← Back Hands-On Activity: Analyze weather data in BigQuery

Practice Quiz • 1h

Congratulations! You passed!

Grade received 100% **To pass** 80% or higher

Go to next item

Hands-On Activity: Analyze weather data in BigQuery

Total points 2

1.

1 / 1 point

Activity overview

Previously, you learned how to use BigQuery to clean data and prepare it for analysis. Now you will query a dataset and save the results into a new table. This is a useful skill when the original data source changes continuously and you need to preserve a specific dataset for continued analysis. It's also valuable when you are dealing with a large dataset and know you'll be doing more than one analysis using the same subset of data.

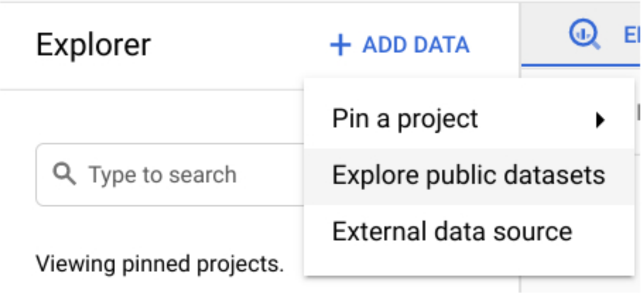
In this scenario, you're a data analyst at a local news station. You have been tasked with answering questions for meteorologists about the weather. You will work with public data from the National Oceanic and Atmospheric Administration (NOAA), which has data for the entire United States. This is why you will need to save a subset of the data in a separate table.

By the time you complete this activity, you will be able to use SQL queries to create new tables when dealing with complex datasets. This will greatly simplify your analysis in the future.

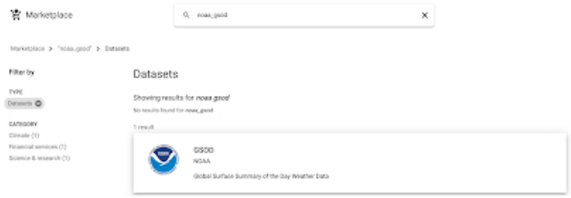
Access the public dataset

For this activity you will need the NOAA weather data from BigQuery's public datasets.

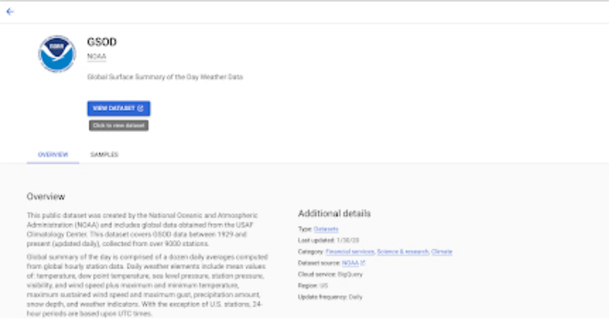
1. Click on the **+ ADD DATA** button in the **Explorer menu pane** and select **Explore public datasets.** This will open a new menu where you can search public datasets that are already available through Google Cloud. If you have already loaded the BigQuery public datasets into your console, you can just search noaa_gsod in your Explorer menu and skip these steps.



2. Type **noaa_gsod** into the search bar. You'll find the GSOD of Global Surface Summary of the Day Weather Data.


Marketplace
noaa_gsod
Marketplace > "noaa_gsod" > Datasets
Filter by
TYPE
Datasets
CATEGORY
Climate (1)
Financial services (1)
Science & research (1)
Datasets
Showing results for noaa gsod
No results found for noaa_gsod
1 result
GSOD
NOAA
Global Surface Summary of the Day Weather Data

3. Click the GSOD dataset to open it. This will provide you with more detailed information about the dataset if you're interested. **Click VIEW DATASET** to open this dataset in your console.



GSOD

NOAA

Global Surface Summary of the Day Weather Data

Overview

This public dataset was created by the National Oceanic and Atmospheric Administration (NOAA) and includes global data obtained from the USAF Climatology Center. This dataset covers GSOD data between 1929 and present (updated daily), collected from over 9000 stations.

Global summary of the day is comprised of a dozen daily averages computed from global hourly station data. Daily weather elements include mean values of: temperature, dew point temperature, sea level pressure, station pressure, visibility, and wind speed plus maximum and minimum temperature, maximum sustained wind speed and maximum gust, precipitation amount, snow depth, and weather indicators. With the exception of U.S. stations, 24-hour periods are based upon UTC times.

