Milestone 6 due Friday, 10/18.

This is the second of two milestones that involves cleansing your main dataset (aka `dataset1`) using Apache Beam.

In the previous milestone, you transformed one of the tables you identified in `TRANSFORMS.txt` using a simple `ParDo`. In this milestone, you will expand this work as follows:

- Transform every table listed in `TRANSFORMS.txt`.
- Apply the appropriate Beam transforms to cleanse the data (e.g. `ParDo, GroupByKey, CoGroupByKey, Flatten`).
- Create two versions of each pipeline, one which uses the Direct Runner and processes a small subset of the source data using the `LIMIT` clause and another that runs the pipeline with the Dataflow Runner and processes the entire source data.

Coding Conventions:

- Each pipeline should transform a different table.
- All of the transforms applied to a table should live in the same Beam pipeline.
- A table should be named `<table>__Beam` if it was produced by a Direct Runner pipeline; it should be named `<table>__Beam_DF` if it was produced by a Dataflow Runner pipeline
- The pipeline scripts should be named `<table>__single.py` or `<table>_cluster.py` where `<table>` is the table being transformed and single versus cluster indicates the compute environment used by the pipeline.
- The code should be commented sufficiently to understand the main logic of the transforms.

CS 327E Milestone 6 Rubric
Due Date: 10/18/19

Create a number of Python scripts, `<table>_single.py` and `<table>_cluster.py` based on the transforms specified in `TRANSFORMS.txt`. The above two files should exist for each transform you make. **-X** for each missing `<table>_single.py/<table>_cluster.py` where X is dependent on the number of transforms you have. If you have 2, -**50** each. 3, -**33** each, and so on. **-10** transform does not work as intended **-10** each transform not using both DirectRunner and DataflowRunner	100

`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**