

DATA SCIENCE

BOOT CAMP

CURRICULUM OVERVIEW

Over the past decade, the explosion of data has transformed nearly every industry. Whether it's marketing, healthcare, government, or activism—the ability to translate data into actionable insights has quickly become a highly in-demand skill by all. **SMU Data Science Boot Camp** is a part-time, 24-week program that will empower students to gain the knowledge and skills to conduct robust analytics on a host of real-world problems.

The program is designed to fit into your life, whether you're employed or attending college full-time, with convenient weekend and evening sessions.

The program is rigorous, fast-paced, and focused on the practical technical skills needed to solve data problems. Throughout the course, students gain proficiencies on numerous marketable technologies, including Excel, Python, JavaScript, SQL Databases, Tableau, and more. Plus, students leave with an impressive professional portfolio and the confidence to succeed in the data-driven economy.

IS THE PROGRAM RIGHT FOR YOU?

Are you a creative, curious, and ambitious professional looking to join the data revolution? If so--or if any of the following describes your situation--enrolling in our Data Science Boot Camp could be a smart career move:

You are currently a professional working with data, but are looking to advance your career by building technical skills.

You are a manager or professional in a business where data can be used to boost your company's bottom line.

You have interests in visualizing social, consumer, or popular trends.

You are looking to enter a new field in healthcare, government, or media and are looking for a way to jump in.

You are a full-time student, hungry to learn more and expand your skill set.

Our most successful learners have basic computer acumen, experience with Excel, and some exposure to at least one programming language.

THE SKILLS YOU'LL GAIN

You will graduate with skills in Data Science, including:*

Intermediate Excel

- Pivot Tables
- VBA Scripting

Fundamental Statistics

- Modelling
- Forecasting

Python Programming

- Python 3
- NumPy
- SciPy
- Pandas
- Matplotlib
- API Interactions

Databases

- PostgreSQL/pgAdmin
- MongoDB
- Extract-Transform-Load (ETL)

Front-End Web Visualization

- HTML
- CSS
- Bootstrap
- Dashboarding
- JavaScript Charting
- Geomapping with Leaflet.js

Business Intelligence Software

- Tableau

Advanced Topics

- R Programming
- Big Data Analytics with Hadoop
- Supervised Machine Learning
- Unsupervised Machine Learning
- Deep Learning

* Note: These topics are subject to change based on local market demand and the input of employer partners.


```

    }
    if(!isset($_SESSION['error'])) {
        $SESSION['error'] = $language->global->error_message->invalid_captcha;
    }
    if(strlen($_POST['username']) < 3 || strlen($_POST['password']) > 32) {
        $SESSION['error'] = $language->register->error_message->username_length;
    }
    if(strlen($_POST['name']) < 3 || strlen($_POST['email']) > 32) {
        $SESSION['error'] = $language->register->error_message->name_length;
    }
    if(Database::exists('user_id', 'users', ['username' => $_POST['username']])) {
        $SESSION['error'] = sprintf($language->register->error_message->user_exists, $_POST['username']);
    }
    if(Database::exists('user_id', 'users', ['email' => $_POST['email']])) {
        $SESSION['error'] = $language->register->error_message->email_exists;
    }
    if(!filter_var($_POST['email'], FILTER_VALIDATE_EMAIL) == false) {
        $SESSION['error'] = $language->register->error_message->invalid_email;
    }
    if(strlen(trim($_POST['password'])) < 6) {
        $SESSION['error'] = $language->register->error_message->short_password;
    }
    $regex = '/^[A-Za-z0-9]+[A-Za-z0-9_.]*[A-Za-z0-9]+$/';
    if(!preg_match($regex, $_POST['username'])) {
        $SESSION['error'] = $language->register->error_message->username_characters;
    }
}

```

Building On The Basics

For those first entering the field of Data Science, knowing where to start can be a daunting task. That's why our curriculum is designed to provide you with a deep foundation on the core technical skills needed to succeed in the field. Throughout the program, expect to learn brand new skills and be challenged in completing difficult real-world problems to demonstrate your new abilities. By the program's end, you will have a strong professional portfolio showcasing your work.

