Data Science Career Track

Course Prework

Springboard

Springboard

Content

Welcome to your first Springboard assignment! **1**

1. Get Started with DataCamp **3**

Set Up Your DataCamp Account (~ 10-15 mins) 3

Introduction to Python for Data Science (4-6 hrs) 3

Intermediate Python for Data Science (4-6 hrs) 3

2. Install Your Python Data Science Stack **4**

Install Anaconda for Data Science (~ 10-20 mins) 4

Anaconda Documentation (~10-20 mins) 4

3. Get Started with Jupyter Notebook (~ 15 mins - 1 hour) **5**

4. Git and GitHub **6**

What is GitHub? (video, ~5 mins) 6

5. Unix Command Line **7**

Introduction to Shell for Data Science (optional) 7

Command Line Challenges (optional) 7

Data Science at the Command Line (optional) 7

Welcome to your first Springboard assignment!

Learning a new skill can be hard, so setting yourself up for success by preparing is essential. Once your cohort begins, you'll receive access to the Data Science Career Track course. Before then, please work through this prework to ensure that you have all the necessary skills and tools you'll need to complete the course.

The five sections of this prework cover important tools that will come up throughout your data science course. These tools include DataCamp, Python, Jupyter Notebooks, GitHub and Git, and Unix command line. This prework is designed to help you become comfortable with these tools so that you can keep pace while working through the course and get the most of the curriculum. The entire prework assignment will take you roughly 20 hours to complete, so please work through it in the days and weeks leading up to the start of your course. ***Please note that if you worked through the entirety of the Data Science Career Track Prep course, some of these resources may look familiar to you — feel free to skip any that you have completed previously.*** Happy learning!

1. Get Started with DataCamp

⭐ **If you took the Data Science Career Track Prep course, you don't need to complete these DataCamp resources unless you'd like to refresh your Python skills.**

Python is the programming language that forms the foundation of the Data Science Career Track. Whether you're new to Python or an experienced programmer, we've collected resources to jumpstart your data science journey. In this section, you'll create a DataCamp account and review some basic Python materials.

Set Up Your DataCamp Account (~ 10-15 mins)

You need to create a DataCamp account to complete the following resources. Springboard has partnered with DataCamp to provide you with free access to interactive Python programming courses while you're actively enrolled in the Data Science Career Track. Please reach out to support@springboard.com directly to request access to DataCamp. You should receive an invitation to join our DataCamp group within 1 business day.

Introduction to Python for Data Science (4-6 hrs)

This workshop reviews Python basics and serves as a great introduction or a quick refresher depending on your comfort level. You'll learn how to store and manipulate data and discover tools to start your own analyses. If you have a strong background in Python, you can use this resource as a way to refresh your knowledge.

Intermediate Python for Data Science (4-6 hrs)

Learn how to visualize real data with matplotlib's functions and get an overview of boolean logic, control flow, and loops. Matplotlib is a plotting library used for Python and is often used in data science to build scatter plots, density plots, and histograms. If you have advanced Python skills, use this as a way to learn a few new tips and tricks.

2. Install Your Python Data Science Stack

Now that you have an understanding of Python, it's time to start working with Anaconda, a collection of Python packages and libraries designed for data scientists. In this section, you'll install Anaconda and learn how it relates to data science.

Install Anaconda for Data Science (~ 10-20 mins)

Now let's dive into Anaconda, a hugely popular collection of Data Science oriented Python packages and libraries. You'll become familiar with packages such as NumPy, SciPy, Jupyter notebook, Nltk, and Scikit-learn. These are just a few of Anaconda's offerings used by the world's best and brightest Data Scientists. **Our mini-projects have all been developed with or updated to Python 3.x, so please ensure that you install Python 3.**

Anaconda Documentation (~10-20 mins)

This Anaconda documentation provides information on packages and functionality. Read through the user guide and installation instructions for additional help during installation.

3. Get Started with Jupyter Notebook (~ 15 mins - 1 hour)

Jupyter Notebook is an application that lets you create and share documents that contain live code, equations, visualizations, and markdown text. Jupyter Notebook allows you to prototype code rapidly and combine your code with useful documentation. Throughout your course, you'll receive Jupyter Notebooks with assignment instructions, and you'll upload completed notebooks for grading. In this section, you'll install Jupyter Notebook.

The only thing you need to do to complete this section is to ***make sure that Jupyter is installed and running on your machine.*** But since any development environment is unfamiliar at first, you're welcome to check out any of the following optional resources to get more comfortable with the environment:

- This video gives you a few awesome tips and tricks to become more adept with Jupyter.

- This video is a neat and concise introduction to Jupyter's many features (***optional***).

- The official Jupyter documentation provides a more in-depth account of Jupyter's features (***optional***).

Psst! If you're struggling to set the default Jupyter kernel to Python 3 due to having multiple versions of Python on your machine, these extra steps should help you smooth things out.

4. Git and GitHub

In addition to Python, Git and GitHub are two tools that programmers use daily. Git is a common system for version control. GitHub is like a social platform for coders built on top of Git. You'll use GitHub to create your portfolio and share your work with both your mentor and potential employers. In this section, you'll become familiar with both tools and create a GitHub account.

Note: GitHub has recently updated it's naming conventions for it's default branch from "master" to "main." This means that, whenever you see the references to a "master" branch in the below resources, you will be using the "main" branch instead. If you are using an older repository, be aware that you might still see references to a "master" branch - make sure that you alter these to "main" instead. For more information on this check out this guidance from Github: https://github.com/github/renaming

What is GitHub? (video, ~5 mins)

Learn how GitHub works. You'll use GitHub to submit and share files with your mentors and peers.

Github for Springboard (video, 10 - 15 mins)

Check out this short video from Springboard mentor Ben Bell, who provides a crash course on setting up your GitHub for this course.

Git and GitHub Tutorial for Beginners (article, ~30-45 mins)

Create your GitHub account using this tutorial and follow along to learn how to commit files, create new repositories, and create pull requests through your terminal.

Git and GitHub for Poets (*optional*)

Discover the relationship between Git and GitHub, an important connection to understand, since you'll use GitHub and Git throughout the course.

Code management with Git (*optional*)

In this DataCamp resource, you'll learn how Git is used for code management while working on a data science project.

Git exercises (*optional*)
For more hands-on practice with Git, check out this site with interactive challenges.

Git and data science (*optional*)
Learn about the GitHub features that are most useful for data scientists.

5. Unix Command Line

While data scientists use a lot of sophisticated tools, sometimes the good ol' Unix command line is the best tool for the job. Many Unix commands are great for data and text manipulation. Unix helps users combine existing programs, automate tasks, and leverage clusters and clouds to run programs. In this section, you'll learn about the most useful commands for data wrangling and cleaning.

Introduction to Shell for Data Science (*optional*)

This resource introduces the key elements of Unix command line. If you already have a strong background in Unix, use this opportunity to refresh your knowledge.

Command Line Challenges (*optional*)

Gain additional practice using the command line by working through these interactive challenges.

Data Science at the Command Line (*optional*)

This ebook serves as an advanced resource for students who are already familiar with shell commands and programming and would like to take their command-line data-crunching skills up a notch. If you completed all the previous resources and feel comfortable with them, we recommend you complete this one too!

Congratulations on completing your first step to becoming a data scientist!

We're looking forward to working together when the course begins.