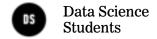# DATA SCIENCE

OVERVIEW

THE FRAMEWORK

Ever wonder how the Netflix recommendation engine works or how Amazon.com determines what items "you may also like?" All of these things are driven by training a computer how to learn using the large datasets.

The data science course is a practical introduction to the interdisciplinary field of data science and machine learning which is at the intersection of computer science, statistics, and business. You will learn to use Python to help you acquire, parse and model your data. A significant portion of the course will be a hands-on approach to the fundamental modeling techniques and machine learning algorithms that enable you to build robust predictive models of real-world data and test their validity. You will also practice communicating your results and insights. By the end of the course, students will be able to:

‣ Perform exploratory data analysis with python.
‣ Build and refine machine learning models to predict patterns from data sets.
‣ Communicate data driven insights to a technical and non-technical audience alike.

STUDENTS

**DATA ANALYSTS OR BUSINESS
INTELLIGENCE ANALYSTS**

This course provides data professional with the skills required to solve problems using computation that involve large data sets such as predicting user behavior on their website, making decisions, or the best way to classify content. Individuals learn how to build the code necessary to be able to make predictions and create models.

PROJECTS

FINAL PROJECT

For the Data Science final project, you will address a data-related problem in your professional field or in a field you interested in. You will acquire a real-world data set, form a hypothesis about it, clean, parse, and apply modeling techniques and data analysis principles to ultimately create a predictive model. Students present their results and each write a report that includes the following:

‣ Clearly articulated problem statement
‣ Summary of data acquisition, cleaning, and parsing stage
‣ Clear presentation of your predictive model and the processes you took to create it
‣ Presentation style appropriate to the audience

Your instructional team will help you scope out your project so that you choose something that is feasible to accomplish given the skills you acquire in the course.

UNITS

UNIT 1: PROGRAMMING BASICS

‣ What is Data Science
‣ Your Development Environment
‣ Foundations of Python

UNIT 2: RESEARCH DESIGN AND EXPLORATORY DATA ANALYSIS

‣ Exploratory Data Analysis in Pandas
‣ Experiments and Hypothesis Testing
‣ Data Visualization in Python
‣ Statistics in Python

UNIT 3: FOUNDATIONS OF DATA MODELING

‣ Linear Regression
‣ Train-Test Split
‣ KNN and Classification
‣ Logistic Regression

UNIT 4: MACHINE LEARNING

‣ Decision Trees and Random Forests
‣ Working with API Data
‣ Natural Language Processing
‣ Time Series Data
‣ Final Project Presentations

Data Science
Units Continued

1 PROGRAMMING BASICS

WHAT IS DATA SCIENCE
‣ Define the workflow, tools and approaches data scientists use to analyze data
‣ Apply the data science workflow to solve a task

YOUR DEVELOPMENT ENVIRONMENT
‣ Navigate through directories using the command line
‣ Use git and GitHub to share repositories

PYTHON FOUNDATIONS
‣ Conduct arithmetic and string operations in Python
‣ Assign variables
‣ Implement loops and conditional statements
‣ Use Python to clean and edit datasets

2 RESEARCH DESIGN AND EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS IN PANDAS
‣ Use DataFrames and Series to read data
‣ Rename, remove, combine, select, and join data
‣ Identify and handle null and missing values

EXPERIMENTS AND HYPOTHESIS TESTING
‣ Determine causality and sampling bias
‣ Test a hypothesis using a sample case study
‣ Validate your findings using statistical analysis (p-values, confidence intervals)

DATA VISUALIZATION IN PYTHON
‣ Define key principles of data visualization
‣ Create line plots, bar plots, histograms and box plots using Seaborn and Matplotlib

STATISTICS IN PYTHON
‣ Use NumPy and Pandas libraries to analyze datasets using basic summary statistics
‣ Create data visualization to discern characteristics and trends in a dataset
‣ Identify a normal distribution within a dataset using summary statistics and visualization

Data Science
Units Continued

3 FOUNDATIONS OF DATA MODELING

LINEAR REGRESSION
‣ Define data modeling and linear regression
‣ Differentiate between categorical and continuous variables
‣ Build a linear regression model using a dataset that meets the linearity assumption using the scikit-learn library

