

DATA SCIENCE (BS)

Program Director: H. David Sheets, PhD (sheets@canisius.edu)

Introduction

The Data Science major is designed to prepare students to be part of the Big Data movement that is transforming our world. Companies, government agencies, and not-for profit organizations have all realized the tremendous value that in depth understanding of their data holds for organizations.

Students in this major will be well prepared to enter the business world as an analyst, to enter graduate programs in data science, and even continue on into Canisius University's Master of Science in Data Analytics (<http://catalog.canisius.edu/graduate/college-arts-sciences/data-analytics/>) degree program.

At the same time students develop skills in technical areas such as mathematics, computer science and business, they will also benefit from the Canisius Liberal Arts Tradition which develops the critical thinking skills, communications abilities and ethical grounding so much in demand by employers. Like all majors at Canisius, class sizes are small and personal attention is key.

Required Minor or Double Major

In addition to the foundational courses in Data Science, Mathematics and Computer Science, students are required to complete **minor or double major**. The required minor or dual major can be in any field, and not just Mathematics (<http://catalog.canisius.edu/undergraduate/college-arts-sciences/mathematics-statistics/>) or Computer Science (<http://catalog.canisius.edu/undergraduate/college-arts-sciences/computer-science/>). Courses in Psychology (<http://catalog.canisius.edu/undergraduate/college-arts-sciences/psychology/>), Finance, Economics or Marketing (<http://catalog.canisius.edu/undergraduate/wehle-school-business/marketing/>) might be especially useful. This second major or minor may be based on career goals, planned graduate studies, and/or other student interests, and adds flexibility to the program in a way that helps students reach their goals. Before a student declares a double major, it is important to meet with the appropriate academic departments for advisement, as well as the Director of the Data Science Program. In order to declare a double major, the student must complete the appropriate double major request form and get the signature of each department chairperson and the appropriate associate dean.

Per college policy, each additional major requires a minimum of 15 credits that do not apply to the student's first or subsequent major. Some double major combinations can be completed within the minimum 120 credit hour degree requirement, but in other cases additional course work may be required. Please note that students will receive only one degree, regardless of the number of majors they complete.

Minors are an important part of the undergraduate curriculum. If students declare a minor by sophomore year, they can usually complete it in a timely manner. Students should work with their advisor to determine if it is possible that the minor can be completed by graduation.

To receive a minor, a student must complete at least 9 credit hours of coursework distinct from their major(s) and from other minors, and students must complete more than 50% of the coursework required for the minor at Canisius. Please note that "ancillary/supporting" courses required for a major may still count as distinct courses as long as the remaining coursework still meets the 30 credit-hours required for a major. For more information about minor policies, please see the Declaring Majors and Minors ([\[catalog.canisius.edu/undergraduate/academics/student-records/declaring-majors-minors/\]\(http://catalog.canisius.edu/undergraduate/academics/student-records/declaring-majors-minors/\)\) page in the catalog.](http://</p>
</div>
<div data-bbox=)

Qualifications

Students must maintain a 2.0 GPA in the major and a 2.0 overall average to graduate with a degree in Data Science.

Advisement

All students should have an advisor in the major and should contact the department directly to have an advisor assigned if they do not already have one. Meetings with academic advisors are required prior to students receiving their PIN for course registration each semester. All majors should work closely with their advisor in discussing career expectations, choosing their major electives, developing their entire academic program and planning their co-curricular or supplemental academic experiences.

Curriculum

An Ignatian Foundation

All undergraduate students must complete either the Canisius Core Curriculum (<http://catalog.canisius.edu/undergraduate/academics/curricular-information/core-curriculum/>) or the All-College Honors Curriculum (<http://catalog.canisius.edu/undergraduate/academics/curricular-information/all-college-honors-program/>). Many schools refer to their college-wide undergraduate requirements as "general education" requirements. We believe that the core curriculum and the honors curriculum are more than a series of required classes; they provide the basis for a Jesuit education both with content and with required knowledge and skills attributes that are central to our mission.

