DATA ANALYTICS (MS OR ADVANCED CERTIFICATE)

Students may choose to pursue this program full time (cohort based) or part time. A graduate certificate program is also available

If you are interested in more information about this program, please contact Sana Spector, spektors@canisius.edu

Program Director:

Sana Spector, Professor of Mathematics & Statistics and Data Analytics

Contacts: 716-888-2845 or spektors@canisius.edu

Overview

The master's in data analytics is offered in a full-time, on-campus cohort system that can be completed in one calendar year, as well as a part-time program taking two years, depending on preparation. A key feature is the incorporation of applied integrative projects that begin early in the program and dovetail with advanced coursework. An accelerated 4+1 program for students completing a bachelor's at Canisius who want to complete a master's in data analytics is also available.

Data analytics is a rapidly developing field driven by the need to effectively utilize big data to inform business decisions. It has the goal of making reliable predictions or inferences from very large collections of data drawn from a particular domain of human endeavor, including a wide range of diverse fields such as business management, science, sports, health-care management, criminal justice and not-for-profit agencies.

The master's in data analytics program at Canisius contains the three standard components of the field: statistics, computer science and a domain area in business, and there are plans to expand the domain areas to include insurance, the health sciences and other areas. Students with a strong interest in other domain areas are encouraged to discuss them with the program director. Domain courses are chosen to meet the student's career goals and needs and their prior academic and/or workplace experience.

In addition to the standard components of data analytics, the program also focuses on developing student capabilities in three crucial soft skills:

- Working collaboratively in multidisciplinary teams.
- Communicating effectively with different audiences.
- Stewarding data ethically.

Consistent with the mission of the university, we have a strong interest in promoting collaborative efforts with the local nonprofit agencies focused on promoting the public good, connecting interested students with analytic opportunities in conjunction with local not-for-profits.

While data analytics programs are rapidly being developed at many institutions, Canisius has a unique history with its focus on ethics, emphasis on the ability to communicate with and understand others, including those on the margins of society who lack access, collaborative learning, and teamwork. The tradition of high levels of personal attention to students at Canisius makes for the ideal environment to foster these soft skills in addition to the technical areas of computer programming and statistical inference.

Admissions Requirements

- Students from any undergraduate major are welcome to apply, as long as they have acquired a bachelor's degree prior to the start of classes.
- Cumulative GPA of 2.8 or higher.
- Successful completion of a college-level Calculus 1 course (comparable to MAT 111 or MAT 115 at Canisius).
- Students may apply at any time. We have rolling admissions.
- Student preparation and background are used to determine if some foundation courses may be waived.

Materials to be Submitted

- Online Application (https://www.canisius.edu/admissions/apply-canisius/), with personal statement
- An official transcript from each college attended
- Resumé
- Official GRE or GMAT score (optional)
- Letters of Recommendations (optional)

Policies

Academic Standing

The Data Analytics program follows the College of Arts and Sciences on students' academic standing. (http://catalog.canisius.edu/graduate/academics/academic-policies/#academicstandingtext)

Matriculation and Continued Program Enrollment

The Data Analytics program follows the Canisius University policy for matriculated students (http://catalog.canisius.edu/graduate/admission-matriculation/#Matriculation) that expects students to maintain a continuous program of academic work.

Registration and Credit Hours

Data Analytics students must be registered for at least 4.5 credits per semester to maintain eligibility for financial aid (if they are eligible). A full load is at least 9 credit hours. No student may register for more than 12 credit hours in any semester.

Curriculum

Foundation Course Help Make Your Program Yours

This program features foundation courses that are designed to work with your level of education and experience, so whether you're completely new to the field or a seasoned pro, this program is designed to work for you.

- **Foundation courses get you up to speed.** If you're a new college graduate or a career changer with little or no experience in the field, you can develop the skills and knowledge that you need for long-term success right at the beginning of your program.
- **Foundation courses can be waived with relevant experience at the discretion of the program director.** If you've mastered the basics and are looking to deepen your knowledge and hone your skills even further, you can get right into the material – and get to degree completion faster.

