



# Reading: 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. This reading will guide you through that process. Once you have finished this reading, you will be ready for the upcoming activities and to deliver key insights to your stakeholders.

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](#).

The screenshot shows the BigQuery homepage. At the top, it says "BigQuery" and describes it as a "Cost-effective, serverless, multicloud enterprise data warehouse to power your data-driven innovation." It mentions "\$300 in free credits" and "10 GB storage and up to 1TB queries free per month." Below this, there are two buttons: "Go to console" and "Contact sales". A list of benefits includes:

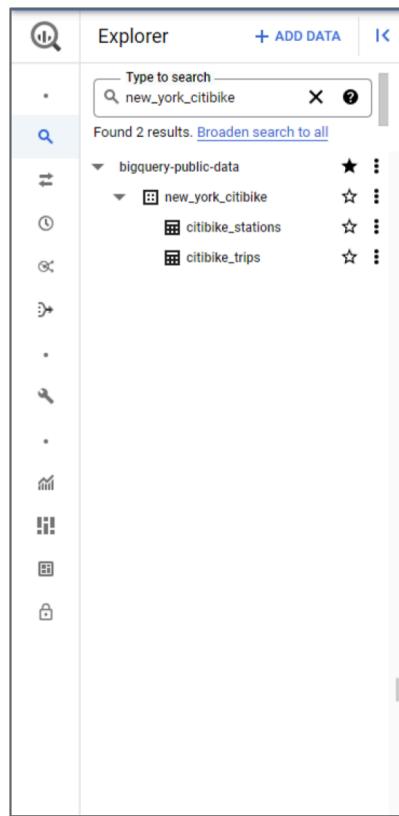
- Run analytics at scale with 27% lower three-year TCO than [cloud data warehouse alternatives](#)
- Democratize insights with built-in business intelligence and machine learning
- Power business decisions from data across clouds with a flexible, multicloud analytics solution
- BigQuery is at the core of Google's unified [data cloud](#) to help you drive data transformation

A "BENEFITS" section lists three items:

- Gain insights with real-time and predictive analytics**: "Query streaming data in real time and get up-to-date information on all your business processes. Predict business outcomes easily with built-in machine learning—without the need to move data."
- Protect your data and operate with trust**: "Rely on BigQuery's robust security, governance, and reliability controls that offer high availability and a 99.99% uptime SLA. Protect your data with encryption by default and customer-managed encryption keys."
- Break data silos and prevent lock-in**: "Embrace the ecosystem of partners, open data standards, and common industry data models to have the choice to work with data across any platform or environment."

To the right, there is a video thumbnail titled "Learn BigQuery in a minute, including how it works and common use cases" with a duration of "01:26".

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

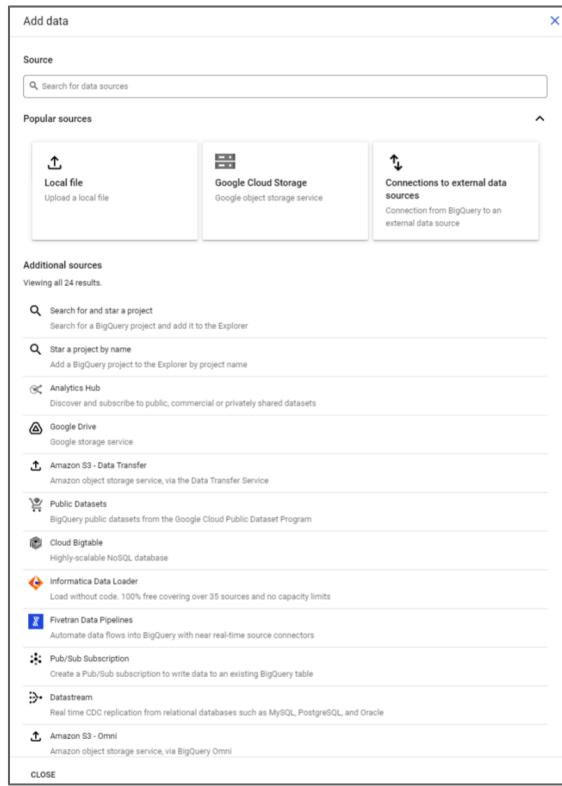


These datasets are already available for you to query, but it can be useful to check out the tables before you start working with them. Find all three datasets by searching the appropriate dataset name in the search bar:

- 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.

The screenshot shows the 'Create table' dialog box with the following settings:

- Source:** Create table from **Upload**. A file named "Cyclistic zip codes - list.csv" is selected.
- File format:** CSV
- Destination:** Project "airy-shuttle-315515" is selected. Under **Dataset**, "cyclistic" is chosen. Under **Table**, "zip\_codes" is entered.
- Schema:**  Auto detect. A note says "Schema will be automatically generated."
- Partition and cluster settings:** Partitioning is set to "No partitioning". Clustering order is specified as "Clustering order determines the sort order of the data. Clustering can be used on both partitioned and non-partitioned tables."
- Advanced options:** A dropdown menu is shown.

At the bottom, there are **CREATE TABLE** and **CANCEL** buttons.

From here, explore the schema, preview the data, and familiarize yourself with this table. Once you have uploaded this dataset, you will be ready to continue with your project!

---