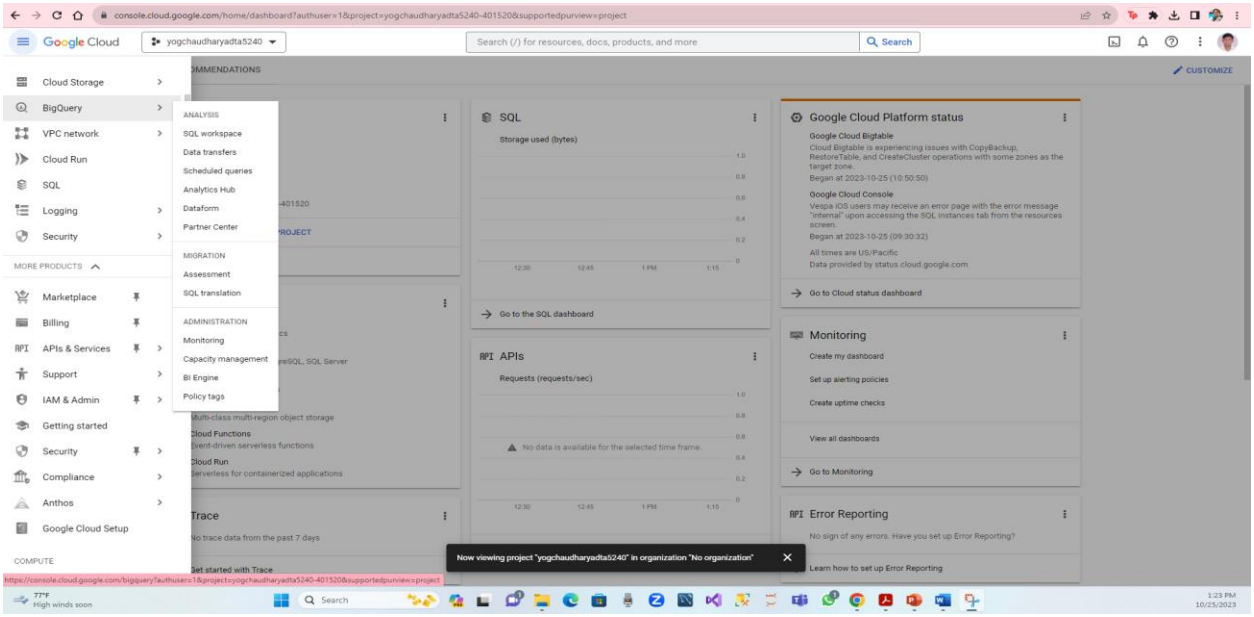
Yog Chaudhary
11727095
ADTA 5240 Week (Harvesting, Storing, And Retrieving Data)
Professor: Dr. Zeynep Orhan
University Of North Texas

