

# DATA SCIENCE, MINOR

Students trained in Data Science will study a blend of topics from many subdomains of communications, philosophy, mathematics, computer science, and information science. A data scientist has a breadth of experience across all of these fields but may not have as much knowledge as a specialist in any particular field. Furthermore, a data scientist trained at William & Mary is equipped to consider the philosophical and moral implications of algorithm development and data collection, and the societal ramifications that new approaches to data manipulation could have. This combination allows William & Mary Data Science students to (a) efficiently conduct computational analyses within their own knowledge domain, (b) manage teams of more specialized individuals to answer far-ranging questions, and (c) communicate technical findings to a wide variety of audiences. Individuals with this knowledge profile are revolutionizing a wide set of domains and are in very high demand not just by faculty researchers at William & Mary, but also by the public and private sectors.

William & Mary offers a Bachelor of Science and a Minor in Data Science, which draw on faculty expertise from many departments. There are four key pedagogic pillars students will be expected to engage with during their time in the program: Computation, Application, Communication, and Deliberation.

- **Computation** - the computer science and mathematics required to responsibly use large datasets to create new knowledge. This is a focus of the introductory coursework, as well as the elective courses using more specialized data
- **Application** - the skills and creative thinking required to identify novel ways to apply computation to new problems. Application is present in all core courses within the Data Science program with goal to promote confidence and creativity.
- **Communication** - the written and oral skills to clearly transmit conclusions and implications derived from data analysis. While many students will naturally receive some communications training as a part of their time at William & Mary, the Data Science program promotes an additional depth of skill due to the challenges in communicating large sets of data. Communication is a strong theme within all Data Science core courses.
- **Deliberation** - the ability to consider the societal, moral, and ethical implications of Data Science. Students are required to take one course examining these topics, but many courses will integrate this type of thinking.

The B.S. in Data Science will require a minimum of 40 credits. The curriculum includes three tracks: Data Application, Algorithms, and Spatial Data Analytics. The degree program culminates in a capstone experience. Each track will further strengthen and deepen students' understanding in data science.

The focus of the core curriculum is to provide students with a solid foundation in Data Science. Students learn data science theory and applications, including critical evaluation of how data can be used to solve novel problems, deliberation (considering the ethical, moral, and societal implications of data science), and communication. Through the core curriculum students learn the basics of programing, modeling, machine learning, data visualization, database structures, and ethics in data science. Students also will take one course in linear algebra and two courses in mathematical statistics. The curriculum provides opportunities for students to use their skills and knowledge to manage

and analyze large data sets efficiently and effectively and to identify and answer novel questions in a variety of settings.

Students will choose a track area to gain knowledge, skills, and abilities that are more specific to particular career aspirations. They are required to take three courses from one of the following tracks: Data Application, Algorithms, or Spatial Data Analytics. Coursework for the Data Application track focuses on teaching additional skills (e.g., data with time dependencies) and providing a more in-depth understanding of analytical and data visualization tools commonly used by data scientists employed by the private industry or government. Coursework for the Algorithms track focuses on expanding students' abilities to develop new software or algorithms for the ingestion or analysis of large sources of frequently near-real-time data. Coursework for the Spatial Data Analytics track focuses on integration of analytical and visualization tools that data scientists typically use when working with data that have spatial dependencies.

In the capstone experience, each student will work closely with a program faculty member to conduct a substantial research project that focuses on synthesis and critical analysis, problem solving in an applied and/or academic setting, creation of original material or original scholarship, and effective communication with diverse audiences.

The curriculum for the Minor in Data Science requires 18 credit hours to complete. A minimum of 120 credits are required to fulfill all degree requirements; information on other degree requirements can be found in the catalog section "Requirements for Degrees." The minor in Data Science is designed to be paired with a wide variety of majors across William & Mary, so there are no restrictions on the primary major pursued in conjunction with the Data Science minor. Under most circumstances, the Data Science minor should be declared no later than the second semester of the Junior year to ensure the minor can be completed. Two courses may be counted toward both your primary major and minor; some courses may be substituted with permission from the director.

