**DBMS (Lab)**

**LAB-01**



**Submitted to**:

Ma’am Darakhshan Abdul Ghaffar

**Submitted by:**

Tayyaba Asif

2019-CE-5

**Date:** 23rd January,2022.

**Question No: 01**

Write your understanding related to BigQuery and GitHub dataset

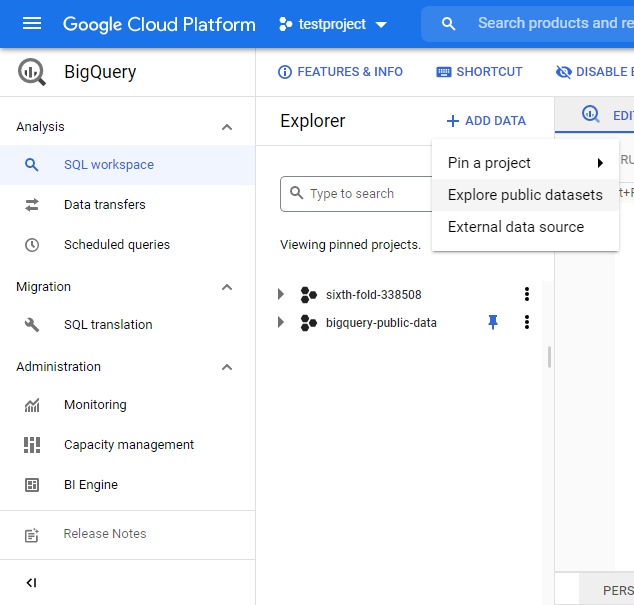
**Answer:**

**BigQuery dataset:**

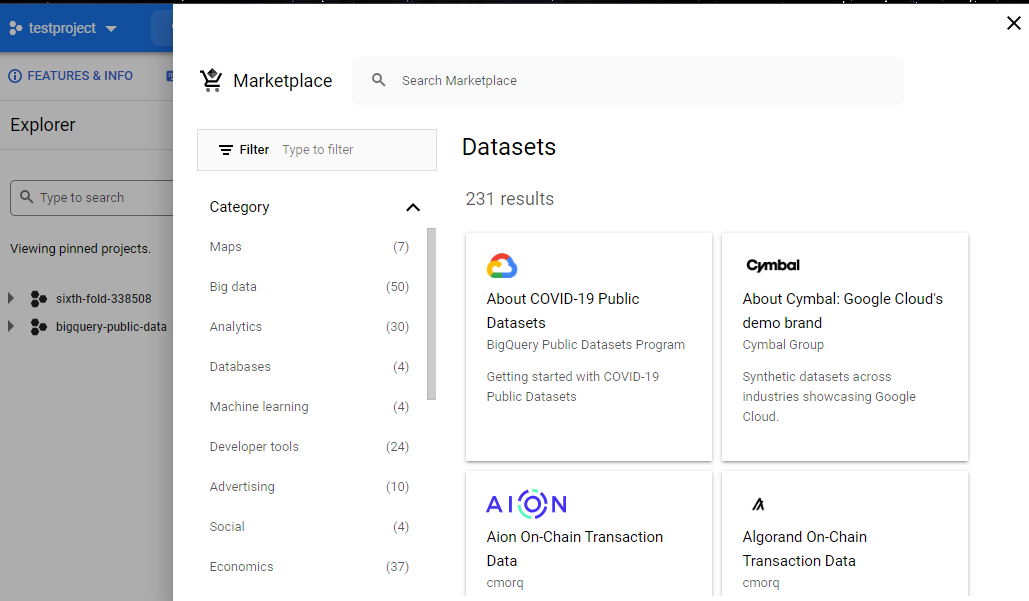
BigQuery is a cloud-based web service for processing very large datasets. It was designed for analyzing the data in form of billions or trillions of rows. This runs on google cloud platform and google cloud storage. Users can easily manage their data in form of objects like tables, views and user defined functions. Users can import data from google storage in different formats like CSV, Parquet, Avro or JSON. Queries are expressed in standard SQL and results are returned in JSON. BigQuery provides access control where users can share datasets with individuals, groups or throughout the world. BigQuery also provides facility to create and execute ML models using SQL Queries. BigQuery provides datasets that can be used in different applications. These datasets are available on cloud storage. BigQuery public datasets can be accessed from Google Cloud Public Dataset Program where datasets related to different topics are available. To explore datasets following steps can be followed:

1- Navigate to <https://console.cloud.google.com/bigquery>

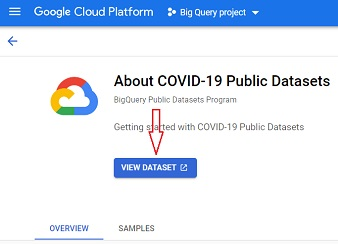
2- Click on **‘+ ADD DATA’** drop down menu and click ‘**EXPLORE PUBLIC DATASETS’**.



