Team Project Load Data

Decide on a dataset and upload it to GCP. I have chosen COVID-19_Public_Therapeutic_Locator from https://healthdata.gov/Health/COVID-19-Public-Therapeutic-Locator/rxn6-qnx8/data

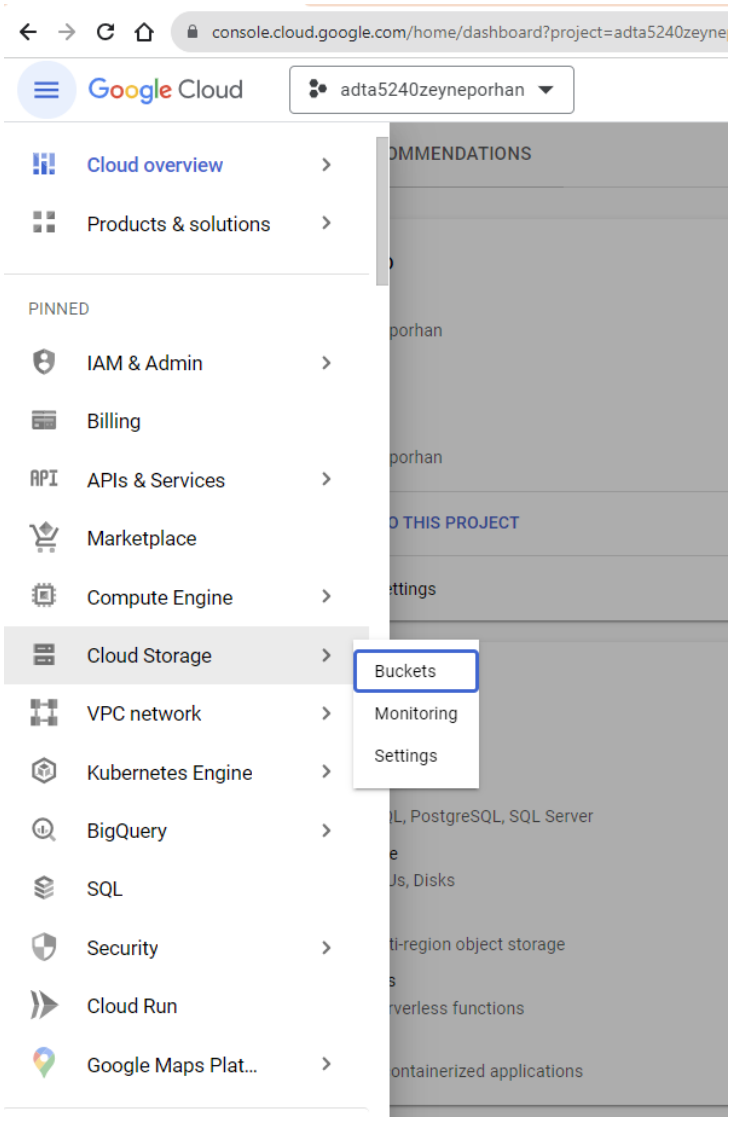
For this first team project, after you have selected your dataset, you will then load the file into GCP Storage and then into BigQuery.

Part I: Load into GCP Cloud Storage

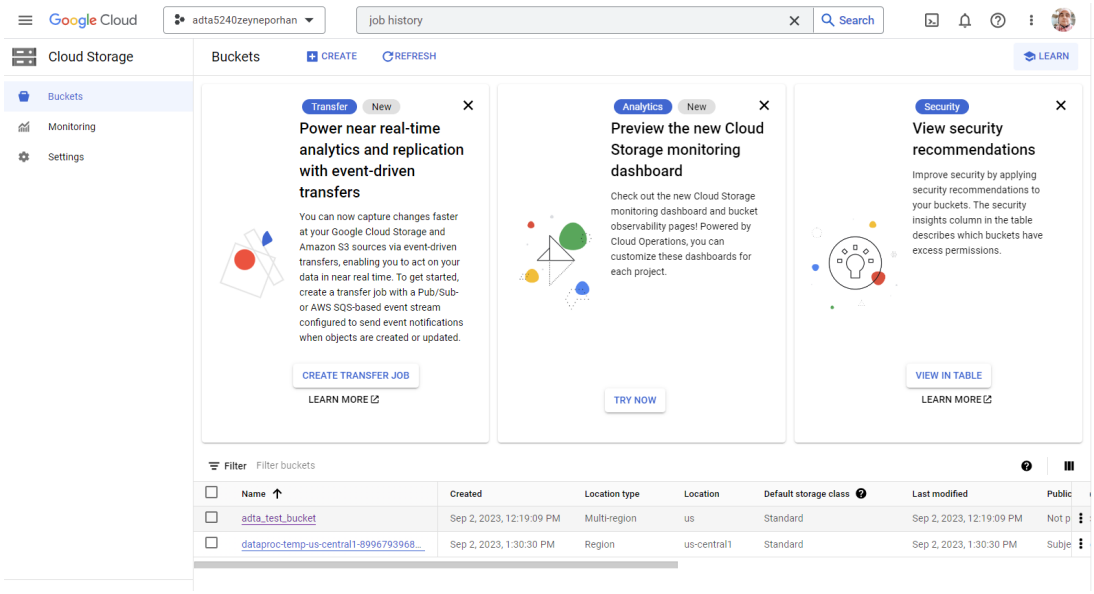
Step 1: Select your project. You only need one project from one team member to finish this part of the final.

