Milestone 6 due Friday, 03/08.

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 identified in *TRANSFORMS.txt* using a simple ParDo. In this milestone, you will build upon this work as follows:

- Transform every table identified 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 that uses the Direct Runner and processes a small subset of the data and another that uses Dataflow Runner and processes the entire table data.

**Coding Conventions:**

- Each pipeline should transform a different table.
- All transforms performed on a table should be in the same Beam 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.
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

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