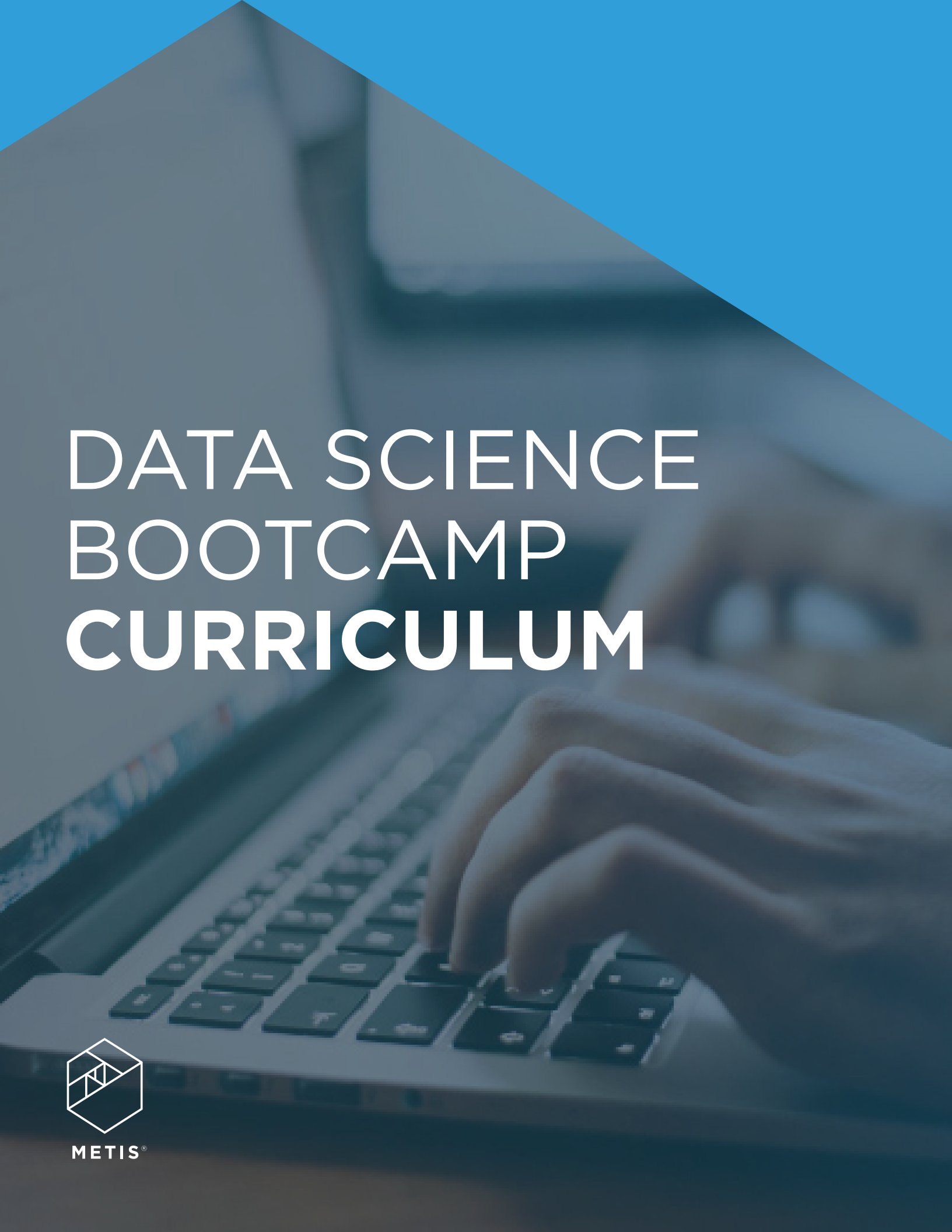
