

DATA SCIENCE BOOTCAMP **CURRICULUM**



METIS®



Introduction

The Metis Data Science Bootcamp is a full-time, twelve-week intensive experience that hones, expands, and contextualizes the skills brought in by our competitive student cohorts, who come from varied backgrounds. Incorporating traditional in-class instruction in theory and technique, students use real data to build a five-project portfolio to present to potential employers. Upon graduating, students have completed rigorous training in machine learning, programming in multiple languages (Python, Unix, JavaScript), data wrangling, project design, and communication of results for integration in a business environment.

Parallel to this core classroom work is a supporting careers curriculum created and implemented by our Careers Team, which works with each student to secure employment rapidly after graduation with a compatible employer.

Each project is a start-to-finish application of the skills needed to be a well-rounded, competitive practitioner in the data science workforce. Each highlights the skills needed in every “facet” of data science: project design, data acquisition and storage, tool selection, analysis, interpretation, and communication. In succession, the projects deepen in both difficulty and independence.



Online Pre-work

Once students are enrolled in the bootcamp, they are granted immediate access to our pre-work materials, a structured program of 25 hours of academic pre-work and up to 35 hours of set-up is designed to get admitted students warmed up and ready to go. All exercises must be completed before the first day of class.

Students are also invited to join their cohort's Slack communication channel, where they meet their TA, get support on pre-work assignments, and will be held accountable to the pre-work schedule of deadlines.

PRE-WORK TOPICS

GitHub

Software & package installation

Code editor selection & familiarity

Command line (OS X/bash)

Python (intermediate & advanced)

Linear Algebra

Statistics

Optional resources for review:

Pandas, SQL, HTML/CSS/JavaScript



Twelve-Week Onsite Bootcamp

After completing pre-work, the cohort convenes on-site for the full bootcamp experience. The first eight weeks are spent learning the theory, skills, and tools of modern data science through iterative, project-centered skill acquisition. Over the course of four data science projects, we “train up” different key aspects of data science, and results from each project are added to the students’ portfolios. In the final four weeks, students build out and complete individual Passion Projects, culminating in a Career Day reveal of their work to representatives from our Metis Hiring Network.

FLOW OF THE DAY

Mornings in the classroom // 9:30am – 12:00pm

- *Pair programming exercises*
 - *Interactive lectures*
-

90-minute lunch // 12:00pm – 1:30pm

- *Long enough to take a coffee meeting, eat a great lunch, and/or just rest your brain*
-

Working afternoons // 1:30pm – 6:00pm

- *Investigation presentations*
 - *Challenges and project work*
 - *Senior Data Scientist instructors and Data Scientist TAs onsite for help*
-

More

- *Careers curriculum*
 - *Guest speakers*
 - *Hosted Meetup events*
 - *Site visits to select hiring partners*
-



The background of the slide features a photograph of students in a classroom or lecture hall. A large, solid blue triangle is positioned in the upper right corner, partially obscuring the image. The text 'CURRICULUM DETAIL' is overlaid on the left side of the image.

CURRICULUM **DETAIL**

WEEK 1

UNIT ONE

Introduction to the Data Science Toolkit

Students jump right in, working with real data as they become acclimated with the core toolset that is used for the remainder of the bootcamp. Starting with a dirty dataset of turnstile entrances and exits from the New York MTA, students use Python, pandas, and matplotlib to find and present patterns in the data. Students create a blog using Jekyll and GitHub Pages to present findings from this and future projects.

TOPICS

Python

Data wrangling and EDA (Exploratory Data Analysis) with Python, pandas, and matplotlib

Git and GitHub workflow: branching and pull requests

Bash shell

GitHub Pages & Jekyll



PROJECT #1:

CODENAME BENSON

Students work in small groups using MTA turnstile data, which they clean themselves, to find patterns in the volume of street traffic. Since no data project exists in a vacuum, each group creates a theoretical client and use case for its findings, brainstorming as a unit and using design thinking principles. Projects are presented to the class and published as posts on each student's new GitHub Pages blog.

WEEK 2

UNIT TWO: PART 1

Fundamentals, Regression, and Web Scraping

The basic workflow is now in place, and we dive into some deeper content. The second project focuses on regression and also touches on fundamental concepts for statistics and probability. For data acquisition, we tackle web scraping (used to gather data for the second project), stored in flat files using fundamental Python input/output. With an eye on our goal to develop well-rounded data scientists, we go over design thinking and the iterative design process, so all efforts have the maximum impact on the intended audience.

TOPICS

Probability theory (discrete, continuous)

Hypothesis testing

Regression & model evaluation in statsmodels and scikit-learn

Web scraping with BeautifulSoup and Selenium

Iterative design and design thinking



CAREER SERVICES

First One-on-One Meeting with Career Advisor

Students have their first of three officially-scheduled meetings with their Career Advisor, all of which take place during and after the bootcamp. Students can discuss topics like resumes, salary negotiation, mock interviews, company introductions, how to craft messages to hiring managers and recruiters, soft skill interviewing, and more.

