CS 327E Project 4 due Thursday, 03/03.

The goals of this project are to familiarize yourself with BigQuery and practice writing group-by and aggregate queries. The dataset you will use is data from Stanford's School Enrollment project, which collected K-12 enrollment data nationwide during 2020-2021. The project is described here and its data is available for bulk download. Note: the data has already been downloaded for you and is available from Google Cloud Storage.

- Pull down the snippets repo and open the Jupyter notebook named `project4.ipynb`.

- Run through all the cells in the notebook to download the dataset and create the school enrollment tables for the various US states. Please note that not all files will load correctly into BQ due to several formatting issues. This is expected and you can ignore the parsing errors you see during the load jobs (e.g. Could not parse 'No school' as INT64, etc.). You should end up with 17 state tables in the `school_enrollment` dataset in BQ.

- Write 8 aggregate queries over the school enrollment tables:
    - All queries should use a GROUP BY clause, one or more aggregate functions, and an ORDER BY clause.
    - At least 4 queries should use a HAVING clause.
    - At least 4 queries should use a WHERE clause.
    - At least 2 queries should use a UNION ALL or UNION DISTINCT.
    - Precede each query with a Markdown comment that describes its function.

- Create data visualizations in Data Studio:
    - Choose 2 of your most interesting queries from the previous section.
    - Create a new BQ dataset for storing your database views. Name the dataset `views`.
    - Create a view for your two chosen queries and assign the view a descriptive name (e.g. `highschool_enrollments_austin`).
    - Open Data Studio and go through this tutorial to familiarize yourself with the tool.
    - Create a Data Source (using the + Create button) which accesses each view.
    - Create a Data Studio chart that visualizes the data in a compelling way.
    - Add both charts to a single Data Studio report (aka dashboard) and add a descriptive title to each chart.
    - Download your dashboard as a pdf and name it `dashboard-v1.pdf`.

CS 327E Project 4 Rubric
Due Date: 03/03/22

Create 8 aggregate queries that use a GROUP BY clause, one or more aggregate functions, and an ORDER BY clause. At least 4 queries must also use a HAVING clause and a WHERE clause. At least 2 queries must use a UNION ALL or UNION DISTINCT. **-80** queries missing from `project4.ipynb`: **-10** for each query missing a GROUP BY, aggregate function or ORDER BY clause **-5** for each query missing a HAVING clause **-3** for each query missing a WHERE clause **-5** for each query missing a UNION clause **-2** each incorrect comment, or comment too similar to query	80
Create data visualizations in Data Studio. Visualizations should display the results from two BQ views. A dashboard in Data Studio should contain the 2 charts with a relevant title for each one describing the data. **-10** `views` dataset missing from BQ project **-5** for each missing view **-10** `./dashboard-v1.pdf` not found in repository **-5** for each missing chart **-5** for each chart created from a BQ table instead of a BQ view **-2** for each missing title	20
`project4.ipynb` and `dashboard-v1.pdf` pushed to your group's private repo on GitHub. Your project **will not** be graded without this submission.	**Required**

`submission.json` submitted into Canvas. Your project **will not** be graded without this submission. The file should have the following schema:
 	
`{`
`  "commit-id": "your most recent commit ID from Github",`
`  "project-id": "your project ID from GCP"`
`}`
 	
Example:
 	
`{`
`  "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9",`
`  "project-id": "some-project-id"`
`}`
	Required

Total Credit:	**100**