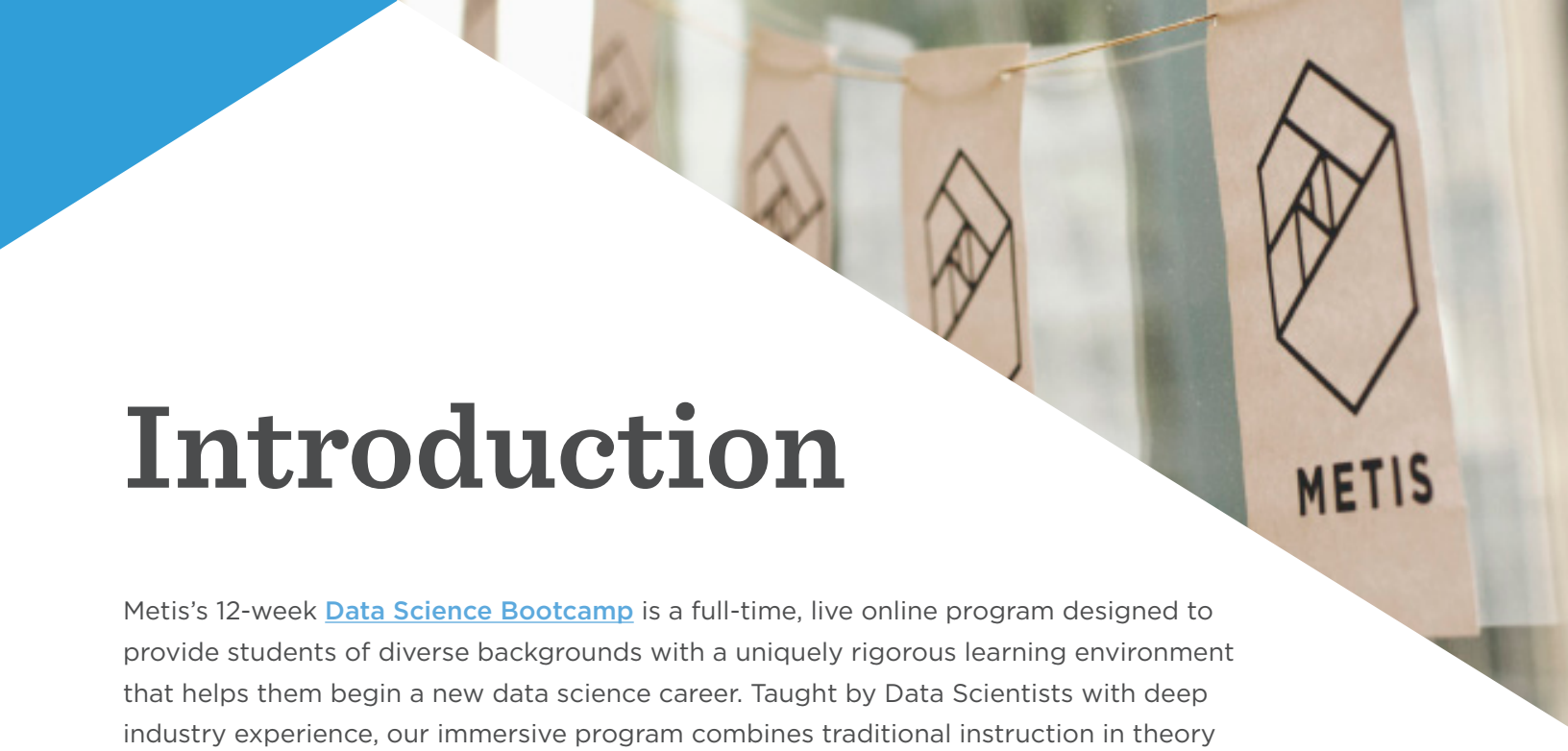


