**Modernizing Data Lakes and Data Warehouses with Google Cloud**

Course · 8 hours

32% complete

[Professional Data Engineer Certification Learning Path](https://partner.cloudskillsboost.google/paths/85) navigate\_next [Modernizing Data Lakes and Data Warehouses with Google Cloud](https://partner.cloudskillsboost.google/paths/85/course_templates/54) navigate\_next Introduction to Data Engineering

**Using BigQuery to do Analysis**

45 minutes No cost

**Overview**

In this lab you analyze 2 different public datasets, run queries on them, separately and then combined, to derive interesting insights.

**What you'll learn**

In this lab, you will:

* Carry out interactive queries on the BigQuery console.
* Combine and run analytics on multiple datasets.

**Prerequisites**

This is a **fundamental level** lab and assumes some experience with BigQuery and SQL.

**Introduction**

This lab uses two public datasets in BigQuery: weather data from the US National Oceanic and Atmospheric Administration (NOAA), and bicycle rental data from New York City.

You will encounter, for the first time, several aspects of Google Cloud Platform that are of great benefit to scientists:

1. **Serverless.** No need to download data to your machine in order to work with it - the dataset will remain on the cloud.
2. **Ease of use.** Run ad-hoc SQL queries on your dataset without having to prepare the data, like indexes, beforehand. This is invaluable for data exploration.
3. **Scale.** Carry out data exploration on extremely large datasets interactively. You don't need to sample the data in order to work with it in a timely manner.
4. **Shareability**. You will be able to run queries on data from different datasets without any issues. BigQuery is a convenient way to share datasets. Of course, you can also keep your data private, or share them only with specific persons -- not all data need to be public.

The end-result is that you will find if there are lesser bike rentals on rainy days.

**Setup and requirements**

**Lab setup**

For each lab, you get a new Google Cloud project and set of resources for a fixed time at no cost.

1. Sign in to Qwiklabs using an **incognito window**.
2. Note the lab's access time (for example, 1:15:00), and make sure you can finish within that time.  
   There is no pause feature. You can restart if needed, but you have to start at the beginning.
3. When ready, click **Start lab**.
4. Note your lab credentials (**Username** and **Password**). You will use them to sign in to the Google Cloud Console.
5. Click **Open Google Console**.
6. Click **Use another account** and copy/paste credentials for **this** lab into the prompts.  
   If you use other credentials, you'll receive errors or **incur charges**.
7. Accept the terms and skip the recovery resource page.

**Note:** Do not click **End Lab** unless you have finished the lab or want to restart it. This clears your work and removes the project.

