

# Data Science (DATASCI)

**Please note: DATASCI courses are only available for Information and Data Science (MIDS) students.**

## DATASCI 200 Introduction to Data Science Programming 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

This fast-paced course gives students fundamental Python knowledge necessary for advanced work in data science. Students gain frequent practice writing code, building to advanced skills focused on data science applications. We introduce a range of Python objects and control structures, then build on these with classes on object-oriented programming. A major programming project reinforces these concepts, giving students insight into how a large piece of software is built and experience managing a full-cycle development project. The last section covers two popular Python packages for data analysis, Numpy and Pandas, and includes an exploratory data analysis.

### Objectives & Outcomes

**Student Learning Outcomes:** Be able to design, reason about, and implement algorithms for solving computational problems.

Be able to generate an exploratory analysis of a data set using Python. Be able to navigate a file system, manipulate files, and execute programs using a command line interface.

Be able to test and effectively debug programs.

Be fluent in Python syntax and familiar with foundational Python object types.

Be prepared for further programming challenges in more advanced data science courses.

Know how to read, manipulate, describe, and visualize data using the Numpy and Pandas packages.

Know how to use Python to extract data from different type of files and other sources.

Understand how to manage different versions of a project using Git and how to collaborate with others using Github.

Understand the principles of functional programming.

Understand the principles of object-oriented design and the process by which large pieces of software are developed.

### Rules & Requirements

**Prerequisites:** MIDS students only

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Formerly known as:** Data Science W200

## DATASCI 201 Research Design and Applications for Data and Analysis 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

Introduces the data sciences landscape, with a particular focus on learning data science techniques to uncover and answer the questions students will encounter in industry. Lectures, readings, discussions, and assignments will teach how to apply disciplined, creative methods to ask better questions, gather data, interpret results, and convey findings to various audiences. The emphasis throughout is on making practical contributions to real decisions that organizations will and should make. Course must be taken for a letter grade to fulfill degree requirements.

### Rules & Requirements

**Prerequisites:** MIDS students only

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Instructor:** Rivera

**Formerly known as:** Data Science W201

## DATASCI 203 Statistics for Data Science 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

This course provides students with a foundational understanding of classical statistics within the broader context of data science. Topics include exploratory analysis and descriptive statistics, probability theory and the foundations of statistical modeling, estimators, hypothesis testing, and classical linear regression. Causal inference and reproducibility issues are treated briefly. Students will learn to apply the most common statistical procedures correctly, checking assumptions and responding appropriately when they appear violated; to evaluate the design of a study and how the variables being measured relate to research questions; and to analyze real-world data using the open-source language R.

### Rules & Requirements

**Prerequisites:** MIDS students only. Intermediate competency in calculus is required. A college-level linear algebra course is recommended

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Formerly known as:** Data Science W203

## DATASCI 205 Fundamentals of Data Engineering 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

Storing, managing, and processing datasets are foundational processes in data science. This course introduces the fundamental knowledge and skills of data engineering that are required to be effective as a data scientist. This course focuses on the basics of data pipelines, data pipeline flows and associated business use cases, and how organizations derive value from data and data engineering. As these fundamentals of data engineering are introduced, learners will interact with data and data processes at various stages in the pipeline, understand key data engineering tools and platforms, and use and connect critical technologies through which one can construct storage and processing architectures that underpin data science applications.

### Rules & Requirements

**Prerequisites:** MIDS students only. Intermediate competency in Python, C, or Java, and competency in Linux, GitHub, and relevant Python libraries. Knowledge of database management including SQL is recommended but not required

**Credit Restrictions:** Students will receive no credit for DATASCI W205 after completing DATASCI 205. A deficient grade in DATASCI W205 may be removed by taking DATASCI 205.

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Instructor:** Crook

**Formerly known as:** Data Science W205

## DATASCI 207 Applied Machine Learning 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

Machine learning is a rapidly growing field at the intersection of computer science and statistics concerned with finding patterns in data. It is responsible for tremendous advances in technology, from personalized product recommendations to speech recognition in cell phones. This course provides a broad introduction to the key ideas in machine learning. The emphasis will be on intuition and practical examples rather than theoretical results, though some experience with probability, statistics, and linear algebra will be important. Course must be taken for a letter grade to fulfill degree requirements.