DATA SCIENCE BOOTCAMP **CURRICULUM**

METIS®



Introduction

Metis's 12-week Data Science Bootcamp is a full-time, live online program designed to provide students of diverse backgrounds with a uniquely rigorous learning environment that helps them begin a new data science career. Taught by Data Scientists with deep industry experience, our immersive program combines traditional instruction in theory and technique with a project-based approach, through which students apply their new knowledge to build a five-project portfolio using real data that they can present to potential employers.

Each project is a start-to-finish application of the skills needed to be a well-rounded, competitive practitioner in the data science workforce. Projects are carefully designed to highlight 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, leading up to a Passion Project that students can leverage during their job search and in their Graduate Directory profiles at the end of the program. This project acts as the final piece in their newly robust online portfolio.

Along with the project work, students also complete challenges and assessments that help reinforce concepts from lecture that are then used in the project work.

Parallel, and equally important, to this core classroom and project work is a supporting careers curriculum created and implemented by our own Career Support Team. Career Advisors work with every student to secure employment rapidly after graduation with a compatible employer.

Upon successful completion of the program, graduates are awarded a Data Science Certificate. They will have completed rigorous training in machine learning, programming in multiple languages (Python, Bash, SQL, PySpark), data wrangling, project design, and communication of results for integration in a business environment. With their new skills and project portfolio, graduates will be ready for the field of Data Science. Job titles range based on the industry and prior experience, but typically include Data Scientist, Data Analyst, Data Engineer, Jr. or Associate Data Scientist, Data Science Consultant, and Machine Learning Engineer.

Online Pre-Work

Once students are enrolled in the bootcamp, they are granted immediate access to our pre-work materials. This structured program of 30 hours of academic pre-work and 15-30 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 pre-work dedicated TA, get support on pre-work assignments, and will be held accountable to the pre-work schedule of deadlines.

PRE-WORK TOPICS

- GitHub and Git
- Software and package installation
- Code editor selection and familiarity
- Command line (OS X/bash)
- Python (intermediate & advanced)
- Linear Algebra
- Statistics

Twelve-Week Live Online Bootcamp

After completing pre-work, the cohort convenes in our live online classroom to kick off their bootcamp experience. The first nine weeks are spent learning the theory, skills, and tools of modern data science through iterative, project-centered work. Over the course of four data science projects, we train students in key aspects of data science and results from each project are added to their portfolios. In the final four weeks, students build out and complete individual Passion Projects, culminating in a recorded presentation of their work that will be shared with representatives from the Metis Hiring Network and beyond.

PROGRAM STRUCTURE & CURRICULUM OVERVIEW

WEEK 1 | PROJECT 1: Exploratory Data Analysis (EDA)

Complete your first data science project from start to finish. Use Jupyter notebooks for writing code, Git and GitHub for version control, the pandas Python package to perform exploratory data analysis, and the Matplotlib package to visualize results.

WEEKS 2-3 | PROJECT 2: Regression & Web Scraping

Begin learning the iterative design process. Use tools for web scraping to gather data, and go in-depth on regression theory and practice using modules from SciKitLearn and Matplotlib. Work on communicating results in the first individual project.

WEEKS 4-6 | PROJECT 3: Classification & Interactive Dashboards

Focus on relational databases and learn additional ways of obtaining, cleaning, and maintaining data. Learn the SQL language for querying relational databases. Explore the concepts of machine learning. Dive deep into algorithms for supervised learning including KNN, logistic regression, SVM, Naive Bayes, decision trees, random forests, and gradient boosting. Then, learn how to build interactive data visualizations using Business Intelligence tools.

(continued on next page)

WEEKS 7-8 | PROJECT 4: Natural Language Processing & Unsupervised Learning

Dig into text data and round out data acquisition methods with APIs and online database servers. Learn about NoSQL databases and start using MongoDB. Learn about NLP algorithms and how large amounts of data are handled, discussing parallel computing and Hadoop MapReduce. Begin to build the conceptual foundation of feature selection and feature extraction. Finally, explore unsupervised learning for clustering algorithms, covering K-means, hierarchical clustering, DBScan, as well as dimensionality reduction techniques such as Principal Component Analysis.

WEEKS 9-12 | PROJECT 5: Passion Project

Work full-time on your Passion Project, which is the capstone of your five-project portfolio. Gain exposure to advanced topics that may be used in the Passion Project, including modern applications of deep learning and big data. Develop a practical approach for applying data science ethics. Practice and record Passion Project presentations with coaching from instructors and career advisors. These will be added to your Graduate Profile and shared with prospective employers.