Step 2: Load the dataset into GCP Cloud Storage

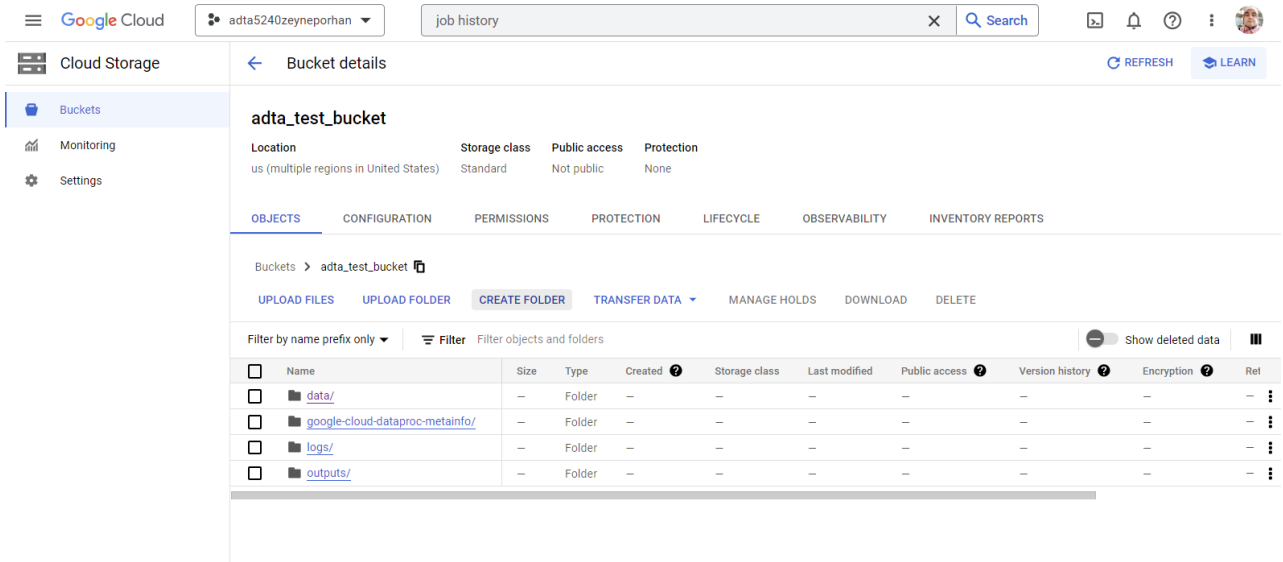
a) Go to STORAGE in GCP, click Buckets

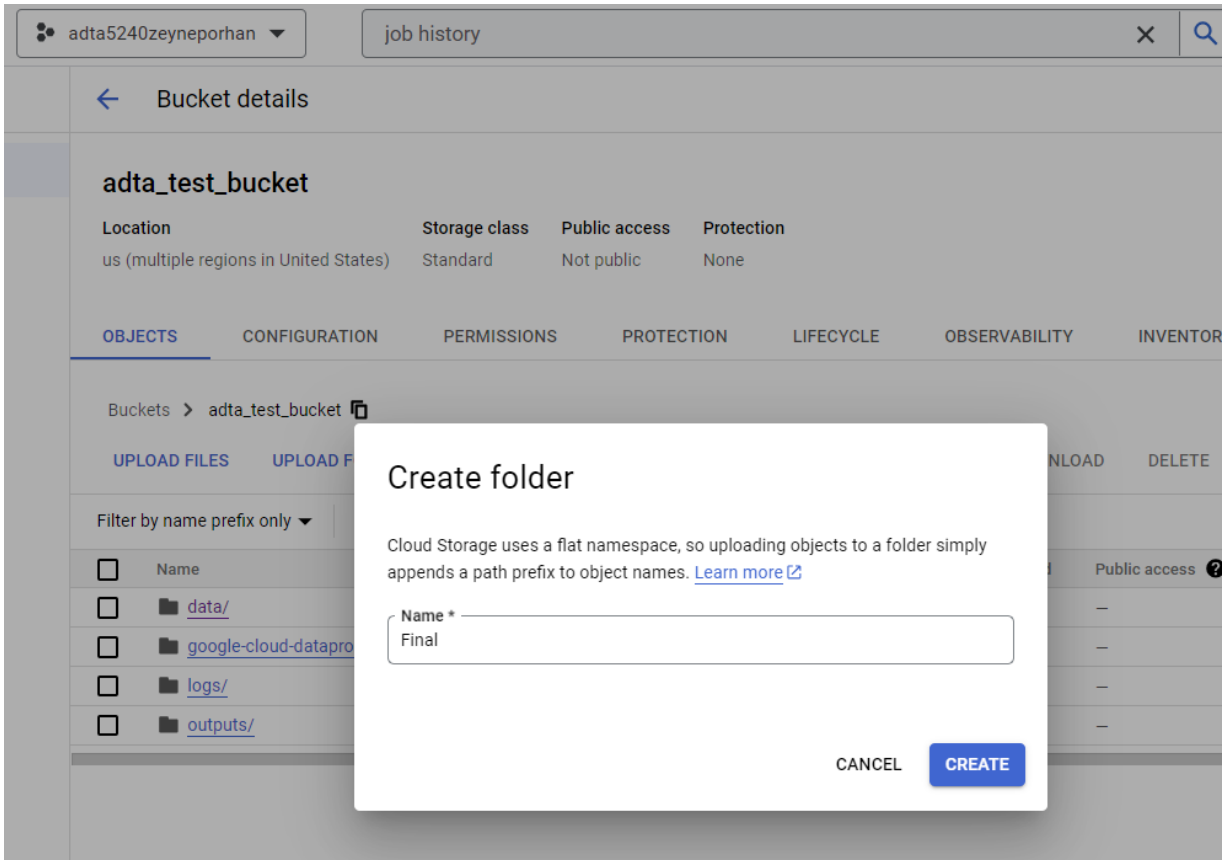