TRAIN-TEST SPLIT
‣ Describe errors of bias and variance
‣ Define overfitting an underfitting
‣ Explore kfolds, LOOCV, and three split methods

KNN AND CLASSIFICATION
‣ Build a K-Nearest Neighbors using the scikit-learn library
‣ Evaluate and tune model by using metrics such as classification accuracy/error

LOGISTIC REGRESSION
‣ Build a Logistic regression classification model using the scikit-learn library
‣ Describe the sigmoid function, odds, and odds ratios and how they relate to logistic regression
‣ Evaluate a model using metrics such as classification accuracy/error, confusion matrix, ROC / AOC curves, and loss functions

4 MACHINE LEARNING

DECISION TREES AND RANDOM FOREST
‣ Define and interpret these models
‣ Build decision trees in scikit-learn
‣ Build random forest models for classification & regression
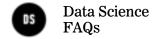
WORKING WITH API DATA
‣ Access public APIs and get information back
‣ Read and write data in JSON
‣ Use the requests library

NATURAL LANGUAGE PROCESSING
‣ Demonstrate how to tokenize natural language text
‣ Categorize and tag unstructured text data
‣ Explain how to build a text classification model using spacy

TIME SERIES DATA
‣ Create rolling means and plot time series data
‣ Perform autocorrelation on time series data

FAQS

WHY IS THIS COURSE RELEVANT TODAY?

Given the large amount of data available, businesses could be making more data driven decisions if this vast amount of data was more deeply analyzed through the use of data science. The data science course provides the tools, methods, and practical experience to enable you to make accurate predictions about data, which ultimately leads to better decision-making in business, and the use of smarter technology.
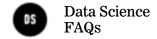
WHAT PRACTICAL SKILL SETS CAN I EXPECT TO HAVE UPON COMPLETION OF THE COURSE?

This course provides you with technical skills in machine learning, algorithms, and data modeling which allow you to make accurate predictions about your data. You'll create your models using Python. Furthermore, you will learn how to programmatically parse and clean your data.

WHO WILL I BE SITTING NEXT TO IN THIS COURSE?

Individuals who have a strong interest in manipulating large data sets, finding patterns in data, and making predictions. Analysts and Business Intelligence Analysts who want to level up their skill set with data modeling. Individuals with a good grasp of data, a solid knowledge of statistics and probability. Pre-work:

‣ CodeAcademy: Learn Python

FAQS

WHAT CAN I EXPECT BY THE END OF THE COURSE?

By the end of the course, you can expect to be able to acquire, parse, clean, and apply various modeling techniques to your data to make predictions. You should also be able to communicate your findings to both a non-technical and technical audience in both written and verbal formats.

WILL THERE BE ANY PRE-WORK?

Yes. You will be required to complete approximately 15 - 20 hours of pre-work.

SHOULD I COME EQUIPPED WITH ANYTHING?

Yes. Please come prepared with a laptop (Mac OSX is preferred but not required).

CONTACT

COURSE APPLICATION

NEW YORK
nyc_admissions@ga.co
1 (877) 348-5665

BOSTON
boston_admissions@ga.co
1 (877) 348-5665

WASHINGTON, DC
dc_admissions@ga.co
1 (877) 348-5665

CHICAGO
chicago_admissions@ga.co
1 (877) 348-5665

ATLANTA
atlanta_admissions@ga.co
1 (877) 348-5665

SEATTLE
seattle_admissions@ga.co
1 (877) 348-5665

AUSTIN
austin_admissions@ga.co
1 (877) 348-5665

SAN FRANCISCO
sf_admissions@ga.co
1 (877) 348-5665

LOS ANGELES
la_admissions@ga.co
1 (877) 348-5665

LONDON
london_admissions@ga.co
+44 (0) 7703 002 046

SYDNEY
sydney_admissions@ga.co
+61 2 0420 438 649

HONG KONG
hk_admissions@ga.co
+852 3952 7209

MELBOURNE
melbourne_admissions@ga.co
+61 1300 779 650

SINGAPORE
singapore_admissions@ga.co