Speaker Series begins (Weeks 2-9)

During the bootcamp, students are exposed to a number of speakers, including ones from our Hiring Network. These speakers provide deep-dives into specific skills and/or career coaching advice and represent excellent opportunities to expand your data science knowledge and network.

WEEK 3

UNIT TWO: PART 2

Advanced Regression and Communicating Results

Continuing with the topics from Week 2, we introduce Bayes Theorem, another fundamental skill in statistical reasoning. Our regression models are refined as we learn about regression model assumptions, transformations, and overfitting. Cross validation and regularization methods help to refine models further, and in preparing for the upcoming Project Luther, we deepen our plotting skills in matplotlib and seaborn.

TOPICS

Machine learning concepts: overfitting and train/test splits

Introduction to Bayes Theorem

Linear Regression: model assumptions, regularization (lasso, ridge, elastic net)

Advanced plotting with matplotlib and seaborn



PROJECT #2:

CODENAME LUTHER

In the second project, we introduce every single facet of data science that will come into play for all future projects, including design, data acquisition, algorithms & analysis, tool selection, and interpretation/communication. Students use regression to predict box office gross, using data they scrape themselves (from web sources of their choice), which they then store in flat files. Students make decisions about regularization and evaluate models using statsmodels or scikit-learn. Each student interprets and presents their individual work to a “client” who would be interested in the findings.

UNIT THREE: PART 1

Machine Learning Concepts, Classification, Databases

The third unit broadens concepts learned in regression by extending to the parent family of supervised learning. Students learn a suite of classification algorithms and concepts of bias-variance tradeoff. Since they work in groups for the upcoming Project McNulty, we create cloud servers to store project data, this time in SQL databases.

TOPICS

Classification and regression algorithms: K-nearest neighbors, logistic regression, support vector machines (SVM), decision trees, and random forest

Databases: SQL

Machine learning concepts: bias-variance tradeoff, classification errors

Other tools: creating and provisioning cloud servers



CAREER SERVICES

LinkedIn Workshop

Learn to build a LinkedIn profile that is specifically suited for data science jobs. Students incorporate their previous work experience and learn how to best position themselves for competitive opportunities.

UNIT THREE: PART 2

Supervised Learning, More Topics in Machine Learning

Continuing from Week 4, students add several more supervised learning algorithms to their arsenals, which they apply to their project data in the afternoons. Machine learning topics taught this week involve deeper use of scikit-learn functionality, introducing automated methods of feature selection, options for estimation including stochastic gradient descent, and advanced metrics for model evaluation. Finally, new probability distributions are added to our growing toolbox of fundamental statistical concepts.

TOPICS

More supervised learning algorithms: Naive Bayes, Categorical MLE, Poisson Regression, Neural networks, and Deep learning

Machine learning: automated feature selection, stochastic gradient descent, advanced model evaluation

Fundamentals: binomial, bernoulli, and poisson distributions



CAREER SERVICES

Networking Workshop

We throw a mock networking event (attended only by members of the cohort) to help students learn how to navigate – and build confidence to attend – industry events and Meetups.

Resume Workshop

Learn how to craft a professional resume that is ready to present to employers by Career Day.

UNIT THREE: PART 3

JavaScript and D3.js

This week provides the final component for the upcoming Project McNulty, in which students finalize their analyses and create interactive dashboards to display results. We diverge from our all-Python diet to take on JavaScript and the data visualization toolkit, D3.js. The end product is an interactive and professional custom dashboard. Creating it presents meaningful exposure to the soup-to-nuts basics of web-based data presentation.

TOPICS

JavaScript and D3.js

Full stack in a nutshell: connecting a front end and a back end with Python Flask

Dashboard design



PROJECT #3:

CODENAME MCNULTY

This time, students get a break from data acquisition, and store data from one of the UCI repository datasets in an SQL database. Using supervised learning, they create a dashboard for a company or data product using D3.js, presenting predictions made on their data. These dashboards pull from a database API they create in Flask to serve data into their interactive visualizations.

UNIT FOUR: PART 1

Unsupervised Learning, NLP, Dimension Reduction, NoSQL

We dive into unsupervised learning and natural language processing (NLP), and go deep into core machine learning concepts like the curse of dimensionality, dimension reduction, vector spaces, and distance metrics. Finally, to support the upcoming Project Fletcher, we introduce NoSQL databases and RESTful APIs, as well as begin culling project data from web APIs to be stored in MongoDB.

TOPICS

Data & databases: RESTful APIs, NoSQL databases, MongoDB, pymongo

Natural language processing: textblob, NLTK, chunking, stemming, POS tagging, tf-idf

Unsupervised learning: overview & introduction, K-means

Machine learning topics: curse of dimensionality, dimension reduction, PCA, SVD, LSI



CAREER SERVICES

Interview Preparation Workshop

Students learn the dos & don'ts of the interview process, including important tips to help achieve successful interviews.