ACADEMIC PROGRAM OBJECTIVES

After completing this course, a student is expected to:

- Have a strong 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 be prepared to identify and dig into new technologies and algorithms needed for the job at hand.
- Know the fundamentals of data visualization and have experience creating static and dynamic visualizations for data and models.
- Have introductory exposure to modern big data tools and architecture such as the Hadoop stack, know when these tools are necessary, and be poised to quickly train up and utilize them in a big data project.

(continued on next page)

PROGRAM SCHEDULE

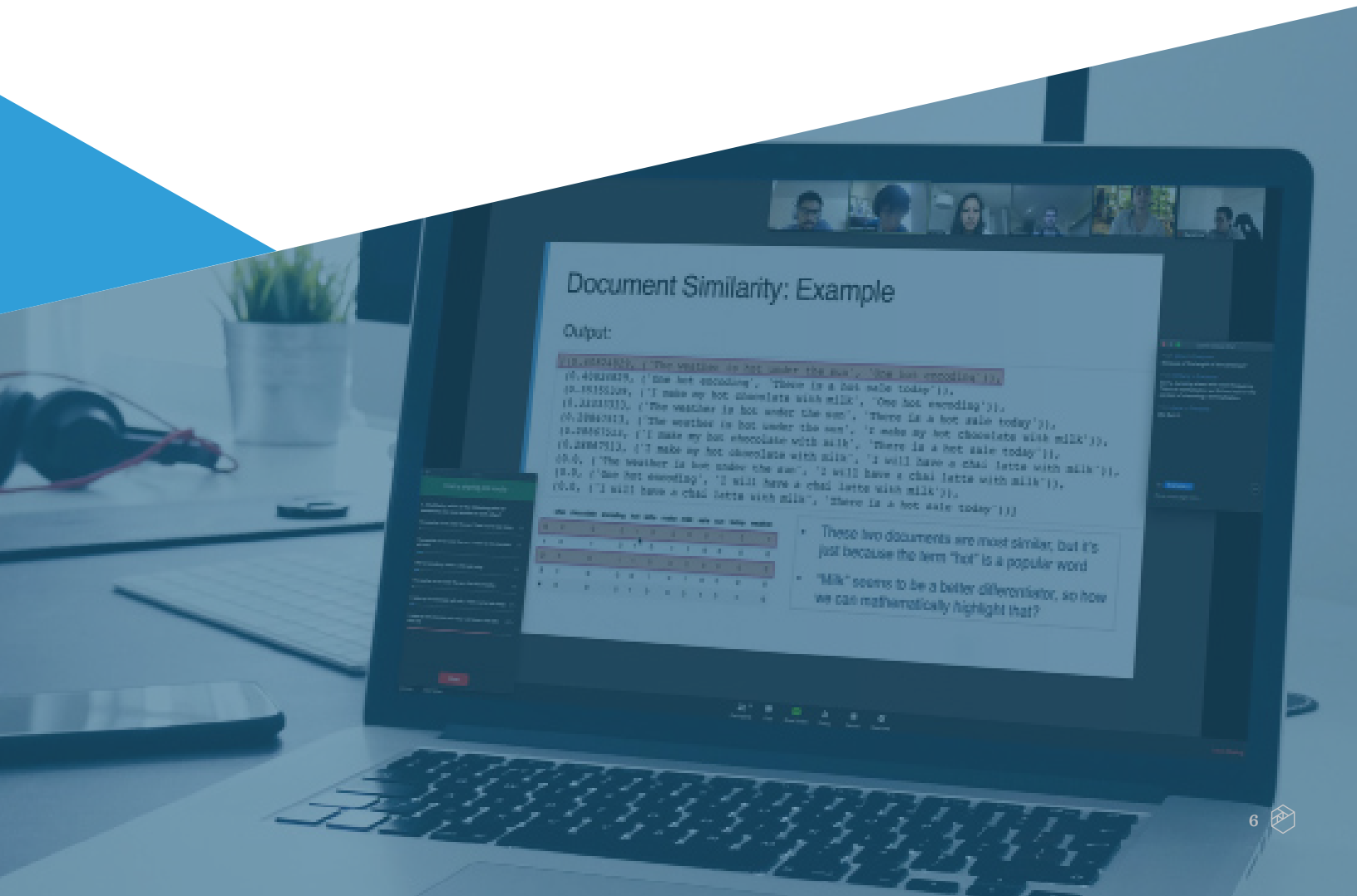
Monday - Friday, with two program schedules to choose from*

