Congratulations! You passed!
Grade received 100%To pass 80% or higher

1. 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.

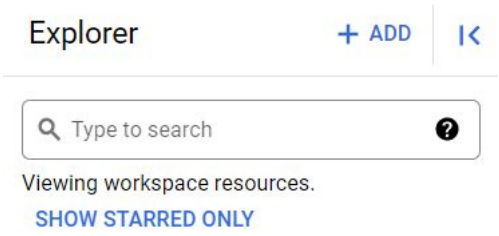
In this scenario, you're a data analyst at a local news station. You have been tasked with answering questions for meteorologists about the weather. You will work with public data from the National Oceanic and Atmospheric Administration (NOAA), which has data for the entire United States. This is why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

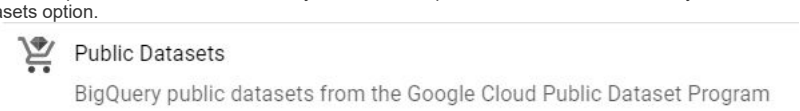
Access the public dataset

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

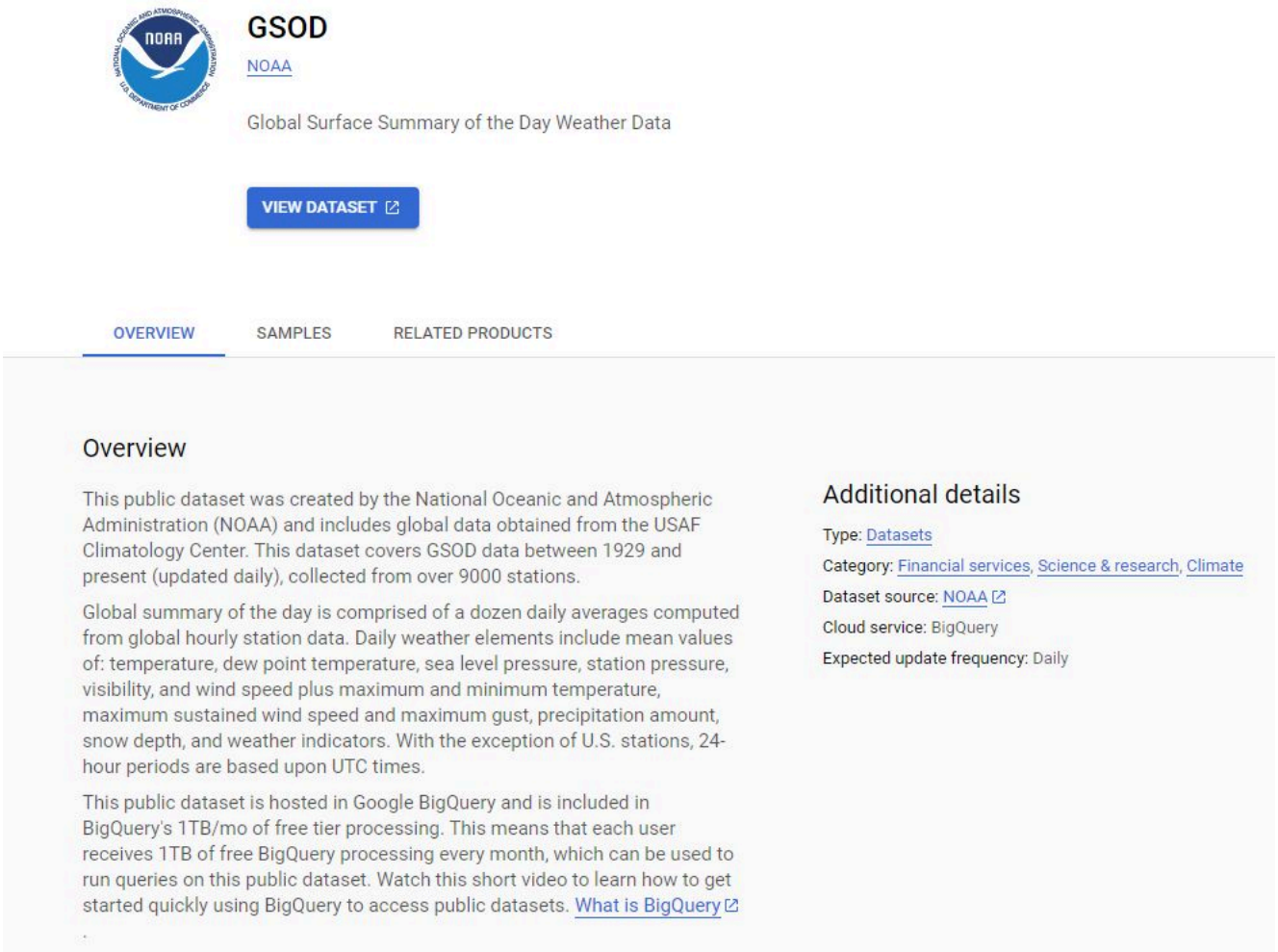
1. Click on the + ADD button in the Explorer menu pane and navigate down the list in the Add window and select Public Datasets.



