

Data Science

CAREER TRACK SYLLABUS

2023

In partnership with



Content



Overview	3
Syllabus Pre-Work	5
Syllabus Core Units	6
Building a Data Science Portfolio	12
Choose Your Specialty Track	14
Career Support	16
Achieving Your Goals with the Springboard Learning Experience	19



Overview

Data science is one of the fastest-growing fields of this decade. The amount of data created daily has surged thanks to the ubiquitous storage of everything from health records to the millions of actions taken on websites and mobile devices. Making sense of this data — and thus driving intelligent, outcomes-focused decisions — is a skill at the heart of being a data scientist.

The **Springboard Data Science Career Track** will prepare you with the core skills needed to succeed in a career in data science through a 500+ hour curriculum designed around three portfolio-worthy capstone projects.

Overview

What You'll Learn

Over the course of six months, you'll learn the six steps of the **Data Science Method** — the process that leads to successful data science projects:

- Problem identification
- Data wrangling
- Exploratory data analysis
- Prep-processing and training data development
- Modeling
- Documentation

How You'll Learn

- **An online curated curriculum** will help you deeply understand key data science concepts through a combination of materials: projects, lectures, theory, coding exercises, reading/viewing exercises, and career-related coursework.
- **Project-based learning** means you'll apply your learning through three capstone projects. You'll hone your skills and nail down each step of the Data Science Method.
- **Specialization tracks** allow you to select one of three areas of expertise where you can learn unique skills that will help distinguish you from other data science professionals.

What You'll Gain

- **1-on-1 mentor support:** You'll be matched with a mentor who will help you tackle the curriculum, provide regular feedback, and answer your questions. Your mentor will keep you accountable and give you an insider's perspective.
- **Career coaching:** You'll work through career-specific units with a career coach guiding you — from defining your strategy, and developing your resume and LinkedIn profile to networking, mock interviews, and salary negotiation.
- **A certificate of completion and our Job Guarantee:** You'll graduate with a certificate from Springboard backed by our Job Guarantee — if you don't land a job after graduating, we'll give you a full refund. **Terms apply.**

Key Program Details



6 months

Program length



20-25

Hours per week



31 total projects

28 mini-projects

3 guided capstone




Prerequisites

- At least 6 months experience in a programming language with 5000 or more lines of code
- Strong background in probability and statistics



Tools & skills learned

 python™

 pandas

SQL

 GitHub

 R

 APACHE Spark



Syllabus

Pre-Work

You'll first work through exercises dedicated to teaching you **Python**, the most popular programming language for data science tasks. We've partnered with DataCamp to bring you resources that go over the foundations of the Python language.

You'll then get a crash course in **statistics** from Khan Academy on data distributions, quantitative data, and scatterplots.

Syllabus

Core Units

1. What Is Data Science?

In this unit, you'll learn what the field of data science involves, including what skills are needed to become a data scientist. You'll then learn about the six steps of the Data Science Method. Each step is examined in detail in one or more units of the bootcamp.

Topics Covered:

- Skills needed to thrive in the data science field
- The Data Science Method overview

2. Problem Identification

This unit covers the first step of the Data Science Method: problem statement. To start a data science project, you need to know the problem you need to solve, clearly define it, and then break it down into manageable pieces. In this unit, you'll identify the correct problem to solve and set goals for a project.

Topics Covered:

- SMART problem statements
- Problem statement worksheet

3. The Python Data Science Stack

Python has become the common language of data science. The Python data science stack refers to the ecosystem within which Python exists. In this unit, you'll learn to program in Python, follow best coding practices, and start using Python-based tools. For cleaning and manipulating data, Pandas will be a potent tool for you to learn and utilize.

Topics Covered:

