Mizzou
University of Missouri

Data Science (DATA_SCI)

DATA_SCI 1030: Foundations of Data Science
This course introduces students to how the Data Science Fundamentals in Mathematics, Statistics, and Computer Science support discovery through data. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: C- or higher in MATH 1100 or MATH 1160 or college algebra placement test score 60% or higher

DATA_SCI 4001: Topic in Data Science and Analytics
This course will act as a placeholder for departmental topics course in Data Science and Analytics. The topics and credits may vary, but will pertain to core instructional or emphasis area topics. Graded on A-F basis only.

Credit Hour: 1-6
Prerequisites: Instructor consent

DATA_SCI 4085: Problems in Data Science and Analytics
Directed study on a topic in data science and analytics.

Credit Hour: 1-6
Prerequisites: Instructor's consent

DATA_SCI 4087: Seminar in Data Science and Analytics
Directed study on a topic in data science and analytics.

Credit Hour: 1-6
Prerequisites: Instructor's consent

DATA_SCI 7001: Topics in Data Science and Analytics
Topics and credit may vary from semester to semester. Can be repeated with departmental approval. Graded on A-F basis only.

Credit Hour: 1-6

DATA_SCI 7002: Python Programming Boot Camp
This course teaches students how to program in Python, including use of auxiliary libraries various Python ecosystems. Students are introduced to the iPython notebooks from the SciPy ecosystem, as well Python's use across the spectrum of Data Science courses and topics. Many activities are focus on data ingestion, cleaning, manipulation, and restructuring (e.g., ETL). Graded on A-F basis only.

Credit Hour: 1
Recommended: Instructor consent

DATA_SCI 7003: Database Basics and SQL Boot Camp
This course covers a core concepts to heterogeneous data management, including relational databases, NoSQL databases, and other data storage systems. The focus is on making students quickly productive in the use of multiple types of database management systems available on the market for data science work. This includes traditional relational databases, NOSQL databases and graph databases. This course is a 1 credit hour course using the JupyterHub learning environment. Graded on A-F basis only.

Credit Hour: 1
Recommended: Instructor consent

DATA_SCI 7004: R Statistical Programming Boot Camp
This course teaches students how to program in R, including use of auxiliary libraries in R focused on various statistical and visualization oriented techniques. Students are introduced to R's use across the spectrum of Data Science courses and topics. Many activities focus on the development of statistical tests, and the use of R for statistical exploration. This course teaches students how to program in R, including use of auxiliary libraries in R focused on various statistical and visualization oriented techniques. Students are introduced to R's use across the spectrum of Data Science courses and topics. Many activities focus on the development of statistical tests, and the use of R for statistical exploration. Graded on A-F basis only.

Credit Hour: 1
Recommended: Instructor consent

DATA_SCI 7005: Introduction to Statistics for Data Analytics Boot Camp
This course explores the use of inferential and predictive statistics for data modeling and analytics. Single--variate and multivariate statistical concepts are discussed, as well as intermediate exposure to statistical modeling. Students learn to evaluate model effectiveness and conduct results driven model selection. Statistical and modeling techniques focus on high dimensional data analytics. Topics related to dimensionality reduction are also covered, such as principal component analysis and factor analysis. Graded on A-F basis only.

Credit Hours: 2
Recommended: Instructor consent

DATA_SCI 7010: Introduction to Data Science and Analytics
(cross-leveled with DATA_SCI 4010). An introductory course in data science and analytics. The objective of the course is to give students a broad overview of the various aspects of data analytics such as accessing, cleansing, modeling, visualizing, and interpreting data. Students will perform hands-on learning of data analytic topics, using technologies such as Python, R, and open source analytic tools. Graded on A-F basis only.

Credit Hours: 3

Recommended: Basic programming and basic database experience including R, Python, and SQL

DATA_SCI 7011: Introduction to Data Science
This course is an introduction to the NGA Program of Study in Data Science (PSDS), the concentration areas, and the role of each concentration area in data science. Participants will learn how to receive, if desired, an accredited Graduate Certificate and/or a Master of Science degree in Data Science and Analytics from the University of Missouri. Participants will receive an introduction to software, tools, and resources to be utilized throughout the program. Participants will learn of systematic methodologies for data science projects and the data science pipeline through review of case studies. Graded on S/U basis only.

