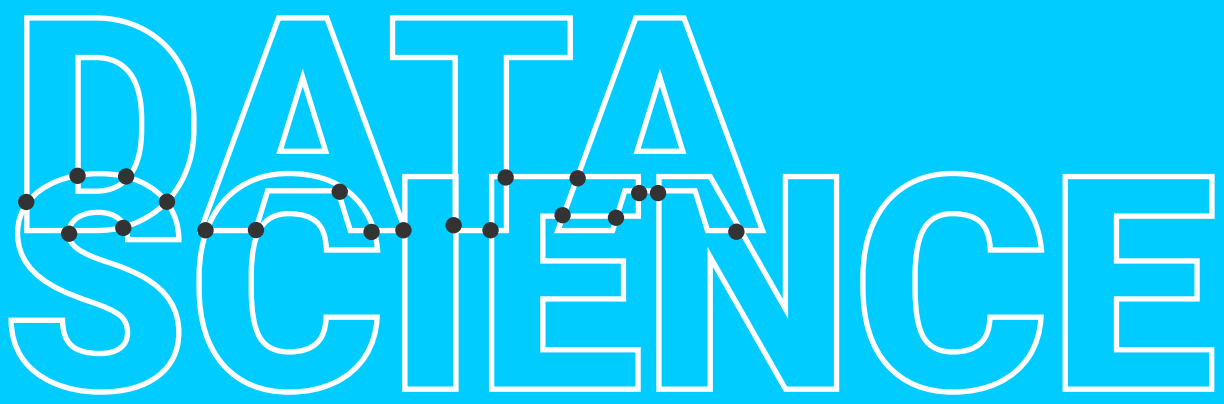
// FLATIRON SCHOOL

Last updated 04.02.18



Immersive Data Science Bootcamp

SYLLABUS & PROGRAM GUIDE

TABLE OF CONTENTS

PROGRAM OVERVIEW

Flatiron School's Immersive Data Science Bootcamp

This program will provide students 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's NYC campus, students will learn how to gather data, apply statistical analysis to answer questions with that data, and make their insights and information as actionable as possible. Our pedagogy ensure not only job readiness for today's market, but the aptitude and skills to keep learning and stay relevant.

What will students 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 proposal 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?

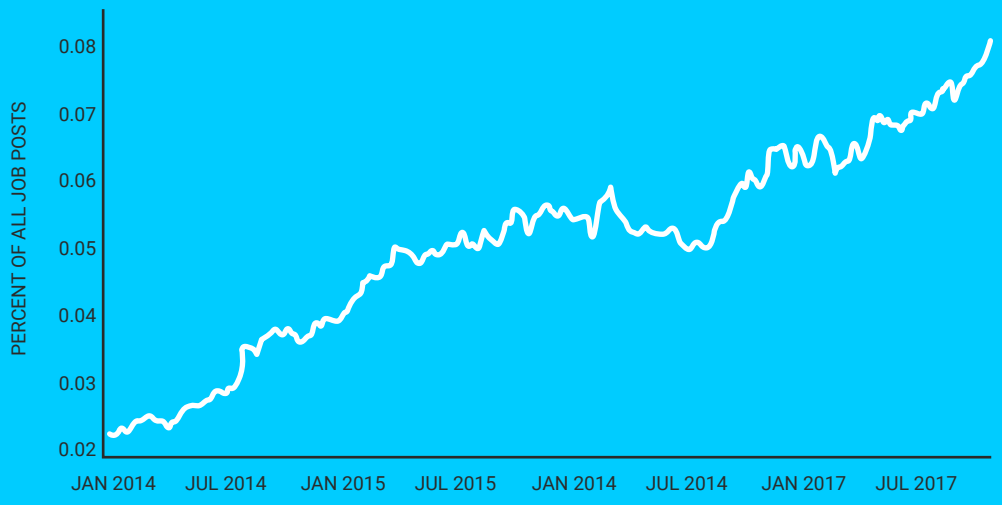
- Classes meet 5 days a week, full-time; Monday-Friday
- Classes will typically begin at 9:00am, with a 90-minute break for lunch around 12:30pm, followed by continued lectures and lab work on-campus through the late afternoon
- Classes meet on campus, in-person at our 11 Broadway location in New York City

WHY DATA SCIENCE?

Why is this course revelant?

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 4 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, so have the positions within the field become more nuanced. After completing our NYC Data Science Bootcamp, students will not only be able to secure a job as a data acientist, 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 students the breadth and depth needed to become well-rounded data scientists. Students also leave with an understanding of how to discover new techniques as their career progresses.

Every 3 weeks students are introduced to a new module that builds off the learnings of the previous section while allowing students enough time to dive into each area for a thorough understanding of the subject matter.

