

Title: **Data Engineering Assignment**

Date: **04/12/2021**

Problem Statement:

Write a Python script to retrieve publicly available data and perform analysis on that data using SQL. Publicly available data is available from the World Bank. One dataset which contains information about countries is available through World Bank API¹ and its documentation provides all the necessary details on how to use it. Another dataset which includes Gross Domestic Product (GDP) data in CSV format is available for download² from World Bank Data Catalog.

Approach:

Python API available and can be called on python script and the dataset can be modelled and data can be presented accordingly.

PostgreSQL DB is used to store and model the data.

Tools:

Jupyter Lab 0.35.4 / Jupyter Notebook 5.7.8

Postgre SQL V6

Steps Followed:

1. Write Python script to retrieve countries' data from a World Bank Country API as well GDP data from World Bank Data Catalog. Then upload the python script on the Github repository and prepare the readme notes and shareable link.
2. Collect the GDP country group and GDP country values data and upload on the PostgreSQL server Database and form table structures.
3. Perform iteration of the data as applicable.
4. Write the required SQL queries as required for the data modelling.
5. Upload the queries as a text or.sql file in the Github repository.