Credit Hours: 2
Recommended: Enrollment in NGA Training Program or instructor consent

---

DATA_SCI 7020: Statistical and Mathematical Foundations for Data Analytics
(cross-leveled with DATA_SCI 4020). An intermediate statistics class designed to build the mathematical foundation for students dealing with Big Data phenomena. Topics include discussions of probability, data sampling, data summarization, sampling distributions, statistical inference, statistical pattern analysis, hypothesis testing, regression, and nonparametric inference over multidimensional data collections. Students will engage in Big Data projects using various publicly available data sets and leveraging modern Data Science tools, techniques, and cyberinfrastructure. Graded on A-F basis only.

Credit Hours: 3
Recommended: Basic understanding of mathematical principles of vectors and matrices, and basic course in probability and statistics

---

DATA_SCI 7030: Database and Analytics
(cross-leveled with DATA_SCI 4030). Covers the Fundamental concepts of current database systems and query methods with emphasis on relational model and non-relational techniques in Big Data environments. Topics include entity-relationship model, relational algebra, indexing, query optimization, normal forms, tuning, security, NoSQL, and data analytics skills in both relational and non-relational environments. Project work involves modern relational DBMS systems and NoSQL environments. Graded on A-F basis only.

Credit Hours: 3
Recommended: Basic understanding of mathematical principles of vectors and matrices, and basic course in probability and statistics

---

DATA_SCI 7040: Big Data Visualization
This course will cover visualization techniques and methods for a broad range of data types prevalent in engineering disciplines, life sciences, media, and business. Theoretical and practical aspects of information visualization and exploratory data visualization will be taught with a hands-on approach to give students experience in handling data with a set of tools and programming environments. Topics will include visual perception and distortions, color theory, preattentive processing, data types and models, visual variables, efficient visualizations, design principles, grammar of graphics, spatial visualization, maps, graph theory network visualization, data storytelling including hands-on programming to create plots, charts, heatmaps, spatial and network visualizations using R and Python libraries. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: Admission to the program, or instructor's consent for non-DSA students. Students are expected to have basic working knowledge of programming in R and Python

---

DATA_SCI 7263: Digital Strategy II
This course provides hands on experience using several digital platforms such as Facebook Insights, Google AdWords, Google Analytics, Adobe Analytics, Clarabridge and Topsy. In this course you'll learn digital advertising terminology and jargon, the importance of digital analytics, the role of analysts, qualities of effective analysts, the digital optimization process, web metrics and key performance indicators, as well as the essentials of collaboration and generating support and buy-in while gaining your executive's attention. Graded on A-F basis only.

Credit Hours: 3

DATA_SCI 8000: Data and Information Ethics
Introduces the ethics related to Big Data in industry, business, academia, and research settings. Students will learn the social, ethical, legal and policy issues that underpin the big data phenomenon. Discussions and case studies will help guard against the repetition of known mistakes and inadequate preparation. The course content will follow the guidelines to be developed by the Council for Big Data, Ethics, and Society. Graded on A-F basis only.

Credit Hour: 1
Prerequisites: DATA_SCI 7010 and DATA_SCI 7040 or instructor's consent

---

DATA_SCI 8001: Advanced Topics in Data Science and Analytics
Topics and credit may vary from semester to semester. Can be repeated with departmental approval. Graded on A-F basis only.

Credit Hour: 1-6

DATA_SCI 8010: Data Analytics from Applied Machine Learning
This course leverages the foundations in statistics and modeling to teach applied concepts in machine learning. Participants will learn various classes of machine learning and modeling techniques, and gain an in-depth understanding how to select appropriate techniques for various data science tasks. Topics cover a spectrum from simple Bayesian modeling to more advanced algorithms such as support vector machines, decision trees/forests, and neural networks. Students learn to incorporate

machine learning workflows into data-intensive analytical processes. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7020 or instructor's consent

DATA_SCI 8020: Big Data Security

This course provides an overview of state-of-the-art topics in Big Data Security, looking at data collection (smartphones, sensors, the Web), data storage and processing (scalable relational databases, Hadoop, Spark, etc.), extracting structured data from unstructured data, systems issues (exploiting multicore, security). Securing sensitive data, personal data and behavioral data while ensuring a respect for privacy will be a focus point in the course Graded on A-F only.

Credit Hours: 3
Prerequisites: DATA_SCI 7010 and DATA_SCI 7030 or instructor's consent

DATA_SCI 8080: Big Data Analysis Case Study

Using a case-study approach, students will engage in discussions on a variety of big data topics relevant to their emphasis area and the realm of Big Data. This course will help students generate ideas and prepare them for the Big Data Capstone. Course work will be performed in small teams, mentored by faculty and/or industry advisors. Teams will research, cultivate, curate, and leverage large data sets. Students will gain hands-on experience applying relevant data science and analytical technology and techniques to gain insight and information from these real-world data sets. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 8410, DATA_SCI 8020, and DATA_SCI 7040 or instructor's consent

DATA_SCI 8085: Problems in Data Science and Analytics

Directed study on a topic in data science and analytics. Graded on A-F basis only.

