

Business Intelligence Project: Cyclistic

End-of-course project for Google Business Intelligence Professional Certificate

PART 2: Cyclistic Datasets

By now, you're getting ready to take the next steps with your Course 2 end-of-course project. To work with the Cyclistic project data, you will need to locate the appropriate public datasets and upload the zip code spreadsheet that your colleague shared into your BigQuery project space.

For this end-of-course project, you will be using two public datasets, which exist in the public data available from the Explorer pane of your console:

- [NYC Citi Bike Trips](#), [Census Bureau US Boundaries](#),
- [GSOD from the National Oceanic and Atmospheric Administration](#)

Additionally, you will need to upload the zip code [spreadsheet](#) your colleague shared with you.

Upload to BigQuery

First, navigate to your BigQuery console. Go to the BigQuery homepage or navigate to the console.

Search and preview the public datasets using the search bar in the Explore pane of your console:

These datasets are already available for you to query, but it can be useful to check out the tables before you start working with them. Find all three datasets by searching the appropriate dataset name in the search bar:

- new_york_citibike
- geo_us_boundaries
- noaa_gsod

After you have familiarized yourself with the public data, upload the zip code dataset. Either save the Google Sheet as a CSV file on your device or download it into your own Drive space. Click on the + ADD DATA button in the Explorer menu pane; this will open the Add Data menu.

From here, select Local file to upload the CSV or Google Cloud Storage to choose the sheet from your personal Drive. However you add the file, you will need to fill out the necessary fields in the Create Table menu. If you haven't already, the Create table menu will also prompt you to create a dataset to house this table.

Select CREATE NEW DATASET and name the dataset appropriately for this project. You can leave the data location set to default. Once you have completed filling out this information, click Create Dataset.

Now, finish filling out the information for your table. Name your table appropriately for your project and select CSV under file type. Finally, select Auto detect for the schema. Once done, select Create Table. The new table should appear under your dataset in the Explorer pane momentarily.

From here, explore the schema, preview the data, and familiarize yourself with this table.

The team at work

As you learned during your previous meeting with Cyclistic, the product development team has begun planning for the next year of Cyclistic's bike-sharing program. Cyclistic's Customer Growth Team is creating a business plan for next year. The team wants to understand how their customers are using their bikes; their top priority is identifying customer demand at different station locations. The Cyclistic team posed an important primary question:

- How can we apply customer usage insights to inform new station growth?

Answering these questions starts with the data from the Cyclistic bikes themselves, which the team has provided you, and the reporting dashboard the team uses to gain insights. In addition to the explicit requests the stakeholders made, you realize a few key things about the team's current processes. First, you realize that there are stakeholders from a variety of different departments accessing and using this data with different levels of technical expertise. There are stakeholders from these teams:

- Product development
- Customer data
- Engineering
- Data analytics
- Data warehousing
- API
- IT

- Cyclistic executive
- Project management

For example, you realize that Earnest Cox, the VP of product development, is often requesting high-level insights into the data and rarely needs detailed overviews of the data. Alternatively, Tessa Blackwell from the data analytics team does explore the data in-depth and spends a lot more time reviewing the dashboard views. As you develop your reporting tools, you will want to find a way to answer both of these stakeholders' needs.

Additionally, one of your coworkers finds out you're working on this project and shares a dataset they created recently for a project of their own that they think might help you: NYC zip codes. This dataset provides the zip codes for the different neighborhoods and boroughs in New York City; this will let you compare the bike data to the weather data more easily since you will be able to match the locations more accurately. It will also help you develop your map visualization later on.