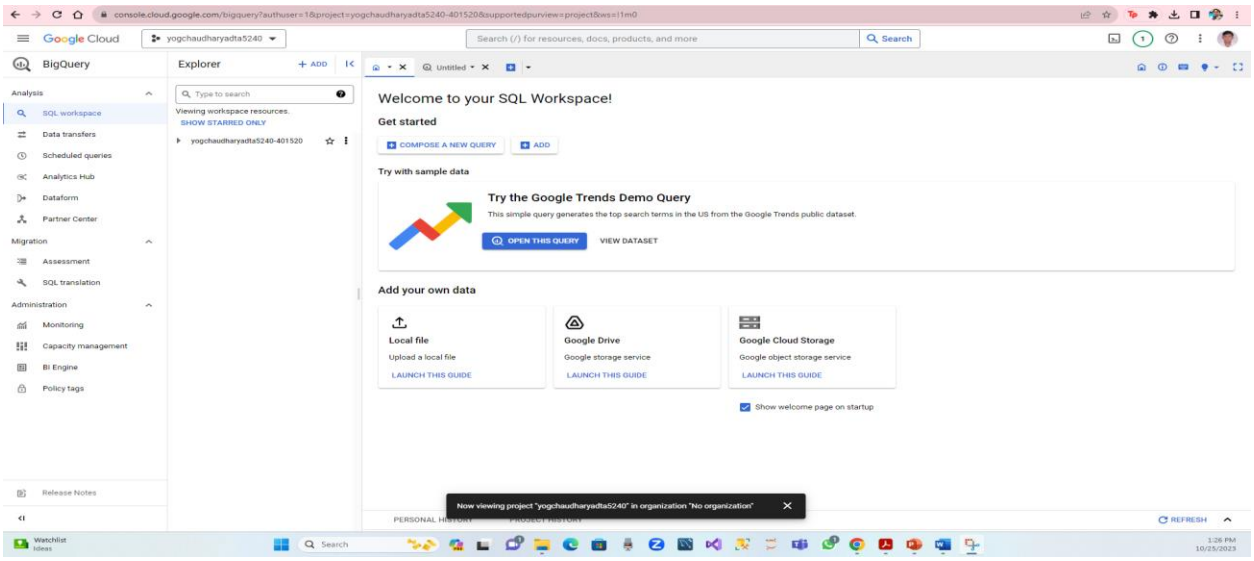
Oct 25, 2023

❖ Joining Data from Two Different Publicly Available Datasets

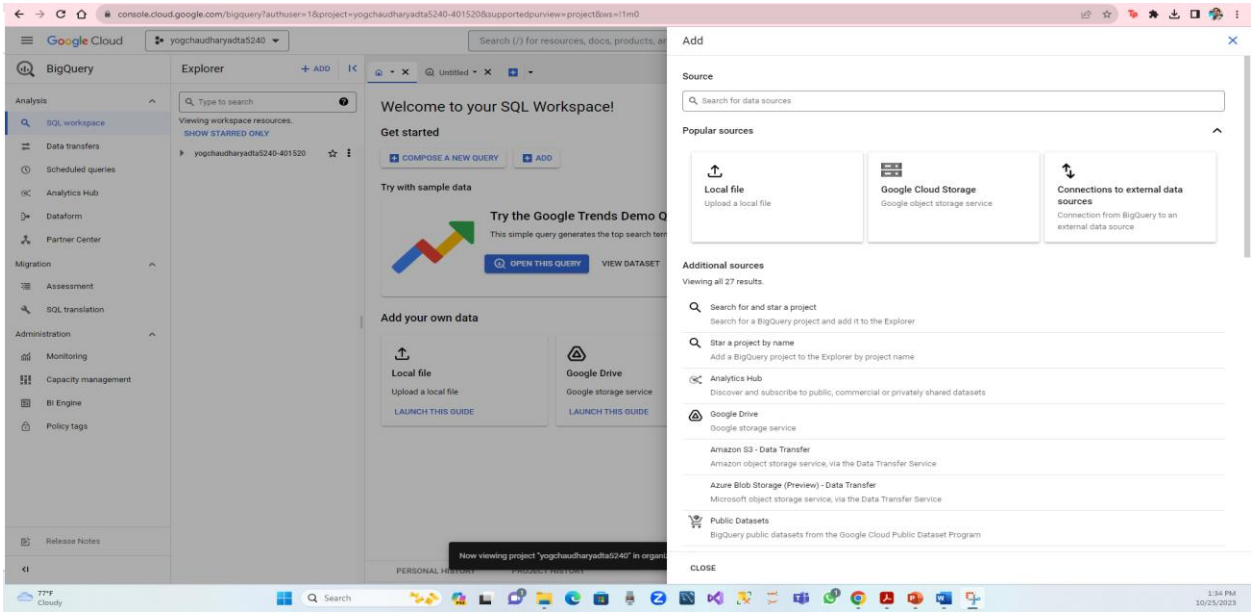
I am going to use 2 public datasets for this. These are Sustainable Development Goals (UN SDG) indicators and World Development Indicators (WDI). These 2 datasets are available through Google Console.

1. The first step is viewing the public datasets through Marketplace in GCP.
   - For this, I have chosen one variable from each dataset.
   - I clicked on the navigation panel and clicked on a big query.
   - After clicking on a big query, I was taken to the page below.