Code	Title	Credits
Foundation Courses (can be waived at the program director's discretion)		
BAN 609	Business Analytics & Forecasting	3
CSC 511 & 511L	Introduction to Programming and Introduction to Programming Lab	3

CSC 512 & 512L	Data Structures and Algorithms and Data Structures and Algorithms Lab	3
Summer		
MAT 500	Topics in Applied Mathematics [1]	4
DAT 500	Interactive Graphical Case Studies in Big Data	1
Elective (Domain specific) [1]		3
Fall		
DAT 511	Data Stewardship: Preparation, Exploration and Handling of Big Data	3
CSC 610 & 610L	Database Management and Database Management Lab	3
DAT 521	Applied Integrative Projects in Data Analytics I	3
Elective (Domain Specific) [1]		3
Spring		
DAT 512	Statistical Approaches to Big Data	3
DAT 514	Data Mining and Machine Learning	3
DAT 515	Visualization and Presentation of Advanced Analytics	3
DAT 522	Applied Integrative Projects in Data Analytics II	3
Total Credits		**41**

[1] Up to 10 credits of coursework (from those noted) may be waived by the program director based on a student's preparation and experience.

Domain Courses

Students will take at least two domain courses drawn from the courses below. Students may apply to the program director to take graduate level courses drawn from other domain areas, or more advanced courses for which they have adequate preparation.

Code	Title	Credits
Business and Finance Domain		
ACC 505	Financial Accounting	
ECO 503	Statistics for Managers with Excel	
FIN 608	Corporate Finance	
FIN 617	Portfolio Analysis	
FIN 619	Financial Modeling	
FIN 620	Investment Management	
FIN 623	Fixed Income Securities	
FIN 628	Derivative Securities	

Roadmap

The following sequences are provided as examples, but students are strongly encouraged to work with the program director to determine the best sequence for the student's background, experience, and interest.

Sample Progression, Full-time Study - Math Background

The following example is for a student with a mathematics degree (assuming at least one course each in statistics and computer programming). Note: this example results in waivers for MAT 500, BAN 609, and CSC 511 & CSC 511L

Code	Title	Credits
Summer		
DAT 500	Interactive Graphical Case Studies in Big Data	1
CSC 512 & 512L	Data Structures and Algorithms and Data Structures and Algorithms Lab	3
One Domain Course		3
Fall		
DAT 511	Data Stewardship: Preparation, Exploration and Handling of Big Data	3
CSC 610 & 610L	Database Management and Database Management Lab	3
One Domain Course		3
DAT 521	Applied Integrative Projects in Data Analytics I	3
Spring		
DAT 512	Statistical Approaches to Big Data	3
DAT 514	Data Mining and Machine Learning	3
DAT 515	Visualization and Presentation of Advanced Analytics	3
DAT 522	Applied Integrative Projects in Data Analytics II	3
Total Credits		**31**

Sample Progression, Full-time Study - Computer Science Background

The following example is for a student with a computer science degree (assuming no statistics or advanced mathematics). Note: this example results in waivers for CSC 511 & CSC 511L and CSC 512 & CSC 512L.

Code	Title	Credits
Summer		
DAT 500	Interactive Graphical Case Studies in Big Data	1
BAN 609	Business Analytics & Forecasting	3
MAT 500	Topics in Applied Mathematics	4
One Domain Course		3
Fall		
DAT 511	Data Stewardship: Preparation, Exploration and Handling of Big Data	3
CSC 610 & 610L	Database Management and Database Management Lab	3
One Domain Course		3
DAT 521	Applied Integrative Projects in Data Analytics I	3
Spring		
DAT 512	Statistical Approaches to Big Data	3
DAT 514	Data Mining and Machine Learning	3
DAT 515	Visualization and Presentation of Advanced Analytics	3
DAT 522	Applied Integrative Projects in Data Analytics II	3
Total Credits		**35**

Sample Progression, Full-time Study - Business Background

The following example is for a student with a business degree (assuming a course in statistics or econometrics). Note: this example results in waivers for BAN 609 and a Domain Course.