- Python data types, foundations, and standard libraries
- Pandas

Core Units

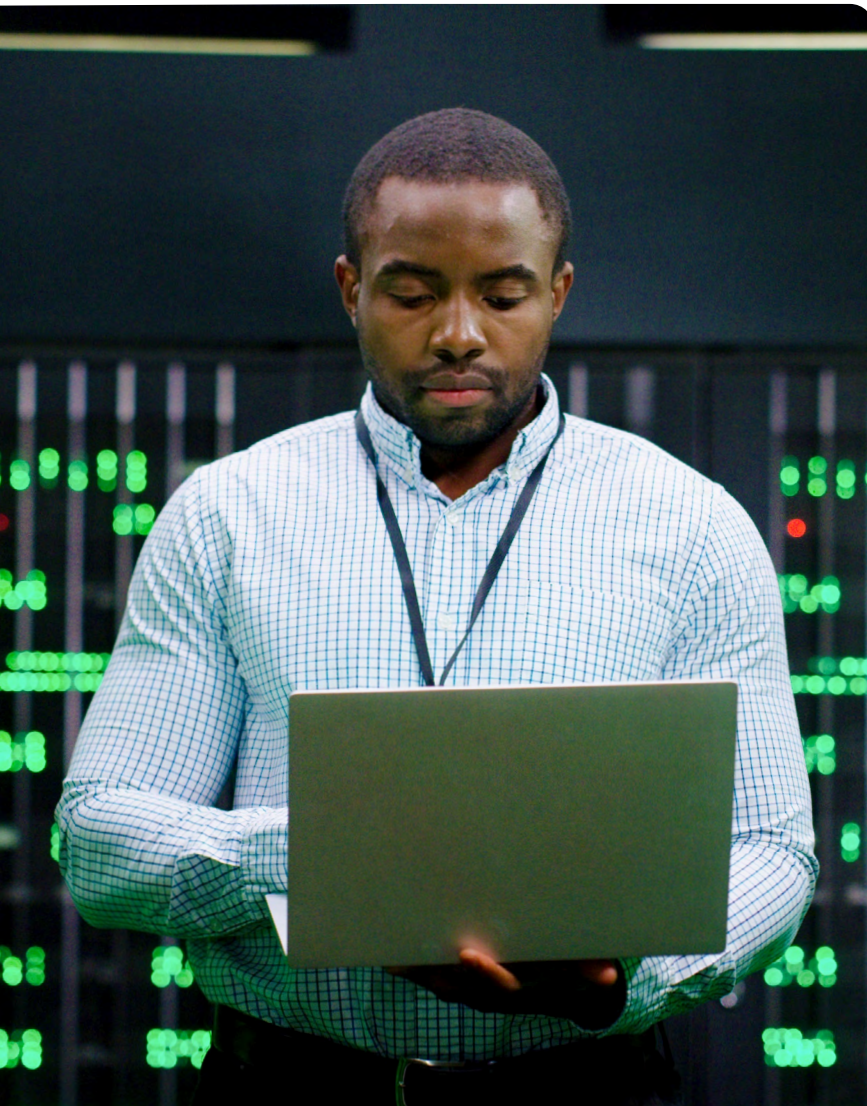
4. Applying the Data Science Method

In this unit, you'll revisit problem identification (the first step of the DSM) and then examine the other steps, applying your new knowledge to hands-on work related to your capstone projects. At the end of the unit, you'll complete a **guided capstone project (Capstone One)** focused on presenting your work to stakeholders, working through problem identification, recommendation and key findings, modeling results and analysis, and summary and conclusion.

Topics Covered:

Six steps of the Data Science Method:

- Problem Identification:
 - › Identify the correct problem to solve and set goals for your project
 - › Create a SMART problem statement and form hypotheses about the problem
- Data Wrangling
 - › Collect, organize and define a dataset or datasets
 - › Compile data, build local file structures, create data profiles, resolve formatting issues, and more
- Exploratory Data Analysis
 - › Create plots and charts to understand the relationship between data and the features of that data
 - › Create data visualizations in Python and use statistics to identify patterns
- Pre-processing and Training Data Development
 - › Standardize and train your dataset
 - › Remove out-of-value ranges and create testing and training subsets of your data
- Modeling
 - › Train and deploy a model to make predictive insights
 - › Use industry-standard algorithms to build models
- Documentation
 - › Document the work you've done and share your findings
 - › Create a project report and present your findings
- Presenting Your Work
 - › Complete a guided capstone project (Capstone One) on presenting to stakeholders



Core Units

5. Data Wrangling

The previous unit gave you some insight and opportunity into data wrangling. This unit explores wrangling — or how to clean, organize, and structure raw data. You'll first submit ideas and a project proposal for your second capstone, then explore data collection, organization, and cleaning through exercises and mini-projects. You'll learn how to compile data, build local file structures, create data profiles, resolve formatting issues, and more. The unit closes by having you wrangle the data you'll use for your second capstone.

Topics Covered:

- Compiling data through data collection
- Reviewing data types, building data profiles, and beginning to develop an understanding of the feature dimensions of your data
- Handling null values, duplicates, and formatting issues

6. SQL and Databases

In this unit, you'll learn how to leverage Structured Query Language (SQL) to query relational database management systems. In other words, you'll use queries to understand the data contained in databases. You'll work through Datacamp courses and then a case study using your new-found knowledge of databases.

