



Capstone Three: Step 0

Estimated Time: 3-6 Hours

Project Overview

For your third capstone, you'll work through the Data Science Method steps again, however, the main difference between this capstone and the second one is that your third capstone should highlight one of the topics or techniques identified in your specialization. Examples include applying a time series model, natural language processing or image processing.

- [Student Example](#)
 - *Problem* - This is a multi-class classification problem where an image of a plant seedling has to be classified as one of the 12 species.
 - [Dataset](#)

Project Steps

1. Explore datasets from [Google Data Search](#). Some other search engines include:
 - [Quandl](#), [US Government Open Data](#), [UCI Machine Learning Repository](#), [Kaggle competitions](#), [Mode Analytics](#)
 - [Data is Plural](#) is an email list that sends newly released and interesting datasets
2. Find **three datasets of interest** and turn those into project idea drafts. You can also combine two or more datasets to answer a single question.

For your initial project ideas, please:

- Include a short description of each idea. The description should briefly discuss the problem and the data you'll use to solve it. At this point, there's no need to outline specific methods and techniques.
- Post your idea, including the title and description, to the community and solicit feedback from both other students.

One word of caution:

Be sure to choose your datasets wisely; check for a reputable source and detailed data documentation or metadata to describe how it was collected and the appropriate level of context required for using it responsibly.

Student Example

Check out Nick Campanelli's [capstone project](#) to get some inspiration.