

# DATA SCIENCE

## Immersive Data Science Bootcamp

---

### SYLLABUS & PROGRAM GUIDE

# TABLE OF CONTENTS

- 3 Program Overview
- 4 Why Data Science
- 5 Curriculum Overview & Prework
  - 6 Module A: Exploratory Data Analysis and Descriptive Statistics
  - 7 Module B: Advanced Data Retrieval and Statistical Analysis
  - 8 Module C: Models for Machine Learning
  - 9 Module D: Advanced Topics, Big Data and Deep Learning
  - 10 Module E: Advanced Data Science Project
- 11 Contact Us

# 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 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 ensures not only job readiness for today's market, but also the aptitude and skills to keep learning and stay relevant.

### What will students learn?

- The foundations of Python as a programming language, and how to use Python for Data Science
- How to retrieve data from outside sources and organize data using Python
- Learn about the foundations of statistics to draw accurate conclusions about your data
- Build powerful machine learning models that create insights into your data
- Application and usage of Big Data
- Use deep learning models to analyze "nontraditional" data sources like images and music
- Presentation techniques to better share conclusions about approach and analysis to key stakeholders

### When and where does the course meet?

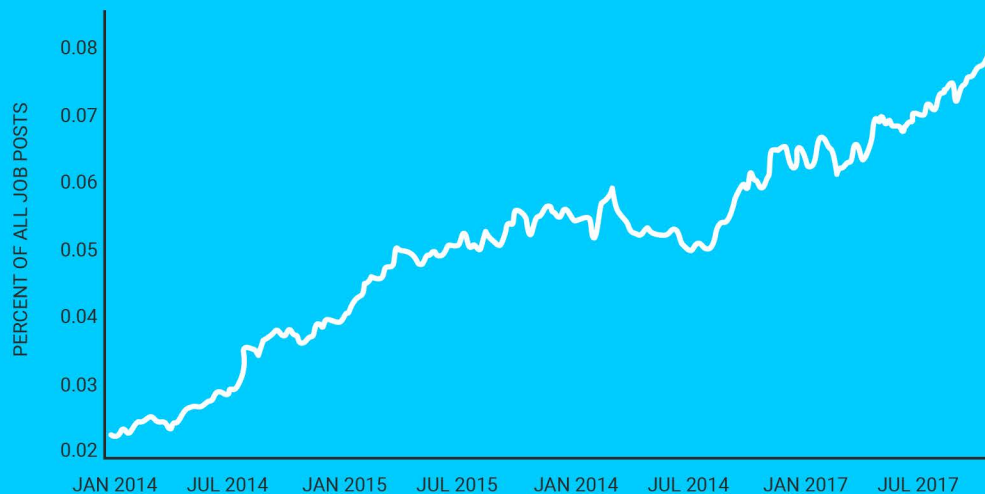
- Classes meet 5 days a week, full-time; Monday-Friday
- Classes will typically begin at 9:00am, with a 90-minute break for lunch around 12:30pm, followed by continued lectures and lab work on-campus through the late afternoon
- Classes meet on campus, in-person

# 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 4 years, the rise of job opportunities for Data Scientists has increased substantially.



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 Immersive Data Science Bootcamp, students will not only be able to secure a job as 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

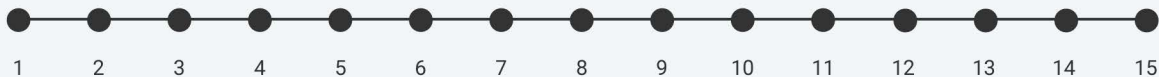
## CURRICULUM OVERVIEW

From Python to Machine Learning, our 15-week data science training program gives 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: Exploratory Data Analysis and Descriptive Statistics	150 Hours (50hrs/week)	15 Days (3 Weeks)
Module B: Advanced Data Retrieval and Statistical Analysis	150 Hours (50hrs/week)	15 Days (3 Weeks)
Module C: Models for Machine Learning	150 Hours (50hrs/week)	15 Days (3 Weeks)
Module D: Advanced Topics, Big Data and Deep Learning	150 Hours (50hrs/week)	15 Days (3 Weeks)
Module E: Advanced Data Science Project	150 Hours (50hrs/week)	15 Days (3 Weeks)
Program Total:	750	

### 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. .

# CURRICULUM

## Module A: Exploratory Data Analysis and Descriptive Statistics



Students will gain an overview of the skills required of a data scientist: understanding data structures, data gathering and data cleaning. We'll cover probability and summary statistics, and presentation of information with visualization libraries. Students will learn the foundations of important Data Science libraries such as Pandas, NumPy and Matplotlib.

