'Course Logo

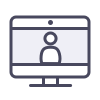
Icon

Description automatically generatedCharito

[Community](https://springboard-2784014.hs-sites.com/community?__hstc=267612241.be780f891a8a47a41ae3bd013af2ae87.1668028322426.1668028322426.1675907646689.2&__hssc=267612241.1.1675907646689&__hsfp=2057162590)

Search

My Progress

[Office Hour Videos](https://www.youtube.com/playlist?list=PLjKxfZpleE6ARDh12WsI12JjdKUN8J2nb)

scroll icon

* 1

[Program overview](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/44839)

  2

[Intro to Data Science & Python](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10671)

  3

[Intermediate Python](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10690)

  4

[Foundations of Probability](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10700)

  5

[Python: Data Structures & Algorithms](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10703)

  6

[Technical Skills Survey](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10712)

  7

[Your Data Science Toolbox](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10720)

  8

[The Apps Project](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10727)

  9

[Congratulations!](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10742)

 10

[What is Data Science?](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18565)

  11

[Problem Identification](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18576)

  12

[The Python Data Science Stack](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18594)

  13

[Creating Your Job Search Strategy](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18610)

* 14

[Applying the Data Science Method](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18619)



* 15

[Data Wrangling](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18648)



* + 15.1

[Capstone Two: Project Ideas & Proposal](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18649) 0/2

* + 15.2

[Data Collection](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18652) 0/5

* + 15.3

[Data Organization](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18659) 0/2

* + 15.4

[Data Definitions](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18662) 0/2

* + 15.5

[Data Cleaning](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18665) 0/3

* + 15.6

[Capstone Two: Data Wrangling](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18669) 0/1

* + 15.7

[Wrap up](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18671)

* 16

[SQL & Databases](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18674)



* 17

[Your Elevator Pitch and LinkedIn Profile](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18694)



* 18

[Statistics for Exploratory Data Analysis](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18708)



* 19

[Python Statistics in EDA](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18751)



* 20

[Effective Networking](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18769)



* 21

[Machine Learning Overview](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18786)



* 22

[Supervised Learning](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18803)



* 23

[Unsupervised Learning](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18856)



* 24

[Feature Engineering](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18892)



* 25

[Informational Interviews](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18911)



* 26

[Machine Learning Applications](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18921)



* 27

[Find the Right Job Title and Companies](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18940)



* 28

[Data Storytelling](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18954)



* [Specializations](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18978)

scroll icon

**Data Wrangling**

**Learning Objectives**

* Learn how to acquire and compile data through data collection
* Develop the skills to keep projects organized by building local file structures and keeping GitHub repo versions updated
* Learn how to review data types, build data profiles, and begin to develop an understanding of the feature dimensions of your data
* Be able to handle problems in your data such as null values, duplicates, and formatting issues

**Work to Complete**

In this unit, you'll:

* Work through three mini-projects:
  + An API mini-project
  + A NASA exercise
  + A text data handling exercise
* Submit ideas and a project proposal for your second capstone
* Wrangle data for your second capstone

Data wrangling is a crucial step in the process of preparing data so that you can identify insights, perform EDA, and build models. By wrangling your data at the beginning of any data science project, you'll set yourself up to be able to explore your data efficiently. You were introduced to data wrangling concepts while working on the guided capstone. This unit takes a deeper dive into each of these concepts and introduces new scenarios and constraints that are important for you to understand. At the end of this unit, you'll put your data wrangling skills to use by kicking off your second capstone. This capstone will feature prominently in your data science portfolio. Employers will expect to see proof of your ability to collect and clean data. Your portfolio will play an integral role in demonstrating your abilities.

'Course Logo

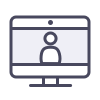
Icon

Description automatically generatedCharito

[Community](https://springboard-2784014.hs-sites.com/community?__hstc=267612241.be780f891a8a47a41ae3bd013af2ae87.1668028322426.1668028322426.1675907646689.2&__hssc=267612241.1.1675907646689&__hsfp=2057162590)

Search

My Progress

[Office Hour Videos](https://www.youtube.com/playlist?list=PLjKxfZpleE6ARDh12WsI12JjdKUN8J2nb)

scroll icon

* 1

[Program overview](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/44839)

  2

[Intro to Data Science & Python](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10671)

  3

[Intermediate Python](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10690)

  4

[Foundations of Probability](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10700)

  5

[Python: Data Structures & Algorithms](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10703)

  6

[Technical Skills Survey](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10712)

  7

[Your Data Science Toolbox](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10720)

  8

[The Apps Project](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10727)

  9

[Congratulations!](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/10742)

 10

[What is Data Science?](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18565)

  11

[Problem Identification](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18576)

  12

[The Python Data Science Stack](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18594)

  13

[Creating Your Job Search Strategy](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18610)

* 14

[Applying the Data Science Method](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18619)



* 15

[Data Wrangling](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18648)



* + 15.1

[Capstone Two: Project Ideas & Proposal](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18649) 0/2