Credit Hour: 1-6
Prerequisites: instructor's consent

DATA_SCI 8090: Big Data Capstone

This course provides an opportunity for participants to tackle a real-world data science project, delivered as a problem-based exercise. Participants will perform the full data science lifecycle methodology on a relevant challenge problem as final learning activity that draws upon all the foundational data science concepts and technologies, as well as specialized technologies and concepts relative to a particular concentration area. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 8080 or instructor's consent

DATA_SCI 8095: Research-Masters Thesis Data Science and Analytics

Investigation and research of a data science thesis topic, including exploratory data analysis, statistical modeling, and machine learning. Outcomes will include data-driven insights that advance science, society, or intelligent automation. Graded on S/U basis only.

Credit Hour: 1-6
Recommended: Successful completion of the Data Science Core courses except DATA_SCI 8000

DATA_SCI 8110: Genomics Analytics

This course will introduce the foundational concepts of genomics and bioinformatics. Genomics is a combination of biological and computational methods that explore the roles of DNA, genes, and proteins on a very large scale. However, understanding how to interpret and understand the results depends (at least) on a basic understanding of biology. The course does not assume a student has a biological background and it will cover the concepts necessary to implement genomics methods. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7010 or instructor's consent

DATA_SCI 8120: Multi-Omics Analytics

The integration of multiple types of omics data set such as genomics, epigenomics, transcriptomic, proteomic and metabolomics are very important to understand the pathophysiology of human complex diseases. This course will describe the basic concepts of Multiple types of Omics datasets and databases. This course will also focus on various tools and its application in knowledge discovery from multi-omics data set and its challenges related to preprocessing, analysis and visualization. Hands-on computer experience will be provided through web resources and Jupyter notebook environment. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 8110 or instructor's consent

DATA_SCI 8130: Data Science for Health Care

This course covers the basic concepts surrounding the analysis of health data. Topics include ethics and regulations of protected health data, healthcare data standards, and statistical analysis and dissemination techniques suitable for health care settings. Project work involves accessing and analyzing real (de-identified) health care data. This course focuses on health data analysis that is done in industry, insurance, hospitals and research. Practical, hands-on course with focus on fundamental data science skillsets such as programming in Python and data carpentry. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7010, DATA_SCI 7030, DATA_SCI 7040, or instructor's consent

DATA_SCI 8140: Advanced Methods in Health Data Science

This course covers advanced topics in health data analysis. Students will learn about research informatics and clinical trials, and advanced statistical methods used in health data analysis. Additionally, students will be exposed to new forms of health data processing such as free text data, image data, and longitudinal data. Students will explore the use of machine learning and AI in health care settings, and applied clinical informatics in the form of decision support. Project work involves accessing and analyzing real (de-identified) health care data. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 8130 or instructor's consent

DATA_SCI 8150: Precision Medicine Analytics

This course will provide a wealth of knowledge about understanding translational research and its application in precision medicine. Students will also learn how to leverage the multi-omics data set to improve the clinical outcome and advance the precision medicine strategies by accounting individuals' biological variability. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 8110 and DATA_SCI 8130 or instructor's consent

DATA_SCI 8160: Population Health Analytics

This course provides an introduction to population health analytics, with a focus on Big Data ecosystem skillsets. Students will gain hands-on experience with large-scale population health data and will prepare original quantitative analysis for presentation. Instructors' lectures are delivered by video and face-to-face interaction. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 8130 or instructor's consent

DATA_SCI 8220: Communication Network Analytics

This course is intended to review theoretical, conceptual, and analytic issues associated with network perspectives on communicating and organizing. The course will review scholarship on the science of networks in communication across a wide array of disciplines in order to take an in-depth look at theories, methods, and tools to examine the structure and dynamics of networks. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7010 or instructor's consent

DATA_SCI 8230: Streaming Social Media Data Management and Analytics

An intermediate data wrangling and analysis class designed to provide students with an in-depth overview of collecting and analyzing Twitter data. Computational topics include composing, sending, and receiving Hypertext Transfer Protocol (HTTP) messages. Data wrangling topics include parsing json files, navigating recursively nested structures, and processing textual data. Analysis methods include machine learning, network analysis, topic modeling, time series, etc. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7010 or Instructor's consent