Code	Title	Credits
Summer (or prior to the start of the cohort)		
CSC 511 & 511L	Introduction to Programming and Introduction to Programming Lab	3
CSC 512 & 512L	Data Structures and Algorithms and Data Structures and Algorithms Lab	3

DAT 500	Interactive Graphical Case Studies in Big Data	1
MAT 500	Topics in Applied Mathematics	4
Fall		
DAT 511	Data Stewardship: Preparation, Exploration and Handling of Big Data	3
CSC 610 & 610L	Database Management and Database Management Lab	3
One Domain Course		3
DAT 521	Applied Integrative Projects in Data Analytics I	3
Spring		
DAT 512	Statistical Approaches to Big Data	3
DAT 514	Data Mining and Machine Learning	3
DAT 515	Visualization and Presentation of Advanced Analytics	3
DAT 522	Applied Integrative Projects in Data Analytics II	3
Total Credits		**35**

Sample Progression, Full-time Study - STEM/ Engineering Background

The following example is for a student with strong mathematics background (STEM/Engineering) including at least one course each in statistics and computer programming. Note: this example includes waivers for MAT 500, BAN 609, and CSC 511 & CSC 511L.

Code	Title	Credits
Summer		
DAT 500	Interactive Graphical Case Studies in Big Data	1
CSC 512 & 512L	Data Structures and Algorithms and Data Structures and Algorithms Lab	3
One Domain Course		3
Fall		
DAT 511	Data Stewardship: Preparation, Exploration and Handling of Big Data	3
CSC 610 & 610L	Database Management and Database Management Lab	3
One Domain Course		3
DAT 521	Applied Integrative Projects in Data Analytics I	3
Spring		
DAT 512	Statistical Approaches to Big Data	3
DAT 514	Data Mining and Machine Learning	3
DAT 515	Visualization and Presentation of Advanced Analytics	3
DAT 522	Applied Integrative Projects in Data Analytics II	3
Total Credits		**31**

Sample Progression, Part-time Study - Math Background

The following example is for a student with a mathematics degree (assuming at least one course each in statistics and computer programming). Note: this example results in waivers for MAT 500, BAN 609, and CSC 511 & CSC 511L.

Code	Title	Credits
Summer 1		
DAT 500	Interactive Graphical Case Studies in Big Data	1
CSC 512 & 512L	Data Structures and Algorithms and Data Structures and Algorithms Lab	3
Fall 1		
DAT 521	Applied Integrative Projects in Data Analytics I	3
One Domain Course		3
Spring 1		
CSC 610 & 610L	Database Management and Database Management Lab	3
DAT 512	Statistical Approaches to Big Data	3
Summer 2		
One Domain Course		3
Fall 2		
DAT 511	Data Stewardship: Preparation, Exploration and Handling of Big Data	3
DAT 514	Data Mining and Machine Learning	3
Spring 2		
DAT 522	Applied Integrative Projects in Data Analytics II	3
DAT 515	Visualization and Presentation of Advanced Analytics	3
Total Credits		**31**

Sample Progression, Part-time Study - Business Background

The following example is for a student with a business degree (assuming a course in statistics or econometrics). Note: this example results in waivers for BAN 609 and 2 Domain Courses.

Code	Title	Credits
Summer 1 (or prior to the start of the cohort)		
DAT 500	Interactive Graphical Case Studies in Big Data	1
CSC 511 & 511L	Introduction to Programming and Introduction to Programming Lab	3
CSC 512 & 512L	Data Structures and Algorithms and Data Structures and Algorithms Lab	3
Fall 1		
DAT 521	Applied Integrative Projects in Data Analytics I	3
One Domain Course		3
Spring 1		
CSC 610 & 610L	Database Management and Database Management Lab	3
DAT 512	Statistical Approaches to Big Data	3
Summer 2		
MAT 500	Topics in Applied Mathematics	4
Fall 2		
DAT 511	Data Stewardship: Preparation, Exploration and Handling of Big Data	3
DAT 514	Data Mining and Machine Learning	3
Spring 2		
DAT 515	Visualization and Presentation of Advanced Analytics	3
DAT 522	Applied Integrative Projects in Data Analytics II	3
Total Credits		**35**

Sample Progression, Part-time Study - Humanities, Social Sciences, or Education Background

The following example is for a student with an undergraduate degree, but no programming background, or mathematics beyond a college Calculus 1 course.