- 9:30am - 5:30pm ET
- 11:30am - 7:30pm ET

*includes independent project work and lunch time.

A typical day for a Metis student includes:

- Lectures
- Career workshops
- Pair programming
- Project work and presentations
- One-on-ones with instructors and career advisors
- Fun online gatherings with classmates



CURRICULUM **DETAIL**

METIS®

PROJECT 1

DURATION: 1 WEEK

Exploratory Data Analysis

Students work in small groups using data from New York's public transportation authority (MTA) on daily usage. You will clean the data and use it to find patterns, incorporating additional analyses and combining datasets as your group sees fit. To simulate real-world problem solving, 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 on each student's new blog.

WEEK 1 | UNIT ONE

Introduction to the Data Science Toolkit

Throughout your first week you will gain confidence in the Jupyter environment, start programming under version control, use commands from the pandas package to perform statistical analysis on your data, and visualize the results using the Matplotlib package.

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

PROJECT 2

DURATION: 2 WEEKS

Regression & Web Scraping

In the second project, we introduce all facets of data science that will come into play for all future projects, including design, data acquisition, algorithms & analysis, tool selection, and interpretation/ communication. You will use regression to predict a phenomenon of your choice, using data you scrape yourself that you will then store in flat files. You will have to make decisions about regularization and evaluate your models using statsmodels or SciKit-learn. At the end of the project, you will present your individual findings to the class.

WEEK 2 | UNIT TWO: PART 1

Linear Regression, Design Process & Web Scraping

After completing your first mini-project, students will gain a deeper understanding of linear models during week 2. You will learn about the iterative design process and how to use it to solve business questions. You will gain a thorough understanding of linear regression, its assumptions, and when to apply it. Additionally, you will learn about web scraping and practice web scraping to gather data for the second project.

TOPICS

- In-depth introduction to linear regression theory and application
- Machine learning concepts: over-fitting, train/test splits and cross-validation
- Regression & model evaluation in statsmodels and SciKit-learn
- Web scraping with BeautifulSoup and Selenium
- Object oriented programming principles

(continued on next page)

PROJECT 2

WEEK 3 | UNIT TWO: PART 2

Statistics Review, Intro Machine Learning & Communicating Results

This unit covers two major elements: diving deeper into machine learning, particularly supervised learning techniques, and developing skills for communicating results. Time series modeling will also be introduced this week. The packages used during this week include regression modules of Scikit-learn and Matplotlib in more depth. Choosing among the analysis methods and approaches to reporting results, you will finish your second project and present your findings.

TOPICS

- Introduction to time-series modeling
- Statistics review
- Hypothesis testing
- Introduction to Bayes' Theorem
- Linear regression regularization (LASSO, Ridge, elastic net)

PROJECT 3

DURATION: 3 WEEKS

Classification & Interactive Dashboards

This time, you will use a dataset of your choice and store it in a relational database to be queried with SQL. You will create a dashboard for a company or data product using interactive data visualizations, presenting predictions made on your data. These dashboards pull from a database API they create in Flask to serve data into their interactive visualizations. You will use a mixture of supervised and unsupervised learning techniques as appropriate for the project.

