Milestone 11 due Friday, 12/06.

Part 1:

1. Implement your 3 cross-dataset queries:
   - Queries should use the modeled tables from `dataset1` and `dataset2`
   - Queries should be wrapped by a SQL view
   - Views should be created in new BQ dataset named `reporting`
   - Add a short comment above each SQL statement to describe the query. Comments should be in Markdown format
   - Queries are written in a Jupyter notebook named `cross_dataset.ipynb`

2. Create 3 data visualizations in Data Studio:
   - Data Sources query the SQL views from Part 1.
   - Charts visualize the data in a compelling way.
   - Add the 3 charts to your existing Data Studio dashboard.
   - Take a screenshot of your dashboard and save it as `dashboard-v3.png`.

Part 2:

Create an Airflow DAG that automates the data pipeline for your `dataset2`.

Required functionality:
- DAG creates a new BQ dataset named `<source>_workflow_staging` to store the staging tables.
- DAG creates a new BQ dataset named `<source>_workflow_modeled` to store the modeled tables.
- DAG loads the CSV files for each dataset into `<source>_workflow_staging` and runs through the series of SQL and Beam transformations, writing the modeled tables to `<source>_workflow_modeled`.
- DAG executes dependent tasks in proper sequence.
- DAG executes independent tasks in parallel.
- DAG is implemented in a standard Python file named `<source>_workflow.py`

What's **not** required:
- Copying the CSV files into GCS.
- Creating the database views used for reporting.

Testing and verification:
- DAG must produce the same end-results as the Beam pipelines from Milestone 10.

CS 327E Milestone 11 Rubric
Due Date: 12/06/19

Part 1 - Create file `cross_dataset.ipynb` that runs the cross-dataset queries. Comment each query with the function they perform. **-20** no `cross_dataset.ipynb` in repository, or missing queries **-5** each missing or erroneous query, up to **-15** **-5** each missing or incorrect comment, up to **-15** **-5** each query not on a transformed table, up to **-15** Create file `dashboard-v3.png` that visually displays the data returned by your queries. **-20** no `dashboard-v3.png` present **-5** each missing query, up to **-15** **-5** each chart without a title, up to **-15**	40
Part 2 - Create file `<source>_workflow.py` implementing your workflow for transforming `dataset2`. The DAG should run operations in a decently efficient manner (operations that don't depend on one another should run at the same time, etc.) **-60** `<source>_workflow.py` does not exist in the repository or missing dependencies such as Beam and Dataflow Python scripts **-10** each task that runs in parallel with its dependency **-10** each task that runs after another task it does not depend on **-20** `<source>_workflow_staging` dataset does not exist in BQ **-20** `<source>_workflow_modeled` dataset does not exist in BQ	60

`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**