Code	Title	Credits
Summer 1		
DAT 500	Interactive Graphical Case Studies in Big Data	1
CSC 511 & 511L	Introduction to Programming and Introduction to Programming Lab	3
CSC 512 & 512L	Data Structures and Algorithms and Data Structures and Algorithms Lab	3
Fall 1		
BAN 609	Business Analytics & Forecasting	3
DAT 521	Applied Integrative Projects in Data Analytics I	3
Spring 1		
CSC 610 & 610L	Database Management and Database Management Lab	3
One Domain Course		3
Summer 2		
MAT 500	Topics in Applied Mathematics	4
One Domain Course		3
Fall 2		
DAT 511	Data Stewardship: Preparation, Exploration and Handling of Big Data	3
DAT 514	Data Mining and Machine Learning	3
Spring 2		
DAT 515	Visualization and Presentation of Advanced Analytics	3
DAT 522	Applied Integrative Projects in Data Analytics II	3
Fall 3		
DAT 512	Statistical Approaches to Big Data	3
Total Credits		**41**

Advanced Certificate in Data Analytics

The Certificate in data analytics will enhance your ability work as a data specialist across a range of industries and organizations. This program is intended for professionals in related fields who want to utilize data analytics methods in their work, extending their existing skills but not completing a graduate degree. Students will learn how to work with and analyze large data sets, and how to apply data knowledge within a field of specialization. You'll gain the effective teamwork and communication skills that are essential to working within diverse, multidisciplinary organizations. And you'll get a solid grounding in the ethics of data stewardship, training that is highly relevant in an age where the accumulation and storage of personal data is ever increasing. Like other programs at Canisius, the Advanced Certification is a foundational preparation for life long learning, not a simple collection of courses.

A total of 5 courses are required for the certificate. A minimum of 12 credits beyond those taken to fulfill the requirements for an undergraduate degree must be completed as part of the certificate.

Code	Title	Credits
Required Courses		
The following three courses are required, students with equivalent undergraduate courses may substitute courses from the electives list below, a total of 5 courses must be completed. Note that the CSC courses do have a required lab associated with them, which carries no graduate credit hours.		
MAT 500	Topics in Applied Mathematics	4
CSC 511 & 511L	Introduction to Programming and Introduction to Programming Lab	3
CSC 512 & 512L	Data Structures and Algorithms and Data Structures and Algorithms Lab	3
Elective Courses		
A minimum of 2 are required, up to five if a student does not need to repeat the material in the required courses		6
CSC 610 & 610L	Database Management and Database Management Lab	
DAT 511	Data Stewardship: Preparation, Exploration and Handling of Big Data	
DAT 512	Statistical Approaches to Big Data	
DAT 514	Data Mining and Machine Learning	
DAT 515	Visualization and Presentation of Advanced Analytics	
DAT 521	Applied Integrative Projects in Data Analytics I	
Total Credits		**16**

Note: Students without prior knowledge of the R programming language should take DAT 511.

Students who begin the program with the equivalent of MAT 500, will take an additional elective in lieu of MAT 500. These students may complete the certificate in 15 credits.

Preparation

Students interested in this degree will need specific skills to be successful. The following courses will help prepare you for the program prior to the start of your formal cohort in the summer. We are providing the following advisement based on university courses. If you are considering taking courses elsewhere, please contact (grad@canisius.edu) about courses at your institution that would provide appropriate backgrounds.

- To satisfy BAN 609 Statistics and Econometrics, student may complete either ECO 455 or ECO 609.
- MAT 500 Topics in Applied Mathematics can be waived for some majors, such as math, physics, or engineering. For other majors, MAT 211 or MAT 219 will NOT satisfy the MAT 500 requirement, but may help students prepare for the course.
- Students may take CSC 111 and CSC 111L to satisfy CSC 511 and CSC 511L.
- Students who have already completed CSC 111, may take CSC 112 and CSC 112L to satisfy the CSC 512 and CSC 512L requirement.
- The following courses could be used to satisfy a domain course requirement: FIN 608, FIN 617, FIN 619, FIN 620, FIN 623, or FIN 628. Other graduate level courses may also satisfy the domain course requirements, please consult the program director.
- Cybersecurity courses can also be used as domain courses within this program.

Learning Goals and Objectives

For the Master of Science degree

Student Learning Goal 1: *Multi-disciplinary analytic capabilities.*

- Objective A: *Domain Knowledge:* Students will be able to apply the computational and statistical methods and analytical tools to strategic and tactical decision making for at least one domain area. In business, for example, this might be: accounting, economics, finance, management, or marketing.

