

Yog Chaudhary

11727095

ADTA 5240 Week (Harvesting, Storing, And Retrieving Data)

Professor: Dr. Zeynep Orhan

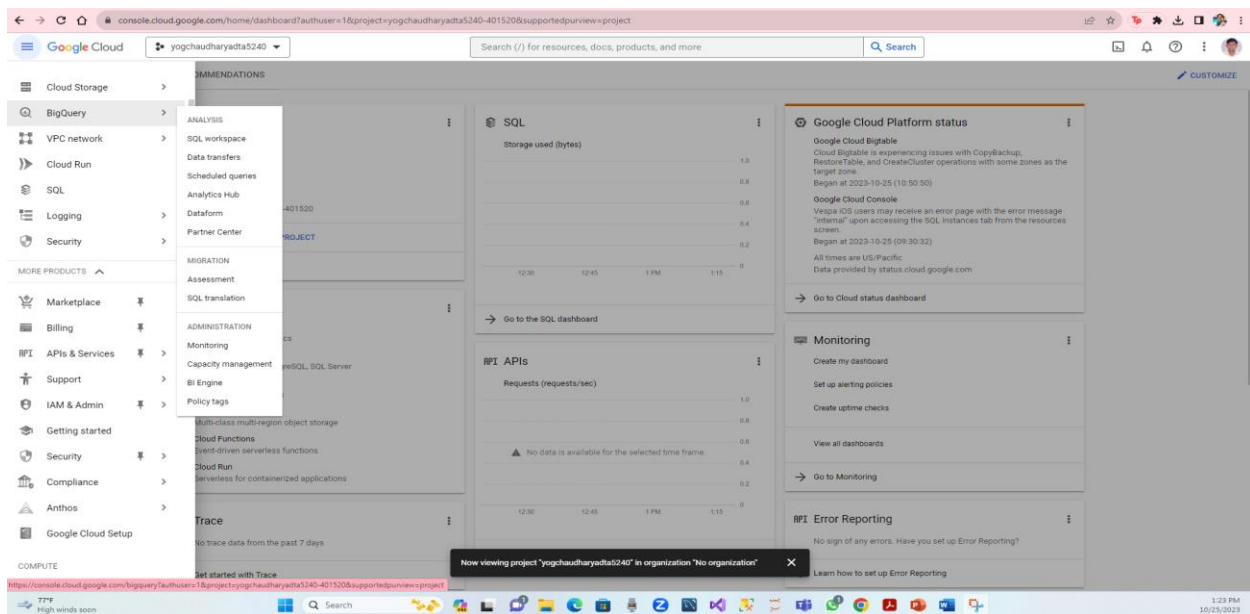
University Of North Texas

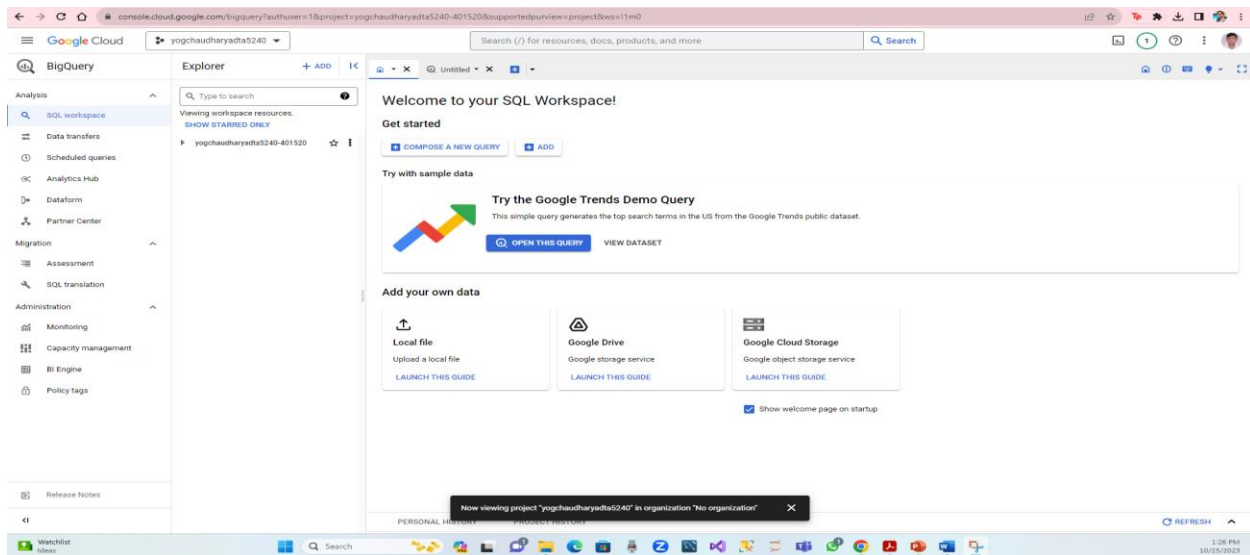
Oct 25, 2023

❖ Joining Data from Two Different Publicly Available Datasets