**Open BigQuery Console**

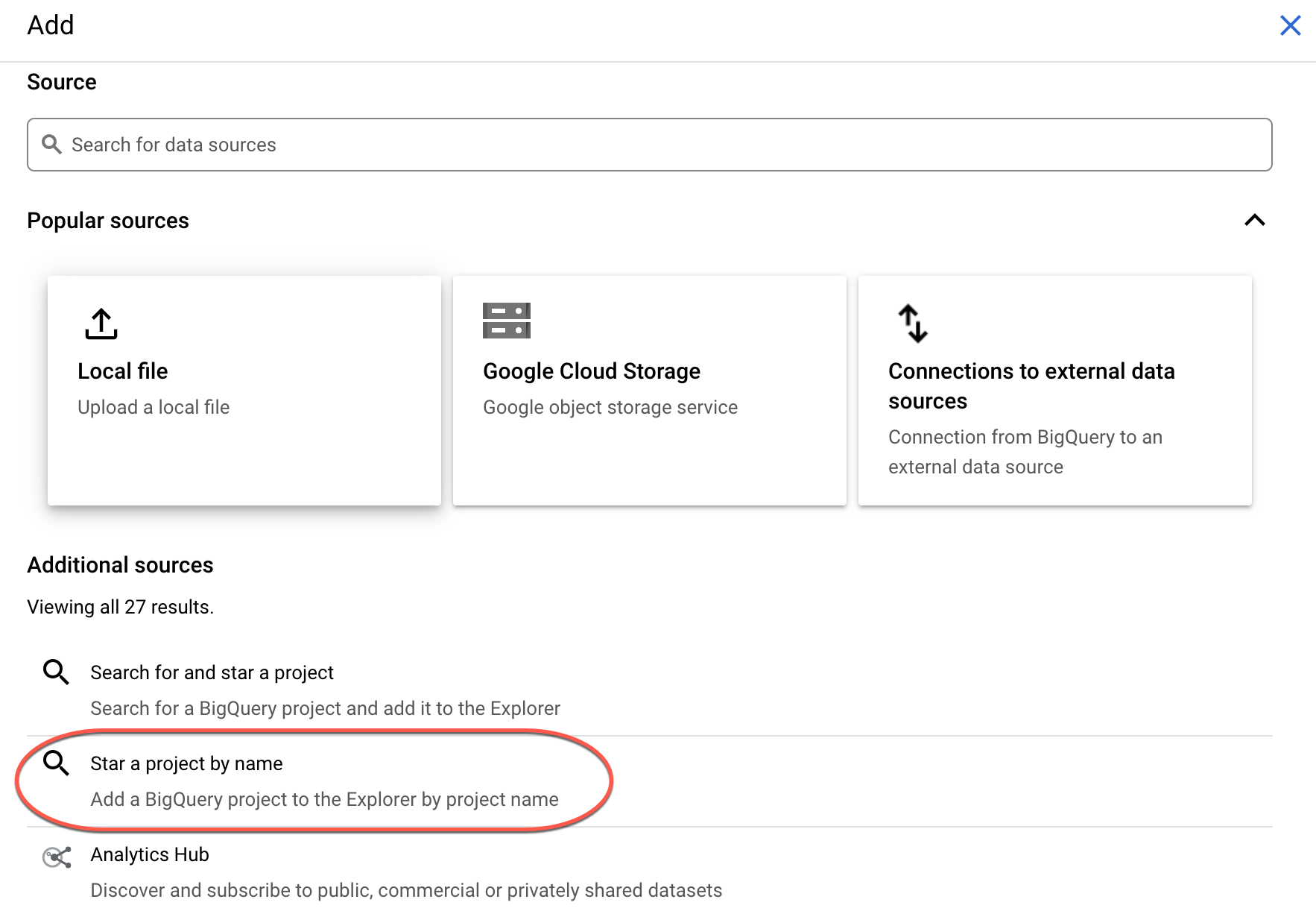
1. In the Google Cloud Console, select **Navigation menu** > **BigQuery**.

The **Welcome to BigQuery in the Cloud Console** message box opens. This message box provides a link to the quickstart guide and lists UI updates.

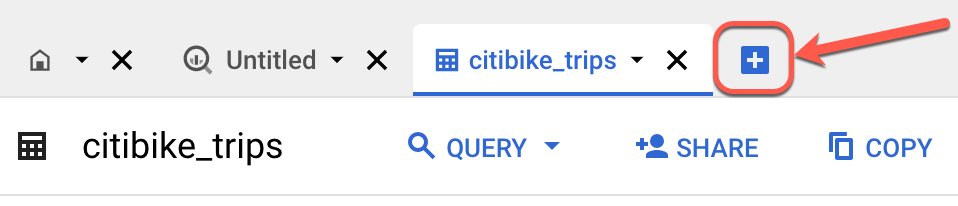
1. Click **Done**.

**Task 1. Explore bicycle rental data**

1. In the left pane, click **+ Add** , then click **Star a project by name**, next in the pop-up window type **bigquery-public-data**, finally click **Star**.



1. In the BigQuery console, you see two projects in the left pane, one named your Qwiklabs project ID, and one named **bigquery-public-data**.
2. In the left pane of the BigQuery console, select **bigquery-public-data** > **new\_york\_citibike** > **citibike\_trips** table.
3. In the Table (citibike\_trips) window, click the **Schema** tab.
4. Examine the column names and the datatypes.
5. Click the **Blue +** button to compose a new query.



Enter the following query:

SELECT MIN(start\_station\_name) AS start\_station\_name, MIN(end\_station\_name) AS end\_station\_name, APPROX\_QUANTILES(tripduration, 10)[OFFSET (5)] AS typical\_duration, COUNT(tripduration) AS num\_trips FROM `bigquery-public-data.new\_york\_citibike.citibike\_trips` WHERE start\_station\_id != end\_station\_id GROUP BY start\_station\_id, end\_station\_id ORDER BY num\_trips DESC LIMIT 10

