Unify data with target tables

As you have been learning, target tables are predetermined locations where pipeline data is sent in order to be acted on in a database system. Essentially, a source table is where data comes from, and a target table is where it's going. This reading provides more information about the data-extraction process and how target tables fit into the greater logic of business intelligence processes.

Data extraction

Data extraction is the process of taking data from a source system, such as a database or a SaaS, so that it can be delivered to a destination system for analysis. You might recognize this as the first step in an ETL (extract, transform, and load) pipeline. There are three primary ways that pipelines can extract data from a source in order to deliver it to a target table:

- **Update notification:** The source system issues a notification when a record has been updated, which triggers the extraction.
- **Incremental extraction:** The BI system checks for any data that has changed at the source and ingests these updates.
- **Full extraction:** The BI system extracts a whole table into the target database system.

Once data is extracted, it must be loaded into target tables for use. In order to drive intelligent business decisions, users need access to data that is current, clean, and usable. This is why it is important for BI professionals to design target tables that can hold all of the information required to answer business questions.

The importance of target tables

As a BI professional, you will want to take advantage of target tables as a way to unify your data and make it accessible to users. In order to draw insights from a variety of different sources, having a place that contains all of the data from those sources is essential.

Activity Overview

In this activity, you will examine and execute a SQL query that moves data into a target table. As you have learned, a target table is the predetermined location where pipeline data is sent in order to be acted on. It's important for BI professionals to design target tables that can hold all of the information required to answer business questions. As a BI professional, you can use target tables to unify your data and make it accessible to users.

You will use this kind of query during the extraction phase of an ETL pipeline. By running this SQL query in BigQuery, you will learn more about using SQL in business intelligence and the role of target tables in data pipelines.

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 scenario that follows. Then complete the step-by-step instructions.

You are a BI professional working for the city of San Francisco. The mayor wants to start an annual tree appreciation program, which will involve decorating trees along streetsides. You will provide the mayor's office with a list of the top 10 addresses where mature trees have been planted along the streets. With this information, they can plan to have trees decorated before the event begins and the decorations cleaned up once it is complete.

In this activity, you will:

- Select this data from the Street Trees dataset and transport it into a target table
- Find the 10 addresses with the most trees planted along the street
- Find the number of trees at each address

The Street Trees dataset* is publicly available on BigQuery and contains more than 190 thousand rows of data about trees planted in San Francisco from 1955 to the present. It has information about each tree maintained by the San Francisco Department of Public Works in the city, including each tree's unique ID, address where it was planted, plot size, geographic coordinates, and more.

**This dataset is publicly available and provided by https://data.sfgov.org//, the San Francisco's Department Publishing Plans portal. This data is provided on an "as-is" basis and is subject to change at any time without notice. Please visit the source to get the latest version.*

Step-By-Step Instructions

Follow the instructions to complete each step of the activity. Then, answer the questions at the end of the activity before going to the next course item to compare your work to a completed exemplar.

