CS 327E Milestone 1, due Sunday 02/02.

1. Find a project partner and choose a name for your team. The name that you choose for your team will be your repo name on GitHub and will also be used to identify your team in project activities such as review sessions and presentations.

2. Request a GitHub private repo for your team by following these steps:
   - Email your names, EIDs, and GitHub usernames for you and your partner and the requested repo name for your team.
   - Email should be addressed to the Prof. and TAs
   - Copy your partner on the email
   - Email subject line should be: [CS 327E] Spring20 Team Info

   Note: You cannot create your own GitHub repo. You must request a private repo under our GitHub organization by following the steps above.

3. If you are new to git & GitHub, please take the time to go through a git tutorial (google: git tutorial) and then set up your git repository. Remember to create a `README` file in your repo with you and your partner's full names, EIDs, and emails.

4. Set up your GCP account.

5. Choose your main dataset for the project:
   If you don't have a specific dataset in mind for your project, start off by signing up for an Enigma Public account and then browse the available datasets on Enigma. You can also use Google's Dataset Search tool to search for data on a particular subject (e.g. sustainability).

   Select a primary dataset (aka *dataset1*) that you and your partner are both interested in exploring and analyzing for your project. The dataset must consist of multiple files and be in tabular format. In the near future, you will also be choosing a secondary dataset (aka *dataset2*) which has connections to your main dataset. For example, if you are interested in analyzing how weather events affect flight on-time performance, your main dataset could be the Airline On-Time Performance Data from BTS and your secondary dataset could be historical weather events by NOAA. See lecture slides for additional dataset examples.

   All the data you collect should be structured, formatted as CSV files, and publicly available.

   Note: You are free to extract the data through an API, but **you** are responsible for writing the API client and formatting the output as CSV.

6. Once you have selected your primary dataset, describe the data in the dataset, making note of the interesting attributes and relationships in the data. Include 3-5 rows of sample data from each file. Very important: explain what insights you hope to gain from analyzing the data. If you don't know *why* you want to analyze the data, you should stop now and look for a different dataset for the project.

   Also, if you have already chosen your secondary dataset, describe this data as well and explain what insights you hope to gain by combining the two datasets.

   Provide the description of your dataset(s) and download links in a `DATASETS.txt` file. If you pulled the data from an API, instead of downloading it, provide the API endpoints. Commit this file to your repo and push it to GitHub.

7. Create a `submission.json` file and submit it through Canvas.

CS 327E Milestone 1 Rubric
Due Date: 02/02/20

Create a `README.md` file in your team's private repository under our cs327e-spring2020 GitHub organization:
 	
`README.md` should contain you and your partner's full names, EIDs, and emails in the following format (not including braces):
 	
`<your full name>, <your UT EID>, <your email>`
`<partner's full name>, <partner's EID>, <partner's email>`
 	
Example:
 	
`William Chia, wc1234, chiaw@example.com`
`Prithvi Chowhan, pc1234, chowp@example.com`
 	
-25 no private repository under the cs327e-spring2020 GitHub organization
-25 no README.md file, file incorrectly named, or incorrect info in file
	25

Create a GCP project and grant access rights to cs327e.spring2020@gmail.com **-25** project name is not equal to team name **-25** cs327e.spring2020@gmail.com does not have permission to access project	25
Find a primary dataset that: ● Is available as CSV format (do NOT upload the actual dataset itself) ● Is comprised of multiple files ● Contains relationships between files **-50** no dataset ● Be explained in a text file named `DATASETS.txt` **-10** each missing explanation **-50** no `DATASETS.txt` file in repository	50

Your repository structure should resemble the following:
 	
`README.md`
`DATASETS.txt`
 	
-10 each misplaced or misnamed file	

`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",`
	Required

`"project-id": "some-project-id"` `}`	
Total Credit:	**100**