- Objective B: *Adaptable grounding in 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.
- Objective C: *Flexible computational skills*. Students will have a strong working knowledge of at least one general purpose programming language, and will be able to work with a range of data structures within those languages. Students will also be familiar with databases and the programming techniques needed to work with Big Data.

Student Learning Goal 2: *Effective teamwork.*

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

Student Learning Goal 3: *Effective business communication.*

- Objective A: 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 B: Students will be able to write effectively to convey data analytic results in business or other domain contexts.
- Objective C: Students will be able to create and deliver effective oral presentations, as well as present ideas in less formal oral settings.
- Objective D: 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 A: Students will have an awareness of the ethic 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.

For the Advanced Certificate

Student Learning Goal 1: To develop multi-disciplinary analytic capabilities of the students.

- Objective A: Students will be able to apply a broad range of computational and statistical methods, including both exploratory methods and predictive analytic tools to large data sets, using both local and distributed computer systems.
- Objective B: Students will possess a broad understanding of the principles of statistical reasoning, which will allow them to understand and assess the utility of new statistical tools as they become available, and to put those tools to practical use. *Adaptable grounding in applied statistics*.
- Objective C: Students will develop flexible computational skills. The students will be proficient in at least one general purpose programming language, and at least one modern statistical package. Students will be able to rapidly learn new languages and packages and maintain their own professional capabilities as new technology and procedures appear.

Student Learning Goal 2: To develop interdisciplinary team-work skills and effective communication skills.

- Objective A: Students will be comfortable and effective working in multidisciplinary teams to address complex real-world problems.
- Objective B: Students will effectively communicate statistical and computational results, to a wide range of different audiences with different needs, using verbal, written, and visual modes of communication.

Student Learning Goal 3: To obtain knowledge and develop skills in ethical use of big data.

- Objective A: Students will have a strong grounding in the ethical use of big data.

Courses

Computer Science

CSC 511 Introduction to Programming **3 Credits**
This foundational course will teach you the basics of computer programming using the Python language. You will design, code, test, and debug computer programs for textual and graphical applications.
Corequisite: CSC 511L.
Offered: every fall & spring.

CSC 511L Introduction to Programming Lab **0 Credits**
Required lab for CSC 511.
Corequisite: CSC 511.
Offered: every fall & spring.

CSC 512 Data Structures and Algorithms **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 used.
Prerequisite: CSC 511 or CSC 111. **Corequisite:** CSC 512L.
Offered: every fall & spring.

CSC 512L Data Structures and Algorithms Lab **0 Credits**
Required lab for CSC 512.
Corequisite: CSC 512.
Offered: every fall & spring.

CSC 530 Operating System Design **3 Credits**
The design of operating system software, including classic OS topics (scheduling, memory management, resource allocation and security) along with newer concepts including cloud computing and virtual machines. Taking CSC 253+lab or equivalent hardware course before this course is highly advised.
Prerequisite: A minimum grade of C in CSC 512 & CSC 512L.
Offered: every fall.

CSC 610 Database Management **3 Credits**
Databases, SQL language, concepts of normalization and database design. Rudimentary discussion of data ethics and security. MySQL or SQLite used.
Prerequisite: CSC 112 or CSC 512; may be taken concurrently.
Offered: every fall.

CSC 610L Database Management Lab **0 Credits**
Required lab for CSC 610.
Prerequisite: CSC 512L. **Corequisite:** CSC 610.
Offered: every fall.

Data Analytics

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 or DAT 211 or MAT 131 or MAT 141 or ECO 255 or ECO 256.
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.

DAT 500 Interactive Graphical Case Studies in Big Data **1 Credit**
Students will be introduced to Data Analytics via the study of a variety of case studies of published studies, or successful commercial applications of methods. Students will also learn to replicate the graphical presentations used in these studies, and develop alternative visual representations of the data used in the studies. The R statistical language will be used, as students learn how to produce publication grade graphics that can be used throughout other courses and in their career.
Offered: every summer.

DAT 511 Data Stewardship: Preparation, Exploration and Handling of Big Data **3 Credits**
Data stewardship refers to the process of managing collections of data in an ethical and effective manner, so that business objectives can be achieved efficiently while respecting the rights of individuals. This course will thus cover the substantial ethical issues related to Big Data, but will also address many technical issues related to working with large data sets. Establishing and maintaining quality data poses surprisingly large challenges and can be very time consuming, so that knowledge of effective data cleaning is a key capability for Data Analytics. Students will learn how to download, clean, and prepare data for future analysis, and document the process, as well as understanding how seemingly harmless actions can pose threats to the information security of others.
Prerequisite: CSC 511 or CSC 111.
Offered: every fall.

