

Milestone 1, due Friday 02/01.

The referenced guides are available from the [wiki section of our snippets repo](#).

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 as follows:
  - 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] Spring19 Team Info
3. Go through a git tutorial, create README file in your repo with you and your partner's full names, EIDs, and emails (follow guide).
4. Sign up for a GCP account (follow guide).
5. Create a GCP project by the same name as your team's name (follow guide).
6. Retrieve the GCP coupon and redeem the coupon code. You should get \$250 in credit (follow guide).
7. Share your GCP project with your partner and instructors. Share with the instructor using the account [cs327e.spring2019@gmail.com](mailto:cs327e.spring2019@gmail.com). Grant **Project Owner** access to both (follow guide).
8. Choose your main dataset for the project:

Start off by signing up for an [Enigma Public account](#) and browse the available datasets. You can also use Google's [Dataset Search tool](#) to search for data on a specific topic.

For now, you'll need to select one dataset that consists of multiple files and will serve as your main dataset in this course (aka *dataset1*). In the near future, you will also need to select an additional dataset which has connections to your main dataset (aka *dataset2*). For example, your main dataset could be historical flight on-time performance from [BTS](#) and your secondary dataset could be historical weather events by [NOAA](#). See guide and lecture slides for more details on dataset requirements for the term project.

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

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.

9. Once you've found your main dataset, describe the data, making note of the interesting attributes and relationships in the data. Explain what insights you hope to gain from analyzing the data. Also, if you already found your secondary datasets, describe those as well and explain what insights you hope to gain by combining the datasets. Be sure to provide the download links to the datasets or API endpoints. Put all these details in a DATASETS.txt file in your repo.
10. Create submission.json and upload submission through Canvas (follow guide).

CS 327E Milestone 1 Rubric

**Due Date: 02/01/19**

<p>Create a repository in Github with the following files:</p> <p>README.md</p> <p>README.md should contain you and your partner's full names, EIDs, and emails in the following format (not including braces):</p> <pre>&lt;your full name&gt;, &lt;your UT EID&gt;, &lt;your email&gt; &lt;partner's full name&gt;, &lt;partner's EID&gt;, &lt;partner's email&gt;</pre> <p>Example:</p> <pre>William Chia, wc1234, chiaw@example.com Prithvi Chowhan, pc1234, chowp@example.com</pre> <p><b>-25</b> no README.md file, file incorrectly named, or incorrect info in file</p>	25
<p>Register a project with GCP, and provide the name to cs327e.spring2019@gmail.com</p> <p><b>-25</b> invalid project name, or none provided</p>	25
<p>Find a main dataset that:</p> <ul style="list-style-type: none"> <li>• Is available as CSV format (do NOT upload the actual dataset itself)</li> <li>• Is comprised of multiple files</li> <li>• Contains relationships between files</li> </ul> <p><b>-50</b> no dataset</p> <ul style="list-style-type: none"> <li>• Be explained in a text file named DATASETS.txt</li> </ul> <p><b>-10</b> each missing explanation</p> <p><b>-50</b> no DATASETS.txt file in repository</p>	50
<p>Your repository structure should resemble the following:</p> <p>README.md</p> <p>DATASETS.txt</p> <p><b>-5</b> each misplaced or misnamed file</p>	
<p>submission.json submitted into Canvas. Your project <b>will not</b> be graded without this submission. The file should have the following schema:</p> <pre>{   "commit-id": "your most recent commit ID from Github",   "project-id": "your project ID from GCP" }</pre> <p>Example:</p> <pre>{   "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9",   "project-id": "some-project-id"</pre>	Required

}	
<b>Total Credit:</b>	<b>100</b>