WEEK 4 | UNIT THREE: PART 1

Databases, Machine Learning Concepts, More Supervised Learning

This unit starts off with querying (relational) databases using SQL and covers more ways of obtaining, cleaning, and maintaining data. We then introduce models used for classification and supervised learning including logistic regression and KNN.

TOPICS

- **Classification and regression algorithms: K-nearest neighbors, logistic regression, support vector machines (SVM), Naive Bayes**
- **Relational databases and writing code to query them (SQL)**
- **Machine learning concepts: bias-variance tradeoff, classification errors, class imbalance**
- **Other tools: creating and provisioning cloud servers**

(continued on next page)

PROJECT 3

WEEK 5 | UNIT THREE: PART 2

Supervised Learning & Visualization

During this unit, we introduce the last of the fundamental supervised learning algorithms before covering more advanced techniques for supervised learning, including addressing class imbalance and using ensembling to improve models. Finally, you will learn how to deploy models to a web app and will create interactive, online visualizations.

TOPICS

- More supervised learning algorithms: Classification and regression trees, Random Forest
- Interactive data visualization using Business Intelligence tools
- Web development essentials including Javascript, HTML, and CSS
- Deploying models in production and full stack in a nutshell: connecting a front end and a back- end with Python's Flask package

WEEK 6 | UNIT THREE: PART 3

Introduction to NLP, NoSQL & Unsupervised Learning

You will finish your third project and further practice communicating the results of a full data science project. We then introduce the theory and practice of using NoSQL databases to capture unstructured data before providing an overview of Natural Language Processing techniques, which leads to the first exposure to an unsupervised learning algorithm. We begin our exploration of unsupervised learning with clustering algorithms, including k-means.

TOPICS

- Natural Language Processing: textblob, NLTK, chunking, stemming, POS tagging, tf-idf
- Data & databases: RESTful APIs, NoSQL databases, MongoDB, pymongo
- Generalized Linear Models and maximum likelihood estimation
- Unsupervised learning algorithms for clustering, including k-means

PROJECT 4

DURATION: 2 WEEKS

Natural Language Processing

For the last (and most lightly) guided project before starting your Passion Project, you will work individually and have very few constraints for the design. You must keep the tenets of good project design in mind as you come up with a project question, procure data, and guide your analysis for a specific audience and use case. The only requirements for the project are that it use unstructured text data, be stored in a NoSQL database (MongoDB), and that it uses some unsupervised learning techniques, such as clustering.

WEEK 7 | UNIT FOUR: PART 1

More NLP, NoSQL & Unsupervised Learning

You will learn the fundamentals of dimensionality reduction and apply this to the NLPspecific approach of topic modeling. Additionally, we wrap up our coverage of clustering unlabeled data and introduce essential deep learning concepts, particularly how deep learning can be used in an NLP context to obtain and use word embeddings. Finally, we introduce recommendation systems.

TOPICS

- **Unsupervised learning: more clustering algorithms, including DBSCAN and hierarchical clustering**
- **Machine learning topics: curse of dimensionality, dimensionality reduction, Principal Components Analysis (PCA), Singular Value Decomposition (SVD), Latent Semantic Indexing (LSI)**
- **Recommender systems**
- **Introduction to deep learning and neural networks, particularly for NLP word embedding tasks**

(continued on next page)

PROJECT 4

WEEK 8 | UNIT FOUR: PART 2

Handling Big Data

You will learn how to handle and process large amounts of data, gaining an understanding of the purpose of big data tools and the landscape. We also introduce designing experiments, especially in an A/B testing context. We introduce Dask, Hadoop and HiveQL for distributed databases.

TOPICS

Design and interpret the results of experiments, including A/B testing

Distributed databases, including Dask, Hadoop and HiveQL

PROJECT 5

DURATION: 3 WEEKS

Passion Project

Building upon all you've learned, you will work to finish your capstone Passion Project.