To gather and clean data, students will learn fundamental concepts in programming using Python and SQL. Topics will include writing, functions, object orientation, and writing your own SQL queries.

From there, students will dive into the first bit of statistical analysis learning about what statistical distributions are, and how to run a linear regression model. By the end of the module, students will be able to run their first data science project using multiple linear regression.

### Outcome of Module A

- A solid foundation in cleaning and gathering data with Python, Pandas, NumPy, and SQL
- How to communicate results using a visualization tool, such as Matplotlib
- Understand how to go from data to insights using a linear regression model

# CURRICULUM

## Module B: Advanced Data Retrieval and Statistical Analysis



Having learned how to gather and explore data with Python and SQL, and linear regression, students can now dive deeper into understanding how regression algorithms project. Students will learn how the Ordinary Least Squares method and gradient descent in a linear regression project.

Students will dive deeper into statistics and probability, learning about Bayes Theorem, the difference between bayesian and frequentist statistics and the central limit theorem.

Next, students will dive a little deeper into SQL, and learn the basics of HTML, XML and JSON to be able to access data from various sources using APIs, and perform Web Scraping. Students will then learn about repeated random sampling, and will revisit experimental design techniques and apply their deeper statistical knowledge to A/B testing a website.

### Outcome of Module B

- Understand the difference between Frequentist and Bayesian, and how to apply each of them
- How to best fit a linear regression to sample data using ordinary least squares, and apply gradient descent to ordinary least squares
- How to build and validate regression models, and improve them
- Understand how Hypothesis testing works and know how to AB test a website



# CURRICULUM

## Module C: Models for Machine Learning



In this module students will move into the wondrous world of machine learning! Students will take their first steps into classification algorithms using logistic regression models, and will work on towards some more hard-core supervised machine learning techniques such as support vector machines, and decision trees. Along with learning about various techniques in decision trees such as pruning, students will learn about ensemble methods such as bagging, boosting and random forests.

Next, students will move onto unsupervised learning techniques such as clustering techniques like k-means and hierarchical clustering, and dimensionality reduction techniques such as Principal Component Analysis.

Along with this, students will spend some time on another important area: time series modeling. This is where students will learn how to conduct time series analysis using Pandas.

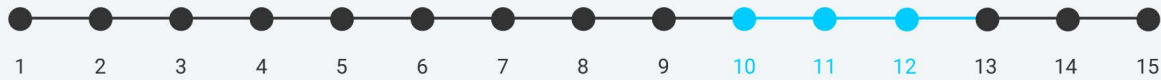
### Outcome of Module C

- How to use supervised learning techniques for classification
- How to used unsupervised learning techniques such as clustering techniques
- How to analyze time series data such as stock market prices, temperature, sales, ... over time



# CURRICULUM

## Module D: Advanced Topics, Big Data and Deep Learning



In this module, students will move on to advanced data science techniques: Natural language processing, PySpark and Deep Learning.

Students will learn how to use regular expressions in Python, and how to manage string values, analyze text and perform sentiment analysis.

Next, students will dive into the world of Big Data, where students will learn how to use PySpark, a distributed programming framework which makes working with huge datasets feasible.

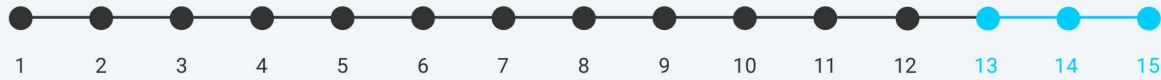
Students will get an in-depth overview of deep learning techniques, learning about densely connected neural networks, for highly performant classifiers, convolutional neural networks for image recognition, and recurrent neural networks, for sequence modeling. Students will learn about techniques to evaluate performance and to optimize and regularize model performance.

### Outcome of Module D

- Deep understanding of unsupervised techniques
- How to use regular expressions and perform Natural Language Processing
- How to use PySpark and understand how the framework is useful when working with big data

# CURRICULUM

## Module E: Advanced Data Science 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 on 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% comes 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](http://www.flatironschool.com) or contact us at [admissions@flatironschool.com](mailto:admissions@flatironschool.com)