Milestone 8 due Friday, 11/30.

Part 1:

Develop the Beam pipelines in support of your cross-dataset queries. The pipelines should read the bad records from a BigQuery table, perform one or more Beam transforms on the input data, and write the clean records back to BigQuery as a new table.

The pipelines should be developed and tested on a small subset of the input data using `DirectRunner`. Once tested, the pipelines should be converted to process all the records from the BigQuery table and run on Dataflow using the `DataflowRunner`.

Programming Style Rules:

For consistency and readability, please follow these programming and naming conventions:

- Each pipeline should transform a different BigQuery table.
- All the transforms performed on a table should be contained in the same Beam pipeline.
- Each Beam pipeline should have two versions, one version that runs on a single machine (CloudShell) using `DirectRunner` and processes a small subset of the input data and the other version that runs on the distributed Dataflow cluster using `DataflowRunner` and processes the entire input data.
- Name the files for each pipeline as `transform_<table>_single.py` or `transform_<table>_cluster.py` where `<table>` is the actual table name being transformed and `single` versus `cluster` indicates the compute environment used by the pipeline.
- Push both versions of each pipeline to your GitHub repo.

Part 2:

Open the file `CROSS-DATASETS.txt` from Milestone 7. For each one of your listed queries, add in the names of the python scripts that implement the supporting Beam pipelines from Part 1 of this milestone.

CS 327E Milestone 8 Rubric

Due Date: 11/30/18

Part 1 - Write the necessary Beam transform scripts needed to transform your data to a new table used by the queries. **-60** no files named `transform_<table>_single.py` or `transform_<table>_cluster.py` i.e `transform_Students_cluster.py` **-10** each missing complementing `cluster/single` file **-10** each broken/incorrectly implemented transform script	60
Part 2 - Add in to your `CROSS-DATASETS.txt` file the names of all transform scripts needed to apply a query (i.e `transform_<table>_single.py`) **-10** each query missing a transform script	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**