* + 15.2

[Data Collection](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18652) 0/5

* + 15.3

[Data Organization](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18659) 0/2

* + 15.4

[Data Definitions](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18662) 0/2

* + 15.5

[Data Cleaning](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18665) 0/3

* + 15.6

[Capstone Two: Data Wrangling](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18669) 0/1

* + 15.7

[Wrap up](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18671)

* 16

[SQL & Databases](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18674)



* 17

[Your Elevator Pitch and LinkedIn Profile](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18694)



* 18

[Statistics for Exploratory Data Analysis](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18708)



* 19

[Python Statistics in EDA](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18751)



* 20

[Effective Networking](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18769)



* 21

[Machine Learning Overview](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18786)



* 22

[Supervised Learning](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18803)



* 23

[Unsupervised Learning](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18856)



* 24

[Feature Engineering](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18892)



* 25

[Informational Interviews](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18911)



* 26

[Machine Learning Applications](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18921)



* 27

[Find the Right Job Title and Companies](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18940)



* 28

[Data Storytelling](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18954)



* [Specializations](https://www.springboard.com/workshops/data-science-career-track-f2c/learn#/curriculum/18978)

scroll icon

**Capstone Two: Project Ideas & Proposal**

To kick off this unit, we'd like you to start planning out your second capstone project. For this project, you'll work through the entirety of the Data Science Method, starting with problem identification. First, you'll identify three possible project ideas. Once you've talked to your mentor about your ideas, you'll expand your idea into a complete project proposal. At the end of this unit, you'll take on the data wrangling step of this capstone.

1

**Capstone Two - Project Ideas**

Save

2 - 4 Hours

55 Points

For the first step of your second capstone, you'll use Google Dataset Search to find three datasets and brainstorm project ideas related to those datasets. Please note that you can also combine two or more datasets to answer a single project question.

Review the[**rubric**](https://docs.google.com/document/d/1_oSYbMy8zeUvld6DSAVtwfev21hYWeb3VdLdtR10of0/edit)to understand how your work will be assessed.

**Submission Steps**

1. Read [**these instructions**](https://www.springboard.com/archeio/download/dff20f432d924ac6a412b03c70c861f9/) to get started.
2. Write a description of three capstone project ideas. Your ideas can be broad and high-level. The descriptions should address the problem and identify the data you’ll use. You don’t need to talk about specific methods and techniques.
3. Submit a Google Doc link here. Please set sharing permissions to “anyone with a link can comment” so that your mentor can provide feedback. Please do not submit .pdfs, .ppts, or markdowns.
4. Review your ideas with your mentor during your next call.
5. (Optional) Post your idea, including a title and description, to the student community for feedback in the #ask-for-feedback Slack channel.
6. After collecting feedback, choose one idea and get ready for the next step: creating a project proposal.

Top of Form

Paste URL(s) to your document(s):

Please ensure that the document url is shared with public accessibility. Not making your submission publicly accessible could increase the time to receive feedback

Top of Form



Bottom of Form

+ Add another link

Bottom of Form

Top of Form

You can submit a project multiple times.

Bottom of Form

2

**Capstone Two - Project Proposal**

Save

2 - 4 Hours

55 Points

With your project ideas narrowed down to one, it's time to write up a project proposal!

Review the [**rubric**](https://docs.google.com/document/d/1w1I2NmG7uIBxD2HNCwwZYZvwRNhD5aGPhpYacaQYnUw/edit) to understand how your work will be assessed.

**Submission Steps**

1. Read [**these instructions**](https://www.springboard.com/archeio/download/edceb89151984c24abff558309b8e503/) to get started.
2. Write your 1-2 page proposal in a Google Doc and submit the link here. Make sure your mentor has permission to comment on the document.
3. Work with your mentor to incorporate any feedback into later drafts of your proposal. Submit a link to your proposal each time you iterate on your idea.
4. Once your mentor has approved your proposal, convert the Google doc to a PDF file.
5. Create a GitHub repository for this project (if you haven’t done so already).
6. Add the PDF to your GitHub repository for this project.
7. (Optional) Share the proposal with your peer community for feedback.

**Please note**: All code and further documentation you write will be added to this repository.

Top of Form

Paste URL(s) to your document(s):

Please ensure that the document url is shared with public accessibility. Not making your submission publicly accessible could increase the time to receive feedback

Top of Form



Bottom of Form

+ Add another link

Bottom of Form

You can submit a project multiple times.

**First, you'll identify three possible project ideas**

**Once you've talked to your mentor about your ideas, you'll expand your idea into a complete project proposal.**

**Google Dataset Search to find three datasets and brainstorm project ideas related to those datasets. Please note that you can also combine two or more datasets to answer a single project question.**

|  |
| --- |
| * A short description of each of the three ideas as a single Google doc has been submitted. |
| * The submission contains three high-level ideas with links to the appropriate data that could be used to support the idea. * The ideas are framed around real-world business problems. |

**The platform is also known for hosting regular competitions where you can go up against other data scientists to solve real-world problems posted by companies.**