Free Electives

Students may graduate with a bachelor's degree with more but not less than 120 credit hours. Free electives are courses in addition to the Canisius Core Curriculum or All-College Honors Curriculum and major requirements sufficient to reach the minimum number of credits required for graduation. The number of credits required to complete a bachelor's degree may vary depending on the student's major(s) and minor(s).

Major Requirements

Code	Title	Credits
Computer Science Courses		
CSC 111 & 111L	Introduction to Programming and Introduction to Programming Laboratory	4
CSC 112 & 112L	Data Structures and Data Structures Laboratory	4
CSC 213	Large Scale Programming	3
CSC 213L	Large Scale Programming Laboratory	1
CSC 310 & 310L	Information Organization and Processing and Information Organization and Processing Laboratory	4
Data Science Courses		
DAT 111-	Intro to Reporting and Analysis	3
DAT 211-	Intro to Statistics with R	3
DAT 411-	Econometrics (cross-list current BUS or DAT course, not a new course)	3
DAT 412-	Machine Learning	3
Mathematics Courses		
MAT 111	Calculus I	4

MAT 112	Calculus II	4
MAT 211	Calculus III	4
MAT 219	Linear Algebra	4
MAT 351	Probability & Statistics I	3
MAT 341	Numerical Analysis	3
Total Credits		50

A minor or double major is required for this program in areas such as Computer Science (<http://catalog.canisius.edu/undergraduate/college-arts-sciences/computer-science/>), Mathematics (<http://catalog.canisius.edu/undergraduate/college-arts-sciences/mathematics-statistics/>), Psychology (<http://catalog.canisius.edu/undergraduate/college-arts-sciences/psychology/>), Business (<http://catalog.canisius.edu/undergraduate/wehle-school-business/business-minor/>), or Marketing (<http://catalog.canisius.edu/undergraduate/wehle-school-business/marketing/#text>). Other areas of studies may be selected based on approval from program director.

Roadmap

Freshman

Fall	Spring
CSC 111 & 111L	CSC 112 & 112L
MAT 111	MAT 112
	Elective

Sophomore

Fall	Spring
CSC 310 & 310L	DAT 111
MAT 211	Minor Course 1
DAT 211	Minor Course 2
	CSC 213 & 213L

Junior

Fall	Spring
MAT 219	MAT 341
Elective	MAT 351
Elective	Minor Course 3
	Minor Course 4

Senior

Fall	Spring
DAT 411	DAT 412
Elective	Elective
Minor Course 5	
Minor Course 6	

Learning Goals & Objectives

Data Science: Program Learning Goals

Student Learning Goal 1: Foundations

- Objective 1A: *A Foundation in Computer Science.* Students will be fluent programs, understand database structures and distributed computing concepts. They will be ready to learn new languages and adapt to rapid changes in software.
- Objective 1B: *Adaptable Foundation applied statistics:* Students will be able to use the basic principles of probability theory in a variety

of contexts, including both classical statistical approaches and computational based methods. Students will be familiar with one modern statistical software platform and will be able to readily adapt to others.

Either:

- Objective 1C: *Substantial knowledge of a domain area.* Students who chose this path will have substantial background in an applications area (or domain) to which they can apply data science tactics. {assessment will be based on the outside major or minor}

Or

- Objective 1D: *Additional mathematical or computational studies-*Students who chose this option will demonstrate substantial additional capabilities in Mathematics or Computer Science, preparing them for research and development efforts in Data Science. {Assessment via the Mathematics or Computer Science major or minor}

Student Learning Goal 2: Effective teamwork.

- Objective 2A: Students will demonstrate the ability to work in multi-disciplinary teams to address real-world problems.
- Objective 2B: Students will understand the current theoretical ideas related to the formation of effective collaborative teams.

Student Learning Goal 3: Effective business communication.

- Objective 3A: Students will be able to identify the needs of different audiences, and effectively present complex information in ways that suit the needs of multiple audiences.
- Objective 3B: Students will be able to write effectively to convey data analytic results in business or other domain contexts.
- Objective 3C: Students will be able to create and deliver effective oral presentations, as well as present ideas in less formal oral settings.
- Objective 3D: Students will be able to create effective graphics, both static and real-time active displays, that convey results to business or other domain audiences.