WEEK 9 | UNIT FOUR: PART 3

Big Data Machine Learning, Deep Learning, Ethics & Begin Passion Project

You will continue to explore the ecosystem of big data tools and will learn how to apply algorithms learned in prior weeks in a distributed environment using Spark. You will learn about state-of-the-art algorithms, including XGBoost and various deep learning models. We will cover both recurrent neural networks and convolutional neural networks. We cover important frameworks for evaluating ethical decisions in the field of data science. Finally, you will begin your final passion project.

TOPICS

- Tools for distributed machine learning, including PySpark
- Deep Learning: Convolutional Neural Networks for images and Recurrent Neural Networks for NLP and time-series modeling
- Frameworks for evaluating the ethics of data science projects

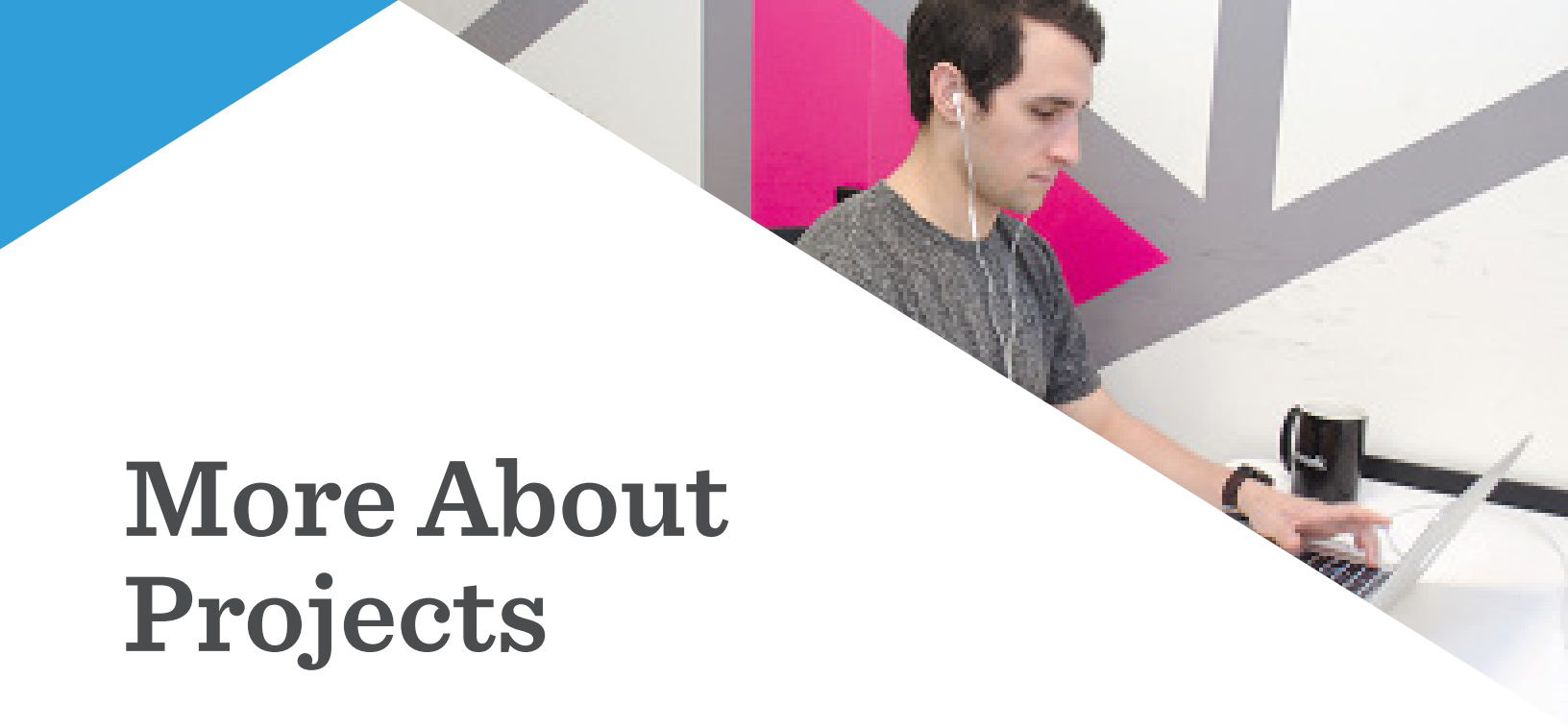
WEEK 10-12 | UNIT FIVE

Final Project

Learn more about cloud computing, system architectures, and feasibility evaluations.