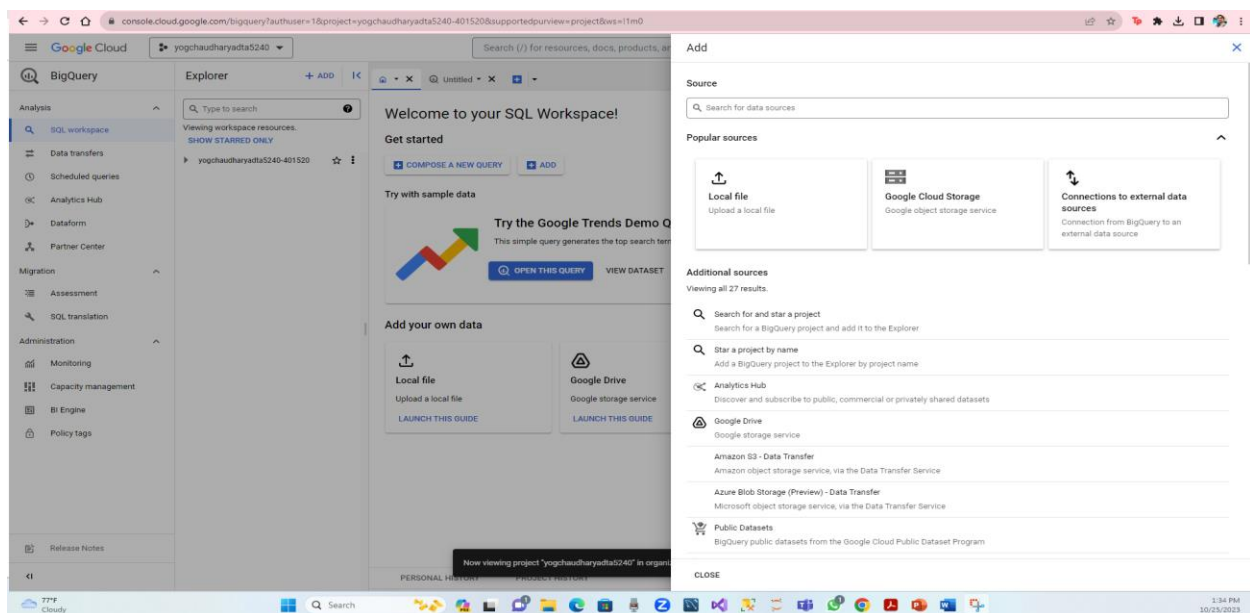
I am going to use 2 public datasets for this. These are Sustainable Development Goals (UN SDG) indicators and World Development Indicators (WDI). These 2 datasets are available through Google Console.

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.