## Course Requirements

| Code                                                                                                                                                                                                                                                                       | Title                                                           | Hours |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|-------|
| <b>Required Credits</b>                                                                                                                                                                                                                                                    |                                                                 |       |
| DATA 101<br>or CSCI 141                                                                                                                                                                                                                                                    | Reasoning with Generative AI<br>Modern Programming Fundamentals | 3     |
| DATA 201                                                                                                                                                                                                                                                                   | Introduction to Data Science & AI                               | 3     |
| DATA 202                                                                                                                                                                                                                                                                   | AI & Data Ethics                                                | 3     |
| DATA 305<br>or DATA 301                                                                                                                                                                                                                                                    | Designing AI Agents<br>Applied Machine Learning                 | 3     |
| <b>Additional Credits</b>                                                                                                                                                                                                                                                  |                                                                 |       |
| Students must select two or more courses overlapping with the pedagogic pillars of the Data Science program (Computation, Application, Communication, Deliberation), which must have a focus on the analysis of temporal, spatial, or numerical data. Approve courses are: |                                                                 | 6     |
| BIOL 325                                                                                                                                                                                                                                                                   | Introduction to Quantitative Biology                            |       |
| BIOL 327                                                                                                                                                                                                                                                                   | Intro to Biostatistics                                          |       |
| BIOL 445                                                                                                                                                                                                                                                                   | GIS for Biologists                                              |       |
| BUAD 350                                                                                                                                                                                                                                                                   | Introduction to Business Analytics                              |       |
| BUAD 351                                                                                                                                                                                                                                                                   | Operations Management                                           |       |
| BUAD 352                                                                                                                                                                                                                                                                   | Decision-making through Visualization and Simulation            |       |
| BUAD 460                                                                                                                                                                                                                                                                   | Big Data Analytics with Machine Learning                        |       |

|                                                                        |                                                                     |
|------------------------------------------------------------------------|---------------------------------------------------------------------|
| BUAD 463                                                               | Supply Chain Analytics                                              |
| BUAD 466                                                               | Developing Business Intelligence                                    |
| BUAD 467                                                               | Predictive Analytics                                                |
| BUAD 468                                                               | Prescriptive Analytics                                              |
| BUAD 469                                                               | Advanced Modeling Techniques                                        |
| or DATA 300 or 400 level courses, excluding DATA 480, DATA 481 and 491 |                                                                     |
| ECON 308                                                               | Econometrics                                                        |
| ECON 331                                                               | Intro Mathematical Economics                                        |
| ECON 407                                                               | Cross Section Econometrics                                          |
| ECON 408                                                               | Time Series Econometrics                                            |
| ECON 412                                                               | Empirical Microeconomics                                            |
| ECON 413                                                               | Applied Macroeconomics                                              |
| ECON 414                                                               | Bayesian Econometrics                                               |
| GIS 201                                                                | Introduction to Geographic Information Systems and Spatial Analysis |
| GIS 405                                                                | Geovisualization & Cartographic Design                              |
| GIS 410                                                                | Introduction to Remotely Sensed Imagery and Analysis                |
| GIS 420                                                                | Advanced GIS Analysis & Programming                                 |
| GOVT 301                                                               | Research Methods                                                    |
| GOVT 302                                                               | Quantitative Methods                                                |
| GOVT 307                                                               | Political Polling and Survey Analysis                               |
| MATH 309                                                               | Theory of Vector Spaces                                             |
| MATH 323                                                               | Operations Research: Deterministic Models                           |
| MATH 332                                                               | Graph Theory and its Applications                                   |
| MATH 351                                                               | Probability and Statistics for Scientists                           |
| MATH 352                                                               | Statistical Data Analysis                                           |
| MATH 353                                                               | Advanced Statistical Data Analysis                                  |
| MATH 408                                                               | Matrix Analysis                                                     |
| MATH 417                                                               | Vector Calculus for Scientists                                      |
| MATH 424                                                               | Operations Research: Stochastic Models                              |
| MATH 451                                                               | Probability                                                         |
| MATH 452                                                               | Mathematical Statistics                                             |
| MATH 455                                                               | Statistical Learning                                                |
| MATH 465                                                               | Mathematics of Financial Economics                                  |
| PSYC 302                                                               | Research Methods/Psy                                                |

**Total Hours****18**

<sup>1</sup> Additional courses may be approved by the Department of Data Science Undergraduate Curriculum Committee; your advisor can assist you in that petition, or you can contact the Data Science Director of Undergraduate Studies directly. Special topics courses in other departments (i.e., courses that may be offered under different names) must be approved through this process.