// FLATIRON
SCHOOL

DATA
SCIENCE

Data Science Course

SYLLABUS & COURSE GUIDE

TABLE OF CONTENTS

COURSE OVERVIEW

Flatiron School's Data Science Course

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, 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 stake-holders

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

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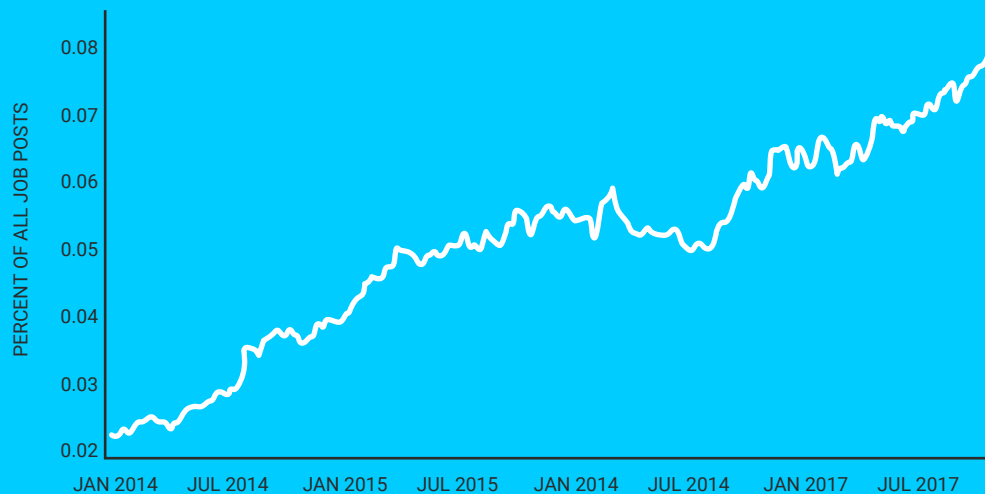
Note: The chart above offers a 7-day rolling mean of all Indeed job posts that featured "data science" or "data scientist" in the title across the world as a percentage of all job posts between January 1, 2014 and November 16, 2017. The data was pulled using Imhotep, Indeed's open source analytics platform.

As this area of expertise has grown, so have the positions within the field become more nuanced. After completing our Data Science Course, 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 2 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 1: Python for Data Science	100 hours	2 weeks
Module 2: Data Engineering for Data Science	100 hours	2 weeks
Module 3: Probability, Samping and AB Testing	100 hours	2 weeks
Module 4: Statistical Modeling	100 hours	2 weeks
Module 5: Machine Learning & Big Data	100 hours	2 weeks
Module 6: Deep Learning & Natural Language Processing	100 hours	2 weeks
Program Total	600 hours + projects + homework	

Getting Started

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

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.

Module 1: Python for Data Science

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

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

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

CURRICULUM

Module 2: Data Engineering for Data Science

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
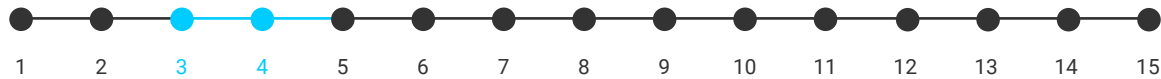
In this course, students will learn about data structures, relational databases and ways to retrieve data. Students will learn about the fundamentals of SQL for data querying for structured databases, as well as NoSQL (and MongoDB) for non-relational databases. Furthermore, they'll learn the basics of HTML, XML and JSON to be able to access data from various sources using APIs, and 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, Sampling and AB Testing

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

A basic course that introduces the fundamentals of probability theory, where students will learn about probability principles such as combinations and permutations. Students will go on and learn about statistical distributions and how to create samples when distributions are known. By the end of this course, students will apply their knowledge 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

CURRICULUM

Module 4: Statistical Modelling

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

In this course, students will learn how and when regression models can be used to transform data into insights. Students will learn about both linear and logistic regression and the algorithm behind regression models. Students will be able to evaluate the result of regression models and extend them to for interaction effects, and polynomial features. To compare the performance of models built, students will 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

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

In this course, students will learn how to build and implement the most important machine learning techniques. Students will take their first steps into classification algorithms through supervised learning techniques such as Support Vector Machines and Decision Trees. Additionally, students will learn how to build even more robust classifiers using ensemble methods such as Bagged and Boosted Trees, and Random Forests. Next, students will move onto unsupervised learning techniques such as Clustering, and dimensionality reduction techniques such as 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: Machine Learning & Big Data

In the final module, students learn how to use regular expressions in Python, and how to manage string values, analyze text and perform sentiment analysis. Additionally, students will get an in-depth overview of deep learning techniques, learning about densely connected neural networks, for high-performing classification performance, 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.

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

Final Project

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

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

Each student pitches three different ideas, from which instructors help determine which will be the best final project. Instructors advise on projects based on difficulty and feasibility given the time constraints of the course. At the end of the course, each student will receive a grade based on various factors.

Upon completion of the project, students know 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