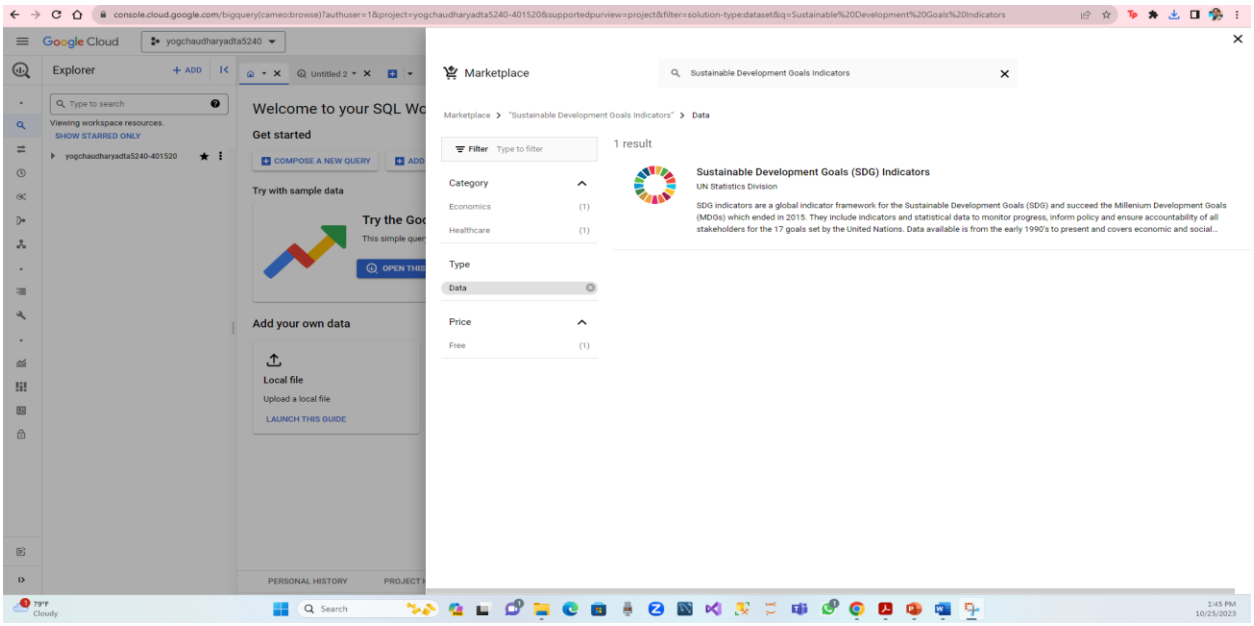




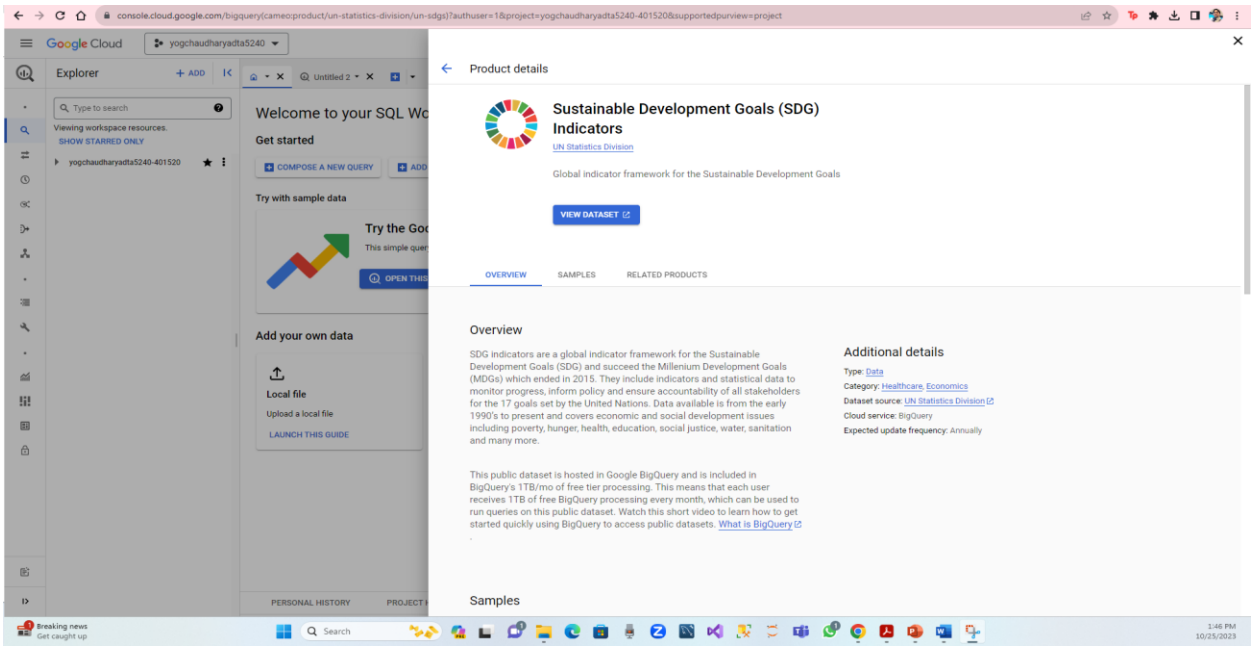
- I clicked on “ADD DATA”.
- Then I was directed to the page that below shows.
- Here I chose public datasets.

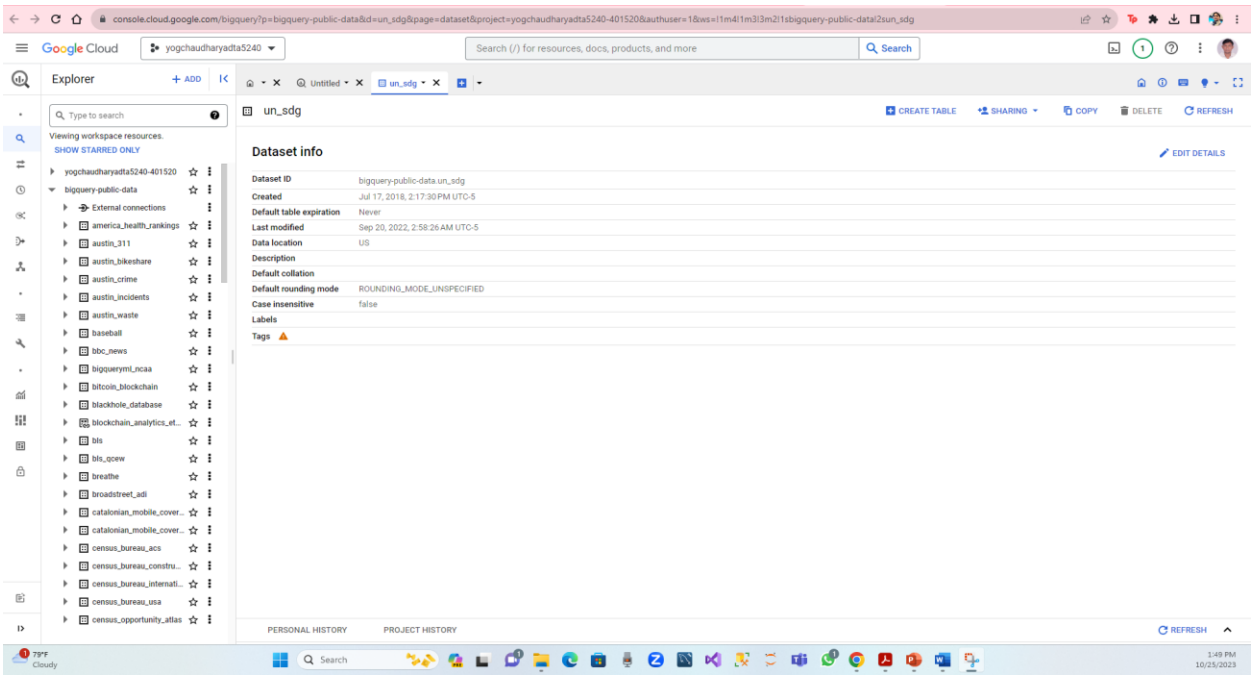


- After clicking on public datasets, I was taken to the marketplace.
- Here I have typed the name of a dataset called Sustainable Development Goals Indicators.