Additional details

Type: Datasets

Last updated: 1/30/20

Category: Financial services, Science & research, Climate

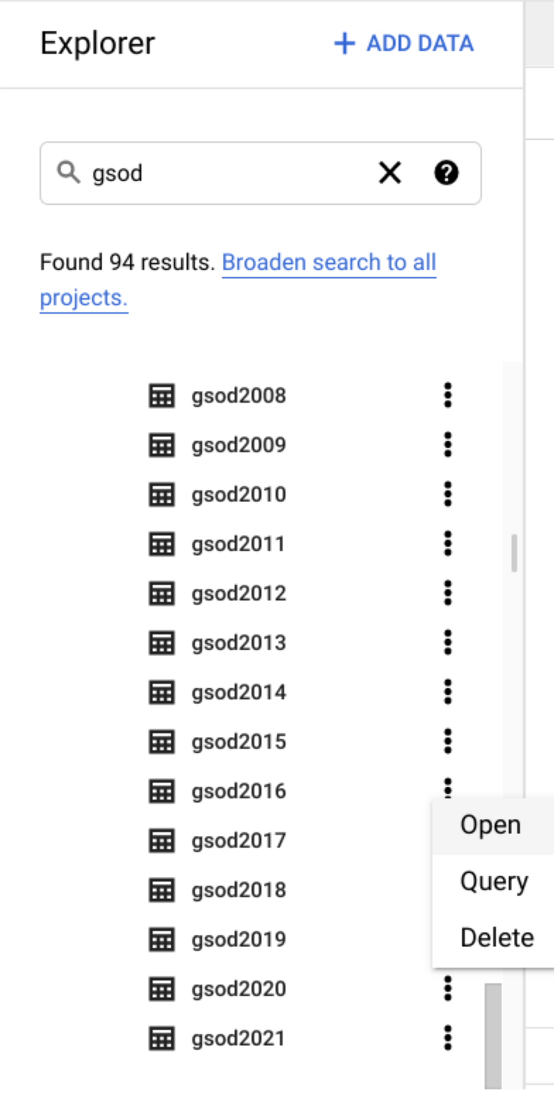
Dataset source: NOAA

Cloud service: BigQuery

Region: US

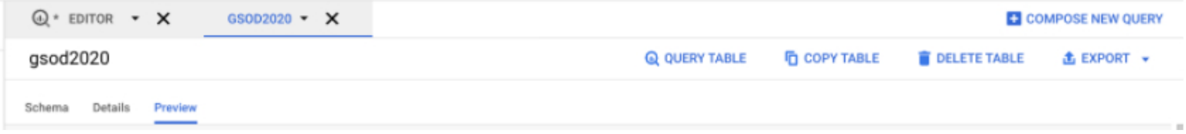
Update frequency: Daily

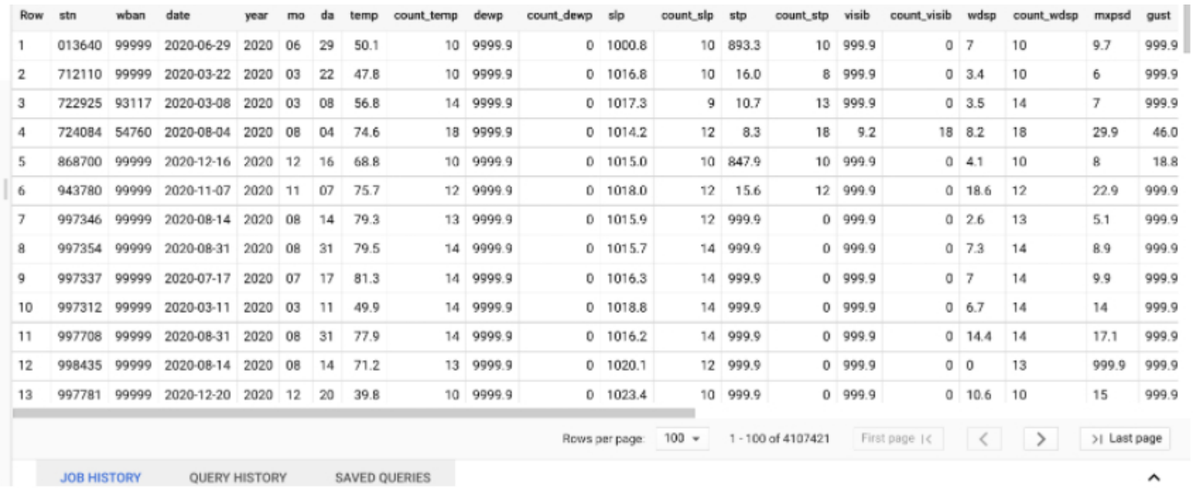
4. Search **noaa_gsod** in your **Explorer menu pane** to find the dataset. **Click the dropdown menu** to explore the tables in this dataset. Scroll down to **gsod2020** and **open the table menu by clicking the three vertical dots.**



