**Homework 9: Joining Data from Two Different Public Datasets**

**Objective: Join two public datasets using GCP**

**Prerequisites:**

* Access to google cloud platform with a Google/Gmail id.
* Project in GCP.
* BigQuery API should be enabled.

A screenshot of a web page

Description automatically generated

**Step 1: Getting public dataset**

* Go to BigQuery Studio and click on “+” button.

A screenshot of a computer

Description automatically generated

* In the search bar type public dataset.

A screenshot of a computer

Description automatically generated

* Click on the Public datasets from the results.
* It takes the user to the Marketplace.
* Type “Sustainable Development Goals Indicators” in the Marktplace search bar.
* The below screen gets displayed.

A screenshot of a computer

Description automatically generated

* Click on “View Data Set”.
* The data set gets open in a new tab.

A screenshot of a computer

Description automatically generated

* Now type “World Development Indicators” in the Marketplace search bar.

A screenshot of a computer

Description automatically generated

* Click on WDI.
* I have starred the dataset for ease of identifying. This step is not mandatory.

A screenshot of a computer

Description automatically generated

**Step 2 : Query data from the datasets**

* We will be using data from these two datasets and join them.
* I am querying the below three fields from un\_sdg dataset for the year 2015 based on Annual Growth Rate of Real GDP per capita (%)
  + Geoareaname
  + Timeperiod
  + Value
* Query:

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 = '2015'

A screenshot of a computer

Description automatically generated

* Now select “world\_bank\_wdi” from the bigquery public data and expand the dropdown.
* Select “Indicators\_data” and hover on the three dots to select Query option.

A screenshot of a computer

Description automatically generated

* I am querying the below three fields from the dataset for the year 2015 and indicator as population.
  + country\_name
  + year
  + value
* Query:

SELECT country\_name, year, value

FROM `bigquery-public-data.world\_bank\_wdi.indicators\_data` as WB\_WDI

WHERE indicator\_name = "Population, total"

AND year = 2015

A screenshot of a computer

Description automatically generated

**Step 3: Joining the two datasets**

* Open a new query tab and use the below query to join the two datasets.
* Query:

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 = '2015' AND WB\_WDI.indicator\_name = 'Population, total'

AND WB\_WDI.year = 2015

* The below screenshot displays the output for the query.

A screenshot of a computer

Description automatically generated

* Click on Save Results to export the data into csv and keep it ready for the next step.

**Step 4: Looker Studio**

* Click on the “Explore Data” option from the right corner and select “Explore with Looker Studio”.

**A screenshot of a computer

Description automatically generated**

* Go to the below url to get into the looker studio.
  + <https://lookerstudio.google.com/u/0/reporting>
* Click on Add Data to and select upload file.

**A screenshot of a computer

Description automatically generated**

* Now, select the file download/saved in the previous step.
* The data gets added to the data section as shown below.

A screenshot of a computer

Description automatically generated

* Select the required fields and charts from the charts section.

A screenshot of a computer

Description automatically generated

* I have selected the Table with bars and combo map for the below visualization.

**A screenshot of a computer

Description automatically generated**

**Conclusion:** We could successfully query two datasets and joined them. Downloaded the joined dataset and used the same to create a visualization using Looker Studio.