REAL PROJECTS, REAL JOBS

Our graduates may be qualified for many different roles, including:

Data Analyst

Database Administrator (Entry Level)

Data Engineer

Big Data Engineer (Entry Level)

Data Scientist (Entry Level)

Business Intelligence Analyst

Data Journalist

Research Analyst

Business Analyst

Software Engineer (Entry Level)

SQL Developer

Computational Scientist

Data Architect

WHAT YOU WILL LEARN

By the time you graduate, you can expect to be able to:

Utilize statistical analysis to characterize and interpret novel datasets

Employ statistical models to predict and forecast trends

Build VBA scripts in Excel to automate tedious manual processes

Utilize real-world data sources to showcase social, financial, and political phenomena

Create Python-based scripts to automate the cleanup, restructuring, and rendering of large, heterogeneous datasets

Interact with RESTful APIs using Python Requests and JSON parsing techniques

Create in-depth graphs, charts, and tables utilizing a wide-variety of data-driven programming languages and libraries

Use ETL process (Extract, Transform, Load) to transform and consolidate data from multiple sources

Use geographic data to create visually exciting, interactive, and informative maps

Build custom interactive data visualizations using JavaScript libraries

Write SQL commands to perform Create, Read, Update, and Delete commands

Use advanced SQL and NoSQL techniques to combine multiple datasets into one so as to create even more impressive and comprehensive databases

Create basic interactive websites and applications to show your work to the entire world

Work with and lead small-scale teams in order to create applications and visual datasets

Scrape information from web pages in order to collect data from a wide variety of online sources

Communicate and glean new business insights using enterprise-grade tools like Tableau

Build data-driven prediction algorithms using machine learning tools and techniques

Work independently or in a group on complex data-mining projects

Understand the basics of troubleshooting and enhancing legacy code

Use version-controlling software such as Git to collaborate on open-source software



COURSE STRUCTURE

Over the course of 24 weeks, you'll attend informative lectures, participate in a variety of individual and team exercises, and work independently inside and outside of class time. Homework assignments provide an opportunity to apply what you've learned and build on it. The goal is to give you a comprehensive learning experience and true insight into a "day in the life" of a data professional.

DISCUSSION



Instructor-led discussions cover the background, history, and use of new technologies or concepts.

PROJECT WORK



You'll work on timed in-class exercises and projects individually and in teams to put classroom teachings into practice.

PORTFOLIO PROJECTS



Your portfolio signals to employers that you are ready for primetime! You'll build a substantial portfolio of projects that demonstrate your abilities across a wide variety of technologies.



WE'RE HERE TO HELP

As you move up the learning curve, you're likely to have questions around some of the concepts covered in class. We're here to help — through office hours, as well as a dedicated Slack channel where you can get assistance from instructors, support staff, and your fellow classmates. In addition to learning to code, you will have access to career services that will help you prepare for technical roles after the program, such as:

Career Content and Practice Sessions

Database of Customizable Tools and Templates

- Multiple Technical Resume Templates
- GitHub Best Practices
- Guidelines to Building a Portfolio
- Creating an Elevator Pitch
- Developing a Bio