MODULE	HOURS	DAYS
Module A: Data Exploration & Analysis	105 Hours (35hrs/week)	15 Days (3 Weeks)
Module B: Probability & Statistics	105 Hours (35hrs/week)	15 Days (3 Weeks)
Module C: Machine Learning - Big Data	105 Hours (35hrs/week)	15 Days (3 Weeks)
Module D: Machine Learning - Deep Learning	105 Hours (35hrs/week)	15 Days (3 Weeks)
Module E: Data Science Advanced Projects	105 Hours (35hrs/week)	15 Days (3 Weeks)

Getting Started

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

CURRICULUM

Module A: Data Exploration and Analysis

Students will gain an overview into skills required of data scientist: data gathering and cleaning, analysis using probability and summary statistics, and presentation of information with visualization libraries. The focus will be on gathering and cleaning data.

To gather and clean data, students will learn fundamental concepts in programming using Python and SQL. Topics will include writing, functions and object orientation, scraping and regular expressions, and SQL.

Students will learn how to go from question requirements to targeted data mining. In doing so, students will learn about using experimental design and issue trees to turn problem requirements into quantifiable pieces of work.

From there, students will learn how to use targeted data mining with tools such as Pandas, probability, and summary statistics to answer their questions. Students will also learn about plotting and communicating results with visualisation tools, such as Seaborn.

Outcome of Module A

- A solid foundation in cleaning and gathering data with Python, Pandas, and SQL
- How to communicate results using a visualization tool, such as Seaborn
- Understand how to go from problem requirements to actionable steps with issue trees and experimental design

CURRICULUM

Module B: Probability and Statistics for Data Science

Having learned how to gather and explore data with Python and SQL students can go deeper into analysing that information with statistics. Students will learn about Bayesian and Frequentist statistics. Students will learn binary classifications and how to evaluate the accuracy classifier system by using confusion matrices, false positives and false negatives, the true positive rate with ROC curves.

Students will learn about repeated random sampling with Monte Carlo simulations and Markov Process. Then students will revisit experimental design techniques and apply their deeper statistical knowledge to A/B testing a website.

Students will go into regression analysis, learning about linear and logistic regression. In building regression models, students will learn about penalization terms, preventing overfitting through regularization and using cross validation to validate regression model.

Finally, students will learn about solving linear regression with gradient descent.

Outcome of Module B

- Understand the difference between Frequentist and Bayesian, and how to apply each
- How to build and validate regression models
- How to best fit a linear regression to sample data using ordinary least squares, and apply gradient descent to ordinary least squares

CURRICULUM

Module C: Machine Learning - Regression Optimization & Big Data

Students will move into supervised learning, non-parametric algorithms like k-nearest-neighbors and support vector machines. Students will also learn about decision tree learning and how it can be applied to classification and regression tree analysis.

Students will learn about various techniques in decision trees such as bagging and boosting to construct more than one decision tree, as well as how to conduct time series analysis using Pandas.

Students will be introduced using threading and multiprocessing to work with big data. In doing so, students will learn about Apache Spark and Apache Spark on AWS.

Outcome of Module C

- How to use disparate large data sets to build classification engines

CURRICULUM

Module D: Machine Learning - Classification and Deep Learning

Students will move onto unsupervised learning techniques such as clustering techniques like k-means and hierarchical clustering. Students will also learn how to simplify their machine learning models by using non-negative matrix factorization and principle component analysis. Students will learn about how to accommodate for imbalanced data in their machine learning models.

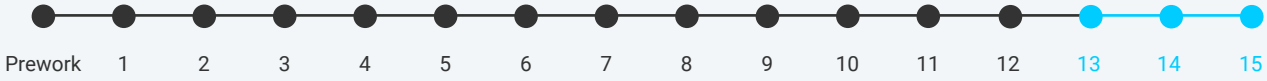
Students will move on to building recommender algorithms using collaborative filtering, matrix decomposition, clustering, and deep learning approaches.

Outcome of Module D

- Deep understanding of unsupervised techniques
- How to implement supervised or unsupervised learning techniques to build recommender systems

CURRICULUM

Module E: Data Science Advanced Project



Students work in groups of two to create a large scale data science and machine learning project. This course provides an in-depth opportunity for students to demonstrate their learning accomplishments and get a feel for what working large scale data science project is really like.

Teams are divided randomly and begin with an initial pitch session. Each team pitches three different ideas, from which instructors will choose one that the team will work on. Instructors choose projects based on difficulty and feasibility given the time constraints of the course.

At the end of the course, each team member will receive a grade based on various factors. Students receive a grade on code quality and project difficulty from their instructors as well as a separate grade from peer evaluations. 70% of the grade comes from instructor evaluations, and 30% come from peer evaluations.

Outcome of Module E

- How to construct a project that gathers, explores, builds statistical or machine learning models to deliver insights and communicate findings with 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