Found 94 results. Broaden search to all projects.

5. **Check the table's schema and preview it** to get familiar with the data. Once you're ready, you can **click COMPOSE NEW QUERY** to start querying the dataset.





Row	stn	wban	date	year	mo	da	temp	count_temp	dewp	count_dewp	slp	count_slp	stp	count_stp	visib	count_visib	wdsp	count_wdsp	mxpsd	gust
1	013640	99999	2020-06-29	2020	06	29	50.1	10	9999.9	0	1000.8	10	893.3	10	999.9	0	7	10	9.7	999.9
2	712110	99999	2020-03-22	2020	03	22	47.8	10	9999.9	0	1016.8	10	16.0	8	999.9	0	3.4	10	6	999.9
3	722925	93117	2020-03-08	2020	03	08	56.8	14	9999.9	0	1017.3	9	10.7	13	999.9	0	3.5	14	7	999.9
4	724084	54760	2020-08-04	2020	08	04	74.6	18	9999.9	0	1014.2	12	8.3	18	9.2	18	8.2	18	29.9	46.0
5	868700	99999	2020-12-16	2020	12	16	68.8	10	9999.9	0	1015.0	10	847.9	10	999.9	0	4.1	10	8	18.8
6	943780	99999	2020-11-07	2020	11	07	75.7	12	9999.9	0	1018.0	12	15.6	12	999.9	0	18.6	12	22.9	999.9
7	997346	99999	2020-08-14	2020	08	14	79.3	13	9999.9	0	1015.9	12	999.9	0	999.9	0	2.6	13	5.1	999.9
8	997354	99999	2020-08-31	2020	08	31	79.5	14	9999.9	0	1015.7	14	999.9	0	999.9	0	7.3	14	8.9	999.9
9	997337	99999	2020-07-17	2020	07	17	81.3	14	9999.9	0	1016.3	14	999.9	0	999.9	0	7	14	9.9	999.9
10	997312	99999	2020-03-11	2020	03	11	49.9	14	9999.9	0	1018.8	14	999.9	0	999.9	0	6.7	14	14	999.9
11	997708	99999	2020-08-31	2020	08	31	77.9	14	9999.9	0	1016.2	14	999.9	0	999.9	0	14.4	14	17.1	999.9
12	998435	99999	2020-08-14	2020	08	14	71.2	13	9999.9	0	1020.1	12	999.9	0	999.9	0	0	13	999.9	999.9
13	997781	99999	2020-12-20	2020	12	20	39.8	10	9999.9	0	1023.4	10	999.9	0	999.9	0	10.6	10	15	999.9

Rows per page: 100 ▾ 1 - 100 of 4107421

Querying the data

The meteorologists who you're working with have asked you to get the temperature, wind speed, and precipitation for stations La Guardia and JFK, for every day in 2020, in descending order by date, and ascending order by Station ID. Use the following query to request this information:

```
SELECT
  stn,

  date,
  -- Use the IF function to replace 9999.9 values, which the dataset description explains is the default
value when temperature is missing, with NULLs instead.
       IF(

        temp=9999.9,
        NULL,
        temp) AS temperature,

  -- Use the IF function to replace 999.9 values, which the dataset description explains is the default
value when wind speed is missing, with NULLs instead.
       IF(

        wdsp="999.9",
        NULL,
        CAST(wdsp AS Float64)) AS wind_speed,

-- Use the IF function to replace 99.99 values, which the dataset description explains is the default
value when precipitation is missing, with NULLs instead.

    IF(

       prcp=99.99,
       0,
       prcp) AS precipitation
FROM
  `bigquery-public-data.noaa_gsod.gsod2020`
WHERE
  stn="725030" -- La Guardia

  OR stn="744860" -- JFK
ORDER BY
  date DESC,
  stn ASC
```

