

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

To pass this practice quiz, you must receive 100%, or 1 out of 1 point, by completing the following activity. You can learn more about the graded and practice items in the [course overview](#).



### Activity Overview

In this activity, you will use your knowledge of SQL and potentially Google Dataflow to combine and move the key datasets you identified for the Cyclistic project into a target table that represents the extraction phase of an ETL pipeline, when data is pulled from different sources and moved to its destination. You will use the table you create in this activity to the final dashboard for stakeholders. As you complete this activity, remember to refer to the previous work you did when [completing the business intelligence project documents for Cyclistic](#) for details about the Cyclistic project, as well as the activity to [create a target table in BigQuery](#) for a refresher on target tables.

Be sure to complete this activity before moving on. The next course item will provide you with a completed exemplar to compare to your own work. You will not be able to access the exemplar until you have completed this activity.

### Scenario

Review the following scenario. Then, complete the activity. As a reminder, the end-of-course project activities are more open to your personal interpretation than other activities in the program. This is to give you an opportunity to practice the skills you have been learning in your own way. If you need help or feel stuck, you can always discuss your work with learners in the discussion forums or review the exemplar to help guide your process.

The product development team at Cyclistic has begun developing their business plan for next year. In order to build a better Cyclistic, the team needs to understand how customers are currently using the bikes, how location and other factors impact demand, and what stations get the most traffic. The Cyclistic team has a few goals:

Understand current customers' needs, what makes a successful product, and how new stations might alleviate demand in different geographical areas

Understand current usage of bikes at different locations

Apply customer usage insights to inform new station growth

Understand how different users (subscribers and non-subscribers) use the bikes

You met with stakeholders to complete project planning documents and uploaded the necessary tables into your BigQuery project space.

### Instructions

Follow the instructions and answer the following question to complete the activity. Then, go to the next course item to compare your work to a completed exemplar.

#### Step 1: Log into your GCP tool

To begin this activity, log into your Google Cloud account and navigate to [the BigQuery console](#). You can complete this activity using the BigQuery Sandbox, which does not require a Google Cloud billing account. You can learn more about enabling the Sandbox from [the BigQuery help guide](#). You can also use Dataflow to execute SQL code as a Job by navigating to [the Dataflow console](#) instead; this will require you to have a Google Cloud account. Both tools are useful for this project, so choose the tool you are more interested in using for this project.

#### Step 2: Querying your data

For this step, keep in mind the key metrics you and your stakeholders have identified, their business questions, and what data you'll need to develop the final dashboard. Previously, you explored the different public datasets your stakeholders provided and uploaded the zip code table your colleague shared with you. For the final dashboard, you will need two target tables: a table to capture the entire year and a table that focuses on summer trends. Here is an example of a query to capture a table with data from the entire year:

```
1  SELECT
2    TRI.usertype,
3    ZIPSTART.zip_code AS zip_code_start,
4    ZIPSTARTNAME.borough AS borough_start,
5    ZIPSTARTNAME.neighborhood AS neighborhood_start,
6    ZIPEND.zip_code AS zip_code_end,
7    ZIPENDNAME.borough AS borough_end,
8    ZIPENDNAME.neighborhood AS neighborhood_end,
9    -- Since this is a fictional dashboard, you can add 5 years to make it look recent
10   DATE_ADD(DATE(TRI.starttime), INTERVAL 5 YEAR) AS start_day,
11   DATE_ADD(DATE(TRI.stoptime), INTERVAL 5 YEAR) AS stop_day,
12   WEA.temp AS day_mean_temperature, -- Mean temp
13   WEA.wdsp AS day_mean_wind_speed, -- Mean wind speed
14   WEA.prcp AS day_total_precipitation, -- Total precipitation
15   -- Group trips into 10 minute intervals to reduce the number of rows
16   ROUND(CAST(TRI.tripeduration / 60 AS INT64), -1) AS trip_minutes,
17   COUNT(TRI.bikeid) AS trip_count
18 FROM
19   `bigquery-public-data.new_york_citibike.citibike_trips` AS TRI
20 INNER JOIN
21   `bigquery-public-data.geo_us_boundaries.zip_codes` ZIPSTART
22   ON ST_WITHIN(
23     ST_GEOPOINT(TRI.start_station_longitude, TRI.start_station_latitude),
24     ZIPSTART.zip_code_geom)
25 INNER JOIN
26   `bigquery-public-data.geo_us_boundaries.zip_codes` ZIPEND
27   ON ST_WITHIN(
28     ST_GEOPOINT(TRI.end_station_longitude, TRI.end_station_latitude),
29     ZIPEND.zip_code_geom)
30 INNER JOIN
31   `bigquery-public-data.noaa_gsod.gsod20*` AS WEA
32   ON PARSE_DATE("%Y%m%d", CONCAT(WEA.year, WEA.mo, WEA.da)) = DATE(TRI.starttime)
33 INNER JOIN
34   -- Note! Add your zip code table name, enclosed in backticks: `example_table`
35   `(insert your table name) zipcodes` AS ZIPSTARTNAME
36   ON ZIPSTART.zip_code = CAST(ZIPSTARTNAME.zip AS STRING)
37 INNER JOIN
38   -- Note! Add your zipcode table name, enclosed in backticks: `example_table`
39   `(insert your table name) zipcodes` AS ZIPENDNAME
```

```
40 ON ZIPEND.zip_code = CAST(ZIPENDNAME.zip AS STRING)
```

