

# 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 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: Principles of Data Science and Analytics

A course covering the array of topics and principles involved with data science and analytics. The objective is to provide students a broad overview of the data science life cycle. Students will use generative AI tools, jupyter notebooks and other technology platforms to perform and support hands-on learning and coding through labs, practices, and exercises related to specific analytical topics such as data accessing, cleansing, exploration, analytics, modeling, visualizing, as well as interpreting analytical results. Discussions will be held to ensure that initiate awareness and knowledge of ethical, bias, and other related

considerations. Specific coding skills will be expanded through these experiences using Python, R, SQL, and other 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

This course is an intermediate statistics class designed to lay the mathematical principles of predictive analytics and provide a foundation for Al/ML. Topics include discussions of probability, data sampling, data summarization, sampling distributions, statistical inference, statistical pattern analysis, hypothesis testing, regression, nonparametric inference, clustering, and linear algebra over multidimensional data collections. Students will engage in hands-on projects using various publicly available data sets and leveraging current data science tools, techniques, and cyberinfrastructure. Graded on A-F basis only.

#### Credit Hours: 3

**Recommended:** Basic working knowledge of programming in R. Basic understanding of mathematical principles of vectors and matrices, and basic course in probability and statistics

#### DATA\_SCI 7030: Applied SQL for Database and Analytics

Concentrates on leveraging well-designed relational databases using the entity-relationship model and normalization to extract valuable insights through basic and advanced SQL queries, data manipulation, modeling, ETL, security, and analytics using real datasets. Extended exploration of complex SQL operations such as query optimization, indexing, subqueries, window functions, triggers, and stored procedures supporting advanced analytics, statistics, and reporting needs is integrated through the course. Hands-on practical examples, exercises, and projects combine the essential elements of SQL with practical knowledge in building efficient and well-structured databases, ensuring data integrity, analytical results, and effective reporting. NoSQL, Network, and Geospatial databases are also examined. Graded on A-F basis only.

#### Credit Hours: 3

**Prerequisites:** DATA\_SCI 7010 or instructors consent required **Recommended:** Basic knowledge of data science principles and familiarity with the Python and R programming languages. No prior experience with database design or SQL is required

#### 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 will be taught with a hands-on approach to give students experience in handling data with a set of tools and programming environments. Visualizations toward exploratory data analysis and storytelling will be implemented. 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 or Corequisites: DATA\_SCI 7010 or instructors consent required

**Recommended:** 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 and basic knowledge of data science principles

#### 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, Information & AI Ethics

Relates ethical concepts, theories, and frameworks to Big Data, presentation of information, and Artificial Intelligence (AI) in industry, business, academia, and research settings through examination of realworld ethical dilemmas and example situations. Students will learn the social, ethical, legal and policy issues that often underpin big data, data science, and AI phenomena contributing to ethical dilemmas and their resolution. Use of discussions, writing, and interpreting case studies will help guard against the repetition of known ethical mistakes, inadequate ethical awareness, and ethical dilemma resolution preparation. Graded on A-F basis only.

Credit Hour: 1

Prerequisites: DATA\_SCI 7010 or instructor's consent required



### Credit Hours: 2

Prerequisites: DATA\_SCI 7040 and DATA\_SCI 8010 or instructor's consent required

### 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: Data Science & Analytics Capstone Project

The Data Science and Analytics Capstone Project is a research or business-driven project performed over an extended course engagement to demonstrate realistic design, management, and execution of a full data science project based on the DSA Project Life Cycle Model, ranging from raw data to business insights, knowledge discovery, or automation. This team project will emphasize realistic data carpentry and exploratory analysis using a variety of complex data sources and target and features. The project will include significant domain appropriate statistical or machine learning modeling for knowledge discovery, predictive analytics, or prescriptive analytics. This complete applied data science project builds upon the knowledge and skills developed throughout the DSA program. Student teams are paired with a faculty mentor for the project. This capstone project experience is an opportunity to refine and polish your skills as a data scientist. The presentation of your final capstone standalone data story/product will appropriately demonstrate your work and capabilities as a professional data scientist. Graded on A-F basis only.

#### Credit Hours: 3

Prerequisites: DATA\_SCI 8080 or instructor's consent required

## 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.

#### 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 with Applied AI and Machine Learning

This course leverages the foundations in statistics and modeling to teach applied concepts in AI and machine learning (AI/ML). Participants will learn various classes of machine learning and modeling techniques and gain an in-depth understanding of how to select appropriate techniques for various data science tasks and predictive analytics. 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 AI/ML workflows into dataintensive analytical processes. Graded on A-F basis only.

#### Credit Hours: 3

**Prerequisites:** DATA\_SCI 7010 or instructor's consent. Working knowledge of programming in python. Recommended DATA\_SCI 7020

#### 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: Data Science and Analytics Case Study Project

The Data Science and Analytics Case Study is an opportunity for students to work in a small team setting to achieve realistic data carpentry, exploratory analysis, and visualization goals. This is the first opportunity for students to apply what they learned in first year courses to a less structured, real-world problem. Student teams are paired with a faculty mentor who serves in the stakeholder-role for the project. Students are expected to work in teams of 3-4. The case study is an opportunity to build your skills as a data scientist while pursuing interests shared by your team and the faculty member. Design of a data science and analytics project guided by the DSA Project Life Cycle Model, applied project management principles, provides the basis for prototyping and communicating a final standalone data story/product using infographics, presentations, R-Shiny, Plotly, or other interactive dashboards and web applications is addressed in this project experience. 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 fundamental concepts surrounding the analysis of health data. Topics include ethics and regulations of protected health data, healthcare data standards, and artificial intelligence & statistical machine learning 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 and DATA\_SCI 7040, or instructor's consent required