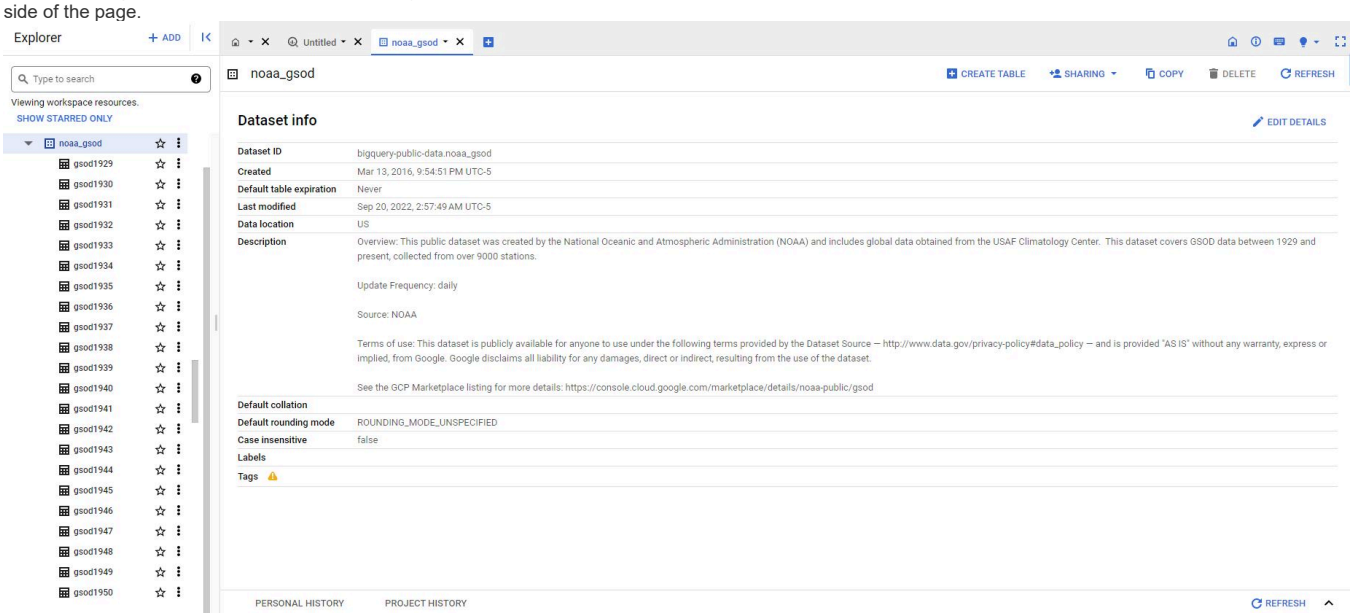
2. This will open a new *Add menu* where you can search public datasets that are already available through Google Cloud. Scroll down the list and select the Public Datasets option.



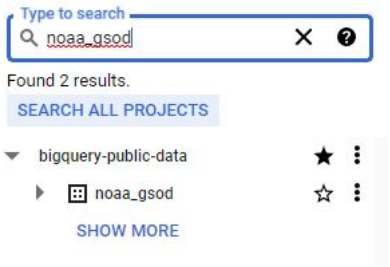
3. This will open up the Marketplace menu. In the search bar at the top of the page, typing the acronym gsod and pressing enter will call up the NOAA global data. You'll find the GSOD of Global Surface Summary of the Day Weather Data. Click anywhere on the graphical search result, and you will receive a detailed description of the dataset.