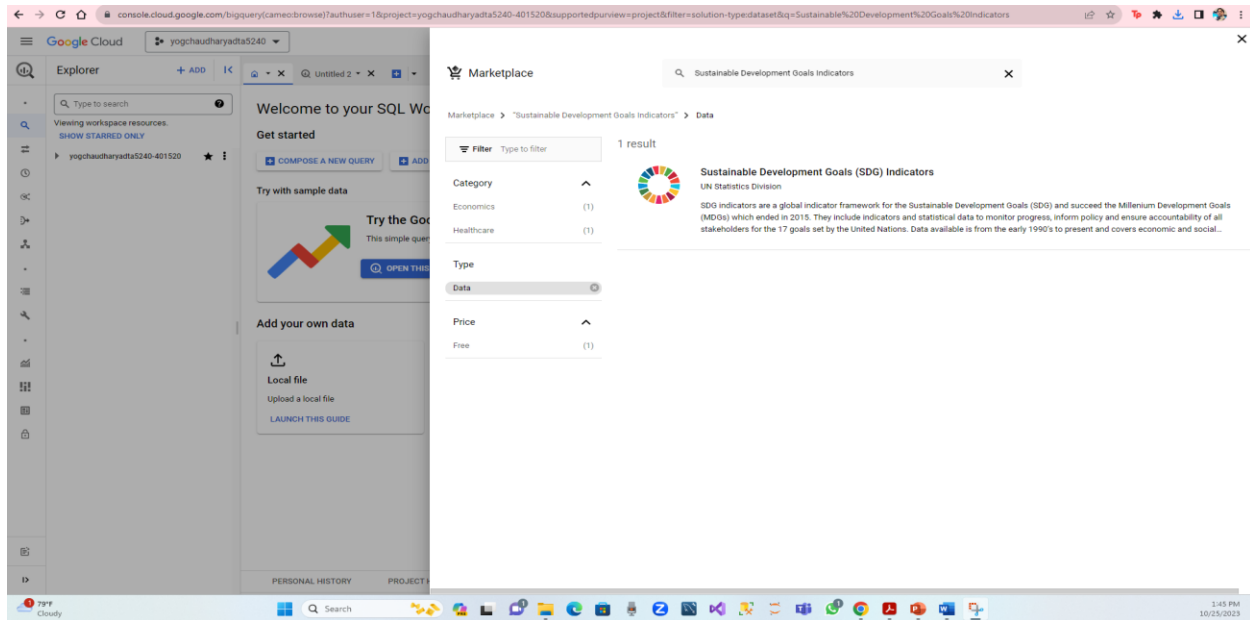




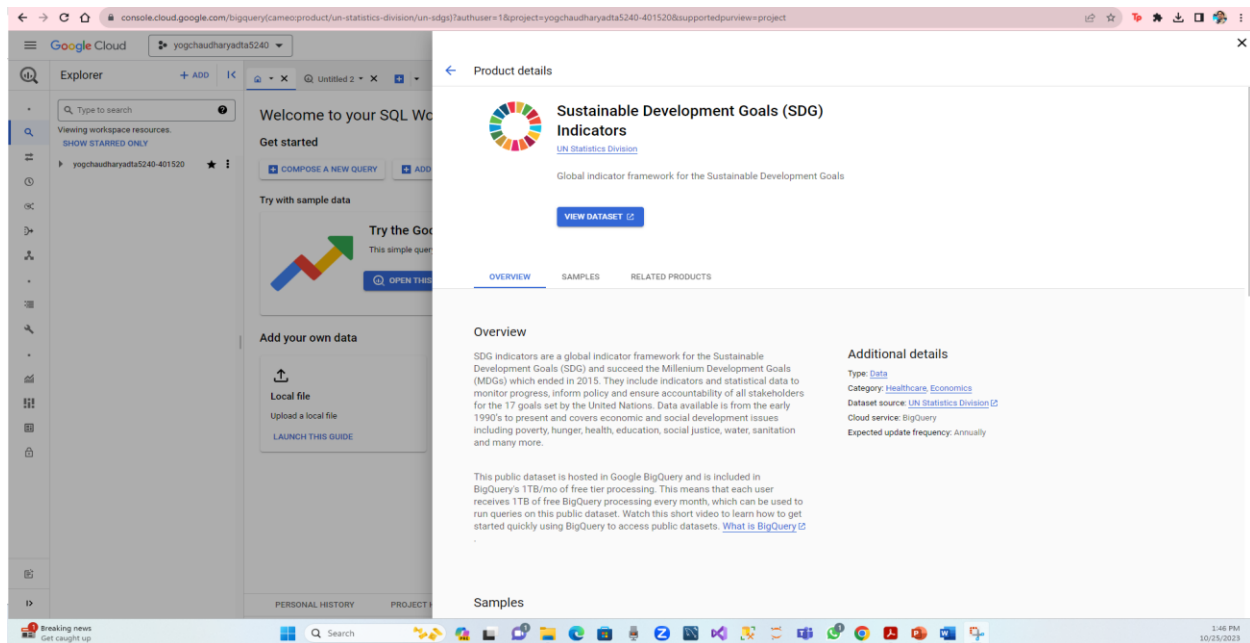
- I clicked on “ADD DATA”.
- Then I was directed to the page that below shows.
- Here I chose public datasets.