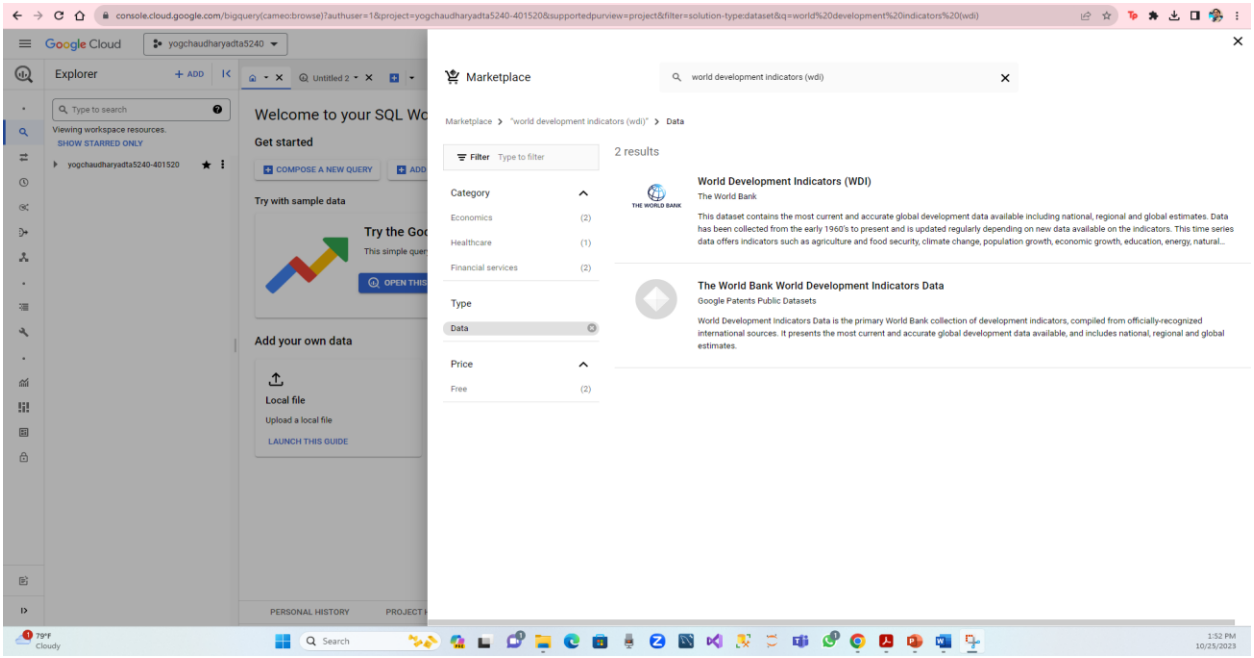


- I clicked on SDG.
- I was taken to the below screenshot.
- I clicked on the view data set.
- By clicking on view data, I have viewed the datasets.