### Rules & Requirements

**Prerequisites:** MIDS students only. DATASCI 201 or DATASCI 201A and DATASCI 203. Intermediate competency in Python, C, or Java, and competency in Linux, GitHub, and relevant Python libraries. Linear algebra is recommended

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Formerly known as:** Data Science W207

## DATASCI 209 Data Visualization 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

Visualization enhances exploratory analysis as well as efficient communication of data results. This course focuses on the design of visual representations of data in order to discover patterns, answer questions, convey findings, drive decisions, and provide persuasive evidence. The goal is to give you the practical knowledge you need to create effective tools for both exploring and explaining your data. Exercises throughout the course provide a hands-on experience using relevant programming libraries and software tools to apply research and design concepts learned.

### Objectives & Outcomes

**Student Learning Outcomes:** Analyze data using exploratory visualization.

Build commonly requested types of visualizations as well as more advanced visualizations using ground-up customization.

Constructively critique existing visualizations, identifying issues of integrity as well as excellence.

Create useful, performant visualizations from real-world data sources, including large and complex datasets.

Design aesthetically pleasing static and interactive visualizations with perceptually appropriate forms and encodings.

Improve your own work through usability testing and iteration, with attention to context.

Select appropriate tools for building visualizations, and gain skills to evaluate new tools.

### Rules & Requirements

**Prerequisites:** MIDS students only. DATASCI 200. Intermediate competency in Python, C, or Java, and competency in Linux, GitHub, and relevant Python libraries. Recommended: experience with HTML, CSS, and JavaScript, or ability to learn new programming languages quickly. If Python is the only programming language you know, you will probably benefit from learning the basics of web development with JavaScript in advance

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Formerly known as:** Data Science W209

## DATASCI 210 Capstone 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

The capstone course will cement skills learned throughout the MIDS program – both core data science skills and “soft skills” like problem-solving, communication, influencing, and management – preparing students for success in the field. The centerpiece is a semester-long group project in which teams of students propose and select project ideas, conduct and communicate their work, receive and provide feedback (in informal group discussions as well as formal class presentations), and deliver compelling presentations along with a Web-based final deliverable. Includes relevant readings, case discussions, and real-world examples and perspectives from panel discussions with leading data science experts and industry practitioners.

### Rules & Requirements

**Prerequisites:** MIDS students only. DATASCI 200, DATASCI 201, DATASCI 203, DATASCI 205, and DATASCI 207. Must be taken in final term of the MIDS program

**Credit Restrictions:** Students will receive no credit for DATASCI W210 after completing DATASCI 210. A deficient grade in DATASCI W210 may be removed by taking DATASCI 210.

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Formerly known as:** Data Science W210

## DATASCI 221 Modern Data Applications 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

This is a multidisciplinary graduate course that synthesizes data management, data economy, and machine learning & AI strategy and research, product innovation, business and enterprise technology strategy, industry analysis, organizational decision-making and data-driven leadership into one course offering. The course provides strategic thinking tools, analytical frameworks, and real-world case examples to help students explore and investigate modern data applications and opportunities in multiple domains and industries. Students are required to participate in weekly sessions and write response pieces as well as a final paper and presentation evaluating one defining application or emerging technology in machine learning/AI end-to-end.

### Objectives & Outcomes

**Student Learning Outcomes:** Anticipate the opportunities and problems likely to be encountered in building and working with any given data application as business and technology requirements as well as secular trends evolve.

Create a strategic business case for a new or emerging data application or data science / machine learning use case.

Develop strategic and business thinking in various data science domains.

Evaluate data science applications and opportunities across a number of situations and domains.

Learn a set of qualitative models and analytical frameworks to evaluate any modern data application and emerging trends in machine learning and AI.

Understand “modern data stacks” and how to manage and use data as an asset in an organization for responsible decision making.

### Rules & Requirements

**Prerequisites:** MIDS students only. DATASCI 205 and DATASCI 207. Students can take DATASCI 205 concurrently with DATASCI 221; students may not drop DATASCI 205 and remain in DATASCI 221. Students can take DATASCI 207 concurrently with DATASCI 221; students may not drop DATASCI 207 and remain in DATASCI 221

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

## DATASCI 231 Behind the Data: Humans and Values 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

Intro to the legal, policy, and ethical implications of data, including privacy, surveillance, security, classification, discrimination, decisional-autonomy, and duties to warn or act. Examines legal, policy, and ethical issues throughout the full data-science life cycle — collection, storage, processing, analysis, and use — with case studies from criminal justice, national security, health, marketing, politics, education, employment, athletics, and development. Includes legal and policy constraints and considerations for specific domains and data-types, collection methods, and institutions; technical, legal, and market approaches to mitigating and managing concerns; and the strengths and benefits of competing and complementary approaches.

### Rules & Requirements

**Prerequisites:** MIDS and MPA students only

**Credit Restrictions:** Students will receive no credit for DATASCI W231 after completing DATASCI 231. A deficient grade in DATASCI W231 may be removed by taking DATASCI 231.

