



Data Science Career Track

Course Pre work



This document is the property of Springboard and is provided to you as a Springboard student. By accessing, downloading or reading through this document you agree to not use (or permit anyone else to use) this document or any materials in it for any purpose other than your personal learning.

Content

Welcome to your first Springboard assignment!	2
1. Get Started with DataCamp	3
Set up your DataCamp account (~ 10-15 mins)	
Introduction to Python for data science (4 hrs)	
Intermediate Python for data science (4 hrs)	
2. Install Your Python Data Science Stack	4
Install Python and Anaconda for data science (~ 20-30 mins)	
Anaconda documentation (~10-20 mins)	
3. Get Started with Jupyter Notebook	5
Jupyter Notebook Tutorial (video, ~30-45 mins)	
Official Jupyter Documentation	
4. Git and GitHub	6
What is GitHub? (video, ~5 mins)	
Git and GitHub tutorial for beginners (article, ~30-45 mins)	
Github for Springboard (video, 10 - 15 mins)	
Introduction to Git for data science (4 hours)	
Git exercises	
Enabling data science with GitHub (30 mins)	
5. Unix Command Line	7
Introduction to shell for data science (interactive, 4 hrs)	
Command line challenges	
Data science at the command line	

Welcome to your first Springboard assignment!

We know that learning a new skill can be hard, so we want to make it as easy and delightful for you as possible. Once your cohort begins, you'll receive access to the data science career track course. To ensure that you have all the skills and tools you need to succeed, we'd like you to review the pre-work to not only complete necessary tasks, but also check your knowledge to see if you need a deep dive into specific topics.

There are five sections, which cover important topics in your data science course, including DataCamp, Python, Jupyter Notebooks, GitHub and git, and Unix command line. Completing this assignment will give you a strong understanding of the material before going into the program, helping you keep pace and get the most out of the curriculum. Use the time until the course starts to complete the pre-work. The entire pre-work assignment will take approximately 20 hours to complete, so please keep that in mind as we get closer to the start of the course.

1. Get Started with DataCamp

Python is a programming language and 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 start the course. 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 to receive access.

[Introduction to Python for data science](#) (4 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, use this as a way to refresh your knowledge.

[Intermediate Python for data science](#) (4 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, histograms. If you have superior 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 dive in with Anaconda, a collection of Python packages and libraries designed for data scientists. In this section, you'll install Python and learn how it relates to data science.

[Install Python and Anaconda for data science](#) (~ 20-30 mins)

Install Anaconda to successfully complete your coursework. You'll use these packages extensively throughout the course, so please carefully follow the steps to ensure that they're correctly installed.

[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

Jupyter Notebook is an application that lets you create and share documents that contain live code, equations, visualizations, and markdown text. It allows you to prototype code rapidly and combine it 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.

[Jupyter Notebook Tutorial](#) (video, ~30-45 mins)

Install Jupyter Notebook, an essential tool that you'll use throughout your course.

[Official Jupyter Documentation](#)

Take a closer look at Jupyter, and read use cases, installation instructions, and [various tips and tricks](#).

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, not only with your mentor but also with potential employers. In this section, you'll become familiar with both tools and create a GitHub account.

[What is GitHub?](#) (video, ~5 mins)

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

[Git and GitHub for poets](#) (video, ~15 mins)

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

[Github for Springboard](#) (video, 10 - 15 mins)

A short video by Springboard mentor Ben Bell giving a crash course on setting up your Github for this course.

[Introduction to Git for data science](#) (4 hours)

Learn how to use Git for code management.

[Git exercises](#)

For more hands-on practice with Git, check out this site with interactive challenges.

[Enabling data science with GitHub](#) (30 mins)

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. This is since many Unix commands are highly efficient 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](#) (interactive, 4 hrs)

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](#)

Gain additional practice on the command line using interactive challenges.

[Data science at the command line](#)

This e-book 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 the first
step to becoming a data scientist!**



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