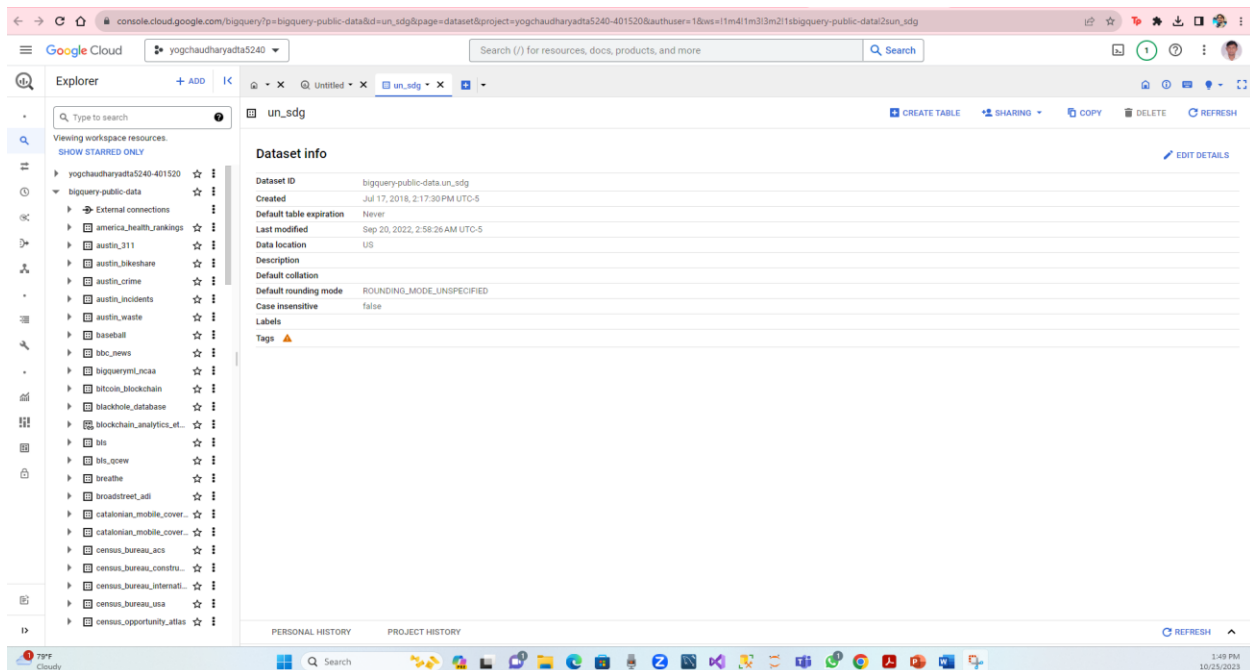


- After clicking on public datasets, I was taken to the marketplace.
- Here I have typed the name of a dataset called Sustainable Development Goals Indicators.