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Instructor:** Morgan

**Formerly known as:** Data Science W231

## DATASCI 233 Privacy Engineering 3 Units

Terms offered: Spring 2025, Fall 2024, Spring 2024

This course surveys privacy mechanisms applicable to systems engineering, with a particular focus on the inference threat arising due to advancements in artificial intelligence and machine learning. We will briefly discuss the history of privacy and compare two major examples of general legal frameworks for privacy from the United States and the European Union. We then survey three design frameworks of privacy that may be used to guide the design of privacy-aware information systems. Finally, we survey threat-specific technical privacy frameworks and discuss their applicability in different settings, including statistical privacy with randomized responses, anonymization techniques, semantic privacy models, and technical privacy mechanisms.

### Rules & Requirements

**Prerequisites:** MIDS students only

**Credit Restrictions:** Students will receive no credit for DATASCI W233 after completing DATASCI 233. A deficient grade in DATASCI W233 may be removed by taking DATASCI 233.

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Formerly known as:** Data Science W233

## DATASCI 241 Experiments and Causal Inference 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

This course introduces students to experimentation in the social sciences. This topic has increased considerably in importance since 1995, as researchers have learned to think creatively about how to generate data in more scientific ways, and developments in information technology have facilitated the development of better data gathering. Key to this area of inquiry is the insight that correlation does not necessarily imply causality. In this course, we learn how to use experiments to establish causal effects and how to be appropriately skeptical of findings from observational data.

### Rules & Requirements

**Prerequisites:** MIDS students only. DATASCI 201 or DATASCI 201A and DATASCI 203

**Credit Restrictions:** Students will receive no credit for DATASCI W241 after completing DATASCI 241. A deficient grade in DATASCI W241 may be removed by taking DATASCI 241.

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Formerly known as:** Data Science W241

## DATASCI 255 Machine Learning Systems Engineering 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

This course provides learners hands-on data management and systems engineering experience using containers, cloud, and Kubernetes ecosystems based on current industry practice. The course will be project-based with an emphasis on how production systems are used at leading technology-focused companies and organizations. During the course, learners will build a body of knowledge around data management, architectural design, developing batch and streaming data pipelines, scheduling, and security around data including access management and auditability. We'll also cover how these tools are changing the technology landscape.

### Objectives & Outcomes

**Student Learning Outcomes:** Construct, measure, and identify metrics relating to performance of a system in order to optimize costs and latency of serving inferences for machine learning models.

Demonstrate understanding of Kubernetes for management of machine learning models.

Describe the difference between a monolithic and microservice architecture, assess and select appropriate use cases for each.

Describe the differences between a development and production system particularly for Machine Learning where the boundaries are blurry.

Know when to leverage a cache for serving machine learning models to reduce load on production systems.

Understand continuous integration and continuous delivery (CI/CD) pipeline for automated code deployment, particularly for ML models.

Understand how stateful systems add complexities to systems engineering.

Understand how to serve machine learning models over an API in real-time.

### Rules & Requirements

**Prerequisites:** MIDS students only. DATASCI 205 and DATASCI 207. We assume you are familiar with generating predictions from a trained machine learning model. Familiarity with command line (Bash), Python, and Git. We assume you have a working knowledge of SSH, Ports, and familiarity with networking concepts such as DNS

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

## DATASCI 261 Machine Learning at Scale 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

This course teaches the underlying principles required to develop scalable machine learning pipelines for structured and unstructured data at the petabyte scale. Students will gain hands-on experience in Apache Hadoop and Apache Spark.

### Rules & Requirements

**Prerequisites:** MIDS students only. DATASCI 205 and DATASCI 207. Intermediate programming skills in an object-oriented language (e.g., Python)

**Credit Restrictions:** Students will receive no credit for DATASCI W261 after completing DATASCI 261. A deficient grade in DATASCI W261 may be removed by taking DATASCI 261.

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Formerly known as:** Data Science W261

## DATASCI 266 Natural Language Processing with Deep Learning 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

Understanding language is fundamental to human interaction. Our brains have evolved language-specific circuitry that helps us learn it very quickly; however, this also means that we have great difficulty explaining how exactly meaning arises from sounds and symbols. This course is a broad introduction to linguistic phenomena and our attempts to analyze them with machine learning. We will cover a wide range of concepts with a focus on practical applications such as information extraction, machine translation, sentiment analysis, and summarization.

### Rules & Requirements

**Prerequisites:** MIDS students only. DATASCI 207

**Credit Restrictions:** Students will receive no credit for DATASCI W266 after completing DATASCI 266. A deficient grade in DATASCI W266 may be removed by taking DATASCI 266.

