

Data Science Self-Study Resources

Galvanize Inc.



Contents

1 - Python	1
How do you know you are ready?	1
Resources	2
Do you need more fluency?	2
2 - SQL Resources	2
How do you know you are ready?	2
Resources	2
3 - Statistics and Probability Resources	3
How do you know you are ready?	3
Resources	3
4 - Comprehensive	4

Have you been thinking about applying to Galvanize? How do you know you have enough of a technical background to be successful in our Data Science Immersive course? Find out by working through this document.

This document was designed to help you to gauge your level of understanding in programming (both Python and SQL) and to test your proficiency in Probability and Statistics. It also references

resources which will help you along the way. If you want to learn more about the Admissions process itself, you will find a full description in DSI-Admission-process.pdf.

1 - Python

Python is an extremely popular and powerful programming language with tools to work in nearly every application area. In particular, Python (and specifically the [anaconda](#) environment that supports it) is the industry standard for machine learning. As such, comfort with Python is a must for any aspiring data scientist.

In the Galvanize Data Science Immersive we teach data science using Python, and only briefly review python topics during the course. We therefore require that students pass a python coding interview.

How do you know you are ready?

You should be ready for the program if you are able to comfortably solve most [CodeWars](#) (<http://bit.ly/2iWF0LL>) Level 6-7 problems in about 30 minutes.

Resources

These are our top 3 online resources. They cover similar topics, but have their own approach to teaching Python.

1. [Learn Python the Hard Way](#) (Exercises 0-39): A combination of explanations and exercises to apply the concepts (<http://bit.ly/2iWHUxT>)
2. [Codecademy's Python Course](#) (Chapters 1-8, 12): An online interactive coding environment that provides lot of guidance and automatic feedback (<http://bit.ly/1GO0Fx1>)
3. [Google's Python course](#) (up to and including dictionaries): Nice resource if you like to have videos to follow along (<http://bit.ly/1bmXyG5>)

Taking the part-time [Python Fundamentals workshop](#) offered at Galvanize is a great option if you want to learn on campus. (<http://bit.ly/1LBc6Ua>)

Even better, if you take the workshop and are accepted into the Galvanize Data Science program, we will credit the amount you paid for the workshop to your tuition balance.

Do you need more fluency?

Practice on coding challenge websites like [CodeWars](#) and [CoderByte](#). Use [Stackoverflow](#) when you're stuck. Keep practicing.

Want to practice on some mock challenges? Checkout the Python section of the self assessment document and work through the Python section of the coding challenges.

2 - SQL Resources

SQL is a building block of many things we do. If you are familiar with programming or data analysis, learning to write basic SQL queries won't take very long.

How do you know you are ready?

You should be familiar with `SELECT`, `FROM`, `WHERE`, `CASE` clauses and know how to use aggregates as well as `JOINS`!

Resources

There are many online resources to learning SQL. Here are some recommendations:

1. [SQLZoo](http://bit.ly/1RAqIc6) (Sections 0 to 7, <http://bit.ly/1RAqIc6>)
2. [Codacademy SQL](http://bit.ly/1QjodIp) (Chapters 1-4, without the pro version, <http://bit.ly/1QjodIp>)

Want to practice on some mock challenges? Checkout the SQL section of the coding challenges.

3 - Statistics and Probability Resources

In the Galvanize Data Science Immersive, we review probability, statistics and classical regression, but then build on these topics as we learn advanced and modern machine learning methodology. It is therefore extremely important that our students be quantitatively and analytically literate, with a solid background in baseline probability and statistical theory.

We evaluate technical analytical skills through a math interview focusing on probabilities, statistics and basic model building.

How do you know you are ready?

The math background we want to see will allow you to (1) comfortably solve probability problems, (2) leverage statistical distributions to solve real world problems, and (3) perform hypothesis testing. A solid understanding of basic tools for linear regression is also appreciated.

Resources

Never taken a stats course?

We recommend you take a full-length course in statistics or probability before applying. [Khan Academy](#) and [Udacity](#) provides some good “entry level statistics videos” and a good introductory course is available on Udacity. Make sure to work through the exercises and take the quizzes.

All of these resources cover the same basic information, just in different ways and formats:

- [Udacity's Intro to Statistics](http://bit.ly/1SnQHUX) (<http://bit.ly/1SnQHUX>)
- [Khan Academy: Probability and Statistics](http://bit.ly/2j1QgXC) (<http://bit.ly/2j1QgXC>)
- Coursera, edX, and other sites also have a large number of introductory statistics classes

Want to refresh your knowledge?

These are areas you should focus on:

- Counting
 - Permutations
 - Combinations
- Probability
 - Conditional probability (Bayes' Theorem)
 - Independent and dependent events
 - Mutually exclusive events
- Probability distributions
 - Binomial
 - Geometric
 - Poisson
 - Uniform
 - Normal
- Hypothesis testing
 - One/two sided tests
 - Confidence intervals
 - Inference for two means and inference for proportions
 - Chi-square tests
- Linear Regression
 - Build simple models and evaluate model performance
 - Have idea of the bias/variance trade-off

There are excellent exercises mentioned in the resources above. Should you wish for some additional learning material online, we would recommend [this Probability Review](http://bit.ly/2i6OLb4) (<http://bit.ly/2i6OLb4>) and [these statistics videos](http://bit.ly/2iWzSap) (<http://bit.ly/2iWzSap>).

Want to go further?

To go beyond the call of duty, and make sure you are in great shape for the program, work your way through chapters 1-4 of [an Introduction to Statistical Learning](http://bit.ly/1Wma81Z) (<http://bit.ly/1Wma81Z>). You could also enroll in [this Stanford online course](http://bit.ly/1IXp8Lg) (<http://bit.ly/1IXp8Lg>) or find the same material in this [Introduction to Machine Learning course](http://bit.ly/2bp2YJX) (<http://bit.ly/2bp2YJX>).

Want to practice on some mock challenges? Checkout the math challenges of the self assessment document.

4 - Comprehensive

[Galvanize's Data Science Primer](http://bit.ly/1TnYbZu) (<http://bit.ly/1TnYbZu>) is the most comprehensive source and is an excellent preparation for the program. It is a substantial packet that may take some time. However, it is a great resource for potential students.