CS 327E Project 4 due Thursday, 09/28.

The goal of this project is to practice writing group-by and aggregate queries in BigQuery (BQ). The dataset you will use is data from Stanford's School Enrollment project, which collected K-12 enrollment data nationwide during the 2020 - 2021 school year. 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 data 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 8 queries should use a GROUP BY clause, one or more aggregate functions, and an ORDER BY clause.
    - At least 4 queries should also use a HAVING clause.
    - At least 4 queries should also use a WHERE clause.
    - At least 2 queries should also use a JOIN clause.
    - Precede each query with a Markdown comment that describes its function.

- Create data visualizations in Looker Studio:
    - Choose 2 of your most interesting queries from the previous section.
    - Open Looker Studio and go through this tutorial to familiarize yourself with the tool.
    - Create a Looker Studio chart that visualizes the data from each of your chosen queries in a compelling way. Add a relevant title for each chart which describes the data. You should end up with two charts, one per query.
    - Download both charts as pdfs and name them `chart1.pdf` and `chart2.pdf`.

CS 327E Project 4 Rubric

Due Date: 09/28/23

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 JOIN clause. **-80** for all queries missing from `project4.ipynb`: **-5** for each query missing a GROUP BY or aggregate function **-1** for each query missing an ORDER BY clause **-2** for each query missing a HAVING clause **-1** for each query missing a WHERE clause **-2** for each query missing a JOIN clause **-1** each incorrect comment, or comment too similar to query	80
Create data visualizations or charts in Looker Studio. Visualizations should display the results from your two chosen queries. Each chart should also have a relevant title describing its data. **-20** for `chart1.pdf` and `chart2.pdf` not found in repository **-8** for each chart created from a table instead of a query -**1** for each chart missing a title	20
`project4.ipynb`, `chart1.pdf`, and `chart2.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**