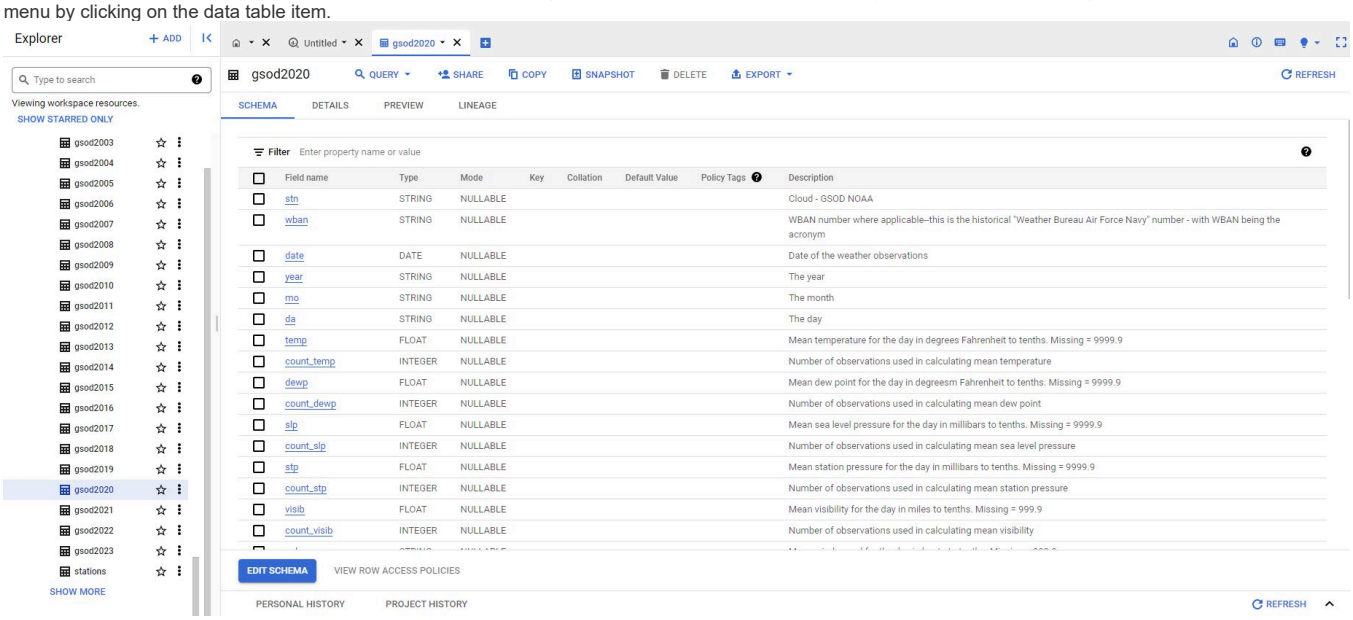
4. After clicking the blue VIEW DATASET button, you then be returned to the main workspace and will see the public dataset populate the Explorer pane on the left side of the page.