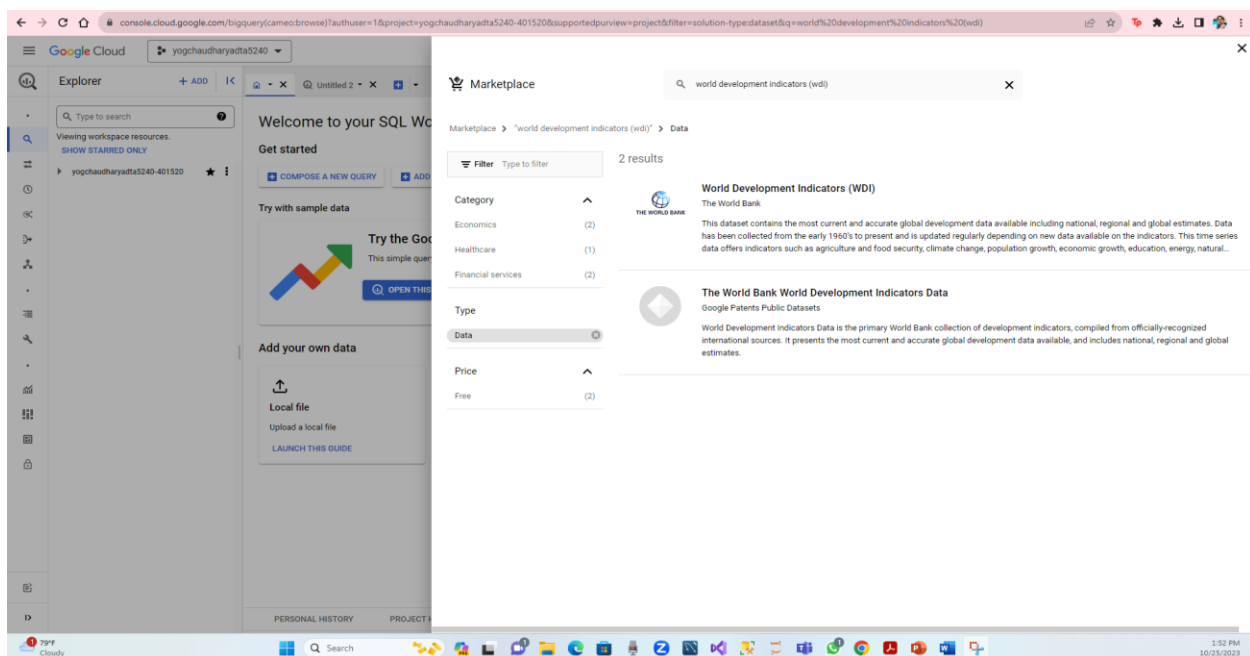


- I clicked on SDG.
- I was taken to the below screenshot.
- I clicked on the view data set.
- By clicking on view data, I have viewed the datasets.





- 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.

console.cloud.google.com/bigquery/comeo/product/the-world-bank/wdi?authuser=1&project=yogchaudharyadta5240-401520&supportedview=project

Google Cloud

Explorer

Viewing workspace resources.

SHOW STARRED ONLY

yogchaudharyadta5240-401520

Product details

World Development Indicators (WDI)

The World Bank

The primary World Bank collection of development indicators

VIEW DATASET (2)

OVERVIEW SAMPLES RELATED PRODUCTS

Overview

This dataset contains the most current and accurate global development data available including national, regional and global estimates. Data has been collected from the early 1960's to present and is updated regularly depending on new data available on the indicators. This time series data offers indicators such as agriculture and food security, climate change, population growth, economic growth, education, energy, natural Resources and many more.

This public dataset is hosted in Google BigQuery and is included in BigQuery's 1TB/mo of free tier processing. This means that each user receives 1TB of free BigQuery processing every month, which can be used to run queries on this public dataset. Watch this short video to learn how to get started quickly using BigQuery to access public datasets. [What is BigQuery](#)

Additional details

Type: [Data](#)
 Category: [Healthcare](#) [Financial services](#) [Economics](#)
 Dataset source: [The World Bank](#)
 Cloud service: [BigQuery](#)
 Expected update frequency: [Annually](#)

Samples

The World Bank World Development Indicators (WDI) are hosted in both [BigQuery](#) and [Data Studio](#). The dataset includes several tables, which can be combined using the [data blending capabilities in Data Studio](#) or [Google Cloud Data Studio](#).

console.cloud.google.com/bigquery/public-data/world_bank_wdi?page=dataset&project=yogchaudharyadta5240-401520&authuser=1&ws=11m41m31m211sbigquery-public-data:2sworld_bank_wdi

Google Cloud

Search (/) for resources, docs, products, and more

Explorer

Viewing workspace resources.

SHOW STARRED ONLY

yogchaudharyadta5240-401520

bigquery-public-data

External connections

america_health_rankings

austin_311

austin_bikeshare

austin_crime

austin_incidents

austin_waste

baseball

bic_news

bigqueryml_ncaa

bitcoin_blockchain

blackhole_database

blockchain_analytics_et

bis

bis_grow

breathe

broadstreet_ad

catalonian_mobile_cover

catalonian_mobile_cover

census_bureau_acs

census_bureau_constru

census_bureau_internati

census_bureau_usa

census_opportunity_atlas

world_bank_wdi

CREATE TABLE SHARING COPY DELETE REFRESH

Dataset info

EDIT DETAILS

Dataset ID	bigquery-public-data.world_bank_wdi
Created	Jun 18, 2016, 3:55:00 PM UTC-5
Default table expiration	Never
Last modified	Sep 20, 2022, 2:58:39 AM UTC-5
Data location	US
Description	World Bank WDI dataset
Default collation	
Default rounding mode	ROUNDING_MODE_UNSPECIFIED
Case insensitive	false
Labels	
Tags	

PERSONAL HISTORY PROJECT HISTORY

REFRESH

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.

The screenshot shows the Google Cloud BigQuery console. The query editor contains the following SQL code:

```
1 SELECT geoareaname, timeperiod, value
2 FROM `bigquery-public-data.un_sdg.indicators` as UN_SDG
3 WHERE seriesdescription = 'Annual growth rate of real GDP per capita (%)' AND timeperiod = '2016'
4
```

The query results are displayed in a table with the following data:

Row	geoareaname	timeperiod	value
1	Philippines	2016	5.26408
2	Europe and Northern America	2016	1.31895
3	Poland	2016	2.97417
4	Montenegro	2016	2.87772
5	Republic of Korea	2016	2.42635
6	Nicaragua	2016	3.54631
7	Northern Africa	2016	1.52536
8	Turkmenistan	2016	4.37991
9	Samoa	2016	5.07684
10	Western Europe	2016	1.27343
11	Uruguay	2016	2.41912

❖ 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.