1. Click **Run**. Look at the result and try to determine what this query does ?

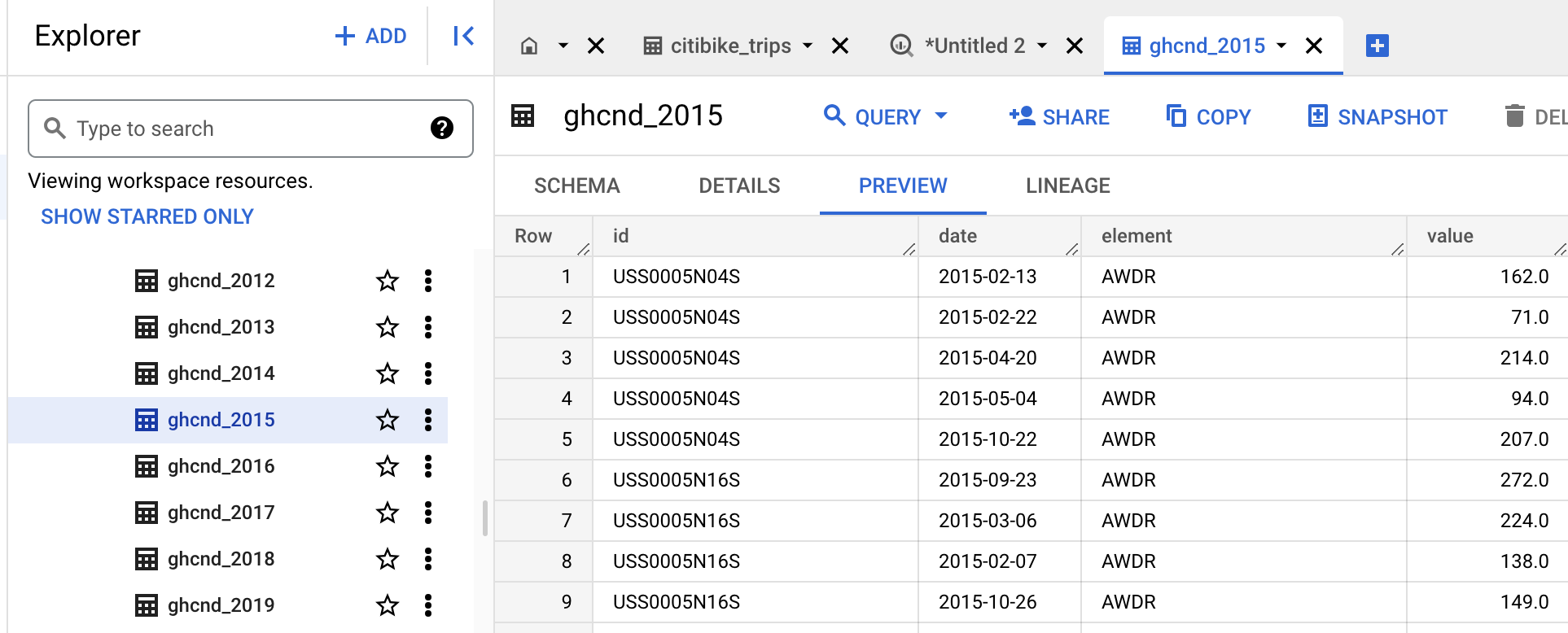
**Hint:** typical duration for the 10 most common one-way rentals)

1. Next, run the query below to find another interesting fact: total distance traveled by each bicycle in the dataset. Note that the query limits the results to only top 5.

WITH trip\_distance AS ( SELECT bikeid, ST\_Distance(ST\_GeogPoint(s.longitude, s.latitude), ST\_GeogPoint(e.longitude, e.latitude)) AS distance FROM `bigquery-public-data.new\_york\_citibike.citibike\_trips`, `bigquery-public-data.new\_york\_citibike.citibike\_stations` as s, `bigquery-public-data.new\_york\_citibike.citibike\_stations` as e WHERE start\_station\_name = s.name AND end\_station\_name = e.name) SELECT bikeid, SUM(distance)/1000 AS total\_distance FROM trip\_distance GROUP BY bikeid ORDER BY total\_distance DESC LIMIT 5 **Note:** For this query, we also used the other table in the dataset called **citibike\_stations** to get bicycle station information.

**Task 2. Explore the weather dataset**

1. In the left pane of the BigQuery Console, select the newly added bigquery-public-data project and select **ghcn\_d** > **ghcnd\_2015**.
2. Then click on the **Preview** tab. Your console should resemble the following:



Examine the columns and some of the data values.