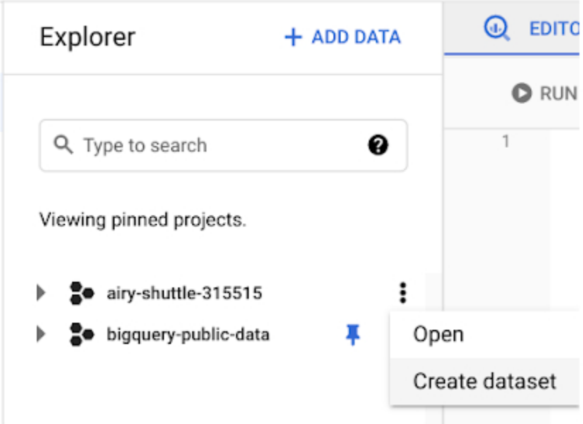
The meteorologists also asked you a couple questions while they were preparing for the nightly news: They want the average temperature in June 2020 and the average wind_speed in December 2020.

Instead of rewriting similar, but slightly different, queries over and over again, there is an easier approach: Save the results from the original query as a table for future queries.

Save a new table

In order to make this subset of data easier to query from, you can save the table from the weather data into a new dataset.

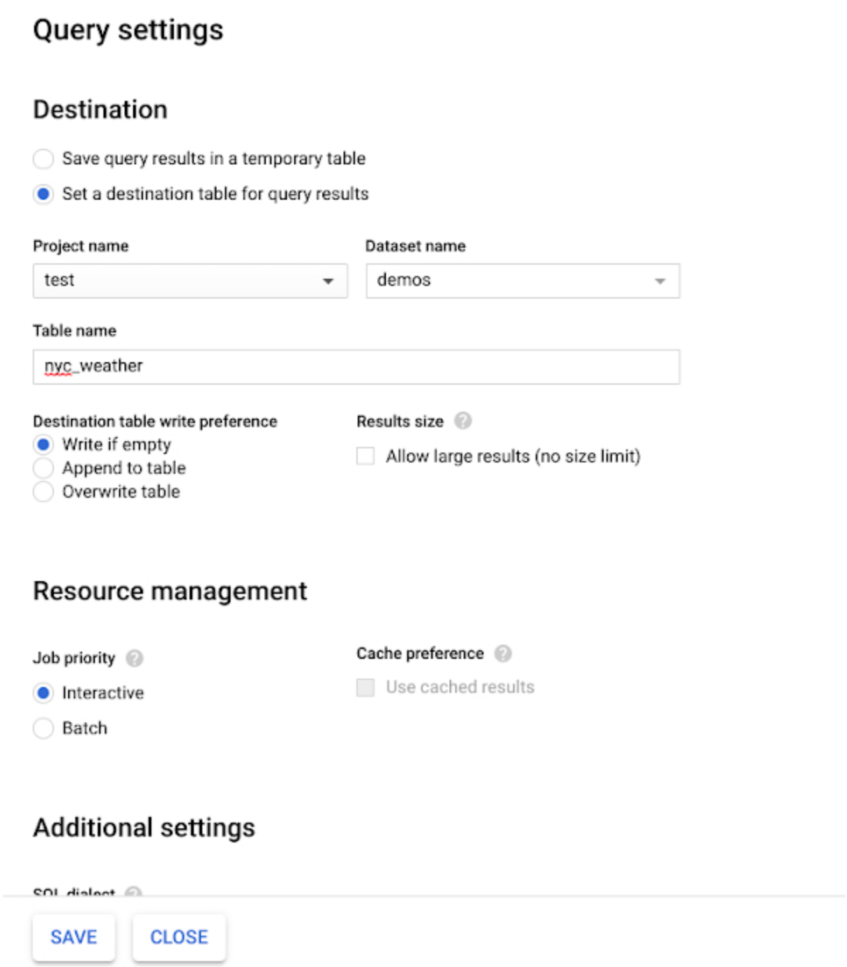
1. From your **Explorer** pane, **click the three vertical dots** next to your project and **select Create dataset.** You can name this dataset demos and leave the rest of the default options. **Click CREATE DATASET.**