The screenshot shows the Google Cloud console interface. On the left is the 'Explorer' pane with a search bar and a list of workspace resources. The main area displays a BigQuery query titled 'Untitled' with the following SQL code:

```
1 SELECT country_name, year, value
2 FROM `bigquery-public-data.world_bank_wdi.indicators_data` AS WB_WDI
3 WHERE indicator_name = 'Population, total' AND year = 2016
```

Below the query editor, the 'Query results' section is visible, showing a table with 11 rows and 3 columns: country_name, year, and value. The results are as follows:

Row	country_name	year	value
1	Namibia	2016	2358044.0
2	Kyrgyz Republic	2016	6079500.0
3	Oman	2016	4479217.0
4	Small states	2016	39198032.0
5	Sub-Saharan Africa (excluding ...)	2016	1022435568.0
6	Ukraine	2016	45004673.0
7	Portugal	2016	10325452.0
8	Yemen, Rep.	2016	27168210.0
9	Cabo Verde	2016	531140.0
10	Benin	2016	10872072.0
11	Turkmenistan	2016	5667971.0

At the bottom of the console, the Windows taskbar is visible with various application icons and the system clock showing 2:08 PM on 10/25/2023.

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
```

The screenshot shows the Google Cloud BigQuery console. The SQL query in the editor is as follows:

