

Milestone 1, due Friday 09/14.

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.
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] Fall18 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.fall2018@gmail.com. Grant Project Editor access to both (follow guide).
8. Start looking for some datasets to use for your project.

Good places to start your search are the BigQuery project `bigquery-public-data` and Google's new [Dataset Search tool](#).

You'll need 2-3 datasets which have some interesting relationships to each other and are produced from different sources. For example, flight on-time performance from BTS and weather events from NOAA.

The data needs to be structured, formatted as a CSV, JSON or BQ table, 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 a CSV or JSON.

9. Once you've found your datasets, briefly describe each one, making note of the interesting attributes and relationships in the data. Provide the link to the download site or APIs, and explain what insights your group hopes to gain by combining the datasets. Put these details in a DATASETS file in your repo.

10. Create submission.json and upload submission through Canvas (follow guide).

CS 327E Milestone 1 Rubric

Due Date: 9/14/18

<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:</p> <pre><your full name>, <your UT EID>, <your email> <partner's full name>, <partner's EID>, <partner's email></pre> <p>Example:</p> <pre>William Chia, wc1234, chiaw@example.com Prithvi Chowhan, pc1234, chowp@example.com</pre> <p>-25 no README.md file, file incorrectly named, or incorrect info in file</p>	<p>25</p>
<p>Register a project with GCP, and provide the name to cs327e.fall2018@gmail.com</p> <p>-25 invalid project name, or none provided</p>	<p>25</p>
<p>Find 2-3 related datasets. Each dataset should:</p> <ul style="list-style-type: none"> • Relate to at least one other dataset • Be available as a CSV, JSON, or BigQuery table <p>-50 no datasets -30 one or more datasets have no relation to other datasets</p> <ul style="list-style-type: none"> • Be explained in a text file named DATASETS <p>-10 each missing explanation -50 no DATASETS file in repository</p>	<p>50</p>
<p>submission.json submitted into Canvas. Your project will not be graded without this submission. The file should have the following schema:</p> <pre>{ "commit-id": "your most recent commit ID from Github" }</pre> <p>Example:</p> <pre>{ "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9" }</pre>	<p>Required</p>
<p>Total Credit:</p>	<p>100</p>