b) Click on the bucket you created in the beginning

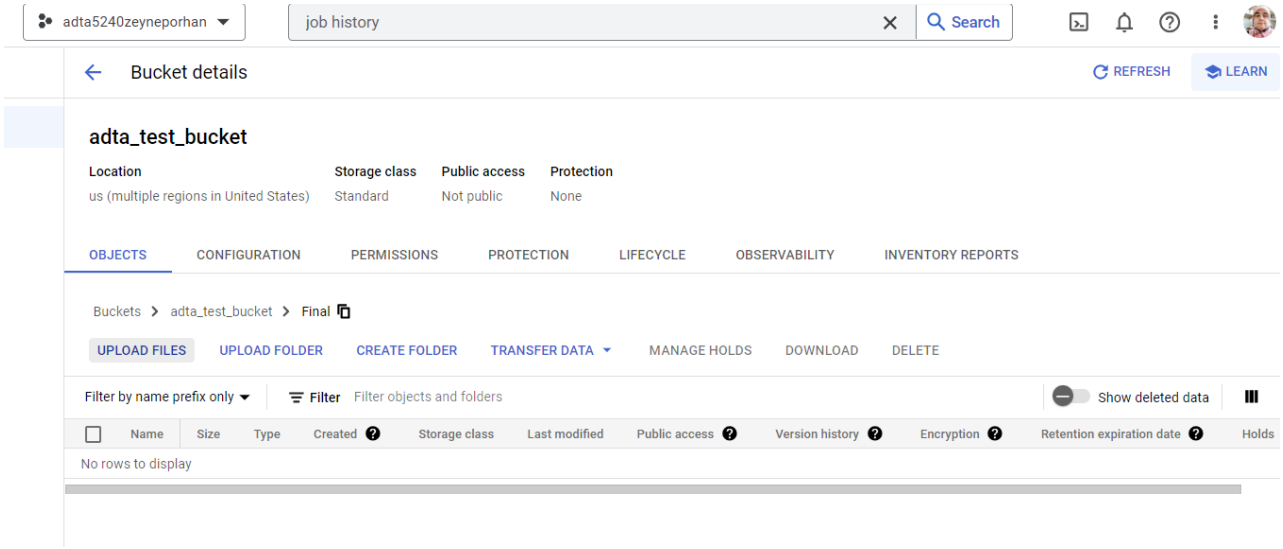


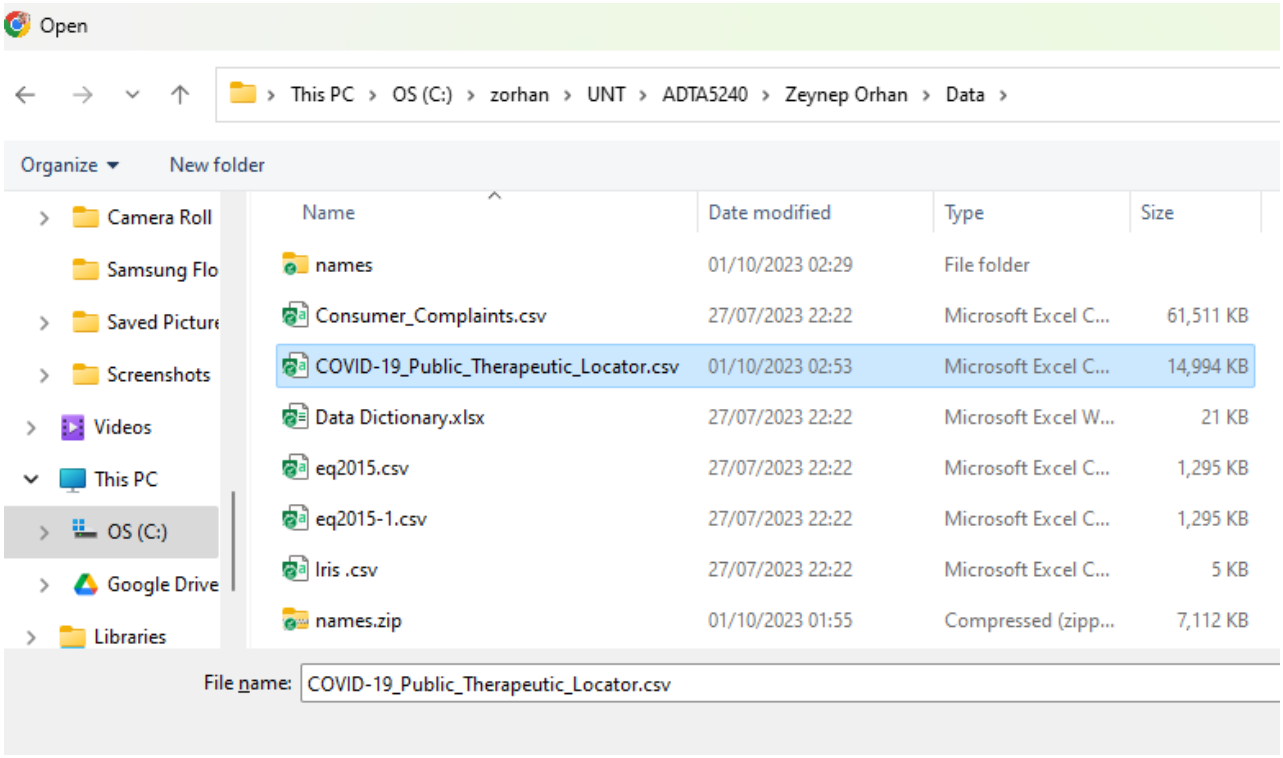