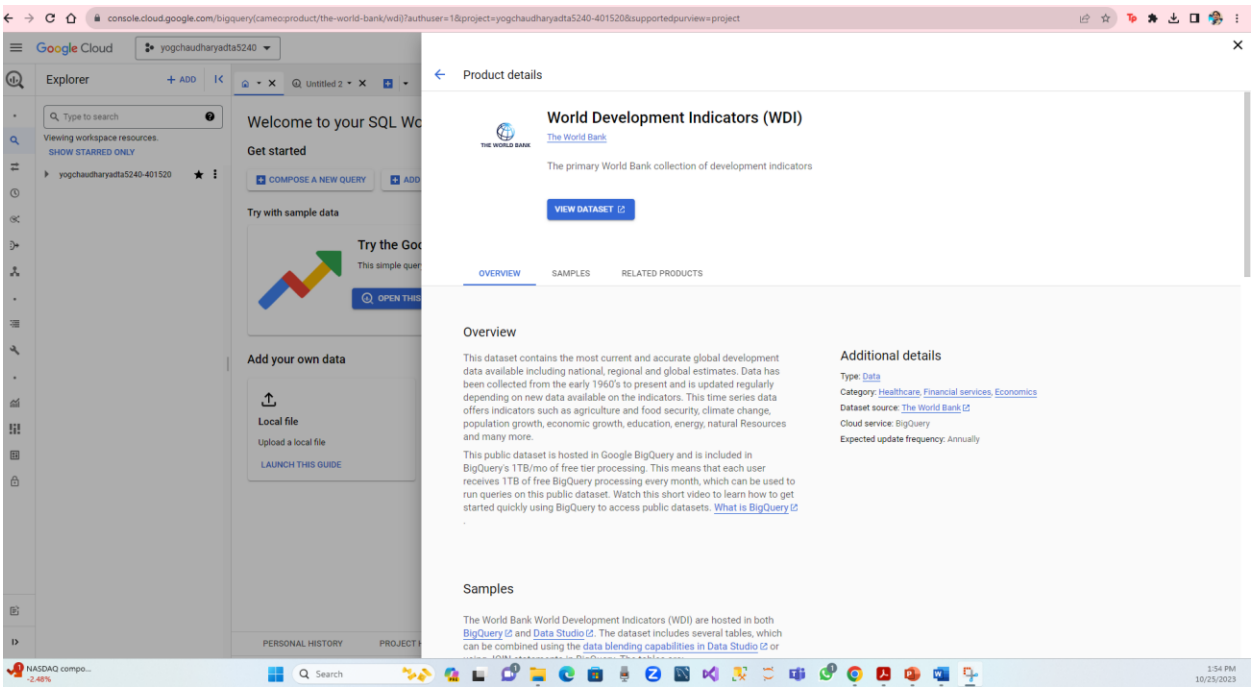


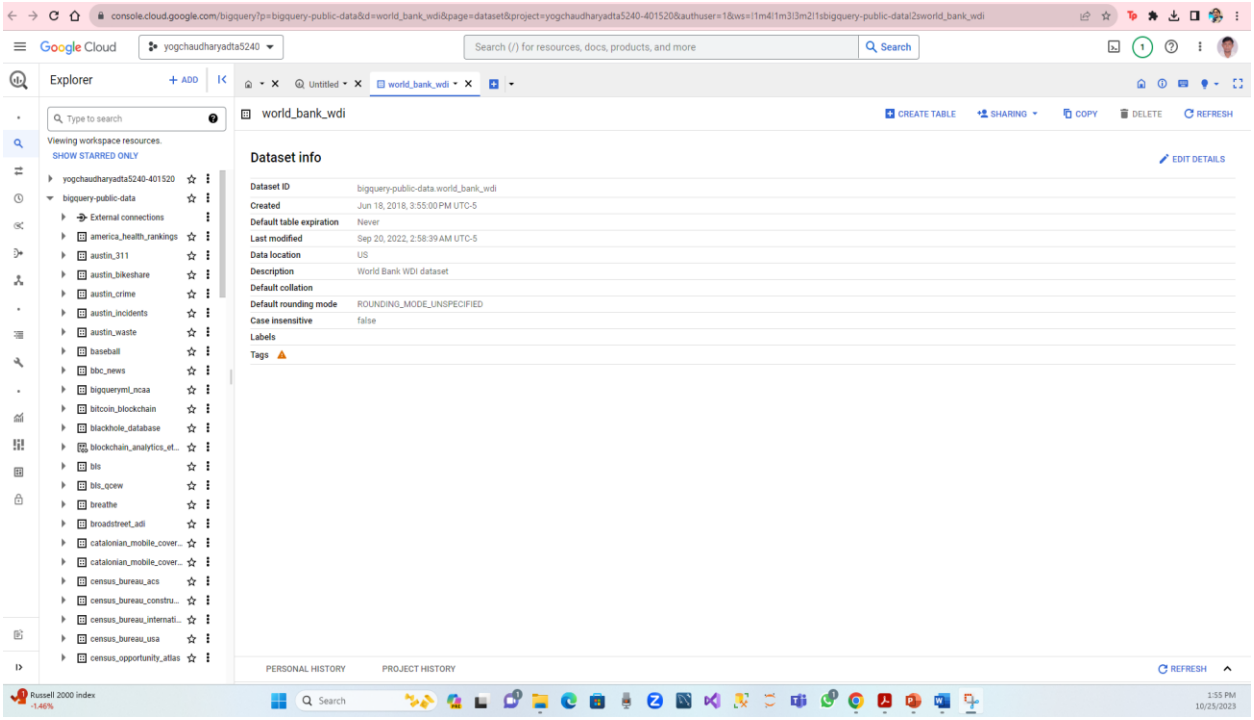


- Again, I have typed the name of a dataset called World Development Indicators.



- After I clicked on WDI, I was taken to the below page.
- Here I clicked on view dataset.
- After clicking on view dataset, I viewed the dataset.





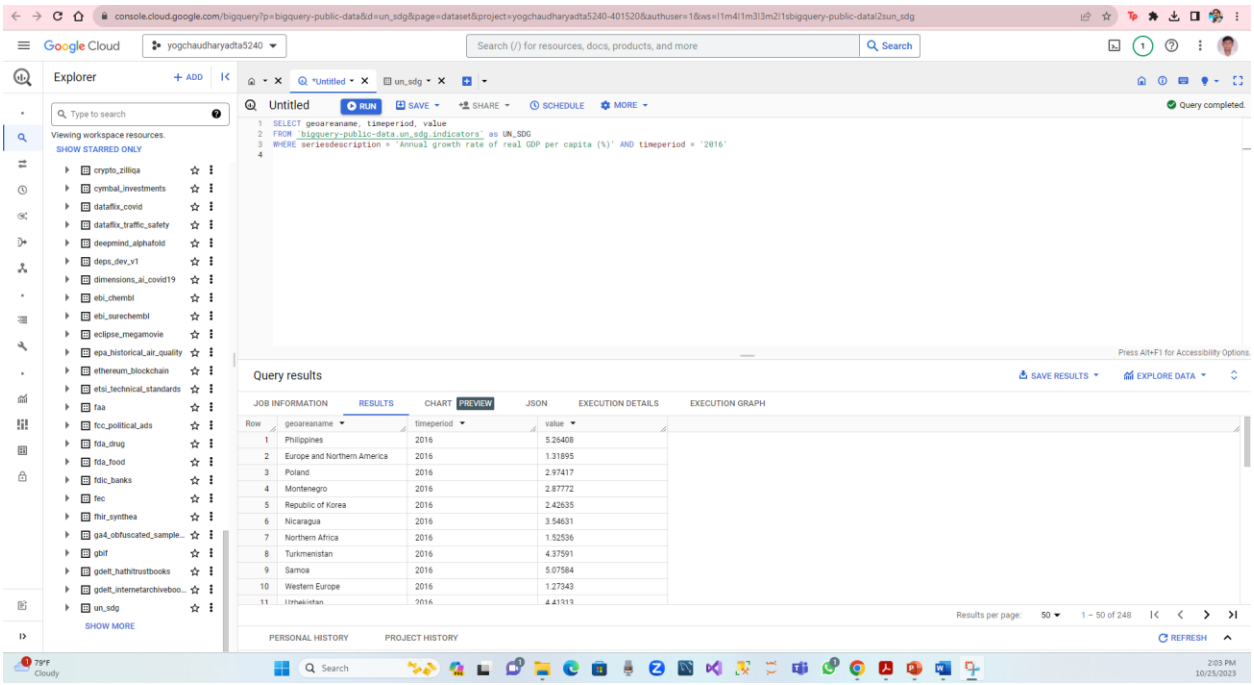
2. **Now I want to ingest from each dataset and create a table with that data.**
   - For this from UN SDG I will use the Annual Growth Rate of Real GDP per capita(%).
   - From WDI I will use population.

❖ **I am creating a table from the SDG using the following SQL commands.**

- For UN_SDG:

  SELECT geoareaname, timeperiod, value
  FROM `bigquery-public-data.un_sdg.indicators` as UN_SDG
  WHERE seriesdescription = 'Annual growth rate of real GDP per capita (%)' AND timeperiod = '2016'

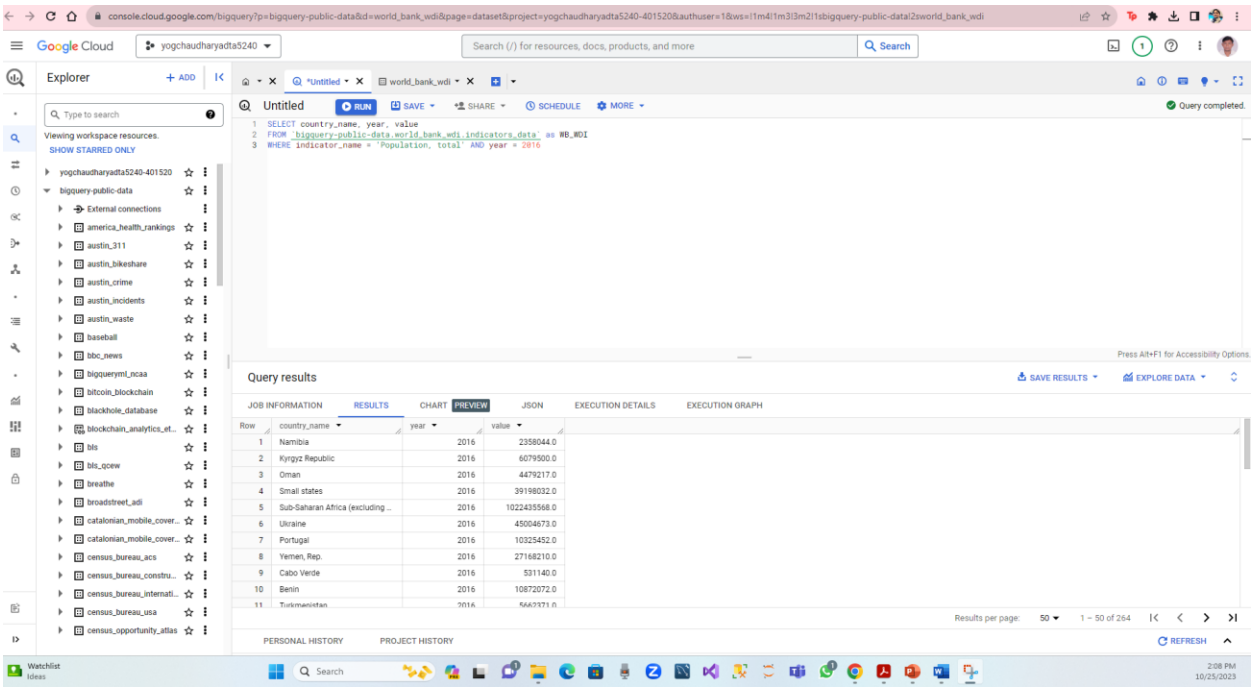
- After giving this command I have viewed the results the page below shows.