Student Learning Goal 4: Ethical Data Stewardship

- Objective 4A: Students will have an awareness of the ethical and moral issues that arise in working with large data-sets, and understand the steps that need to be taken to protect the rights and privacy of the individuals involved.

Courses

Computer Science

CSC 111 Introduction to Programming **3 Credits**
Algorithms, programming, computers, and applications to problem solving in Python.

Corequisite: CSC 111L.

Fulfills College Core: Field 7 (Mathematical Sciences)

Offered: every fall, spring, & summer.

CSC 111L Introduction to Programming Laboratory **1 Credit**
Required lab for CSC 111.

Corequisite: CSC 111.

Offered: every fall, spring, & summer.

CSC 112 Data Structures 3 Credits

Introduction to object-oriented programming, recursion, and data structures, including lists, stacks, queues, trees and maps. Rudimentary discussion of analysis of algorithms. Python language is used.

Prerequisite: minimum grade of C in CSC 111 & CSC 111L. **Corequisite:** CSC 112L.

Offered: every fall, spring, & summer.

CSC 112L Data Structures Laboratory 1 Credit

Required lab for CSC 112.

Prerequisite: minimum grade of C in CSC 111 & CSC 111L. **Corequisite:** CSC 112.

Offered: every spring.

CSC 310 Information Organization and Processing 3 Credits

Databases, SQL, and NOSQL systems, along with concepts of normalization and database design. Rudimentary discussion of data ethics and security. MySQL and MongoDB used.

Prerequisite: minimum grade of C in CSC 112 & CSC 112L or a minimum grade of C in CSC 213 & CSC 213L. **Corequisite:** CSC 310L.

Offered: every fall & spring.

CSC 310L Information Organization and Processing Laboratory 1 Credit

Required lab for CSC 310.

Prerequisite: minimum grade of C in CSC 112 and CSC 112L or minimum grade of C in CSC 213 & CSC 213L. **Corequisite:** CSC 310.

Offered: occasionally.

CSC 330 Operating System Design and Distributed Computing 3 Credits

The design of operating system software, including processor scheduling, memory management, storage and resource allocation, and security issues. Taking CSC 253/L before this course is preferred.

Prerequisite: Either minimum grade of C in CSC 112 & CSC 112L or minimum grade of C in CSC 213 & CSC 213L. **Corequisite:** CSC 330L.

Offered: every fall.

CSC 330L Operating System Design and Distributed Computing Laboratory 1 Credit

Required lab for CSC 330. Students will write parallel programs using MPI and a High Performance Computing Cluster.

Prerequisite: CSC 281L and CSC 112/L (min grade of C) or CSC 213/213L (minimum grade of C). **Corequisite:** CSC 330.

Offered: every fall.

Data Science**DAT 111 Introduction to Reporting and Analysis** 3 Credits

An introduction to the methods and tools for reporting quantitative data for decision support in a wide range of fields. This course is meant as an introductory course in the Data Science program, and for students in other disciplines preparing for decision support roles in a range of commercial, educational or research roles. Both the general theories and approaches to the presentation of data for decision support in tabular and graphic forms, and practical technical methods will be covered in the course. Most of the course time will be spent using Excel for these tasks, but Tableau and/or PowerBI as well as some basic SQL queries will also be covered. Whenever possible, "real-world" data drawn from a wide range of fields and disciplines will be used to illustrate problems and approaches to reporting of data.

Fulfills College Core: Field 7 (Mathematical Sciences)

Offered: every spring.

DAT 211 Advanced Statistics with R 3 Credits

This course is designed to introduce students to the programming language R. We will begin by talking about the benefits of R from a practical to an ethical level. Students will learn to install R and load packages. Students will then identify a data set they want to work with over the semester and preregister their project rationale, hypotheses, and analytic plan with OSF. Students will spend the majority of their time learning to execute their analytic plan in R. Students will present their project at Ignatian Scholarship Day. After their ISD presentation, students will archive their materials on OSF and update their preregistration to reflect any modifications made to the plan as they conducted their research, changes they would make if they were going to do the project again, and future analyses they would like to conduct with the data set.

Offered: once a year.