c) Create a new folder and title it “Final.”

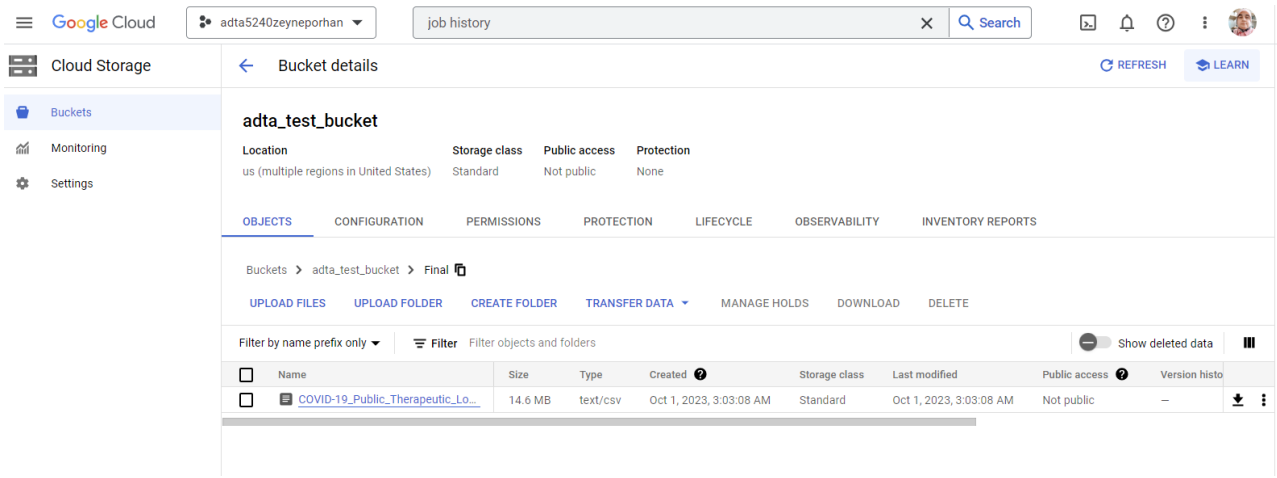




d) Click on "UPLOAD FILES" and choose the file that you have downloaded.



