

DATA SCIENCE (DS)

DS 110. Introduction to Data Science. 3 Credits.

This course introduces students to the foundations of data science and the impact that data science has had on modern society. Topics include the history of data science, descriptive statistics, data collection, an introduction to algorithms and algorithmic thinking, and the ethics of data science. No prior experience in statistics or programming is required.

Prerequisites: Take MA 107 or placement level of 3.

Offered: Every year, Fall

UC: Breadth Elective

DS 201. Introduction to Python. 1 Credit.

This course is designed to bridge the gap between CSC106 or CSC110 and the CAS Data Science major. CSC106 and CSC110 provide an important introduction to concepts in computer science and programming. For new programmers, it can be difficult to apply these concepts to a new language. As such, DS201 applies these important concepts to the Python programming language.

Prerequisites: Take CSC 106 or CSC 110.

Offered: Every year, Spring

DS 210. Algorithms for Data Science. 4 Credits.

Algorithms and computation are at the core of data science. This course introduces students to the underlying principles behind digital computation and algorithmic development for scientific purposes. Students also learn foundational algorithms for four common tasks: solving linear systems, determining least squares solutions, implementing unconstrained optimization, and using random number generation for simulation and statistical inference. Throughout the course, the advantages and disadvantages of each algorithm are explored, particularly as they relate to a dataset's properties.

Prerequisites: Take DS 201 DS 110 MA 229 MA 151 MA 285 or EC 272.

Offered: Every year, Fall

DS 215. Communicating with Data. 3 Credits.

Perhaps the most important skill a data scientist can have is the ability to effectively communicate about data. This is especially true in the context of many industries and academic disciplines relying on data to help people make decisions. This course is designed to guide students through the foundations of information visualization and communication using data. In some ways, it differs from many of the courses in the QU Data Science major - students will spend time thinking about color theory, decision-making, and storytelling. In other ways, it is quite similar, as we will investigate methods for dimensionality reduction in the context of visualizations. Importantly, we will also explore tools like Tableau that help practitioners to enrich their analysis.

Prerequisites: Take DS 210;

Offered: Every other year, Spring

DS 299. Independent Study Data Science. 1-6 Credits.

This individual study in a specialized area is open to juniors and seniors by special arrangement with the department chairman. This is a structured program of reading, problem solving and experiments established through conferences with a member of the mathematics faculty. Graded by examination or term project.

Offered: As needed, All

DS 300. Tools for Data Science. 3 Credits.

The course is designed to give students the tools to hit the ground running in a data science/computation production environment. Topics include collaboration and version control systems (git), Linux (Unix) command line programming, cloud computing (AWS), SQL/databases, and more. Some programming experience is required.

Prerequisites: Take CSC 110 DS 201 CIS 245 or BAN 300;

Offered: As needed

DS 325. Database Systems. 3 Credits.

Databases are central to modern data workflows in almost every industry and increasingly in academic settings where "big data" problems arise. In addition, their use extends to many other environments and domains where large amounts of data must be stored for efficient update, retrieval, and analysis. The purpose of this course is to provide a comprehensive introduction to the use of data management systems for data science. Some traditional database topics covered are relational data models, SQL, transactions, and parallel data processing. Some non-traditional topics covered are dynamic hierarchical database design (NoSQL), and JSON data structures. Python, along with its many database bindings, will be used.

Prerequisites: Take DS 210.

Offered: Every other year, Fall

DS 350. Big Data. 3 Credits.

The term "Big Data" means many different things to many different people. For the purposes of this class, it will take on the following meaning - problems and data that are constrained by local memory. This is perhaps one of the largest problems for the modern data scientist. In industry, companies collect petabytes of data on customers and wish to extract value from it. Social sciences such as economics and political science, increasingly use large-scale micro-data to analyze individual behavior. And, of course, hard sciences like physics and genomics have struggled with massive datasets for decades. We will take both a theoretical and practical approach to addressing this issue. Specifically, this course will cover parallel algorithms, distributed computing, databases, cloud resources and computing, and "big data" technologies like Apache Spark.

Prerequisites: Take DS 210;

Offered: Every other year, Fall

DS 380. Data Mining. 3 Credits.

This course introduces students to the concepts and techniques of data mining (the extraction of patterns from large data sets). Topics include data preprocessing, data warehouse architecture, online analytics processing (OLAP), frequent pattern mining, association rules mining, classification analysis.

Prerequisites: Take EC 365.

Offered: Every year, Fall

DS 385. Machine Learning. 3 Credits.

This course introduces students to the theory of machine learning and practical applications. Topics include supervised learning, unsupervised learning, learning theory, regularization models, validation and models.

Prerequisites: Take DS 380.

Offered: Every year, Spring

DS 399. Applied Time Series Analysis and Forecasting. 3 Credits.

Most real-world phenomena involve systems that change through time (e.g., stock market indices, population sizes, salinity levels in a body of water, etc.). Often, the past behavior of these phenomena provides predictive power for future behaviors. This course introduces students to time series analysis, a field of statistics that centers on mathematical relationships between the past and future and how to predict the future through forecasting. Topics include autocorrelation and partial autocorrelation functions, autoregressive models, moving average models, ARMA, ARIMA, and ARMAX. Additional topics may include seasonal models, state-space models and Kalman filtering, volatility models, and machine learning models for forecasting. In all cases, real-world applications and computer implementations in R or similar software are explored.

Prerequisites: Take MA 285 or EC 272; Minimum grade C- , TR

DS 480. Data Science Capstone. 3 Credits.

This course serves as a culminating experience for the Data Science major. Students work on an independent project that will allow them to integrate knowledge from their previous courses in the major and apply that knowledge to a problem in a domain of their interest.

Prerequisites: Take DS 385.

Offered: Every year, Spring