Topics Covered:

- The landscape of SQL and databases
- Writing queries in SQL
- Working with relational databases in Python

7. Statistics for Exploratory Data Analysis

Statistics is the mathematical foundation of data science. It allows you to draw useful conclusions from data. In this unit, you'll learn concepts from David Spiegelhalter's book, "The Art of Statistics." While this unit is heavier on the reading side, the key learnings will help distinguish you. You'll read through one or two chapters in each sub-unit, work on an exercise, test your knowledge with a quiz, and review takeaway notes.

Topics Covered:

- Learning from data
- Models and algorithms
- Assessing uncertainty through resampling
- Probability theory and hypothesis testing
- Advanced statistics

Core Units

8. Python Statistics in EDA

Within statistics, inferential statistics is a set of techniques that helps you identify significant trends and characteristics of a data set. Not only is it useful to explore the data and tell a good story, but it also paves the way for deeper analysis and actual predictive modeling. In this unit, you'll review several important inferential statistics techniques in detail. You'll then take your learnings and apply the Exploratory Data Analysis (EDA) step to your second capstone.

Topics Covered:

- Theory of inferential statistics
- Statistical significance
- Parameter estimation
- Hypothesis testing
- Correlation and regression
- Exploratory data analysis

9. Machine Learning Overview

Machine learning combines aspects of computer science and statistics to extract useful insights that allow us to make recommendations and predictions from complex data sets.

In this unit, you'll begin to learn the foundations of machine learning and understand best practices and common challenges when working on machine learning applications.

Topics Covered:

- Intro to supervised vs. unsupervised learning
- Batch vs. online learning
- Instance-based vs. model-based learning

10. Supervised Learning

In this unit, you'll receive a crash course in supervised learning, one of the most commonly used forms of machine learning. In supervised learning, you give the machine your labeled training data and encode procedures for the machine to learn to assign those labels itself. You'll learn the most popular algorithms, including linear and logistic regression, support vector machines, decision trees, clustering, time series and forecasting, ensemble learning with random forests, and gradient boosting.

Topics Covered:

- Supervised learning and its common applications
- Regression and classification techniques
- Decision trees, random forests, and gradient boosting
- Time series analysis and forecasting
- Support Vector Machine (SVM) and kernels

Core Units

11. Unsupervised Learning

Unsupervised learning requires minimal human supervision. Unlike supervised learning, the machine looks for patterns in a dataset with no pre-existing labels. In this unit, you'll learn the most commonly used clustering techniques, complete exercises on distance metrics and cosine similarity, and complete a case study about customer segmentation using K-means clustering.

Topics Covered:

- Euclidean and Manhattan distances
- K-means clustering
- Agglomerative hierarchical clustering
- Cosine similarity and principal components analysis
- Singular value decomposition



Core Units

12. Machine Learning Applications

Furthering your understanding of machine learning, this unit takes you behind the scenes of modeling metrics and hyperparameter tuning. You'll complete exercises on model evaluation metrics and learn which model metric to use based on the business problem you're trying to solve.

Choosing your model metric helps you communicate the impact of predictions to stakeholders. Hyperparameter tuning is a vital step in the machine learning process, where you'll set the parameters of your model. You'll finish the unit by working on the modeling stage for your second capstone.

Topics Covered:

- Evaluation metrics for regression and classification tasks
- Model optimization
- Capstone Two: Modeling

13. Data Storytelling

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 that requires creativity. This final core unit before moving on to the technical career tracks will get your creative juices flowing by suggesting some interesting questions you can ask of your dataset. The unit also covers a few plotting techniques you can use to reveal insights. You'll then execute the last step of the DSM, Documentation, by developing a final project report.

Topics Covered:

- Data storytelling 101
- Presenting your work
- Applying your storytelling skills
- Capstone Two: Documentation



Building a Data Science Portfolio

Capstone projects are a great way for you to practice the skills you'll need during your first data science job and demonstrate your knowledge and experience to potential employers. While working through this bootcamp, you'll complete three capstone projects to showcase your talents.

Building a Data Science Portfolio

Capstone One

Guided Capstone

Your first capstone project comes up fairly early in the program. For this project, you'll be given a lightweight introduction to each step of the Data Science Method. You'll then be guided through those steps with helpful tips and instructions. This first capstone is designed to build your foundational knowledge and practice these important steps before applying your knowledge to the second capstone.

Capstone Two


Your second capstone project follows the same Data Science Method steps as the first capstone, but with less guidance. You'll:

- Come up with a project idea and proposal
- Find and wrangle data
- Use exploratory data analysis techniques to understand that data
- Pre-process and create a training dataset
- Build a working model
- Document and present your work

Each of these steps will be their own submission.

Capstone Three

For your third capstone, you'll again work through each step of the Data Science Method, but this time you'll choose a project idea that aligns with the specialization track that you choose (see the specialization track descriptions on the next pages).