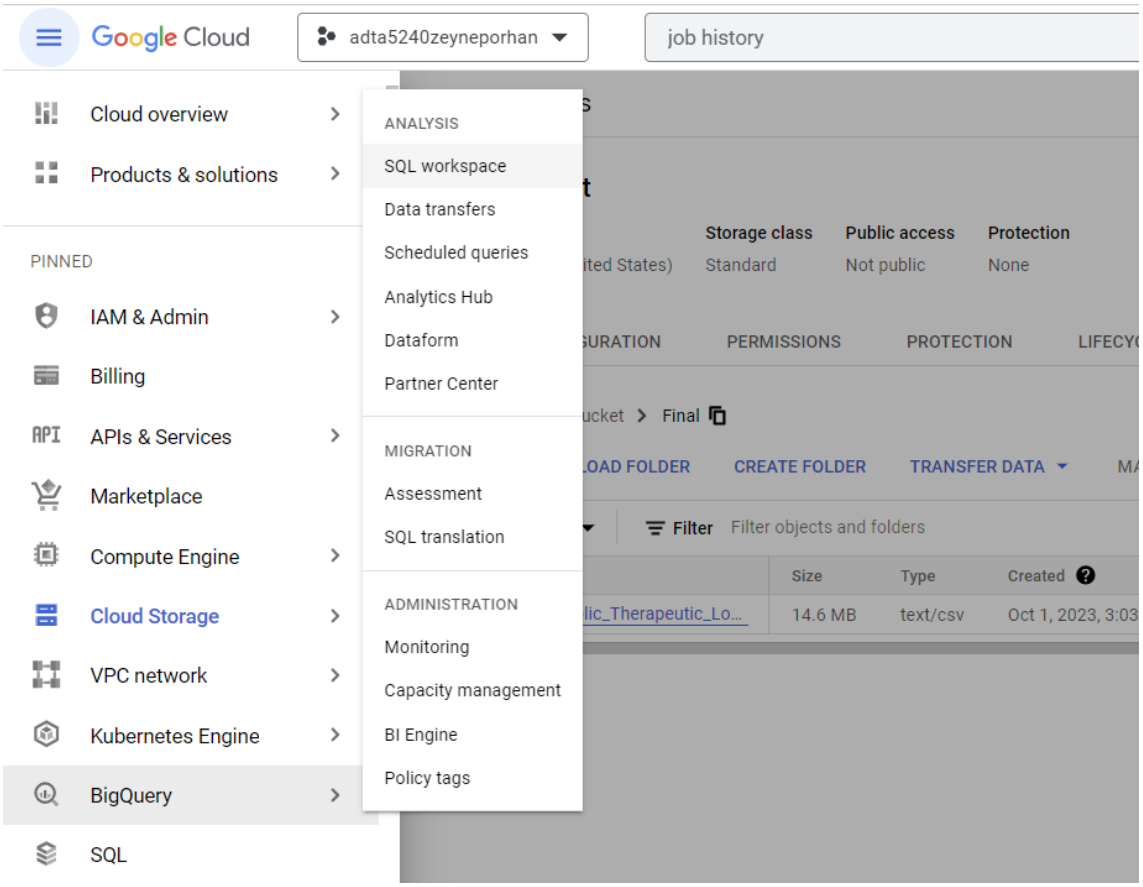




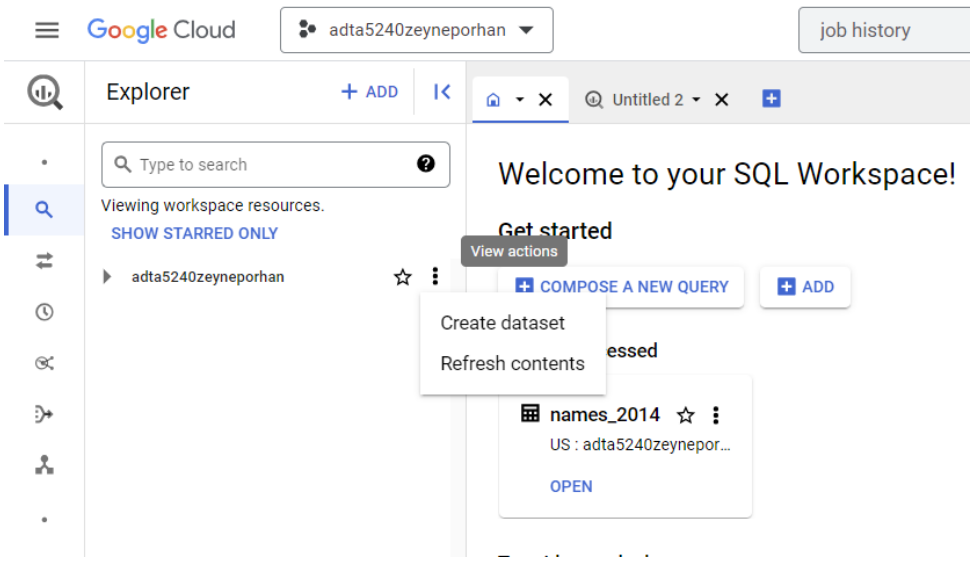
e) The file is now loaded into GCP Storage

Part II: Load into BigQuery

Step 1: In the Cloud Console, double click BigQuery – SQL Workspace.



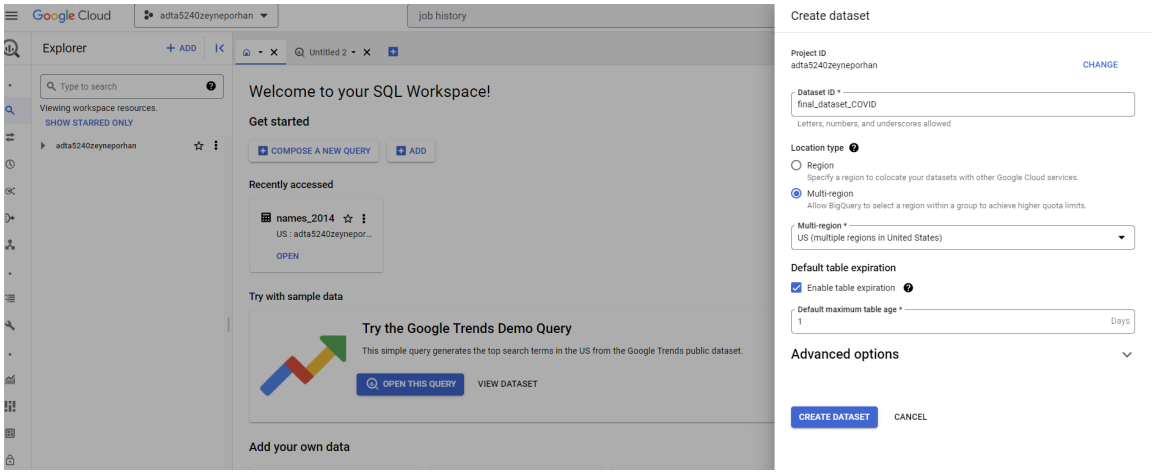
Step 2: In the **Explorer** panel of your projects (3 dots), click on "Create a Dataset"