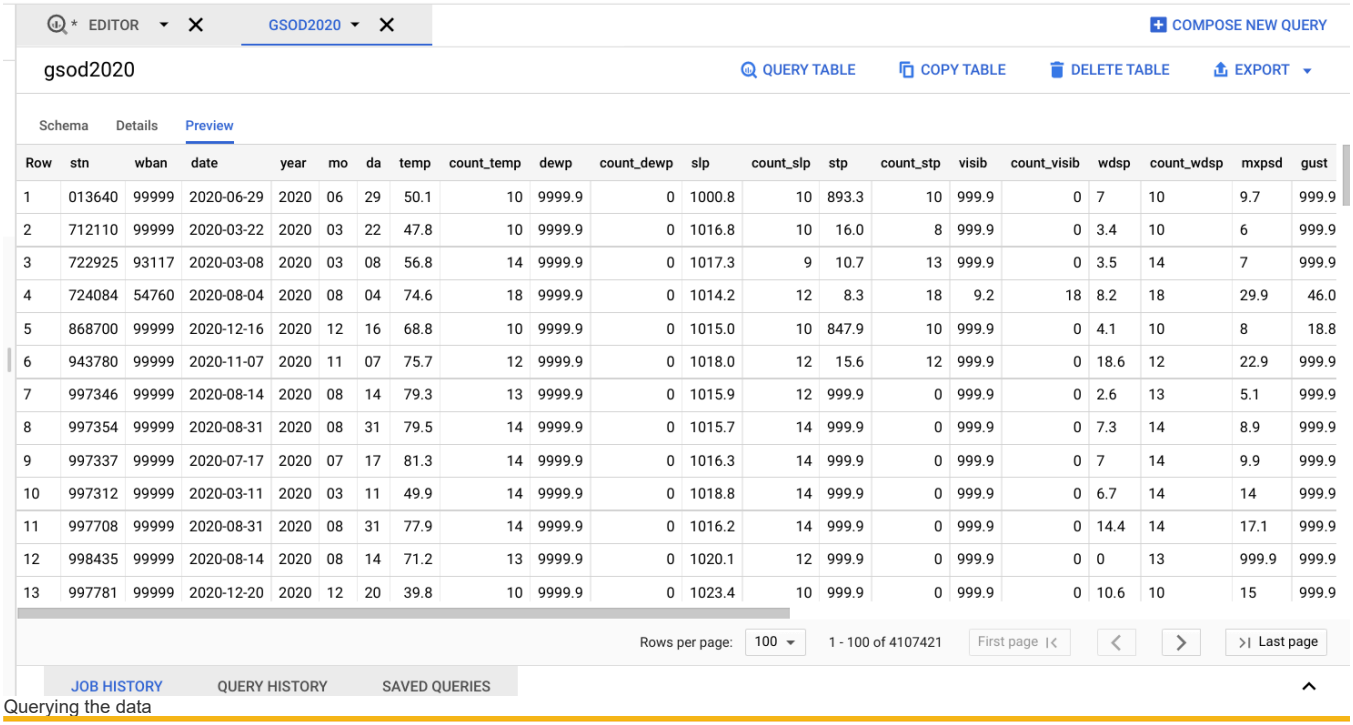
5. If the BigQuery public dataset has already been starred in your Explorer pane, you may also simply type the word noaa_gsod in the search bar to obtain the same results.



6. Expand by clicking the drop-down arrow to the left of the noaa_gsod dataset item and expand the yearly subsets. Scroll down to gsod2020 and open the table menu by clicking on the data table item.



7. 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

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. Click on the blue QUERY button in the row of tab functions, select the In split tab option in the drop-down menu, and type the following query in the new workspace to request this information:

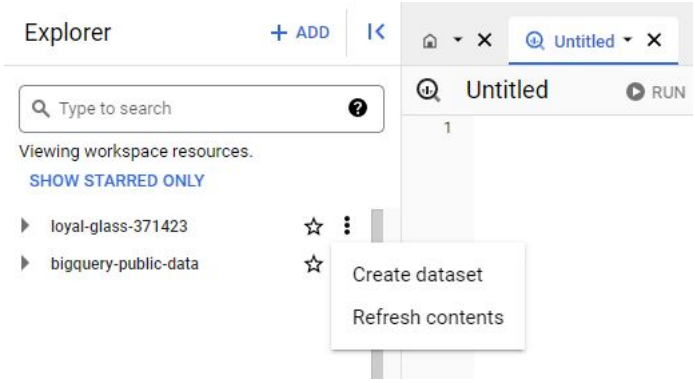
```
SELECT
  stn,
  date,
-- Use the IF function to replace 9999.9 values, which the dataset description explains is the default value when temperat
  IF(
    temp=9999.9,
    NULL,
    temp) AS temperature,
-- Use the IF function to replace 999.9 values, which the dataset description explains is the default value when wind spee
  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 precipita
  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.
Note: Depending on the type of web browser you are using, you may have issues conducting a direct *copy + paste* of the above syntax directly from this assignment to your BigQuery workspace. If you are experiencing this issue, it is recommended to create two separate windows on your screen to manually type the above information into the BigQuery workspace area.

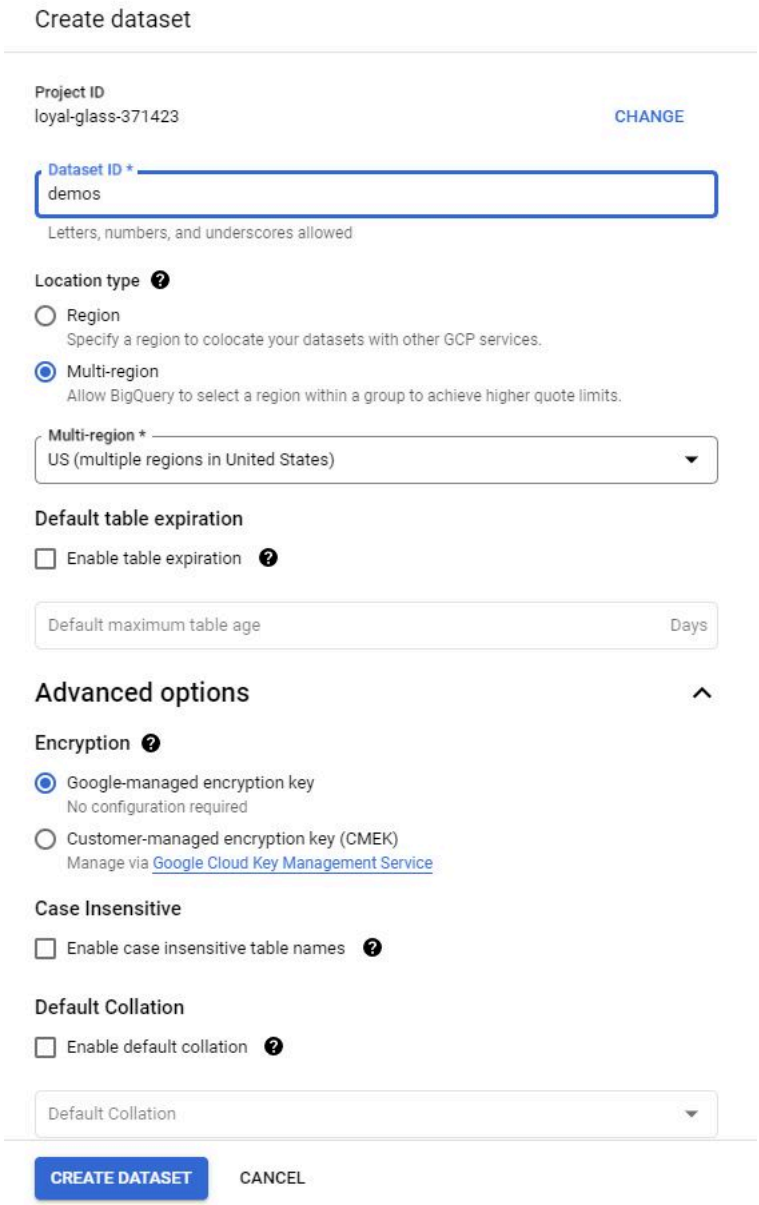
Save a new table

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.



2. Type the name demos into the Data ID box, set the *Location type* to Multi-region(US), and leave the rest of the Advanced default options. Once you have completed this task, click the blue button CREATE DATASET at the bottom of the page.

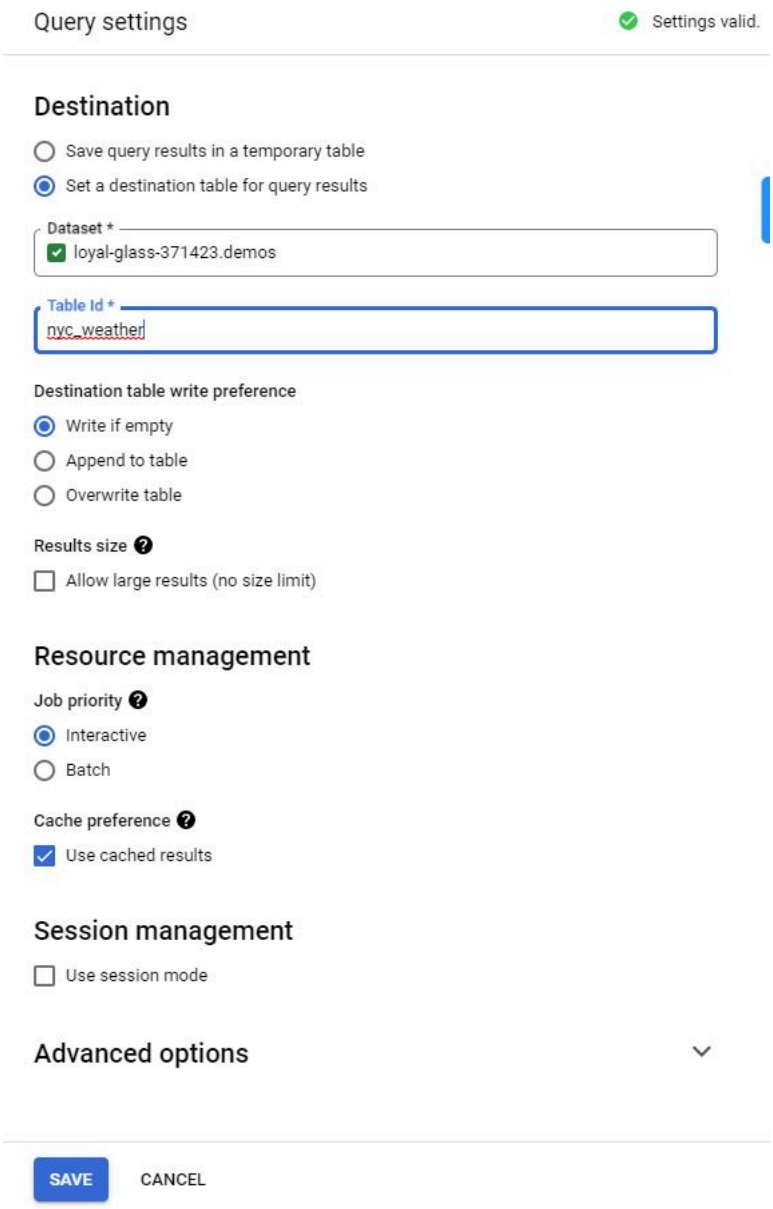


3. Open your new dataset and click on the blue + button to the right side of the demos tab. Input the following query to get the average temperature, wind speed, visibility, wind gust, precipitation, and snow depth La Guardia and JFK stations for every day in 2020, in descending order by date, and ascending order by Station ID:

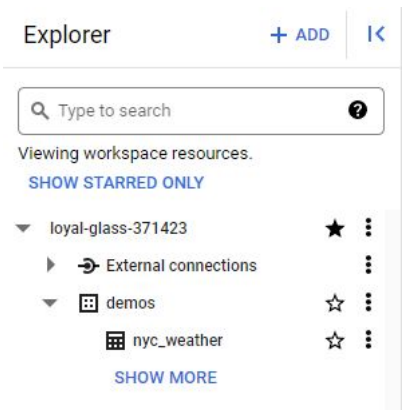
```
SELECT
  stn,
  date,
-- Use the IF function to replace 9999.9 values, which the dataset description explains is the default value when temperat
    IF(
      temp=9999.9,
      NULL,
      temp) AS temperature,