DAT 512 Statistical Approaches to Big Data **3 Credits**
This course is a Core course in the Data Analytics program. It starts with a brief review of univariate statistics and then covers selected topics usually taught in courses in multivariate statistical analysis and regression analysis. It is assumed that every student in this course has completed at least one college-level statistics course. The theoretical knowledge and analytical skills gained in this course are an essential component of the Data Analytics program.
Prerequisite: BAN 609 or equivalent, CSC 512 or equivalent, & MAT 500 or equivalent.
Offered: every spring.

DAT 514 Data Mining and Machine Learning **3 Credits**
This course is a Core course in the Data Analytics program. It starts with a brief introduction to Data Mining and Statistical Learning, includes a brief summary of relevant methods covered in a much greater detail in other courses in this program, such as Data Stewardship and Statistical approaches to Big Data, and then covers a number of methods essential in the modern Data Mining and Statistical Learning.
Prerequisites: MAT 500, CSC 511, and CSC 512 or equivalents.
Offered: every spring.

DAT 515 Visualization and Presentation of Advanced Analytics **3 Credits**
Students will develop the ability to present complex results from Data Analytics to a range of audiences. The course will cover both real time interactive displays and tools, such as graphic user interface and dashboard design, as well as written, oral and graphical communication of analytic results. Students will complete a range of projects in each of these areas.
Prerequisites: DAT 511 & DAT 521 (courses may be taken concurrently) and the ability to program in Python.
Offered: every spring.

DAT 517 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 512 and CSC 512L.
Offered: once a year.

DAT 521 Applied Integrative Projects in Data Analytics I **3 Credits**
In this course, students would learn SAS. Since the focus is on hands-on, all lectures would be conducted in a computer lab. Students learn how to input various types of data into SAS, such as text, csv, binary and sas7bdat. How to clean data is an important skill students are expected to master. Students learn how to deal with missing variables and run basic sample statistics such as mean, standard deviation, minimum and maximum. Many visualization techniques would be taught. In addition, students learn how to run some basic statistical functions, such as linear regression. Since this course is a preparation for the next course (DAT 522) titled "Applied Integrative Projects in Data Analytics II", students could start to think about their next big projects.
Offered: every fall.

DAT 522 Applied Integrative Projects in Data Analytics II **3 Credits**
This course is supervised internship or project course. Students may chose to apply for a competitive internship position in Data Analytics with a local corporation, government or not-for-profit agency, or may apply to carry out a data analytics project with an employer or on-campus research sponsor.
Prerequisites: DAT 500, DAT 514, DAT 521.
Offered: every fall, spring, & summer.

DAT 555 Seminar on Deep Learning **1 Credit**
Deep Learning is a computational and mathematical approach to building "deep" or many layer neural network architectures for solving complex machine learning tasks, such as image processing, audio processing, complex time series, natural language processing and other big data problems. This course would teach students to build and training deep learning models using current state of the art tools.
Prerequisite: CSC 112 or CSC 512 and MAT 500 or MAT 211 or MAT 219.
Offered: occasionally.

Mathematics

MAT 500 Topics in Applied Mathematics **4 Credits**
This course provides a brief overview of the basic tools from Linear Algebra and Multivariable calculus, with particular attention given to topics that are needed in Data Science. To facilitate students' understanding of the concepts, rigor and proofs will be de-emphasized while numerous examples will be discussed, including the use of computer software like MATLAB.
Prerequisites: One semester of Calculus (MAT 111 or MAT 115 at Canisius, or equivalent).
Offered: every summer.

MAT 519 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: By permission of the instructor.
Offered: every spring.

MAT 551 Probability & Statistics I **3 Credits**
Introduction to the mathematical aspects of modern probability theory and the theory of statistics.
Prerequisite: By permission of the instructor.
Offered: every spring.

MAT 591 Discrete Mathematics **3 Credits**
Fundamental topics with computer science applications. Sets and logic, propositional and predicate calculus, elements of combinatorics and counting, elementary discrete probability, functions and relations, and graphs.
Offered: every fall.