a) Follow the prompts

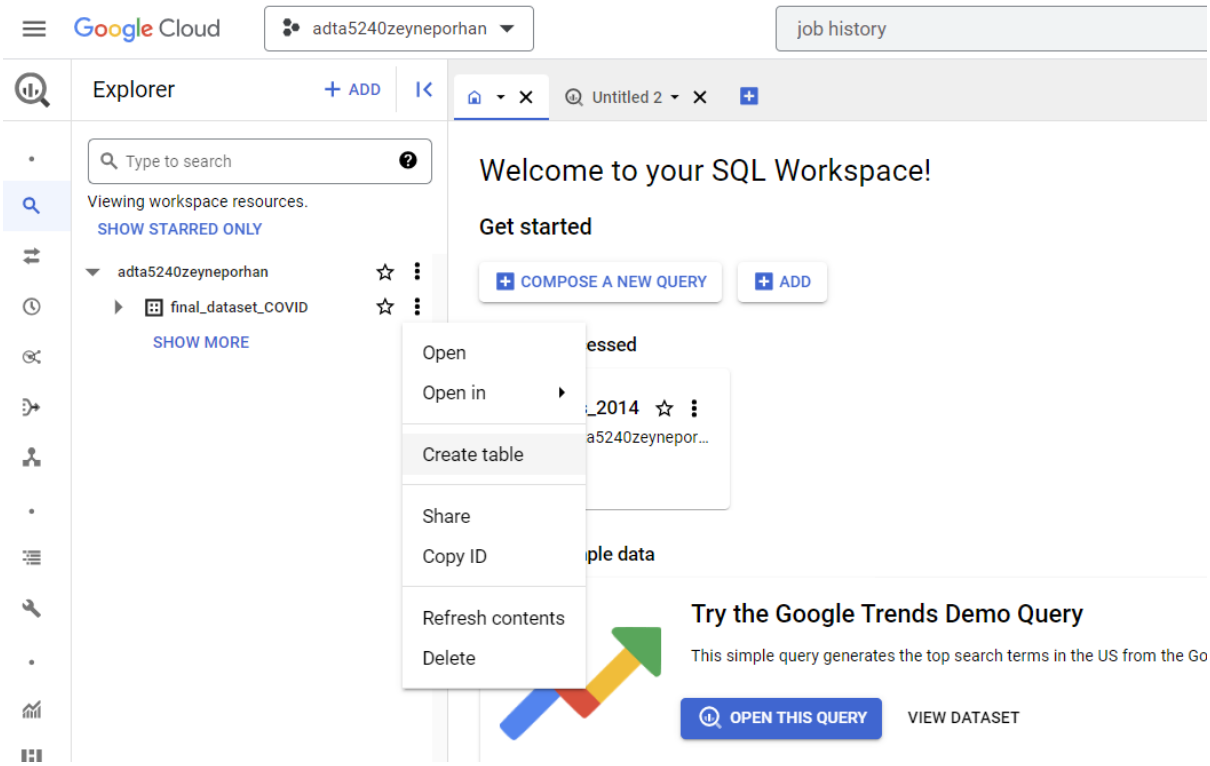
b) Add the dataset name and location. You must use the same location of the bucket for the dataset. I used multiple regions in United States. You may need to go back to your notes to see your location.

c) You will also set a default table age. Set this to 1 so you will not be charged.



d) Click "CREATE DATASET"

Step 3: Expand the dataset you created by clicking on the chevron next to your project name.