THE GOAL OF THIS UNIT IS TO:

- Enable students to make their own decisions related to algorithms, software tools, and visualizations.
- Record their Passion Projects for Graduate Directory.



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.

Career Services & Support

1 INDIVIDUALIZED SUPPORT & ONE-ON-ONE MEETINGS

Throughout the bootcamp, as well as during post-graduation support, you have opportunities to meet with your Career Advisor. The purpose and outcome of each meeting will be different for every student, but some of the most frequent topics we discuss and/or review are resumes, LinkedIn profiles, salary negotiation, mock interviews, company introductions, crafting messages to hiring managers and recruiters, and soft skill interviewing.

2 SPEAKER & EVENTS

During the bootcamp, you 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 knowledge, gain insider career information, meet potential hiring managers, and expand your data science network.

3 WORKSHOPS

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

LINKEDIN
We teach you how to build a LinkedIn profile that is specifically suited for data science jobs. Learn how to incorporate your previous work experience and how to best position yourself for competitive opportunities.

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

RESUMES
You learn how to craft a professional resume that is ready to present to employers during your search.

SALARY NEGOTIATION
We share the latest data scientist salary information and walk you through salary negotiation best practices.

4 MOCK INTERVIEWS

Throughout the bootcamp you will participate in mock interviewing. This includes pair problems, a technical round-robin interview with instructors and careers and a one-on-one non-technical mock interview post graduation.

5 PRESENTATION PREPARATION

Leading up your final project presentation, you have multiple opportunities to demo your final project to Metis staff, students, and instructors, and you'll receive personalized feedback to help you refine your project recording.

6 GRADUATE DIRECTORY

During your final week, you will record your final project presentation. Then we'll add your profile, including your resume and recording, to our graduate directory. This will allow you to share with potential employers, and provides our hiring network with access to review our graduating students. Participating companies have included Capital One, Spotify, Booz Allen Hamilton, the NBA, Amazon, and HBO.

7 POST-GRADUATION SUPPORT

After you graduate — and until you are employed — you can expect to receive:

- ✓ *An invitation to join MADE - the Metis Alumni Dashboard for Employment - that allows Metis graduates to connect with job opportunities.*
- ✓ *An invitation to join the Metis Alumni Network, which puts you in touch with all Metis alumni.*
- ✓ *Access to the Alumni Resources folder.*
- ✓ *Links to job openings, events, and job searching articles and resources. (These are shared via the messaging service Slack.)*
- ✓ *Information on specific job opportunities that we think will be a fit for you based on geographic or industry preference, or because of the employer's preference for certain skills or work experience.*
- ✓ *An optional one-on-one meeting with your Career Advisor.*

Curriculum Updates

Keeping our curriculum up to date is a critical part of ensuring our students graduate with the in-demand skill sets required in a constantly evolving field. To that end, topics in our curriculum are expanded upon regularly by our team of expert Data Scientists. Latest revisions include:

EXPANSION OF TOPICS

- In-demand business decision-making skills: Lessons have been revised and expanded to adapt to the constantly evolving skills students will need both during interviews and on the job.
- Software development: Lessons have been revised and expanded to emphasize the coding component of data science, including a lesson on Object Oriented Programming (OOP) principles.
- Sequencing of lessons has been optimized to ensure that students get the information they need for each project as efficiently as possible.

UPDATES TO PROCESS

- Each lesson has been through a new, rigorous set of standards and a peer review process to ensure they are up to date.
- Lessons incorporate new, important ideas from the field of Learning Sciences, including how we set learning outcomes and evaluate the quality of lessons.
- We now utilize consistent notation and explanations to unify concepts across the disparate fields of statistics, machine learning, and computer science.

METIS®