❖ I clicked on compose a new query for adding the data from WDI.

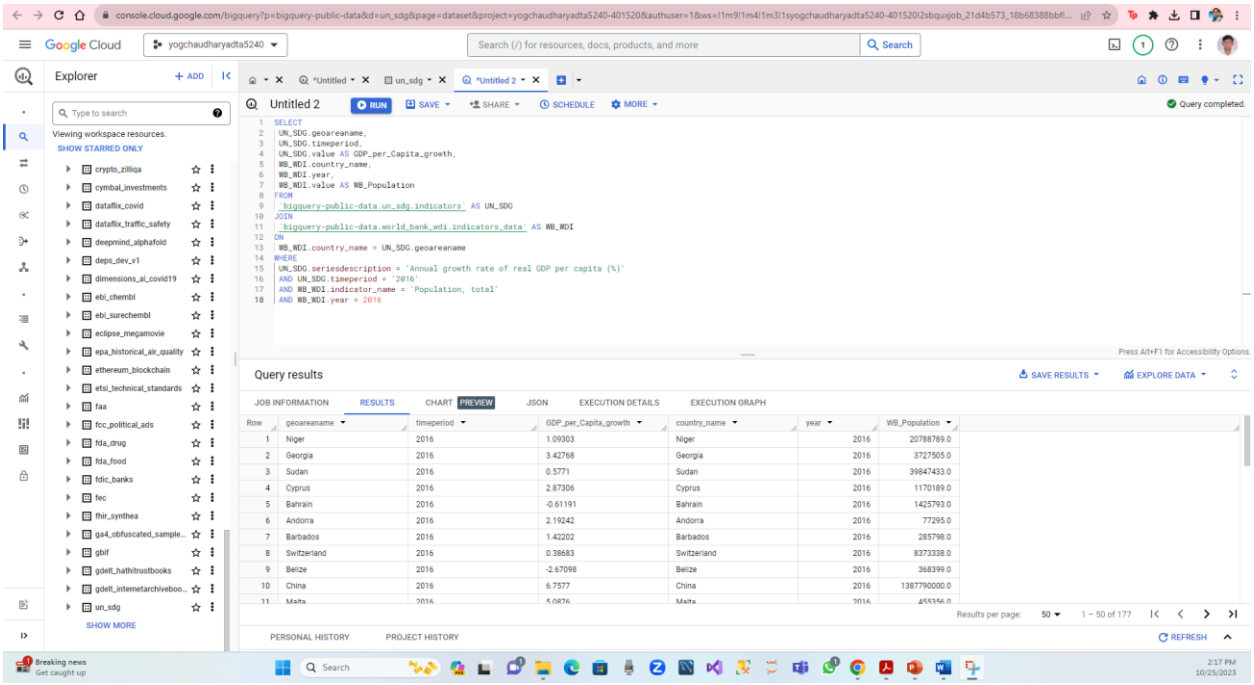
- For WDI:
  SELECT country_name, year, value
  FROM `bigquery-public-data.world_bank_wdi.indicators_data` as WB_WDI
  WHERE indicator_name = 'Population, total' AND year = 2016

- After giving this, I have viewed the results the page below shows.

