

DATA SCIENCE IMMERSIVE

ONBOARDING TASKS

Congratulations on joining the Data Science Immersive program at General Assembly! We can't wait to get started, so we've outlined a few assignments below to get everyone on the same page. These onboarding tasks will prepare you to hit the ground running, and are due **by the first day of class**. As always, let us know if you have any questions. Have fun and we'll see you soon!

GA + the DSI Instructional Team

WHAT CAN I EXPECT FROM THE ONBOARDING?

Your onboarding tasks include a number of required and optional tutorials to help you review the fundamentals of stats, python, SQL, git, and the command line. These are accompanied by a set of multiple-choice challenges that assess whether you've fully grasped the concepts that you've learned.

There are time estimations under each module but depending on your background and familiarity with these topics, you may spend more or less time on each part.

If you find yourself thinking...

"This stuff is way too easy, is this course actually going to challenge me?"

The purpose of these tasks is to make sure that everyone starts the course on the same page. If you are speeding through this material, you're in a great place to start the course, and we've included bonus material for you to work through. Don't worry, you'll be challenged in this course - we like to take it slow in the beginning but will ramp up fast!

"This stuff is way too hard, am I going to be able to keep up?"

That's why you're doing these assignments in the first place - so that you can get up to speed! We'll be reviewing these concepts in the first week as well so if you don't get it now, you'll have a chance to fully catch up by the end of the first week of class.

A NOTE ABOUT TECHNOLOGY

You'll need an account with [Khan Academy](#) and [Code Academy](#) to complete these assignments.

You'll also need to create a [Github account](#). Go ahead and take 5 minutes to create these accounts if you don't have them already.

Additionally, before the course starts, we'll provide you with directions on how to install the following technologies:

- [Python 2.7](#) - We will be using Python & its packages as our primary language
 - [Anaconda](#) - We will be using Anaconda as our primary development environment
 - [Postgres](#) - We'll be using Postgres for local SQL-based data storage
 - [Slack](#) - We'll be using Slack on a daily basis to communicate with each other
-

STEP 1: COMPLETE THESE REQUIRED MODULES

35 -40 hrs

Complete the following **five** modules. If you've already done these in the past, then try reviewing the "bonus" activities instead:

I. Statistics

- Get up to Speed: [Probability & Statistics Review](#)
 - Bonus: [Additional Practice](#)

II. Python

- Get up to Speed: [Python Syntax Review](#)
 - Bonus: [Additional Practice](#) (Exercises 1-20)

III. Git

- Get up to Speed: [Git Tutorial](#)
 - Bonus: [Additional Practice](#)

IV. SQL

- Get up to Speed: [SQL Tutorial](#)
 - Bonus: [Additional Practice](#)

V. Command Line

- Get up to Speed: [Command Line Overview](#)

Bonus: [Additional Practice](#)

STEP 2: SUBMIT THESE REQUIRED EXERCISES

1-2 hrs

Once you have completed Step 1, please complete the Onboarding Exercises below. Follow the directions in the form to submit your answers:

[Data Science Onboarding Exercises](#)

STEP 3: TRY SOME OPTIONAL BONUS MATERIALS

1 hr - however long you want!

Up for a challenge? These advanced exercises will help you sharpen your wits when applying statistics and programming:

Bonus Materials

- Try some [Project Euler problems](#)
 - Read [Chapters 1-6 in “Think Stats”](#) and try some exercises
 - Get comfortable with [technical documentation](#)
 - Review [fundamental Git concepts](#)
 - Supplement [your learning experience](#)
-

STEP 4: DO SOME OPTIONAL READING

1 hr - however long you want!

Time to get excited about data! Check out the following resources for a brief look at how data science is changing the world:

Articles

- [Six Categories of Data Scientist](#)
- [Surprising Factors behind Successful Movies](#)
- [Computers Solving Illnesses with Poker](#)

Blogs

- Check out Randy Olson's blog where he shares his data tinkering problems and discoveries like how [he computed the optimal way to find Waldo](#).
- [FiveThirtyEight](#) is Nate Silver's brain child and a data nerd's playground. Browse around for some interesting and inspiring uses of statistics and data modeling in the real world.

Videos & Podcasts

- [Math of Crime and Terrorism](#): A video on crime, data, and the Poisson distribution
- [Data Skeptic Podcast](#): A podcast that explores basic data concepts and interesting topics related to data, with an eye towards skepticism

Textbooks

- [Data Science From Scratch: First Principles with Python](#) (\$32)
- [Dive into Python](#) (Free)