



DATA SCIENCE

Due June 10, 2014

PreWork ~ 5-7 hrs

Introduction to Data Science ~ 2.2 hrs

You'll need to activate your Front Row subscription for the videos.

- General Assembly Front Row (watch the videos, not case study) - <https://generalassemb.ly/online/videos/introduction-to-data-science-and-analysis>
- General Assembly Front Row (watch the videos, not case study) - <https://generalassemb.ly/online/videos/introduction-to-data-science>
- Harvard Business Review - <http://blogs.hbr.org/2013/11/how-to-start-thinking-like-a-data-scientist/>

Data Science Overview ~ .5 hr

- Python and Scala vs. Others - <http://strata.oreilly.com/2013/12/data-scientists-and-data-engineers-like-python-and-scala.html>
- Data Science Workflow - <http://cacm.acm.org/blogs/blog-cacm/169199-data-science-workflow-overview-and-challenges/fulltext>

Get Familiar with Python ~ 2.5 hrs

- Codecademy - <http://www.codecademy.com/tracks/python>
- Learn Python - <http://www.learnpython.org>
- Git Tutorial by Atlassian - <https://www.atlassian.com/git/>
- Learn Git by CodeSchool - <http://try.github.io/levels/1/challenges/1>

Machine Learning in Python ~ .5 hr

- Scikit-Learn: Machine Learning in Python - <http://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>

Big Data ~ .5 hr

- What's the current obsession with Big Data - <https://www.quora.com/Big-Data/Why-the-current-obsession-with-big-data>
- ETL 2.0: Redefining Data Integration - <http://www.youtube.com/watch?v=uG2J4XWTtts>

Bonus: Front-End Web Development

- Learn to make awesome websites - <https://dash.generalassemb.ly>

Python and Data Science Resources

The class will be taught with Python 2.7. There's an ecosystem of Python tools we'll be using, listed below.

- Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. This book is a great place to start.

Intro to Python

- Python at Code Academy: Browser-based interactive introduction
- Free E-Book: Python for Informatics: Exploring Information by Charles Sverance - This is a remix of How to Think Like a Computer Scientist
 - Companion Videos to Python for Informatics
- Free E-Book: How to Think Like a Computer Scientist / Learning with Python by Allen Downey, Jeff Elkner and Chris Meyers

iPython and iPython Notebook

- iPython: A nicer interactive interface for Python
- iPython Notebook: To create Python reports

Pandas (Python Data Analysis Library)

- Pandas Official Page
- Pandas Video Intro
- Pandas Tutorial by Hernan Rojas

Scikit-Learn (Python Machine Learning Libraries)

- Scikit-Learn: Python Machine Learning Libraries

Code Editors

- Sublime Text works on Mac / Windows / Linux
- TextMate version 2 works on Mac only.
- Pycharm version 3 This is a full-blown IDE (integrated development environment), and might be overkill. Works on Mac / Windows / Linux

Intro to SQL and NoSQL Technologies

- SQL Zoo Browser-based intro to SQL
- Try MongoDB Browser-based intro to MongoDB
- Try Redis Browser-based intro to Redis