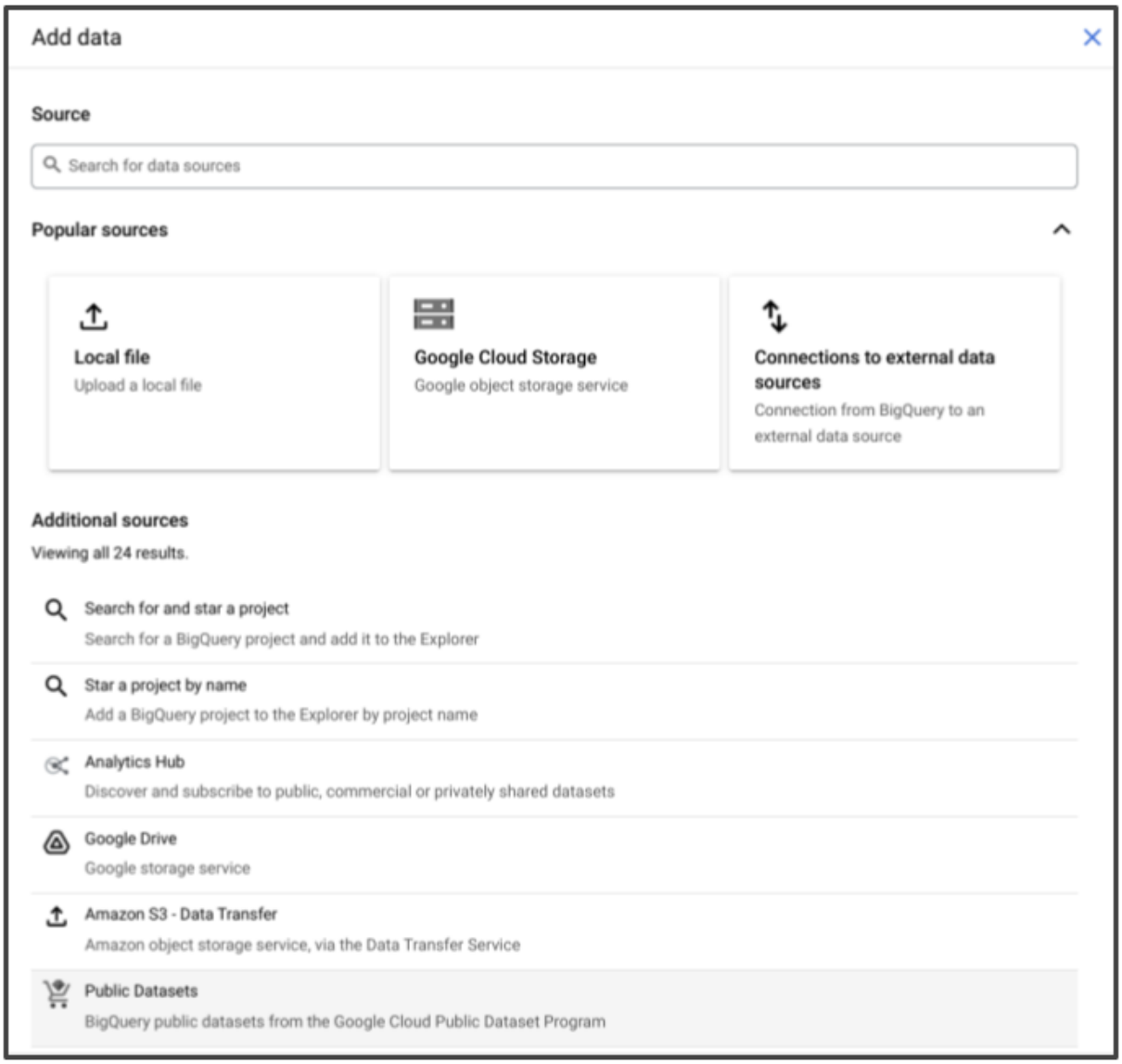
Step 1: Open the BigQuery console

To begin this activity, navigate to your BigQuery console.

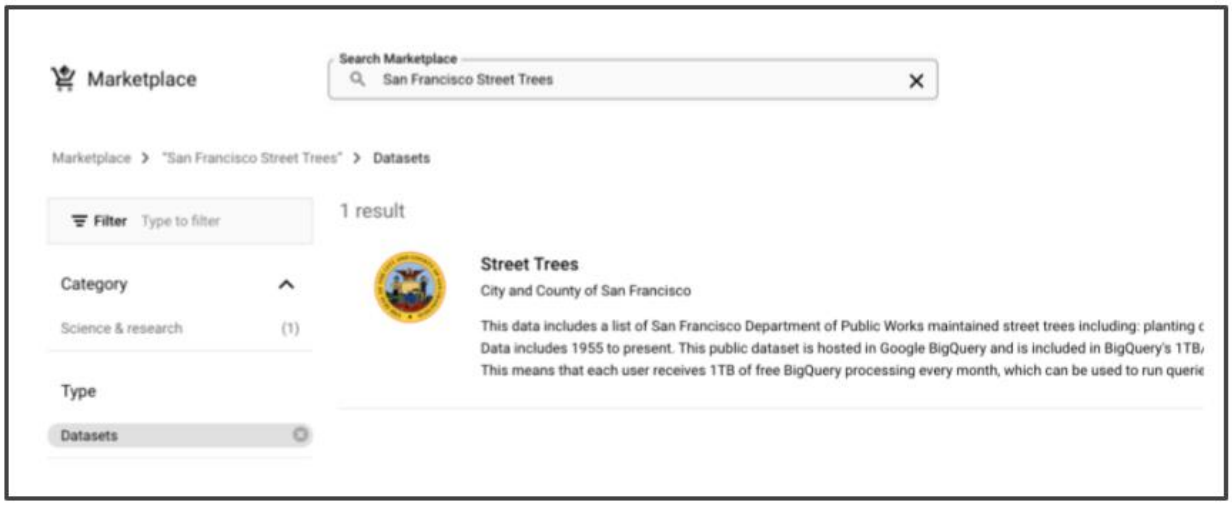
Step 2: Open the San Francisco Street Trees dataset

Go to the Explorer menu and Click **+ ADD DATA**.

