// FLATIRON SCHOOL

SYLLABUS

On Campus
Data Science Immersive

Last Updated 05/10/2019

Table of Contents

Overview

FLATIRON SCHOOL'S DATA SCIENCE IMMERSIVE

This program will provide you with the knowledge, skills, and experience to get a job as a data scientist – which requires a mix of software engineering, statistical understanding, and the ability to apply both skills in new and challenging domains.

Over 15 challenging weeks at Flatiron School, you'll learn how to gather data, apply statistical analysis to answer questions with that data, and to make your insights and information as actionable as possible. Our pedagogy ensures not only job readiness for today's market, but the aptitude and skills to keep learning and stay relevant in the industry in the years ahead.

What will you learn?

- How to retrieve data from outside sources and organize data using Python
- Organize data into at least three different tables or equivalent grouping
- Explore data and write down multiple hypotheses for data, and write proposals to use subset of algorithms to analyze the data
- Build machine learning API that outputs results of an analysis
- Application and usage of Big Data
- Presentation techniques to better share conclusions about approach and analysis to key stakeholders

When and where does the course meet?

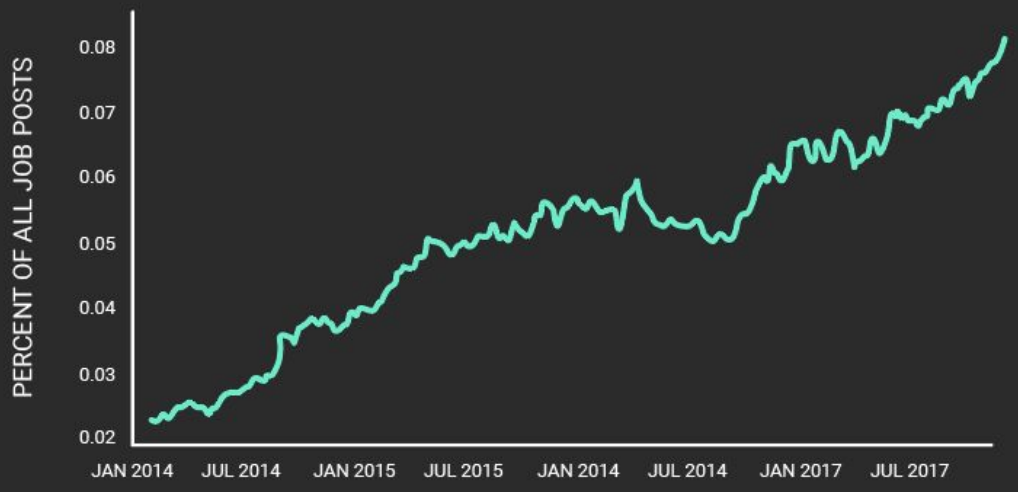
- Classes meet 5 days a week, full-time; Monday-Friday
- Classes will typically begin at 9 a.m., with a 90-minute break for lunch around 12:30 p.m., followed by continued lectures and lab work on-campus through the late afternoon

Why Data Science?

WHY IS THIS COURSE RELEVANT?

More than ever before, industries are capturing data on a variety of topics, behaviors, and trends. Without data science, this information stays stuck, without a story to tell or insights to share. In order to determine business goals, more and more companies are looking to data scientists to fill in the gaps and find opportunities never before considered.

Over the last four years, the rise of job opportunities for Data Scientists has increased exponentially.



Note: The chart above offers a 7-day rolling mean of all Indeed job posts that featured "data science" or "data scientist" in the title across the world as a percentage of all job posts between January 1, 2014 and November 16, 2017. The data was pulled using Imhotep, Indeed's open source analytics platform.

As this area of expertise has grown, the positions within the field have become more nuanced. After completing our immersive Data Science bootcamp, you'll not only be able to secure a job a Data Scientist, but can also consider pursuing any of the following related positions:

- Data Engineer
- Machine Learning Engineer
- Big Data Engineer
- Back-End Engineer
- Natural Language Processing

Curriculum Overview

From Python to Machine Learning, our 15-week data science training program gives you the breadth and depth needed to become a well-rounded data scientist. You'll learn the languages, skills, and processes used by data scientists today and graduate with an understanding of how to discover new techniques as your career progresses.

Every 2 weeks, you'll be introduced to a new module that builds off the learnings of the previous section while allowing you enough time to dive into each area for a thorough understanding of the subject matter.

Module	*Student Hours*	*Weeks*	*Projects*
Module 1: Python for Data Science	100	2 weeks	Mod 1 Project
Module 2: Data Engineering for Data Science	100	2 weeks	
Module 3: Probability, Sampling & AB Testing	100	2 weeks	Mod 3 Project
Module 4: Statistical Modeling	100	2 weeks	Mod 4 Project
Module 5: Machine Learning & Big Data	100	2 weeks	Mod 5 Project
Module 6: Deep Learning & Natural Language Processing	100	2 weeks	Final Project

Program Total: 600+ hours, projects, and homework

CURRICULUM OVERVIEW

Getting Started

Prework	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

The Data Science program moves quickly and our passionate students embrace that challenge. While no experience is necessary to apply, we require you to demonstrate some data science knowledge prior to getting admitted, then complete a prework course before Day 1. To help you prepare for our bootcamp, we provide a free introductory course. This prework ensures that you come in prepared and able to keep pace with the class.