-- Use the IF function to replace 999.9 values, which the dataset description explains is the default value when wind spee
    IF(
      wdsp="999.9",
      NULL,
      CAST(wdsp AS Float64)) AS wind_speed,
```

```
14  -- Use the IF function to replace 99.99 values, which the dataset description explains is the default value when precipita
15      IF(
16         prcp=99.99,
17         0,
18         prcp) AS precipitation
19  FROM
20    `bigquery-public-data.noaa_gsod.gsod2020`
21  WHERE
22    stn="725030" -- La Guardia
23    OR stn="744860" -- JFK
24  ORDER BY
25    date DESC,
26    stn ASC
```

3. Before you run the query, select 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.



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
    `your project name.demos.nyc_weather`

--remember to format the beginning syntax to your project name before running this query. You can view the full Table ID by

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?

- ○ 92.099
- ○ 87.671
- ◉ 72.883
- ○ 74.909

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:

- 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?
Creating tables from queries allows you to preserve specific subsets of data, making future analyses faster and more efficient, especially when working with large, frequently updated datasets. This reduces the need to repeatedly query the original data and ensures consistency for ongoing analyses.

Why is being able to view specific subsets of a dataset important?
Viewing specific subsets of a dataset lets analysts focus on relevant data, reducing noise and enhancing clarity. It is essential for answering targeted questions, optimizing processing resources, and minimizing errors in complex datasets where only certain parts of the data are relevant to the analysis.

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.