



# Machine Learning Engineering Career Track

## Course Prework

**Content**

**Welcome to Your First Springboard Assignment! 2****1. Programming Refresher 3**

Python for Programmers (2 hrs)

Git - the Simple Guide (10-15 mins)

Collaborating with Git - Syncing, Pull Requests & Branching Workflows (20 - 30 mins)

Set Up GitHub for Springboard (10 mins)

Git Exercises (30-45 mins)

The SQL Tutorial for Data Analytics - Mode (30-45 mins)

Intermediate SQL - Mode (30-45 mins)

**2. Tech Refresher 4**

Algorithms & Data Structures - EdX (15 hrs)

Data Structures (1-2 hrs)

Fundamentals of Algorithms (1-2 hrs)

**3. Math & Algorithms Refresher 5**

Essence of Calculus (18 mins)

Essence of Linear Algebra (2 hrs)

**4. Statistics Refresher 5**

Probability and Bayes Theorem (45 minutes)

Basics of EDA (45 minutes)

Distributions (45 minutes)

**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 Machine Learning Engineering Career Track course. To ensure that you have all the skills and tools you need to succeed, we'd like you to review this prework to refresh your knowledge and determine whether you'll need to take a deeper dive into specific topics.

There are four sections to this prework, each of which covers important topics you'll need to know to successfully work through the course. These topics include Python, Git and GitHub, algorithms, data structures, the math behind machine learning, and statistics. Completing this assignment will give you a strong understanding of the material before going into the program so that you can stay on track and get the most out of the curriculum.

Use the time until the course starts to complete the prework. The entire prework assignment will take approximately 25 hours to complete, so please keep that in mind as you get closer to the start of the course. The same resources will also be added throughout the course so that you can review them as needed.

## 1. Programming Refresher (4-5 hrs total)

### [Python for Programmers](#) (2 hrs)

Python is an essential programming language for machine learning and contains special libraries that help to execute machine learning methods. The language itself is also great for working with machine learning algorithms.

By the time you begin the course, you should be comfortable with the fundamentals of Python. Unlike the other resources listed here that will reappear in the course for your reference, Python for Programmers is only hosted here in this prework. We highly recommend that you review the tutorial now and bookmark it for future reference.

In order to go through this tutorial, you'll need a code editor or IDE such as [PyCharm](#), [Eclipse](#) or [Atom](#). You're welcome to use your favorite editor or IDE here. If you're looking for one, we recommend you try [PyCharm](#), since it's free and available on all platforms.

#### References

[Python Documentation](#)

[Pep 8: The Python coding style guide](#)

[Top 5 IDEs and code editors for Python](#)

### [Git - the Simple Guide](#) (10-15 mins)

Take a look at this quick and simple guide that covers all you need to know for getting started with Git, including downloading Git for your OS system, creating your repository, branching, and merging.

### [Collaborating with Git - Syncing, Pull Requests & Branching Workflows](#)

(20 - 30 mins)

This tutorial from BitBucket will show you how to collaborate on projects using Git. This is an important skill to have, as you'll use Git to work with your mentor and TA. This resource also includes some tutorials that cover advanced Git functions.

### [Set Up GitHub for Springboard](#) (10 mins)

This short video from Springboard mentor Ben Bell will give you a crash course in setting up your Github for this course. If your Git skills are at all rusty, you should definitely check out this resource.

### [Git Exercises](#) (30-45 mins)

For more hands-on practice with Git, check out this resource!

### [The SQL Tutorial for Data Analytics - Mode](#) (30-45 mins)

This tutorial on the basic syntax of SQL is meant to be used with Mode, an analytics platform that you'll also use during the course. To effectively complete this tutorial, try running the example queries and analyzing the results.

### [Intermediate SQL - Mode](#) (30-45 mins)

This tutorial focuses on intermediate SQL. If you skipped the previous one, you might want to take some time reviewing this resource. Keep in mind that there are multiple ways of generating the correct results, so be sure to examine the results carefully to fully understand what the code is doing.

## **2. Tech Refresher (17-19 hrs total)**

You're strongly encouraged to review the topics included here, as you'll use them throughout the course and during your career as a machine learning engineer.

### [Algorithms & Data Structures - EdX](#) (15 hrs)

This course from edX will help you combine your programming experience with fundamental ML concepts like algorithms and data structures. You'll also review famous algorithms and equations and look through practical examples and real-life scenarios.

### [Data Structures](#) (1-2 hrs)

This resource contains a wealth of information on data structures. Start your exploration of this resource by checking out the links listed under "Topics" and then go from there!

### [Fundamentals of Algorithms](#) (1-2 hrs)

For this resource, you might also want to start with the links under "Topics." If you feel that your knowledge of algorithm fundamentals is still fresh, try taking the quiz found in this resource first and then working backward to identify areas and topics that you need to review.

### 3. Math & Algorithms Refresher (2.5 hrs total)

In preparation for the machine learning unit of the course, we recommend that you review the resources that focus on the math behind machine learning, including calculus and linear algebra.

#### [Essence of Calculus](#) (4 hours, 45 mins)

This visual and easy-to-follow series of videos starts with a simple geometry question and goes into integrals, derivatives, and, finally, the fundamental theorem of calculus. For the entire series on calculus, take a look [here](#).

#### [Essence of Linear Algebra](#) (2 hrs)

In the words of Skyler Speakman, Technical Lead for A.I. at IBM Research, "Linear Algebra is the mathematics of the 21st century." Indeed, linear algebra comes up frequently in both machine learning and deep learning. With that in mind, use this resource to help you visualize the fundamental concepts of linear algebra.

### 4. Statistics Refresher (2 hrs total)

The following provides a review of probability statistics, exploratory data analysis (EDA), and distributions. Feel free to circle back to these resources as needed.

#### [Probability and Bayes Theorem](#) (45 minutes)

Bayes theorem is essentially what powers the entire concept of Bayesian statistics. This resource hosts Chapter 1 of Allen Downey's *Think Bayes* and is available as a [free PDF](#).

#### [Basics of EDA](#) (45 minutes)

This resource focuses on exploratory data analysis (EDA) and examines ways to combine data and methods when working with uncertainty. Featured in this resource is Chapter 1 of Allen Downey's *Think Stats*, also available as a [free PDF](#).

#### [Distributions](#) (45 minutes)

Continuing from the previous resource, this resource is all about distributions and ways to visualize and comprehend data with histograms. This resource hosts Chapter 2 of Allen Downey's *Think Stats*, also available as a [free PDF](#).

**Congratulations on completing the first step to becoming a machine learning engineer!**



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