#### 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 handson 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 indepth 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

This course covers the fundamentals of advanced, interactive visualizations by applying theoretical aspects of visualization and communications within an interactive programming environment to create web applications and dashboards to teach interactive design principles to design and implement interactive predictive and exploratory interfaces for data analytics. Conceptual topics will include cognition and visual perception, cognitive bias, sensory-level mechanisms in visual perception, human-centered design and evaluation, interaction techniques, evaluation of efficiency, user interface styles. Students will also have reading assignments chosen from published papers. Extensive hands-on programming practices and exercises will include learning



the Shiny platform for creating interactive web applications to display, manipulate, and explore data with efficient, interactive visualizations for purposes of predictive analytics, and a brief introduction to HTML and Javascript. Graded on A-F basis only.

#### Credit Hours: 3

Prerequisites: DATA\_SCI 7040 or instructor's consent required

#### 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 (DM) and information retrieval (IR) systems. The content covers various data mining topics and methods to extract hidden and predictive patterns from large data collections. Furthermore, theory and techniques for supervised, semi-supervised, and unsupervised, indexing, and retrieving relational, non-relational, text-based, vector, and multimedia databases are considered. Topics include an introduction to the ensemble leaning and feature importance, data mining process, mining frequent patterns, and pattern analysis, as well as different information retrieval models and evaluation, query languages and operations, indexing/searching methods, and recommender systems. Discussions and application to different domain contexts provide use case scenarios; emerging trends in the consideration of artificial intelligence and language models are also considered. Graded on A-F basis only.

#### Credit Hours: 3

Prerequisites: DATA\_SCI 7010 and DATA\_SCI 8010 or instructor's consent required

#### DATA\_SCI 8420: Cloud Computing for Data Analytics

This course introduces students to the use of cloud computing for Artificial Intelligence/Machine Learning (AI/ML) and big data analytics. Topics include a survey of cloud computing platforms, architectures, and use-cases, including concepts such as cloud compute, cloud storage, cloud containers and applications, as well as serverless data flow designs for AI/ML and advanced analytics. Students will examine data science scaling techniques and algorithms using a variety of cluster and cloud paradigms, built atop cloud computing concepts in multiple commercial cloud platforms. Graded on A-F basis only.

#### Credit Hours: 3

Prerequisites: DATA\_SCI 8010 or instructor's consent required

#### 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 and Geodatabase Development

This course provides a deeper dive into the theoretical, conceptual, foundational, and practical issues encountered when working with geospatial data (both vector and raster). A focus on integrating and leveraging geospatial data into a data science database and project as well as the concept of 'thinking spatially'. Data discovery, access, evaluation of use, retrieval, projection, datum, loading, and other technical and data carpentry concepts are investigated. Important aspects of geospatial database design and storage paradigms (enterprise versus desktop) are explored along with addressing Geospatial Big Data. Core issues in geospatial data storage, management, exploitation, feature engineering, multi-data set entity resolution / correlation, and dealing with special data types such as elevation (3D) and time-series are also examined. Graded on A-F basis only.

#### Credit Hours: 3 Prerequisites: DATA\_SCI 7030 or instructor's consent required



#### DATA\_SCI 8520: Spatial Analytics and Geostatistical Analysis

This course provides an overview of key issues encountered when working with and analyzing the various forms of spatial data (raster, vector, 3-D, temporal) as well as an overview of major spatial analysis tools, application areas, and analytical approaches. Simple geostatistical measures of centrality, advancing to spatial autocorrelation, geographic weighted regression, and various forms of interpolation. Laboratory, practice, and exercise work will focus on implementation, geostatistical analysis, and derived analytical spatial measures to inform context. Discussions will focus on interpretation 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 required

#### DATA\_SCI 8530: Geospatial AI and Image Analysis

This course provides an overview of key principles of Artificial Intelligence/Machine Learning (AI/ML) as applied to imagery analysis and advanced geospatial analytics. This will include classification of imagery (land cover), change detection, anomalies identification, and forecasting / prediction. The course also delves into and discusses both theoretical and practical issues associated with dynamic spatial systems and techniques such as digital twins, smart cities, as well as touching on spatial simulation methods (i.e., agent-based modeling and cellular automata). Labs, practices, and exercises cover standard geospatial AI processing techniques, including preprocessing and normalization, pixellevel feature extraction, information extraction, classification, data fusion, downscaling, and image understanding. Graded on A-F basis only.

#### Credit Hours: 3

Prerequisites: DATA\_SCI 8510 or instructor's consent required

#### DATA\_SCI 8615: Atomistic Materials Analytics

(same as MAE 8615, CHEM 8615, CH\_ENG 8615, CMP\_SC 8615, ECE 8615, BIOL\_EN 8615). Introduction to the principles of materials characterization and analysis leading to information extraction from experimental and computational data sets. Examines theoretical and practical issues associated with understanding the relationships between atomic/molecular structure and material properties. Describes encoding and visualizing atomic structure information. Measurement techniques covered include Ultraviolet-Visible (UV-Vis) spectroscopy, (Fourier Transform) Infrared Spectroscopy (FTIR), X-ray diffraction (XRD), and electronic conductivity. Covers standard data processing techniques, including sparse vs. abundant data, preprocessing and normalization, fixed length vs. variable length data, heterogeneous data streams, feature extraction, and the role of simulated data. Graded on A-F basis only.

#### Credit Hours: 3

Prerequisites: DATA\_SCI 7010, DATA\_SCI 8010, Instructor consent