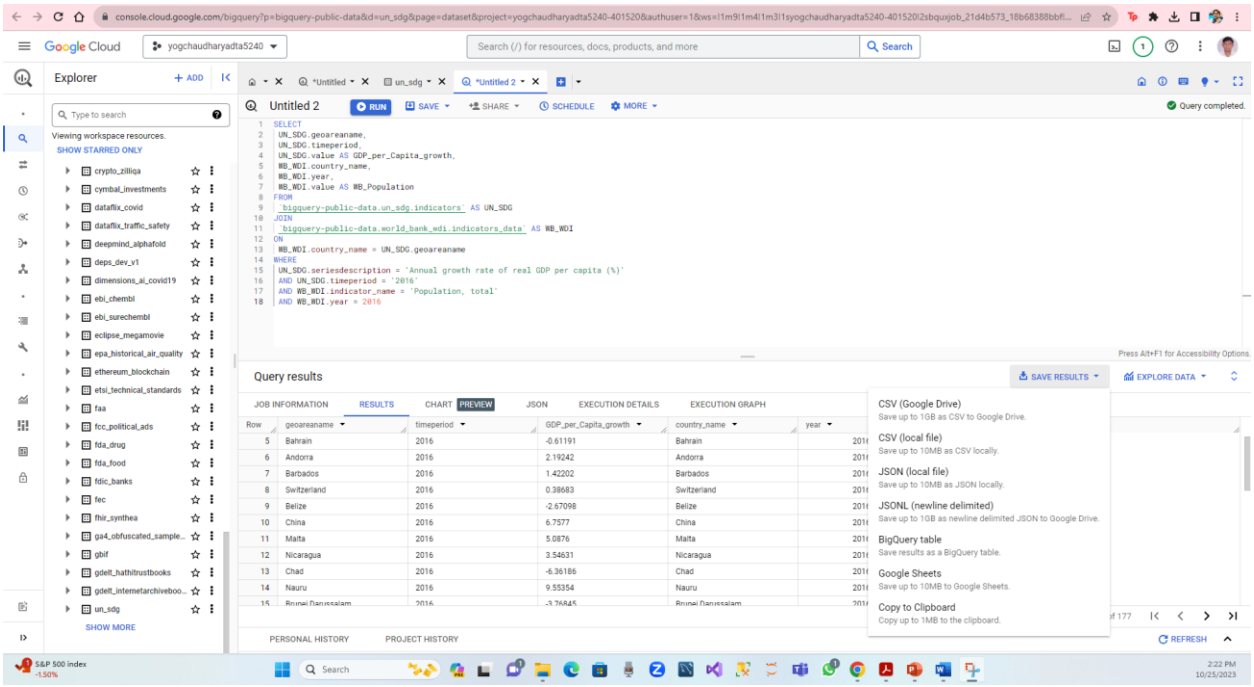


3. Now I am going to join the two datasets together based on the previous selection by using this query.
   - I clicked on “Compose New Query” to add the data from WDI.

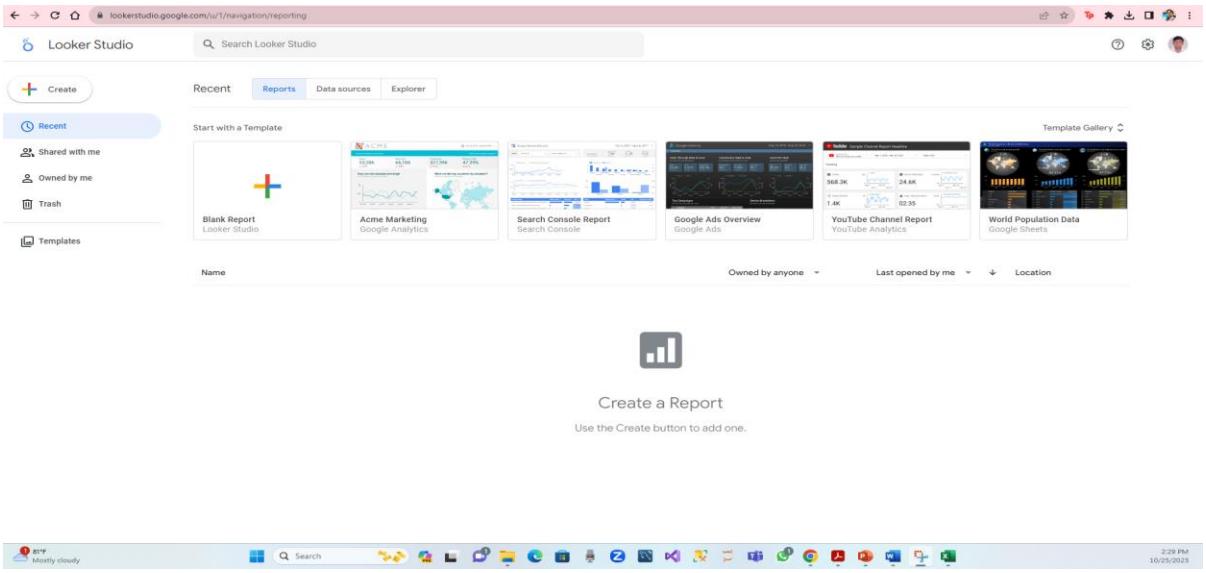
```
SELECT
 UN_SDG.geoareaname,
 UN_SDG.timeperiod,
 UN_SDG.value AS GDP_per_Capita_growth,
 WB_WDI.country_name,
 WB_WDI.year,
 WB_WDI.value AS WB_Population
FROM
 `bigquery-public-data.un_sdg.indicators` AS UN_SDG
JOIN
 `bigquery-public-data.world_bank_wdi.indicators_data` AS WB_WDI
ON
 WB_WDI.country_name = UN_SDG.geoareaname
WHERE
 UN_SDG.seriesdescription = 'Annual growth rate of real GDP per capita (%)'
 AND UN_SDG.timeperiod = '2016'
 AND WB_WDI.indicator_name = 'Population, total'
 AND WB_WDI.year = 2016
```