Online Career Events with Industry Professionals

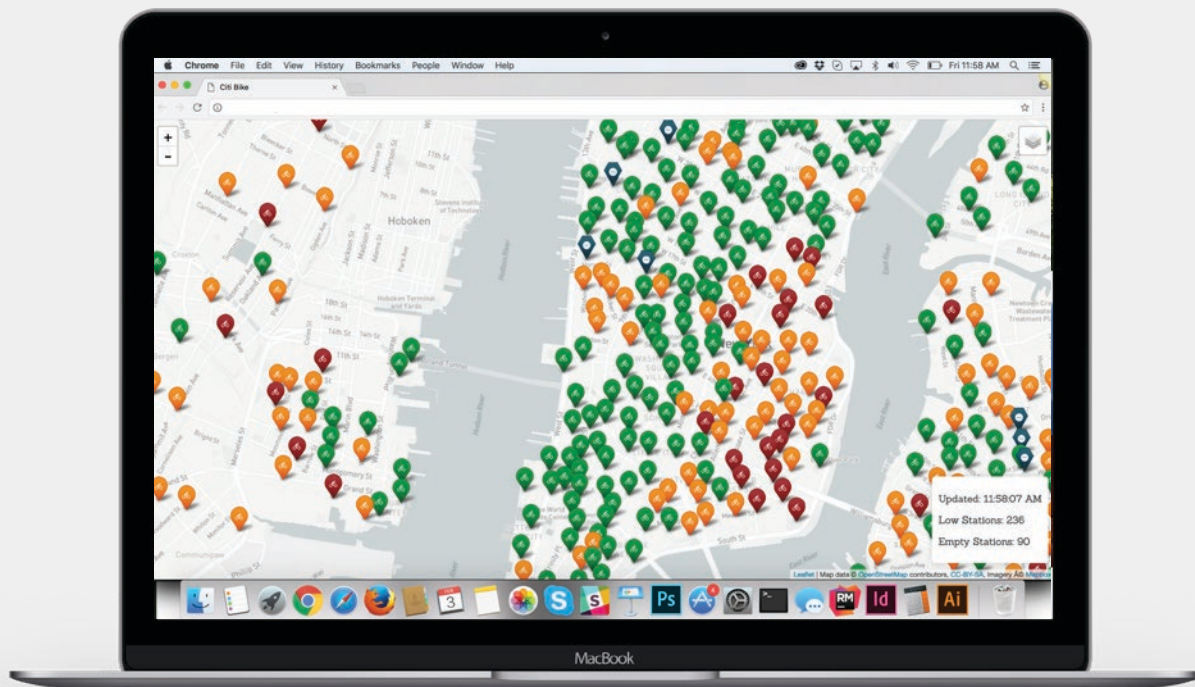
Soft Skills Training

One-on-One Career Coaching

BUILDING

YOUR PORTFOLIO

It's a fact: companies care about what you can do, not what you say you can do. For that reason, our curriculum teaches you how to put what you've learned to work. We cover real-world data projects, ranging from visualizing bike sharing data in New York City to mapping worldwide earthquakes in real-time.



BUILDING YOUR PORTFOLIO

Bank Deserts

Social economists have long noted a trend that in geographic areas with higher poverty rates, there is often a dearth of reputable banks or financial services. The shortage leads to higher rates of financial victimization in these areas. But how could we show this trend using data? In this activity, you'll learn how to combine data from the U.S. Census, Google Maps, and Google Places to visualize the relationship between various socioeconomic factors and the number of banks in a given zip code.

Skills Needed

- Python
- Pandas
- Google Maps
- Google Places
- Matplotlib
- APIs

Objectives

- Utilize the Python Requests library to make hundreds of API calls to the U.S. Census and Google Maps datasets
- Utilize the Python pandas library to organize the retrieved information by zip code and socioeconomic factors
- Build scatter plots to easily communicate the Banking Desert phenomena
- Design statistical models to quantify relationships between factors

Earthquake History

Data isn't just about finance and numbers. It can also be used for good as well. In this activity, you will create an interactive visualization of historic earthquakes over time using Leaflet.js, a popular JavaScript geomapping library. Your final application will provide a near-live feed of global earthquakes and their relative magnitudes.

Skills Needed

- HTML
- CSS
- Javascript
- Leaflet.js
- APIs
- JSON

Objectives

- Harness the power of APIs and JSON to gather earthquake data from USGS datasets
- Utilize Leaflet.js library to create visually compelling, animated maps
- Embed the created map onto a live web page using HTML and CSS

Non-Profit Investment Analysis

Using a dataset with more than 34,000 non-profit organizations that have received funding from Alphabet Soup over the years, you will develop a binary classifier neural network model to predict whether applicants will be successful if funded by Alphabet Soup. You will preprocess data, design network structure, and train, evaluate, and optimize a neural network model.

