Milestone 9 due Friday, 04/12.

Part 1:

- Search for a secondary dataset in CSV format that meets our dataset requirements.
- Add a description of your dataset to the file `DATASETS.txt`.
- Import the data from your secondary dataset into BQ. Ensure that you import the data into a new BQ dataset.
- Model the data in your secondary dataset by applying the design principles from Milestone 4.
- Update your ERD to include the tables in your new dataset. Be sure to denote in the diagram the relationships between the tables within the new dataset as well as **across** datasets. Name your updated ERD file `ERD-v4.pdf`.

Part 2:

Think of 6 interesting queries that span your main and secondary datasets (e.g. `dataset1` and `dataset2)`. These queries should use a join to combine the data. In addition, we also want these queries to require some prior data transformations in order to join the datasets. These transformations will be done through Apache Beam in the next milestone.

For each of your 6 queries:

- Briefly describe what results the query is expected to produce and what SQL operations the query will use to produce those results (1-2 sentences).
- Briefly describe what type(s) of data transforms are required to successfully implement the query (1-2 sentences).

Create a file `CROSS-DATASETS.txt` and add your descriptions and explanations to this file.

CS 327E Milestone 9 Rubric
Due Date: 04/12/19

Part 2 - Create a file `./CROSS-DATASETS.txt` containing query and transformation information for 6 queries, as described in the outline. Keep in mind that you do not actually have to *write* the query, just a description of one and transformations required to make the query work. **-60** `./CROSS-DATASETS.txt` does not exist **-10** for each missing pair of query description and required transformation(s) description, up to **-60**	60
Part 1 - Create a new ERD titled `ERD-v4.pdf` which also includes data in your new dataset. Diagram their relationships as you have in previous milestones - this does include adding potential relationships between tables from both datasets. **-30** `ERD-v4.pdf` is missing. **-10** each incorrectly labeled keys **-10** each incorrect relationship **-5** each incorrectly labeled data type Edit the file `./DATASETS.txt` to include information on your new dataset. **-10** no description of new dataset in `DATASETS.txt`	40

`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**