MODULE 1

Python for Data Science

Prework	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Our first module introduces the fundamentals of Python for data science. You'll learn basic Python programming, how to use Jupyter Notebooks, and will be familiarized with popular Python libraries that are used in data science, such as Pandas and NumPy. Additionally, you'll learn how to use Git and Github as a collaborative version control tool. At the end of this module, you'll also learn how to build a basic linear regression model and how to evaluate the results. Finally, we'll conclude with a heavy focus on visualizations as a way to go from data to insights.

What is Covered in Module 1?

- Variables
- Booleans and Conditionals
- Lists
- Dictionaries
- Looping
- Functions
- Data Cleaning
- Pandas
- NumPy
- Matlotlib/Seaborn for Data Visualization
- Git/Github
- Linear regression

MODULE 2

Data Engineering for Data Science

Prework	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

In this module, you'll learn about data structures, relational databases, ways to retrieve data, and the fundamentals of SQL for data querying for structured databases, as well as NoSQL (and MongoDB) for non-relational databases. Furthermore, we'll cover the basics of HTML, XML, and JSON in order to access data from various sources using APIs, as well as perform Web Scraping.

What is Covered in Module 2?

- Data structures
- Relational Databases
- SQL
- Object-Oriented Programming
- NoSQL databases
- MongoDB
- JSON
- HTML/XML
- Accessing Data Through APIs
- CSS
- Web Scraping

MODULE 3

Probability Statistics & AB Testing

Prework	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

This is a basic module that introduces the fundamentals of probability theory, when you'll learn about probability principles like combinations and permutations. You'll go on to learn about statistical distributions and how to create samples when distributions are known. By the end of this course, you'll apply your knowledge by running Monte Carlo simulations and AB tests.

What is Covered in Module 3?

- Combinatorics
- Probability Theory
- Statistical Distributions
- Bayes Theorem
- Naive Bayes Classifier
- Sampling Methods
- Monte Carlo Simulation
- Hypothesis Testing
- AB Testing

MODULE 4

Statistical Modeling

Prework	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

We'll cover how and when regression models can be used to transform data into insights. You'll learn about both linear and logistic regression and the algorithm behind regression models. By the end, you'll be able to evaluate the result of regression models and extend them to interaction effects and polynomial features. To compare the performance of models built, you'll dive deeper into model evaluation and the bias-variance trade-off.

What is Covered in Module 4?

- Linear Algebra
- Linear Regression and extensions
- Polynomials
- Interaction effects
- Logistic regression
- Optimization Cost Function
- Gradient Descent
- Maximum Likelihood Estimation
- Time Series Modeling
- Regularization and Model Validation

MODULE 5

Machine Learning & Big Data

Prework	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

In Module 5 you'll learn how to build and implement machine learning's most important techniques. You'll take your first steps into classification algorithms through supervised learning techniques such as Support Vector Machines and Decision Trees. Additionally, you'll learn how to build even more robust classifiers using ensemble methods like Bagged and Boosted Trees, as well as Random Forests. Next, you'll move on to unsupervised learning techniques such as Clustering, and dimensionality reduction techniques like Principal Component Analysis.

What is Covered in Module 5?

- Distance Metrics
- K Nearest Neighbors
- Clustering
- Decision Trees
- Ensemble Methods
- Dimensionality Reduction
- Pipeline Building
- Hyperparameter Tuning
- Grid Search
- Scikit-Learn

MODULE 6

Deep Learning & Natural Language Processing

Prework	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

In the final module, you'll learn how to use regular expressions in Python and how to manage string values, analyze text, and perform sentiment analysis. Additionally, you'll get an in-depth overview of deep learning techniques, densely connected neural networks for high-performing classification performance, convolutional neural networks for image recognition, and recurrent neural networks for sequence modeling. You'll also learn about techniques to evaluate performance and to optimize and regularize model performance.

What is Covered in Module 6?

- Neural Networks
- Convolutional Neural Networks
- Ngrams
- POS Tagging
- Text Vectorization
- Context-Free Grammars
- Neural Language Toolkit
- Regular Expressions
- Word2Vec
- Text Classification

CURRICULUM OVERVIEW

Final Project

Prework	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

In our final project, you'll work individually to create a large-scale data science and machine learning project. This final project provides an in-depth opportunity for you to demonstrate your learning accomplishments and get a feel for what working on a large-scale data science project is really like.

You and your fellow students will each pitch three different ideas and then decide on your final project with your instructors. Instructors advise on projects based on difficulty and feasibility given the course's time constraints. At the end of the project, you'll receive a grade based on various factors.

Upon project completion, you'll know how to construct a project that gathers and builds statistical or machine learning models to deliver insights and communicate findings through data visualisation and storytelling techniques.

Contact Us

For more information, please check out our website at **www.flatironschool.com** or contact us at **admissions@flatironschool.com**

t

// FLATIRON SCHOOL