Note that this query includes a DATE\_ADD function to add five years to the data. The public data you are using to create this dashboard is from 2014 and 2015, so this is a way your dashboard appear more recent. Normally, you would not change the dates in a dataset, but because this is a fictional project, you can include this in your own query. This query is optional; however, your exemplar will appear differently if you don't include it. You will need to develop a similar query to capture a table with data from July through September to explore summer trends specifically.

### Step 3: Finish the job

Once you execute the code, it will take a few moments to process. After the query has finished running, you will be able to download the tables as CSV files by using the Save dropdown and selecting the appropriate file type.

The screenshot shows a BigQuery interface with a query execution result. The query is a SELECT statement that joins data from 'bigquery-public-data.new\_york\_citibike.citibike\_trips' and 'bigquery-public-data.geo.us\_boundaries.zip\_codes'. The query results are displayed in a table with columns: user\_type, zip\_code\_start, borough\_start, zip\_code\_end, and borough\_end. The 'SAVE RESULTS' dropdown menu is open, showing options: CSV (Google Drive), CSV (local file), JSON (local file), JSONL (newline delimited), BigQuery table, Google Sheets, and Copy to Clipboard. Red arrows point to the 'SAVE RESULTS' button and the 'CSV (Google Drive)' option.

Row	user_type	zip_code_start	borough_start	zip_code_end	borough_end
1	Subscriber	10024	Manhattan	10024	Manhattan
2	Subscriber	11221	Brooklyn	11221	Brooklyn
3	Subscriber	11206	Brooklyn	11206	Brooklyn
4	Subscriber	10028	Manhattan	10028	Manhattan
5	Customer	10028	Manhattan	10028	Manhattan
6	Subscriber	10028	Manhattan	10028	Manhattan
7	Subscriber	11206	Brooklyn	11206	Brooklyn
8	Subscriber	11222	Brooklyn	11222	Brooklyn
9	Subscriber	11206	Brooklyn	11206	Brooklyn
10	Subscriber	10024	Manhattan	10024	Manhattan
11	Subscriber	10024	Manhattan	10024	Manhattan

This might take a few minutes. Once you have downloaded the table, you will be ready to upload it to Tableau to create your dashboard!

### What to Include in Your Response

Be sure to address the following criteria:

Necessary tables are successfully combined into summary tables

Appropriate tables are downloaded and ready to upload to Tableau

1. Did you complete this activity?

- ☒ Yes  
☐ No

✓ Correct

Thank you for completing this activity! Target tables are an essential part of the extraction phase of an ETL pipeline. Go to the next course item to compare your work to a completed exemplar.