2. Open your new dataset and select **COMPOSE NEW QUERY**. Input the following query to get the average temperature, wind speed, visibility, wind gust, precipitation, and snow depth La Guardia and JFK stations for every day in 2020, in descending order by date, and ascending order by Station ID:

```
SELECT
 stn,

 date,
 -- Use the IF function to replace 9999.9 values, which the dataset description explains is the default value when temperature is missing, with NULLs instead.
    IF(

     temp=9999.9,
     NULL,
     temp) AS temperature,

 -- Use the IF function to replace 999.9 values, which the dataset description explains is the default value when wind speed is missing, with NULLs instead.
       IF(

        wdsp="999.9",
        NULL,
        CAST(wdsp AS Float64)) AS wind_speed,

-- Use the IF function to replace 99.99 values, which the dataset description explains is the default value when precipitation is missing, with NULLs instead.

    IF(

       prcp=99.99,
       0,
       prcp) AS precipitation
FROM
  `bigquery-public-data.noaa_gsod.gsod2020`
WHERE
  stn="725030" -- La Guardia
  OR stn="744860" -- JFK
ORDER BY
  date DESC,
  stn ASC
```

3. Before you run the query, s**elect the MORE menu** from the Query Editor and **open the Query Settings menu**. In the Query Settings menu, select **Set a destination table for query results**. Set the dataset option to **demos** and name the table **nyc_weather**.



Query settings

Destination

○ Save query results in a temporary table

● Set a destination table for query results

Project name: test

Dataset name: demos

Table name: nyc_weather

Destination table write preference

● Write if empty

○ Append to table

○ Overwrite table

Results size

□ Allow large results (no size limit)

Resource management

Job priority

● Interactive

○ Batch

Cache preference

□ Use cached results

Additional settings

SQL dialect

4. Run the query from earlier; now it will save as a new table in your demos dataset.

5. Return to the **Query settings menu** by using the **MORE dropdown menu**. Reset the settings to **Save query results in a temporary table**. This will prevent you from accidentally adding every query as a table to your new dataset.

Query your new table

Now that you have the subset of this data saved in a new table, you can query it more easily. Use the following query to find the average temperature from the meteorologists first question:

```
SELECT

AVG(temperature)

FROM

`airy-shuttle-315515.demos.nyc_weather` --remember to change the project name to your project before
running this query

WHERE
```

```
date BETWEEN '2020-06-01' AND '2020-06-30'
```

You can also use this syntax to find the average wind_speed or any other information from this subset of data you're interested in. Try constructing a few more queries to answer the meteorologists' questions!

The ability to save your results into a new table is a helpful trick when you know you're only interested in a subset of a larger complex dataset that you plan on querying multiple times, such as the weather data for just La Guardia and JFK. This also helps minimize errors during your analysis.

Confirmation and reflection

What was the average temperature at JFK and La Guardia stations between June 1, 2020 and June 30, 2020?

○ 87.671

● 72.883

○ 74.909

○ 92.099

✓ **Correct**
The average was 72.883. To find out the average temperature during this time period, you successfully created a new table using a query and ran another query against that table. Going forward, you will be able to use this skill to create tables with specific subsets of your data to query. This will help you draw insights from multiple data sources in the future.

2. In the text box below, write 2-3 sentences (40-60 words) in response to each of the following questions:

1 / 1 point

- How can creating tables from queries help you perform data analysis in the future?
- Why is being able to view specific subsets of a dataset important?

How can creating tables from queries help you perform data analysis in the future?
It makes it easier to save the new tables with the information.
Why is being able to view specific subsets of a dataset important?
Because it is much harder to analyze the whole set of data.

✓ **Correct**
Congratulations on completing this hands-on activity! In this activity, you explored two public datasets and created a new table using a query. A good response might include that creating tables using your queries allows you to work with a subset of data without changing the original.

For instance, now you can query weather data from just the weather stations in New York that you need. This is important for finding trends within a subset of data. In upcoming activities, you will continue analyzing data like this.