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Instructor:** Gillick

**Formerly known as:** Data Science W266

## DATASCI 267 Generative AI 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

This course focuses on the practical aspects of LLMs to enable students to be effective and responsible users of generative AI technologies.

The course has three parts. Introduction section covers the historical aspects, key technical ideas and learnings all the way to Transformer architectures and various LLM training aspects. The Practical Aspects and Techniques section, students learn how to train, deploy, and use LLMs; and discuss core concepts like prompt tuning, quantization, and parameter efficient fine-tuning, and explore use case patterns. Finally, a discussion of challenges & opportunities offered by Generative AI, which includes highlighting critical issues like bias and inclusivity, fake information, safety, and some IP issue

### Objectives & Outcomes

**Student Learning Outcomes:** Appreciate the history of the path towards Large Language Models (LLMs) and Generative AI approaches.

Be able to understand key use case patterns of Generative AI approaches and know how to think about incorporating them into applications.

Become aware of critical issues such as bias, inclusivity problems, hallucinations, and IP questions

Become conversant in PyTorch and key neural net coding strategies.

Know how to approach improving the results obtained from LLMs through prompt-tuning, instruction-based fine-tuning, and Reinforcement Learning with Human Feedback.

Understand the foundations of LLMs, how they are trained, and how to deploy and use them, for and beyond text-focused problems.

### Rules & Requirements

**Prerequisites:** MIDS students only. DATASCI 207. Students need to know what gradient descent is. Simple linear classifiers and softmax are reviewed in the course at a high level, but students should have at least heard of these terms. Linear algebra required, which is used for vector representations and deep learning in the course. Intermediate competency in Python required. Experience in PyTorch recommended

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

## DATASCI 271 Statistical Methods for Discrete Response, Time Series, and Panel Data 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

A continuation of DATASCI 203, this course trains data science students to apply more advanced methods from regression analysis and time series models. Central topics include linear regression, causal inference, identification strategies, and a wide-range of time series models that are frequently used by industry professionals. Throughout the course, we emphasize choosing, applying, and implementing statistical techniques to capture key patterns and generate insight from data. Students who successfully complete this course will be able to distinguish between appropriate and inappropriate techniques given the problem under consideration, the data available, and the given timeframe.

### Rules & Requirements

**Prerequisites:** MIDS students only. DATASCI 203 taken in Fall 2016 or later and completed with a grade of B+ or above. Strong familiarity with classical linear regression modeling; strong hands-on experience in R; working knowledge of calculus and linear algebra; familiarity with differential calculus, integral calculus and matrix notations

**Credit Restrictions:** Students will receive no credit for DATASCI W271 after completing DATASCI 271. A deficient grade in DATASCI W271 may be removed by taking DATASCI 271.

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

**Formerly known as:** Data Science W271

## DATASCI 281 Computer Vision 3 Units

Terms offered: Summer 2025, Spring 2025, Fall 2024

This course introduces the theoretical and practical aspects of computer vision, covering both classical and state of the art deep-learning based approaches. This course covers everything from the basics of the image formation process in digital cameras and biological systems, through a mathematical and practical treatment of basic image processing, space/frequency representations, classical computer vision techniques for making 3-D measurements from images, and modern deep-learning based techniques for image classification and recognition.

### Objectives & Outcomes

**Student Learning Outcomes:** Be able to read and understand research papers in the computer-vision literature.

Build computer vision systems to solve real-world problems.

Properly formulate problems with the appropriate mathematical and computational tools.

Understand the building blocks of classical computer vision techniques.  
Understand the building blocks of modern computer vision techniques (primarily artificial neural networks).

Understand the process by which images are formed and represented.

### Rules & Requirements

**Prerequisites:** MIDS students only. DATASCI 207. We assume you are familiar with machine learning techniques. You should also be comfortable with linear algebra, which we'll use for vector representations and when we discuss deep learning. Intermediate programming skills in an object-oriented language (e.g., Python). This course will use Python for all examples, exercises, and assignments

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.

## DATASCI 290 Special Topics 3 Units

Terms offered: Summer 2024, Spring 2024, Fall 2023

Specific topics, may vary from section to section, year to year.

### Rules & Requirements

**Prerequisites:** MIDS students only

**Repeat rules:** Course may be repeated for credit when topic changes. Students may enroll in multiple sections of this course within the same semester.

### Hours & Format

**Fall and/or spring:** 14 weeks - 3 hours of lecture per week

**Summer:** 14 weeks - 3 hours of lecture per week

### Additional Details

**Subject/Course Level:** Data Science/Graduate

**Grading:** Letter grade.