DATA SCIENCE

Data Science is an emerging field that aims to extract actionable insights from vast arrays of information. Drawing on techniques and theories from statistics, computer science, and mathematics, the program has applications in business, engineering, natural sciences, social sciences, humanities, and health care. Developed through a collaborative effort between the Departments of Statistics (http://statistics.columbian.gwu.edu), Mathematics (http://math.columbian.gwu.edu), Physics (http://physics.columbian.gwu.edu), Economics (http://economics.columbian.gwu.edu), Geography (http://geography.columbian.gwu.edu), and Political Science (http://politicalscience.columbian.gwu.edu), the Data Science program, which offers the Master of Science in Data Science and a graduate certificate in data science, teaches students to understand data and contribute important insights that will change the way people live, work, and communicate. Through a structured curriculum that provides foundational knowledge as well as application skills, students learn how to confront the most complex problems facing government and private industry.

GRADUATE

Master's program

• Master of Science in the field of Data Science (http://bulletin.gwu.edu/arts-sciences/data-science/ms)

CERTIFICATES

• Graduate certificate in data science (http://bulletin.gwu.edu/arts-sciences/data-science/certificate)

COURSES

Explanation of Course Numbers

• Courses in the 1000s are primarily introductory undergraduate courses
• Those in the 2000-4000s are upper-division undergraduate courses that can also be taken for graduate credit with permission and additional work
• Those in the 6000s and 8000s are for master’s, doctoral, and professional-level students
• The 6000s are open to advanced undergraduate students with approval of the instructor and the dean or advising office

DATS 6101. Introduction to Data Science. 3 Credits.
The class covers the basic techniques of data science, algorithms for data mining, and basics of statistical modeling. The class focuses on concepts and abstractions, as well as a set of concrete, practical techniques that every data scientist should know. It is an Introduction class hence it is less focused on the details of specific software tools and statistical methodologies. Restricted to Designed primarily for students in the Data Science program, however other students with appropriate backgrounds can register for the course with permission of the instructor. Prerequisites: Basic knowledge of statistics -- STAT 2118 Regression Methods or a similar course that covers analysis of research data through simple and multiple regression and correlation. Exceptions can be made with permission of instructor. Recommended background: An undergraduate degree with a strong background in science, mathematics, or statistics. (Same as STAT 6289).

DATS 6102. Data Warehousing. 3 Credits.
This course is a practical approach to fundamentals of data warehousing specifically for the IT professional who wants to get into the field. It covers all significant topics, including planning requirements, infrastructure, design, and maintenance. Restricted to Designed primarily for students in the Data Science program, however other students with appropriate backgrounds can register for the course with permission of the instructor. Prerequisites: Basic knowledge of statistics -- STAT 2118 Regression Methods or a similar course that covers analysis of research data through simple and multiple regression and correlation. Exceptions can be made with permission of instructor. Recommended background: An undergraduate degree with a strong background in science, mathematics, or statistics.

DATS 6103. Introduction to Data Mining. 3 Credits.
This course surveys concepts, principles and techniques related to data mining. You will become acquainted with both the strengths and limitations of various data mining techniques like Classification, Association analysis and Cluster analysis. Restricted to Designed primarily for students in the Data Science program, however other students with appropriate backgrounds can register for the course with permission of the instructor. Prerequisites: Basic knowledge of statistics -- STAT 2118 Regression Methods or a similar course that covers analysis of research data through simple and multiple regression and correlation. Exceptions can be made with permission of instructor. Recommended background: An undergraduate degree with a strong background in science, mathematics, or statistics.
DATS 6201. Numerical Linear Algebra and Optimization. 3 Credits.
This course is a study of linear and quadratic programming, nonlinear equations, global and unconstrained optimization, and general linearly and nonlinearly constrained optimization as used in data science. Restricted to Designed primarily for students in the Data Science program, however other students with appropriate backgrounds can register for the course with permission of the instructor. Prerequisites: Basic knowledge of linear algebra - Math 2184 or Math 2185. Exceptions can be made with permission of instructor. Recommended background: An undergraduate degree with a strong background in science, mathematics, or statistics.

DATS 6202. Machine Learning I: Algorithm Analysis. 3 Credits.
This course is a practical approach to fundamentals of algorithm design associated with machine learning. Topics include techniques of statistical and probability theory, combinatorial optimization, and factor graph and graph ensemble as used in machine learning. Restricted to Designed primarily for students in the Data Science program, however other students with appropriate backgrounds can register for the course with permission of the instructor. Recommended background: An undergraduate degree with a strong background in science, mathematics, or statistics. (Same as PHYS 6620).

DATS 6203. Machine Learning II: Data Analysis. 3 Credits.
This course is a practical approach to fundamentals of machine learning with an emphasis on data analysis; i.e., how to extract useful information from different datasets. Topics include linear models, error and noise, training and testing methods, and generalization as used in machine learning. Restricted to Designed primarily for students in the Data Science program, however other students with appropriate backgrounds can register for the course with permission of the instructor. Prerequisites: DATS 6101 - Introduction to Data Science. Recommended background: An undergraduate degree with a strong background in science, mathematics, or statistics. (Same as PHYS 6720).

DATS 6401. Visualization of Complex Data. 3 Credits.
This course is a practical approach to fundamentals of data visualization specifically for data science professional. It covers all significant topics, including graphics, discrete and continuous variables, clustering and classification. Restricted to Designed primarily for students in the Data Science program, however other students with appropriate backgrounds can register for the course with permission of the instructor. Prerequisites: Data Science program core courses: 6101, 6102, and 6103. Exceptions can be made with permission of instructor. Recommended background: The Master’s in Data Science core courses: 6101, 6102, 6103. Exceptions can be made with the permission of the instructor.

DATS 6402. High Performance Computing and Parallel Computing. 3 Credits.
This course is a practical approach to high performance computing specifically for the data science professional. It covers topics such as parallel architectures and software systems, and parallel programming. Restricted to Designed primarily for students in the Data Science program, however other students with appropriate backgrounds can register for the course with permission of the instructor. Prerequisites: Data Science program core courses: 6101, 6102, and 6103. Exceptions can be made with permission of instructor. Recommended background: An undergraduate degree with a strong background in science, mathematics, or statistics.

DATS 6450. Topics in Data Science. 3 Credits.
The purpose of DATS 6450 being a topics course is to respond to new ideas and issues in the rapidly evolving fields of Data Science and Big Data. Possible topics may include new application areas in Big Data, emerging new languages and development systems, and policy issues arising from impacts of Big Data on individuals and society. Restricted to Intended primarily for students in the Data Science Master’s and Certificate programs. Prerequisites: DATS 6101 Introduction to Data Science or permission of instructor. Recommended background: Enrollment in a Data Science graduate program.

DATS 6501. Data Science Capstone. 3 Credits.
The course is a final practical application of the knowledge and skills acquired during the data science curriculum. The topics of the capstone team projects will be chosen in consultation with the Capstone Course instructor and the members of the teams. The course is designed to help students transition into the data science profession. Restricted to Designed for students in their last semester of the Data Science program as their final required core course. Prerequisites: Eight courses in the Data Science program, including the core courses 6101, 6102, and 6103 plus five approved courses from the categories Intermediate Analytics, Advanced Analytics, and Electives. Recommended background: Completion of the required courses in the Data Science Master’s program.