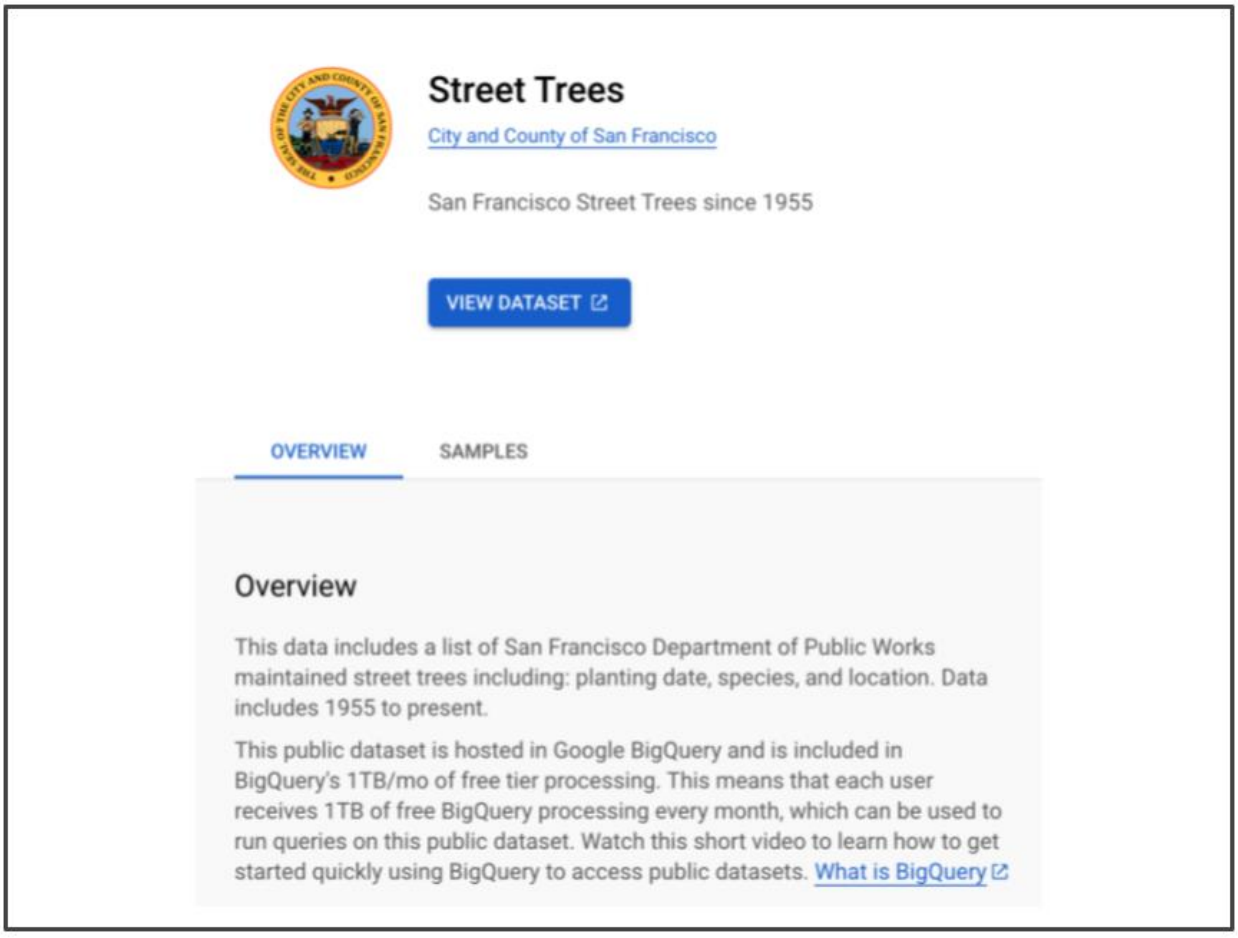
Select **Public datasets**.



This will open a new menu, where you will search for “San Francisco Street Trees” and click the first result.



Click **VIEW DATASET** to open the dataset in BigQuery. This will open a new tab. Return to the **Editor** tab.



Step 3: Move relevant data into a target table

This code will extract data from the San Francisco Street Trees BigQuery Dataset and place it into its own table.

Copy and paste the following query into the BigQuery editor window:

```
SELECT
    address,
    COUNT(address) AS number_of_trees
FROM
    bigquery-public-data.san_francisco_trees.street_trees
WHERE
    address != "null"
GROUP BY address
ORDER BY number_of_trees DESC
LIMIT 10;
```

Step 4: Save your query as a .CSV

After you run your query, you will receive a target table with two columns. It will list the 10 addresses with the most trees planted along the street, as well as the number of trees at each address. This helps the mayor and their team plan how many trees to decorate and where to send teams to clean up after the tree appreciation event has concluded.

Now, save this target table as a .csv file so you can compare it to your exemplar.

To do this, click **SAVE RESULTS** in the Query results window. Select **CSV (Google Drive)** to automatically save it to your Drive or **CSV (local file)** to save a version of it to your local files. You'll need this .csv file in an accessible place to compare your result to the upcoming exemplar.

What to Include in Your Response

You should confirm the following for your completed .csv file:

- It has two columns
- Column 1 is the address of the trees
- Column 2 is the number of trees at the address

In addition to this criteria, in a business role you might consider including a presentation to the mayor's team with your findings related to the tree appreciation event. You could include the following:

- A dashboard that includes likely addresses and trees suitable for decoration
- A recommendation for planning the supplies needed for the event
- A recommendation for planning an efficient decoration route before the event begins, and an efficient cleaning route after the event has concluded