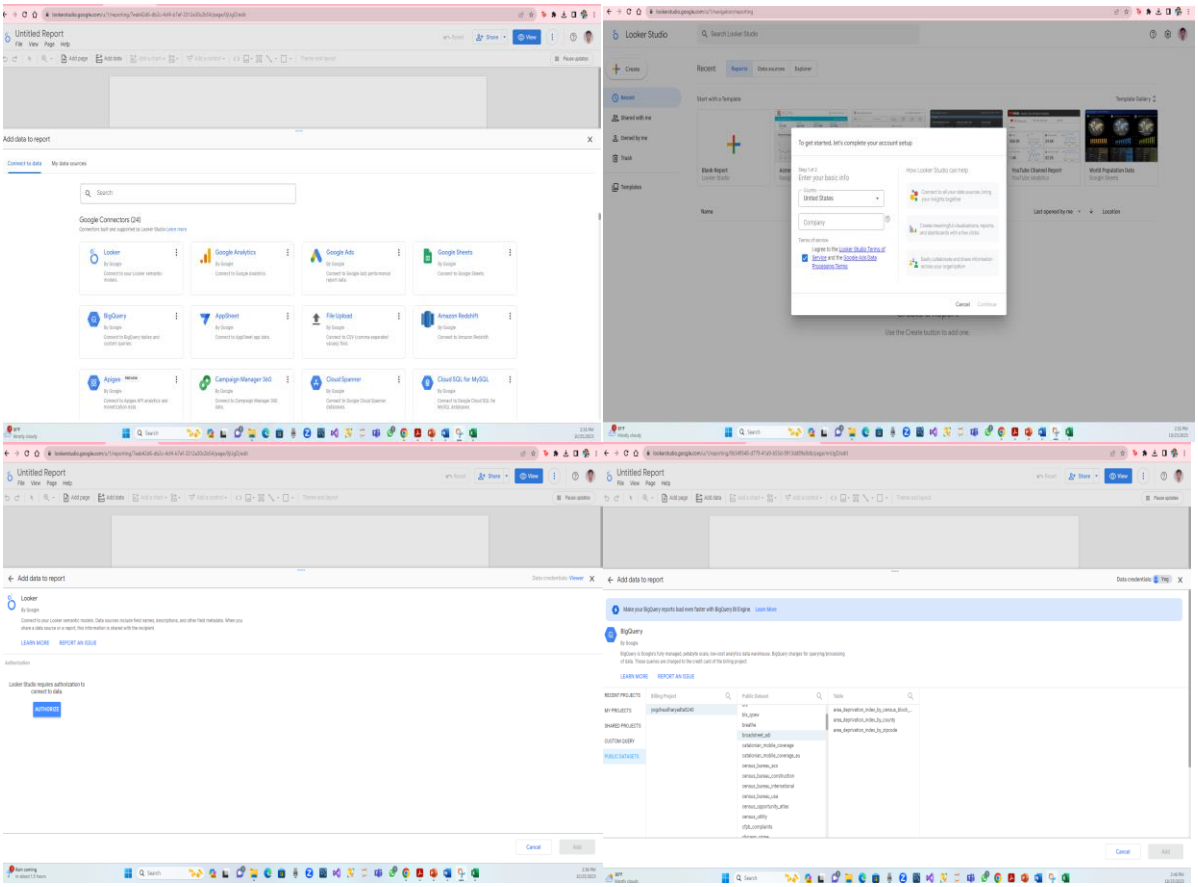
4. Then to make a visualization of data I clicked on save results CSV (local file).



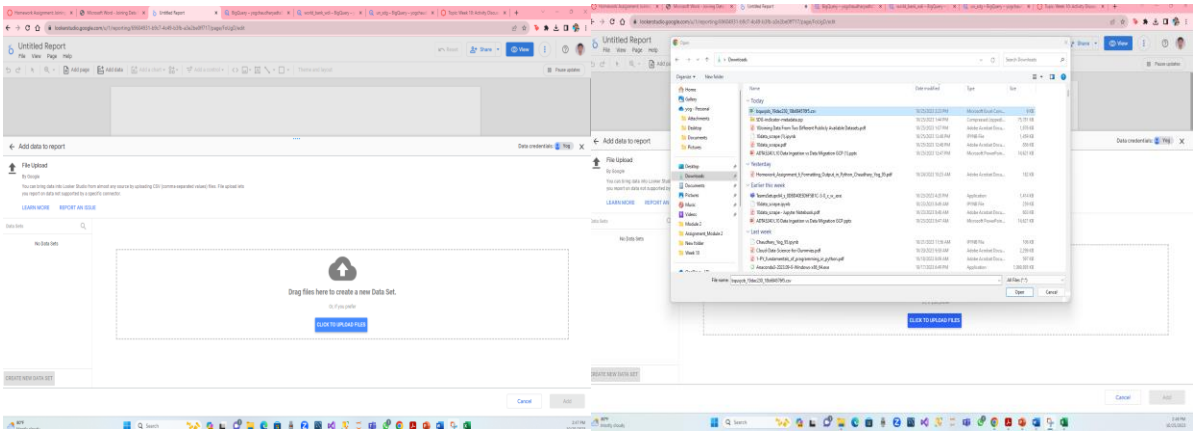
- After that I clicked on explore data in that I clicked on explore with looker studio.

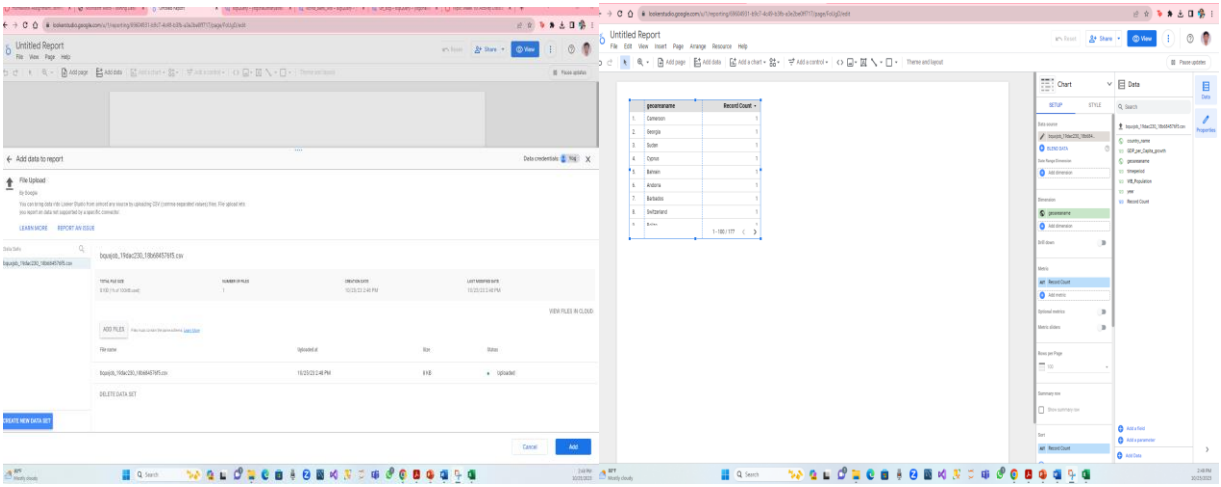


- After that I was taken to the below page.



- Hare screenshot upload file CSV





- **Finally, successfully visualize the data using Google Data Studio.**