UNIT FOUR: PART 2

Natural Language Processing (NLP)

Finishing up the main 8-week content component of the bootcamp, the final week of Project Fletcher continues with Natural Language Processing (NLP) tools including topic modeling, latent dirichlet allocation, and word2vec. We add several more unsupervised learning algorithms to our arsenal, and learn formally about varieties of, and considerations in, choosing distance metrics.

TOPICS

NLP: Topic modeling, LDA, word2vec

Unsupervised learning: hierarchical clustering, DBSCAN, Mean Shift, Spectral

Machine learning topics: distance metrics



PROJECT #4:

CODENAME FLETCHER

For the last (and most lightly) guided project, students work individually and have very few constraints for the design. They must keep all facets of a data science project in mind, however, including designing their analysis thinking of a specific audience and use case, choosing and collecting their data (which must include text data and data sourced from an API), storing (at least some of) it in a NoSQL database, using NLP and unsupervised learning techniques in their analysis, and interpreting and presenting their findings in a way that makes sense for their use case.

UNIT FIVE

Big Data Tools and Passion Project

During the final four weeks, students transition into full-time focus on their final, passion projects. Week 9 includes final lectures and challenges for big data tools and techniques, but for the rest of Unit 5, they work with instructors to build out their passion project for Career Day. Each hones their presentation over many iterations to showcase the work in its best light when it counts the most!

TOPICS

Big data tools: Hadoop, Hive, Pig, Spark, Cloud servers 2

Algorithms: MapReduce

Project and time management: iterative design, minimum viable products



CAREER SERVICES

WEEK 9: Data Science Career Paths Workshop

Students get their burning questions answered around differences in job titles, how skills vary by industry, the impact of an advance degree, and more.

Salary Negotiations Workshop

Learn the latest data scientist salary information and walk through salary negotiation best practices.

WEEK 10: Second One-on-One Meeting with Career Advisor

Mock Interviews

Toward the end of the bootcamp, students participate in a mock technical interview conducted by data scientists from the Metis Hiring Network.

They have the opportunity to “whiteboard” and respond to typically asked data science questions. Afterward, they get feedback on their performance.

WEEK 11: Career Day Preparation

Leading up to Career Day, students have multiple opportunities to demo their final project in front of Metis staff, students, and instructors – and all receive personalized feedback to help them better prepare for Career Day.

WEEK 12: Career Day

During the final week of the bootcamp, we host Career Day, at which students are introduced to companies actively hiring for data scientists. Each presents their final project and networks with attendees throughout the event. Participating companies have included Capital One Labs, Booz Allen Hamilton, Spotify, Zynga, and HBO.



PROJECT #5:

CODENAME KOJAK (AKA, PASSION PROJECT)

For the last (and most lightly) guided project, students work individually and have very few constraints for the design. They must keep all facets of a data science project in mind, however, including designing their analysis thinking of a specific audience and use case, choosing and collecting their data (which must include text data and data sourced from an API), storing (at least some of) it in a NoSQL database, using NLP and unsupervised learning techniques in their analysis, and interpreting and presenting their findings in a way that makes sense for their use case.



CAREER SERVICES

Post-Graduation Support

Upon graduating, students get access to the Metis Alumni Network on Slack (including our exclusive Job Postings Channel), the Employ hiring app, and our Alumni Resources folder. Until employed, you also receive tailored information regarding open job opportunities that fit your interests and skills. Additionally, within four weeks of graduation, you can schedule another meeting with your Career Advisor.



More About Projects

Data science projects can be divided into the useful dimensions of domain, design, data, algorithms, tools, and communication. Each unit covers certain content from several domains, which are reinforced in that unit's project.

The rigor with which we attack the topics covered in the bootcamp allows us to sleep soundly at night. We feel confident in saying that our graduates haven't simply learned about the tools data scientists use, but rather, by the time they leave our classroom, they are data scientists. They are ready to approach the problem space in their new careers, assemble the suite of tools and methods to answer insightful questions, and communicate comprehensible results. They are competent, capable, and confident – and they are ready to work.





Objectives

Upon graduating from the Metis Data Science Bootcamp, students are prepared for positions on teams hiring for data scientists or data analysts. This means a student will:



Have a fluid understanding of, and practical experience with, the process of designing, implementing, and communicating the results of a data science project.



Be a capable coder in Python and at the command line, including the related packages and toolsets most commonly used in data science.



Understand the landscape of data science tools and their applications, and will be prepared to identify and dig into new technologies and algorithms needed for the job at hand.



Know the fundamentals of data visualization and will have experience creating static and dynamic data visuals using JavaScript and D3.js.



Have introductory exposure to modern big data tools and architecture, such as Hadoop and Spark – and they will know when these tools are necessary and will be poised to quickly train up and utilize them in a big data project.