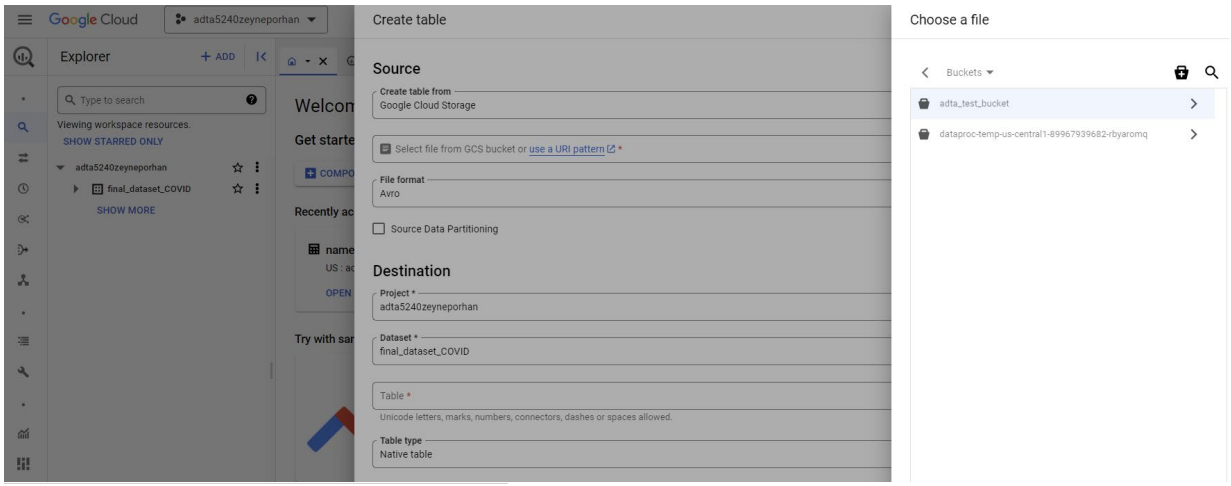


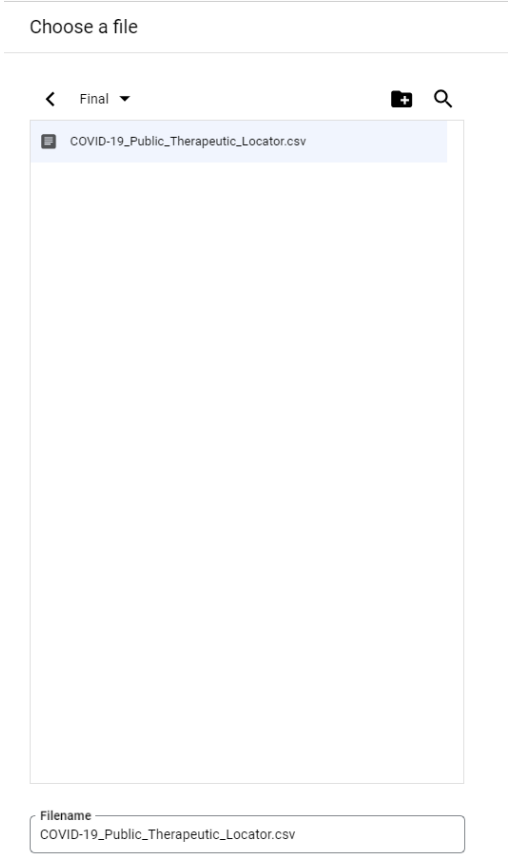
Step 4: Click on the 3 dots next to your newly created data set
Step 5: Click "**Create table"**

Step 6: In the **Create table** page, follow the prompts

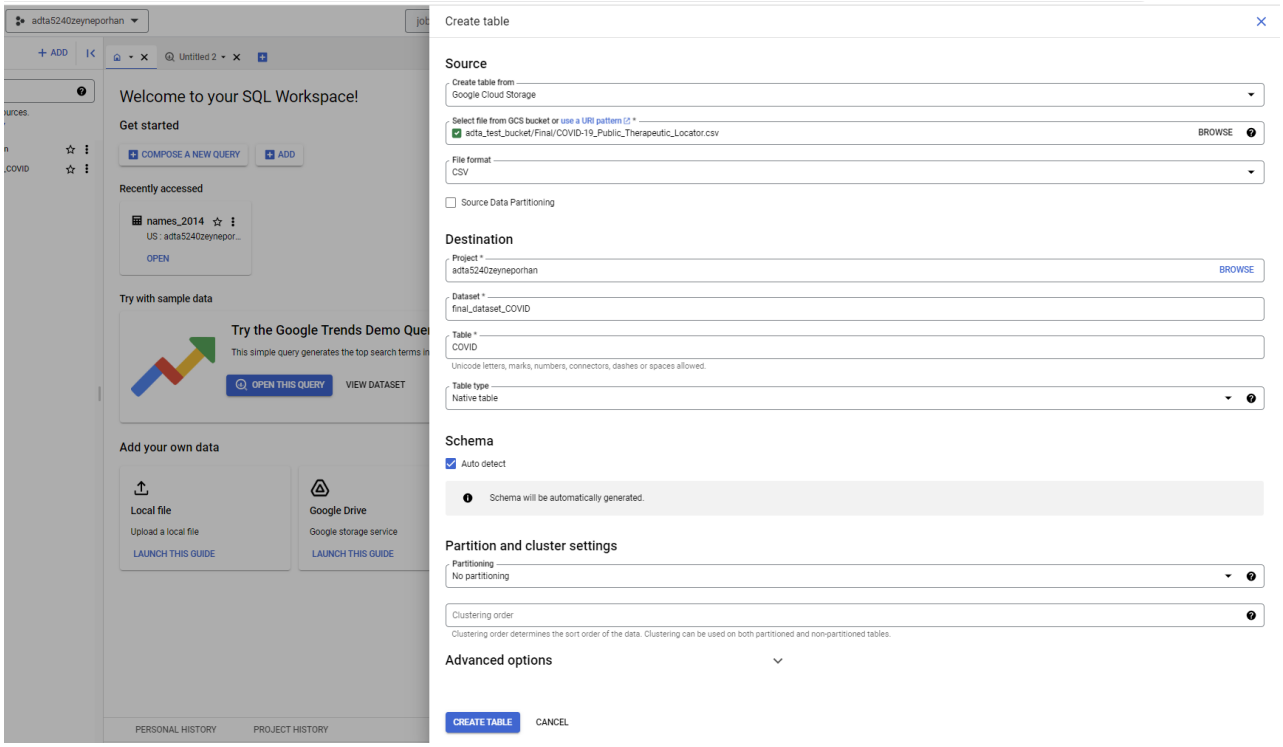
a) in the **Source** section:

1) Click Google Cloud Storage
2) Choose your bucket and the file you loaded from Part 1. The Cloud Storage bucket must be in the same location as the dataset that contains the table you're creating.
3) Click "Select"
4) For **File format**, select **CSV** (if it does not auto-populate)



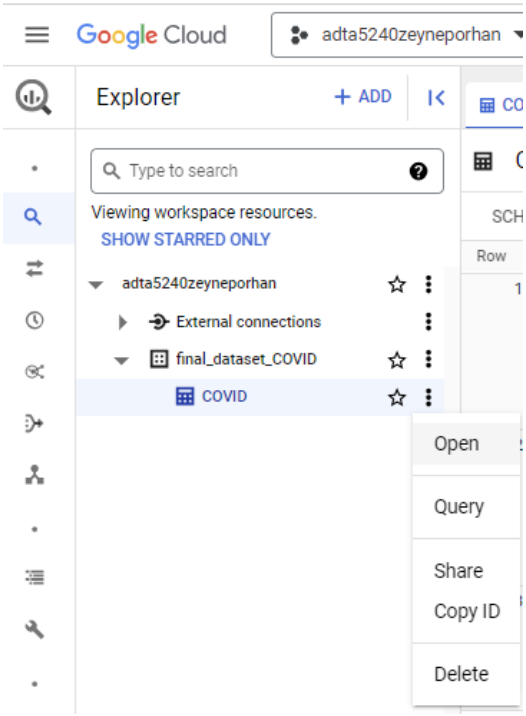


b) In the **Destination** section:
    1) For **Dataset name**, choose the appropriate dataset
    2) Verify that **Table type** is set to **Native table**.
    3) In the **Table name** field, enter the name of the table you're creating in BigQuery.
c) In the **Schema** section, for **Auto detect**, check **Schema and input parameters** to enable schema auto detection.
d) Create table

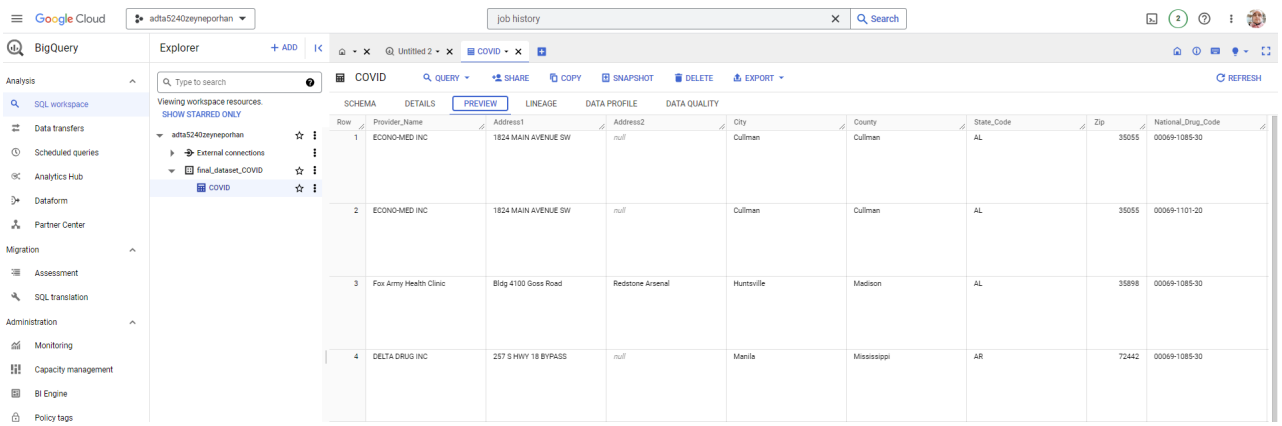


Step 7: Be sure that your table was created

1. On the BigQuery menu, select the name of your Google Cloud project from the list of resources and expand it.
2. Find the data set you created, expand it by clicking on the chevron
3. Click on the 3 dots next to the table you created and click “OPEN”



4. Now click on the "PREVIEW" tab to see the table you created. This is the second screenshot that you must include for this submission.



IMPORTANT: Remember, you have only 10GB of free storage in BigQuery. So, in order to keep from being charged, delete your table and dataset. This is very easy to duplicate when it is time to query your data. It always helps to do an activity more than once.