Following list will appear:



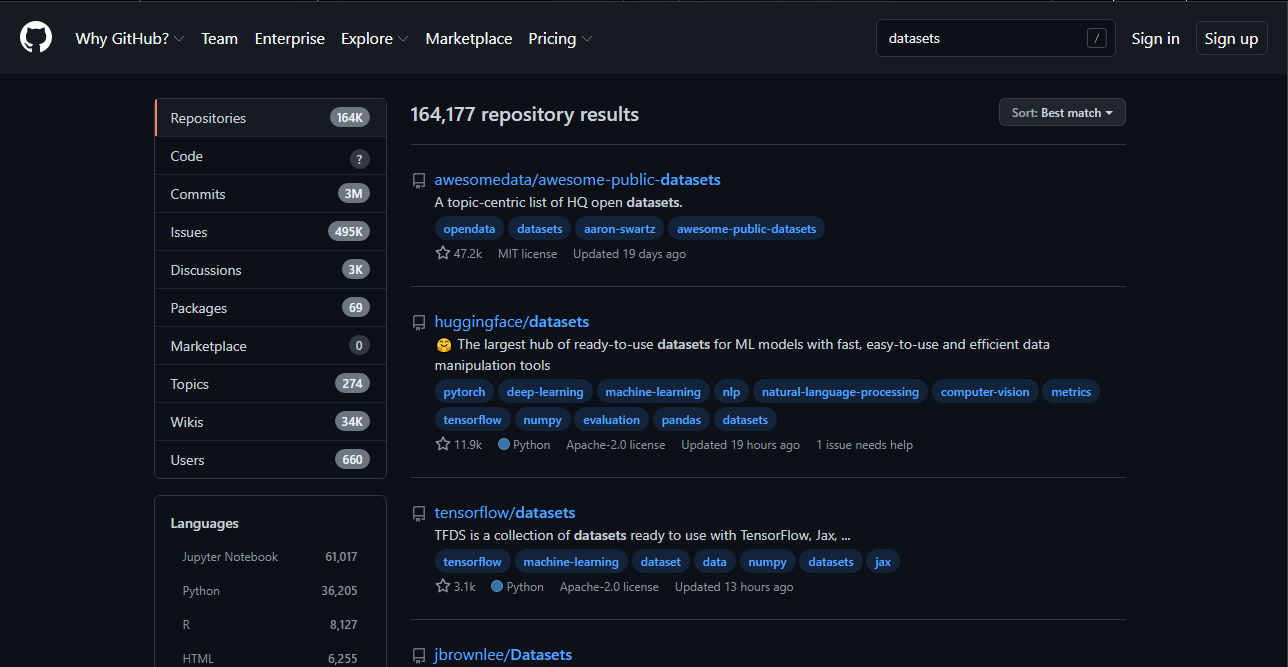
3- Select the dataset category you want to explore and then click **View Dataset** button.



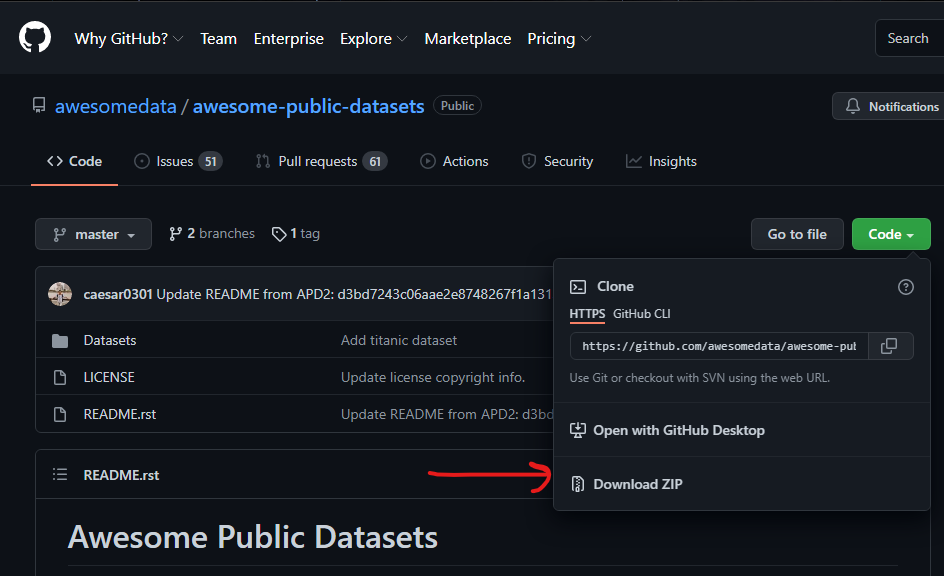
**GitHub Dataset:**

GitHub is a code hosting platform for collaboration with other people. It helps you to work with other people on different projects from anywhere. On GitHub various datasets are available that have been uploaded by different GitHub users. We can easily access these datasets using the following steps:

1- Search for datasets on GitHub. Following page will be displayed:



2- Click on your required dataset and download it.



3- After clicking Download ZIP, a ZIP folder will be downloaded.

4- Extract the folder.

5- Use the dataset wherever you want.

GitHub itself doesn’t provide any dataset whereas different users upload their datasets on their accounts that are easy to access for public. Other sites like Kaggle provides huge datasets to its users.