1. Click the **Blue +** button to compose a new query and enter the following:

SELECT wx.date, wx.value/10.0 AS prcp FROM `bigquery-public-data.ghcn\_d.ghcnd\_2015` AS wx WHERE id = 'USW00094728' AND qflag IS NULL AND element = 'PRCP' ORDER BY wx.date

1. Click **Run**.

This query will return rainfall (in mm) for all days in 2015 from a weather station in New York whose id is provided in the query (the station corresponds to NEW YORK CNTRL PK TWR ).

**Task 3. Find correlation between rain and bicycle rentals**

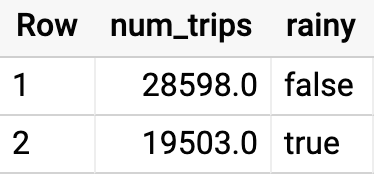
How about joining the bicycle rentals data against weather data to learn whether there are fewer bicycle rentals on rainy days?

1. Click the **Blue +** button to compose a new query and enter the following:

WITH bicycle\_rentals AS ( SELECT COUNT(starttime) as num\_trips, EXTRACT(DATE from starttime) as trip\_date FROM `bigquery-public-data.new\_york\_citibike.citibike\_trips` GROUP BY trip\_date ), rainy\_days AS ( SELECT date, (MAX(prcp) > 5) AS rainy FROM ( SELECT wx.date AS date, IF (wx.element = 'PRCP', wx.value/10, NULL) AS prcp FROM `bigquery-public-data.ghcn\_d.ghcnd\_2015` AS wx WHERE wx.id = 'USW00094728' ) GROUP BY date ) SELECT ROUND(AVG(bk.num\_trips)) AS num\_trips, wx.rainy FROM bicycle\_rentals AS bk JOIN rainy\_days AS wx ON wx.date = bk.trip\_date GROUP BY wx.rainy

1. Click **Run**.

Now you can see the results of joining the bicycle rental dataset with a weather dataset that comes from a completely different source:



Running the query yields that, yes, New Yorkers ride the bicycle 47% fewer times when it rains.

**Summary**

In this lab you did ad-hoc queries on two datasets. You were able to query the data without setting up any clusters, creating any indexes, etc. You were also able to mash up the two datasets and get some interesting insights. All without ever leaving your browser!

**Congratulations!**

You learned how to run some very interesting queries on BigQuery!

**End your lab**

When you have completed your lab, click **End Lab**. Google Cloud Skills Boost removes the resources you’ve used and cleans the account for you.

You will be given an opportunity to rate the lab experience. Select the applicable number of stars, type a comment, and then click **Submit**.

The number of stars indicates the following:

* 1 star = Very dissatisfied
* 2 stars = Dissatisfied
* 3 stars = Neutral
* 4 stars = Satisfied
* 5 stars = Very satisfied

You can close the dialog box if you don't want to provide feedback.

For feedback, suggestions, or corrections, please use the **Support** tab.

Copyright 2022 Google LLC All rights reserved. Google and the Google logo are trademarks of Google LLC. All other company and product names may be trademarks of the respective companies with which they are associated.

* [Overview](https://partner.cloudskillsboost.google/course_sessions/7288730/labs/448219#step1)
* [Prerequisites](https://partner.cloudskillsboost.google/course_sessions/7288730/labs/448219#step2)
* [Introduction](https://partner.cloudskillsboost.google/course_sessions/7288730/labs/448219#step3)
* [Setup and requirements](https://partner.cloudskillsboost.google/course_sessions/7288730/labs/448219#step4)
* [Task 1. Explore bicycle rental data](https://partner.cloudskillsboost.google/course_sessions/7288730/labs/448219#step5)
* [Task 2. Explore the weather dataset](https://partner.cloudskillsboost.google/course_sessions/7288730/labs/448219#step6)
* [Task 3. Find correlation between rain and bicycle rentals](https://partner.cloudskillsboost.google/course_sessions/7288730/labs/448219#step7)
* [Summary](https://partner.cloudskillsboost.google/course_sessions/7288730/labs/448219#step8)
* [Congratulations!](https://partner.cloudskillsboost.google/course_sessions/7288730/labs/448219#step9)
* [End your lab](https://partner.cloudskillsboost.google/course_sessions/7288730/labs/448219#step10)

Top of Form

Bottom of Form

Top of Form