increased every year since 2013. Further, for the years 2014, 2015, and 2016, weekly ridership at *Atlantic Av - Barclays Ctr* exceeded the overall average for the period 2013-2016 (listed in the average column).



This kind of data is useful for managing public transportation. It can help you determine which stations or routes to expand due to increased ridership. Using basic calculations in your query allows you to quickly discover significant patterns in your data.

## Confirmation and reflection



Write a query to find the total ridership from the years 2016-2018. What is the total ridership for the Atlantic Av - Barclays Ctr station in this timeframe?

1 point

128,017

125,256

147,765

171,060

2.

Question 2

In this activity, you wrote queries with single and multiple calculations to find patterns in subway ridership data. In the text box below, write 2-3 sentences (40-60 words) in response to each of the following questions:

* What other patterns can you discover in the subway ridership dataset using calculations with multiple operators?
* How can basic calculations in SQL help you make sense of a large amount of data?