Skills Needed

- TensorFlow
- Keras
- Machine Learning
- Deep Learning

Objectives

- Use Python and Pandas to preprocess the data for training a deep neural network model
- Compile, train, and evaluate a deep neural network model with TensorFlow and Keras
- Communicate results of the model in comparison to other machine learning models

Web Scraping Application

Sometimes, data is just out of reach. Whether it's a social media website that is guarding its information, a government agency that has poorly organized records, or a cookbook website filled with secret recipes -- data isn't always accessible by external applications. This is where data scraping comes in. Utilizing Python libraries like BeautifulSoup, you will learn to convert data straight from raw HTML into a queryable and storable form, opening up troves of data for your future applications.

Skills Needed

- Python
- BeautifulSoup
- HTML
- CSS
- MongoDB

Objectives

- Scrape your favorite social media website for otherwise inaccessible data
- Parse through the retrieved information and store it into a MongoDB database
- Create new representations of the data using HTML and CSS

Game Studio Analytics

Congratulations! You have landed a job as the Lead Analyst for an independent game company and for your first assignment you have been given the difficult task of analyzing data and creating a report for their latest smash hit release. You will be using the Python Pandas Library and Jupyter Notebook to create demographic and financial reports.

Skills Needed

- Python
- Jupyter Notebook
- Pandas Library

Objectives

- Use Python and the Pandas library to create a report containing a vast amount of data
- Make the data viewable using Jupyter Notebook
- Find, analyze, and write up descriptions of observable trends in the data

Classifying Yelp Reviews

A Nielsen report concluded that 82% of visitors to Yelp intended to make a purchase, so it's no surprise that companies take online customer reviews and ratings seriously. In this section of the course, you'll build an application that can analyze reviews and tell you through Natural Language Processing whether it's negative or positive. This means you don't have to have a human read every review that gets posted and respond accordingly. You can instead have a machine flag negative reviews for you so you can trigger an action like outreach and more.

Skills Needed

- PySpark
- Machine Learning
- Natural Language Processing

Objectives

- Perform Natural Language Processing with PySpark-ML
- Establish a big data processing pipeline to clean and process data
- Train and validate a Naive Bayes machine learning model that can make predictions from customer reviews

COURSE CURRICULUM BY MODULE

Module	Description	What You'll Learn
Module 1: Excel Crash Course	Learn to do more with Microsoft Excel. In this module we'll cover advanced topics like statistical modeling, forecasting and prediction, pivot tables, and VBA scripting. You'll even learn to model historic stock trends -- and hopefully, learn to beat the market!	<ul style="list-style-type: none">• Microsoft Excel• VBA Script• Statistics Modeling
Module 2: Python Data Analytics	Gain a strong foothold in one of today's fundamental programming languages. In the course of this module, you'll gain deep proficiencies with core Python, data analytic tools like NumPy, Pandas, Matplotlib, and specific libraries for interacting with web data like Requests and BeautifulSoup.	<ul style="list-style-type: none">• Python• APIs• JSON• NumPy• Pandas• Matplotlib• BeautifulSoup• SciPy
Module 3: Databases	Work with PostgreSQL and MongoDB to organize data into well-structured and easily retrievable data formats.	<ul style="list-style-type: none">• SQL• PostgreSQL• MongoDB• ETL Process
Module 4: Web Visualization	Building visualizations is of little benefit without a way to communicate the message. In this module, you'll be learning the core technologies of web development (HTML, CSS, and JavaScript) to create new, interactive data visualizations that you can share with everyone on the web.	<ul style="list-style-type: none">• HTML• CSS• JavaScript• AJAX• Leaflet
Module 5: Advanced Topics	By program's end, you'll be immersed in new and in-demand topics like Tableau, Hadoop, and Machine Learning.	<ul style="list-style-type: none">• Tableau• Hadoop• Supervised Machine Learning• Unsupervised Machine Learning• Deep Learning
Module 6: Final Project	Bring everything that you have learned in class altogether to create an impressive data-visualization application with a small team. Get creative and come up with something cool to show off to the whole world!	<ul style="list-style-type: none">• Dreaming up something fantastic and understanding the bounds of reasonable and achievable