DATA_SCI 8310: Advanced Visualization I

Covers the fundamental concepts of current visualization concepts and technologies, adding in Infographic and Interactive Visualization Design. Unlike many data visualization courses, this one focuses on principles of visualization design and the grammar of graphics as they can be applied to combining art and technology to tell data stories. These principles are then implemented in popular contemporary visualization technologies. Students will develop an advanced knowledge of the appropriate selection, modeling, and evaluation of data visualizations. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7010 or instructor's consent
Recommended: Instructor consent

DATA_SCI 8320: Advanced Visualization II

Covers the fundamental concepts of animated visualization design that build on Infographic and Interactive Visualization Design techniques. Unlike many data visualization courses, this one focuses building animations and highly interactive representations of data. These principles are then implemented in popular contemporary visualization technologies. Students will develop an advanced knowledge of the appropriate selection, modeling, and evaluation of data visualizations. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 8310 or instructor's consent

DATA_SCI 8330: Usability Evaluation for Data Science

Usability is concerned with how well a person can use a designed system to accomplish the goals for which that system is designed. This course provides an overview of methods for usability testing of data science applications through readings, examples and discussions. Students will work in groups to develop and present a usability test plan for a data science application or system. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7010 or instructor's consent

DATA_SCI 8410: Data Mining and Information Retrieval
The course introduces the main concepts and techniques of data mining and information retrieval. It covers a variety of data mining topics and methods to extract hidden and predictive patterns from large data collections. Furthermore, theory and techniques for the modeling, indexing, and retrieval of relational, nonrelational, textbased and multimedia databases is covered. Topics include introduction to data mining process, mining frequent patterns, and pattern analysis, as well as different information retrieval models and evaluation, query languages and operations, and indexing/searching methods. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7010 and DATA_SCI 7030 or instructor's consent

DATA_SCI 8420: Cloud Computing for Data Analytics
This course introduces students to cluster and cloud computing big data ecosystems. Topics include a survey of cloud computing platforms, architectures, and use-cases. Students will examine scaling data science techniques and algorithms using a variety of cluster and cloud paradigms, such as those built atop Hadoop (Map-Reduce) concepts, and others. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7020 and DATA_SCI 7030 or instructor's consent

DATA_SCI 8430: Parallel Computing for Data Analytics
This course will provide in-depth treatment of the evolution of high performance, parallel computing architectures and how these architectures and computational ecosystems support data science. We will cover topics such as: parallel algorithms for numerical processing, parallel data search, and other parallel computing algorithms which facilitate advanced analytics. To reinforce lecture topics, learning activities will be completed using parallel computing techniques for modern multicore and multi-node systems. Parallel algorithms will be investigated, selected, and then developed for various scientific data analytics problems. Programming projects will be completed using Python and R, leveraging various parallel and distributed computing infrastructure such as AWS Elastic Map Reduce and Google Big Query. Students will research emerging parallel and scalable architectures for data analytics. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7020 and DATA_SCI 7030 or instructor's consent

DATA_SCI 8510: Geospatial Data Engineering
This course provides an overview of theoretical and practical issues encountered when working with geospatial data for both the vector and raster data models with a focus on incorporating geospatial data into the data science lifecycle. Data access, indexing, retrieval, and other technical concepts are investigated. Important data storage paradigms such as enterprise geospatial databases and desktop GIS systems are explored along with scalable computational tools beyond desktop computing for Geospatial Big Data. Core issues in geospatial data storage, management, exploitation, and multi-data set entity resolution / correlation are examined. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 7030 or instructor's consent

DATA_SCI 8520: Spatial and Geostatistical Analysis
This course will provide a practical overview of key issues encountered when working with and analyzing spatial data as well as an overview of major spatial analysis approaches. Discussions and laboratory work will focus on implementation, analysis, and interpretive issues given constraining factors that commonly arise in practice. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 8510 or instructor's consent

DATA_SCI 8530: Remote Sensing Data Analytics
Introduction to the principles of remote sensing of the environment leading to information extraction from remote sensing geospatial raster data sets. Examines theoretical and practical issues associated with digital imagery from spacecraft and airborne systems, thermal imaging, and microwave remote sensing. Covers standard processing techniques, including preprocessing and normalization, pixel-level feature extraction, information extraction, and structural/object extraction. Graded on A-F basis only.

Credit Hours: 3
Prerequisites: DATA_SCI 8510 or instructor's consent