```

1 SELECT
2   UN_SDG.geoname,
3   UN_SDG.timeperiod,
4   UN_SDG.value AS GDP_per_Capita_growth,
5   WB_MDI.country_name,
6   WB_MDI.year,
7   WB_MDI.value AS WB_Population
8 FROM
9   `bigquery-public-data.un_sdg.indicators` AS UN_SDG
10 JOIN
11   `bigquery-public-data.world_bank_mdi.indicators_data` AS WB_MDI
12 ON
13   WB_MDI.country_name = UN_SDG.geoname
14 WHERE
15   UN_SDG.seriesdescription = 'Annual growth rate of real GDP per capita (%)'
16   AND UN_SDG.timeperiod = '2016'
17   AND WB_MDI.indicator_name = 'Population, total'
18   AND WB_MDI.year = 2016

```

The query results are displayed in a table with the following columns: geoname, timeperiod, GDP_per_Capita_growth, country_name, year, and WB_Population. The results show data for 11 countries in 2016.

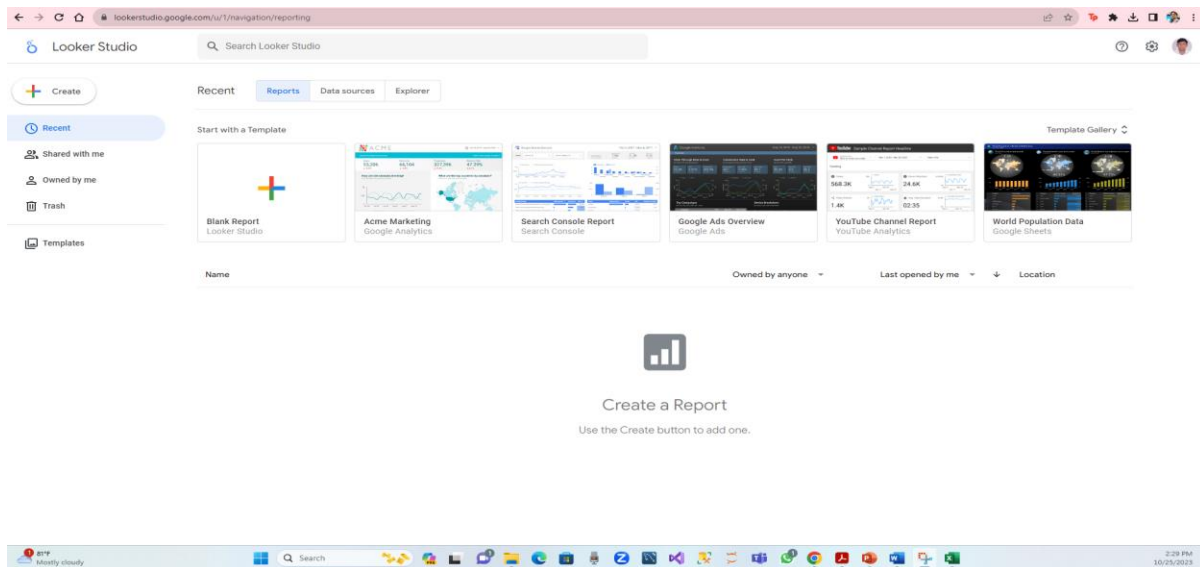
Row	geoname	timeperiod	GDP_per_Capita_growth	country_name	year	WB_Population
1	Niger	2016	1.09303	Niger	2016	20786789.0
2	Georgia	2016	3.42768	Georgia	2016	3727505.0
3	Sudan	2016	0.5771	Sudan	2016	39847433.0
4	Cyprus	2016	2.87306	Cyprus	2016	1170189.0
5	Bahrain	2016	-0.61191	Bahrain	2016	1425793.0
6	Andorra	2016	2.19242	Andorra	2016	77295.0