DAT 411 Econometrics 3 Credits

Econometrics is the science in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic phenomena. Econometric modeling is an important research tool in Economics, Finance, and many other academic disciplines. The goal of this course is to provide you with a basic understanding of Econometric theory and practice. We will focus on model specification, estimation, and testing, using a "hands on" approach. Throughout the course, we will use EXCEL, R, and SAS. We will cover most of Chapters 1-10 of the textbook, followed by some selected special topics as time permits. You should read through each chapter as we cover it. Special emphasis will be placed on conceptual understanding and application of econometric methods. For those who are interested in more involved discussions of the theoretical framework and/or the statistical or mathematical derivation behind any of the ideas discussed in class, feel free to meet with me outside of class.

Prerequisite: MAT 111 and CSC 111.

Offered: once a year.

DAT 412 Machine Learning 3 Credits

A foundational development of the core ideas and concepts in machine learning, with emphasis on the statistical foundations of machine learning but also applied work in Python, or a comparable language. Topics covered will include feature engineering and basis sets, gradient descent model fitting, kernel methods, Model selection methods, bootstrapping and other permutation methods, model inference and averaging, tree based methods with boosting and bagging, neural nets and deep learning and graph based methods.

Prerequisite: MAT 219 and CSC 112.

Offered: once a year.

DAT 417 Machine Learning for Natural Language Processing 3 Credits

This course is on constructing, training and using Machine Learning tools (neural networks) for Natural Language Processing, covering the fundamentals of operation of ChatGPT and other tools for generative language applications, translation, theme detection, text summarize, question answering and a range of other applications. This is a programming driven course, in which students will construct and evaluate a number of machine learning applications. Students will construct NLP processing models (neural networks) using the Pytorch and/or TensorFlow frameworks within the python programming language using the Jupyter notebook system. The course will also cover text encoding, tokenization, embedded and other reduced space representations, string and sentence transformations and related topics. Basic predictive models will be covered in the introduction to PyTorch and TensorFlow. Data storage in the Apache Arrow and HuggingFace datasets systems will also be discussed. Students may need to subscribe to the Google Colab Pro platform at a modest cost if they do not have regular access to a computer with an Nvidia GPU. Cost of the subscription is comparable to that of a typical electronic textbook.

Prerequisite: CSC 112 and CSC 112L.

Offered: once a year.

DAT 499 Independent Study Course in Data Science 1-3 Credits

Study and work with a faculty supervisor. Project to be determined by faculty agreement. Independent studies require an application and approval by the associate dean.

Prerequisite: DAT 211.

Offered: every fall & spring.

Mathematics

MAT 111 Calculus I 4 Credits

For science and mathematics majors. Calculus of functions of single variable. Functions, limits, differentiation, continuity, graphing, logarithm, exponential and inverse trigonometric functions, related rates, optimization problems, mean value theorem, l'Hospital's rule, anti-differentiation, definite integral. Credit not allowed if student already has credit for MAT 109 and MAT 110 or for MAT 115.

Fulfills College Core: Field 7 (Mathematical Sciences)

Offered: fall & spring.

MAT 112 Calculus II 4 Credits

Applications of integration, integration techniques, improper integrals, sequences, series, convergence tests, Taylor's series, applications; parametric and polar curves.

Prerequisite: minimum grade of C- in one of the following MAT 109 & MAT 110, MAT 111 or MAT 115.

Offered: fall & spring.

MAT 211 Calculus III 4 Credits

Continuation of MAT 111 and MAT 112. Analytic geometry of 3-dimensional space and calculus of functions of several variables.

Prerequisite: minimum grade of C- in MAT 112.

Offered: every fall.

MAT 219 Linear Algebra 4 Credits

Vector spaces and inner product spaces. Linear transformations and matrices. Eigenvectors, eigenvalues, and applications. Orthogonal transformations. Quadratic forms and quadric surfaces.

Prerequisite: MAT 112 or permission of instructor.

Offered: spring.

MAT 341 Numerical Analysis 3 Credits

The methods used to obtain numerical solutions of functional and differential equations. Polynomial interpolation. Various approximation algorithms.

Prerequisite: MAT 219 and CSC 111 & CSC 111L.

Offered: spring of even-numbered years.