Choose Your Specialization Track

Hone your skills in a specific area of expertise by **choosing one** of our three specialization track options.



Option 1

The Generalist

This track will prepare you to take on versatile data science roles across a wide variety of business domains and geographical locations. You'll build on the foundational skills you learned in the core units and tackle more advanced topics like working with big data and software engineering best practices.

Capstone Project

You'll be asked to develop your third capstone around one of the topics covered in this track, including time series analysis, image processing, or natural language processing. You can choose to use machine learning algorithms you learned in the core units and apply them to big data at-scale approaches like Spark or through the use of a cloud machine learning platform. Implementing data science methods on these advanced platforms is one way to demonstrate your advanced data science knowledge.

Topics Covered

- Advanced time series analysis
- Machine learning topics (please note that you will be asked to choose one of the topics listed below. The other topics will be optional)
 - > Natural Language Processing (NLP)
 - > Image processing
 - > Recommendations systems
 - > Network analysis
- Data science at scale, including:
 - > Hadoop
 - > Spark and PySpark
 - > Neural networks
- Machine learning in the cloud
- Software engineering for data scientists



Option 2

Business Insider

The goal of this track is to teach you advanced data visualization and business analytics skills to extract actionable business insights. While you will have the ability to build predictive machine learning models, you'll primarily focus on learning how to identify insights and effectively communicate recommendations.

Capstone Project

Your third capstone should focus on a business problem that requires your advanced analytics, visualization, and SQL skills to solve. Remember, this track is all about identifying actionable insights that have a significant impact on a company. While you are not yet working for a company, choosing a topic that solves a business problem will help you demonstrate your skills to potential employers.

Topics Covered

- Structured thinking, including the use of:
 - > Issue trees
 - > Value Driver Trees
- Business analytics, including:
 - > A/B testing
 - > Customer segmentation
 - > Data-driven metrics
- Advanced data visualizations, including learning to use tools like:
 - > Tableau
 - > Plotly
 - > Bokeh
 - > D3.js
- Advanced SQL

Option 3

Advanced Machine Learning

This track aims to teach you advanced machine-learning skills and concepts, including deep learning and the deployment of machine-learning models on standard industry platforms. If you want to broaden your machine-learning skills, this track may be right for you.

Capstone Project

You'll be asked to focus your capstone on one of the specialized topics covered in this unit, including (but not limited to) deep learning, image processing, or time series analysis. We encourage you to plan to implement advanced machine learning methods on an advanced platform like Paperspace or an API application.

Topics Covered

- Advanced time series analysis
- Deep learning, including:
 - › Deep learning implementation
 - › Neural network architectures
- Production machine learning methods
 - › Deploying models
- Advanced machine learning topics:
 - › Natural Language Processing (NLP)
 - › Image processing
 - › Recommendations systems
 - › Network analysis
- Data science at scale, including:
 - › Hadoop
 - › Spark and PySpark
 - › Neural networks



Career Support

Career units throughout the bootcamp will help you create a tailored job search strategy based on your background and goals. Learn to craft a resume that stands out from the pack, evaluate companies and roles, ace interviews, and negotiate the best possible salary.

Your career coach will be with you every step of the way, offering feedback and providing personalized tips based on your goals.

Topics Covered

- Types of industry roles
- Job search strategies
- Building a network and using it to land interviews
- Creating a high-quality resume, linkedin profile, and cover letter
- Preparing for technical and non-technical interviews
- Successful negotiation

Build the Skills and Confidence to Transform Your Career

Learn through projects. Work 1-on-1 with a mentor and career coach.
Land a job or your money back.



Springboard Students Achieve Life-Changing Outcomes

NUMBER OF ENROLLED STUDENTS

3,523

Enrolled students in the Data Science Career Track since December 2016. ¹

September 2022

12 MONTH JOB PLACEMENT RATE

90.6%

Of job-qualified individuals who reported an offer, received it within 12 months of graduation. ²

September 2022

AVERAGE SALARY INCREASE

\$25,911

Average salary increase of Data Science students who provided pre- and post-course salaries. ³

September 2022

¹ Number of students refers to all students who enrolled in the career track excluding any that were refunded due to cancellation in the first 7 days following course start.

² Job-qualified individuals defined as all graduates who maintained Job Guarantee eligibility (terms are from the Data Science Career Track Job Guarantee) throughout their job search ("Job-Qualified Graduates"), or Job Guarantee-eligible students who receive a job regardless of completion status ("Early Offerees").

³ Data on compensation was not self-reported by 445 students who reported receiving offers.

Ready for the next step?

Learn more and apply [here](#)



Questions? We're here to help

Email us at hello@springboard.com
or call [+1.415.966.2533](tel:+14159662533)