Units – What you'll learn

Each module will cover a key aspect of Data Science and have a combination of materials: lectures, theory, coding exercises, reading/viewing exercises, and career-related coursework. The recommended time allocation is based on a total of 500 hours of work, and can be scaled according to student needs.

1. The Python Data Science Stack (21+ hours)

Python has become a lingua franca of data science. In this module, you'll learn to program in Python, how to follow best coding practices, and start using an ecosystem of useful and powerful Python-based tools.

Topics covered:

1. Python
2. Matplotlib, Seaborn—visualization tools in Python
3. Writing clear, elegant, readable code in Python using the PEP8 standard

---

2. Data Wrangling (54+ hours)

Data scientists spend a lot of time on data wrangling (i.e. acquiring raw data, cleaning it, and getting it into a format amenable for analysis), usually with the help of semiautomated tools. In this module, you'll learn the most common tools and workflows in Python that make this normally onerous task a snap.

Topics covered:

1. Deep dive into Pandas for data wrangling
2. Data in files: Work with a variety of file formats from plain text (.txt) to more structured and nested formats files like csv and JSON
3. Data in databases: Get an overview of relational and NoSQL databases and practice data querying with SQL
4. APIs: Collect data from the internet using Application Programming Interfaces (APIs)

---

3. Data Story (30+ hours)

If there's one thing that most data scientists would have loved to know before they entered the field, it's that data science is not just about the math, the algorithms and the analysis, it's also about telling a good story. In real life, data scientists don't work in a vacuum - there's always a client, internal or external, waiting on the results of their work.

A data story is a powerful way to present insights to your clients, combining visualizations and text into a narrative. But storytelling is an art, and needs creativity. This section will try to get your creative juices flowing by suggesting some interesting questions you can ask of your dataset, and a few plotting techniques you can use to reveal insights.

---

4. Statistical Inference (46+ hours)

Statistics is the mathematical foundation of data science. Within statistics, inferential statistics is a set of techniques that helps us identify significant trends and characteristics of a data set. Not only is it useful to explore the data and tell a good story, it also paves the way for deeper analysis and actual predictive modeling. In this module, we cover several important inferential statistics techniques in detail.

Topics covered:

1. Theory of inferential statistics
2. Statistical significance
3. Parameter estimation
4. Hypothesis testing
5. Correlation and regression
6. Exploratory data analysis
7. A/B testing

---

5. Machine Learning (120+ hours)

Machine learning combines aspects of computer science and statistics to extract useful insights and predictions from data. Machine learning is what lets us make useful predictions and recommendations, or automatically find groups and categories in complex data sets.

In this module, we'll cover the major kinds of machine learning algorithms (supervised and unsupervised), with several techniques within each of them. You'll learn when these algorithms are useful, the assumptions they incorporate, the tradeoffs they involve, and the various metrics you can use to evaluate how well your algorithm performs.

Topics covered:

1. Scikit-learn
2. Supervised and unsupervised learning
3. Top machine learning techniques: Linear and logistic regression, naive bayes, support vector machines, decision trees, clustering
4. Ensemble learning with random forests and gradient boosting
5. Best practices
6. Evaluating and tuning machine learning systems

---

6. Career Resources (35+ hours)

You'll receive career material at strategic points both in the curriculum as well as via calls with our career support coaches. We'll help you create a tailored job search strategy based on your background and goals, teach you how to evaluate companies and roles, show you how to effectively get and ace interviews, and explain how to negotiate an above-market salary.

Topics covered:

1. Anatomy of a tech company
2. Job search strategies that top candidates use
3. How to build your network and effectively use it to land interviews

4. Create a high-quality resume, LinkedIn profile and cover letter
5. Interview coaching and practice, including mock interviews for both technical and non-technical topics
6. Negotiation success tips
7. Practice interview questions for each technical topic
8. Algorithms and data structures to ace your coding interviews

---

Capstone Project: Building a Data Product

133+ hours

The capstone project is a key part of our curriculum that every student must complete. The projects are designed to provide you with the experience of working in a realistic data science scenario. Working with your mentor, you'll pick a data set and a problem of interest. From start to finish, your project will be targeted to a specific client (real or imaginary). Using the data science techniques, you've learned, you'll not only come up with a reasonable solution to the problem but learn to present it to them as a compelling story.

You will work on two capstone projects that involve the following:

1. Formulating a problem based on exploratory data analysis;
2. Building a model and transforming data so that it can be input to an algorithm;
3. Iteratively evaluating performance, and adapting model/data input to figure out if more data or a different algorithm is needed to best solve the problem.

If you choose one of our specialization tracks, your second capstone project will be related directly to the specialization of your choice.