7	Barbados	2016	1.42202	Barbados	2016	285798.0
8	Switzerland	2016	0.38683	Switzerland	2016	8373338.0
9	Belize	2016	-2.67098	Belize	2016	368399.0
10	China	2016	6.7577	China	2016	1387790000.0
11	Malta	2016	5.0876	Malta	2016	454366.0

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

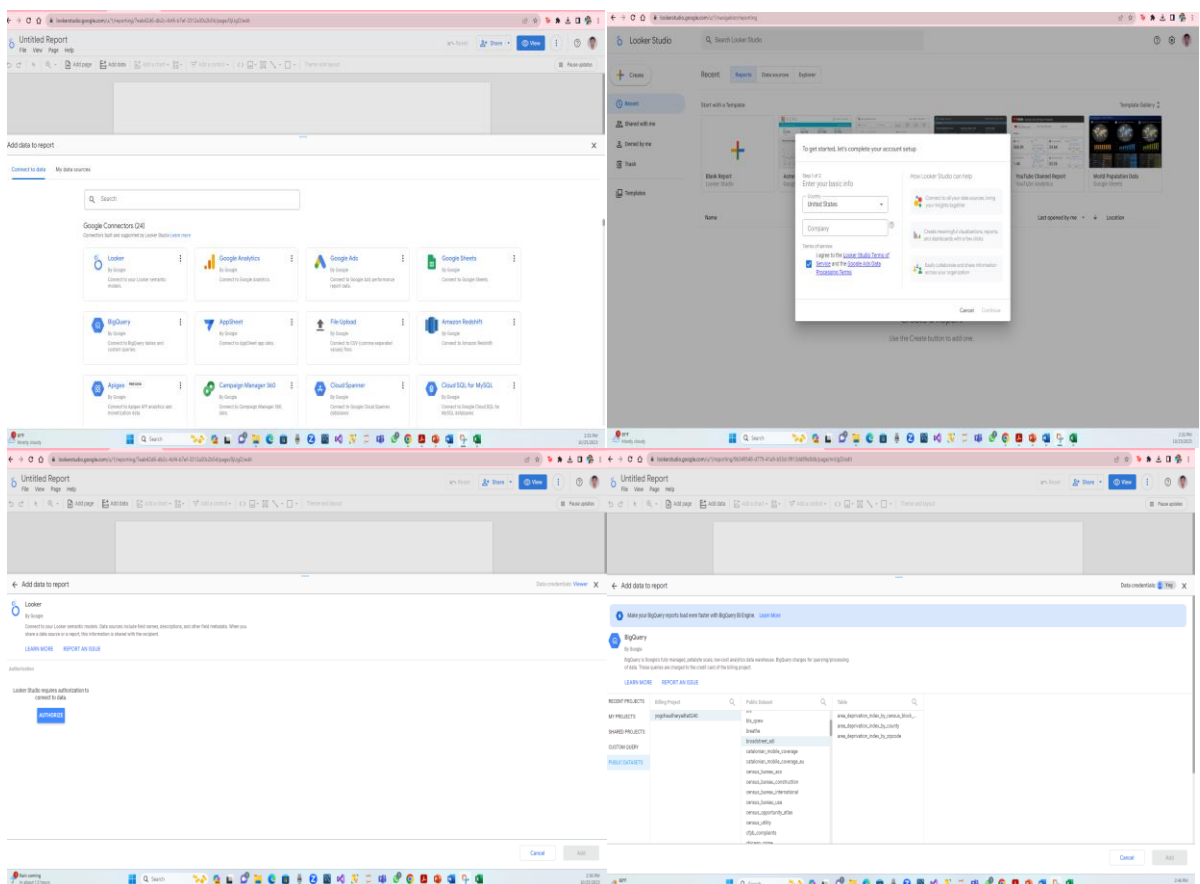
The screenshot shows the Google Cloud BigQuery console with the 'Query results' dropdown menu open. The menu options are:

- CSV (Google Drive): Save up to 1GB as CSV to Google Drive.
- CSV (local file): Save up to 10MB as CSV locally.
- JSON (local file): Save up to 10MB as JSON locally.
- JSONL (newline delimited): Save up to 1GB as newline delimited JSON to Google Drive.
- BigQuery table: Save results as a BigQuery table.
- Google Sheets: Save up to 10MB to Google Sheets.
- Copy to Clipboard: Copy up to 1MB to the clipboard.

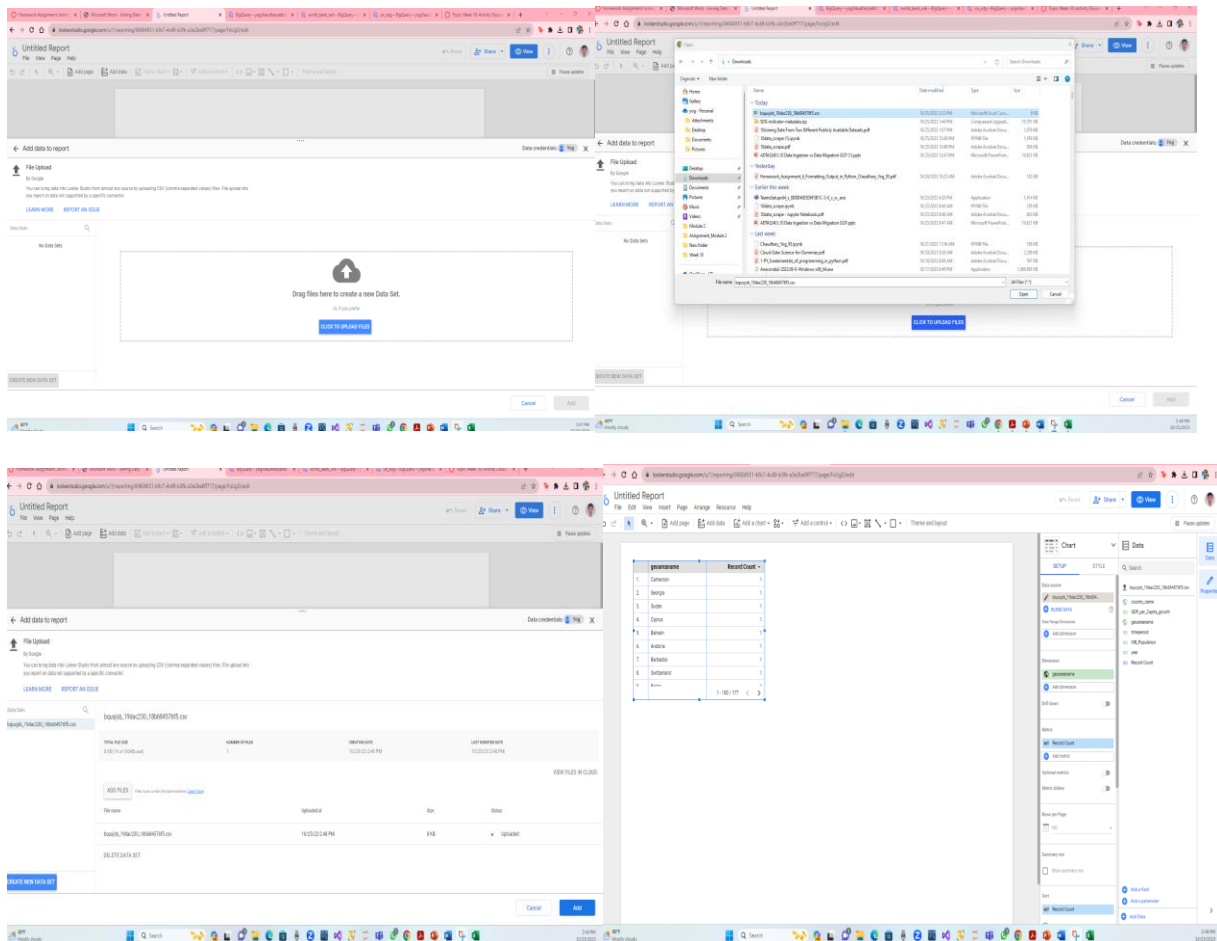
- After that I clicked on explore data in that I clicked on explore with looker studio.



- After that I was taken to the below page.



- Here screenshot upload file CSV



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