

Undergraduate Certificate in Data Science Essentials

Program outcomes: Data scientists will have essential competencies in several areas related to analysis of data. In particular, a data scientist should: have strong programming ability in a language popular in data science (e.g., Python, R, Julia); be able to extract, manipulate, and visualize data; have an understanding of probability and statistics in order to quantify uncertainty; be able to build complex models for finding patterns and explaining data. This certificate should provide students with essential skills for introductory data science. Additional training related to database management, high-performance computing, and modeling would likely be necessary for advanced data science analysis.

Course requirements:

Programming In order to ensure adequate programming skills for data science, students should take a course that develops strong programming skills in a programming language popular in data science (e.g., Python, R, Julia). The list of currently approved courses includes:

Math 1376 Programming for Data Science
Math 4650 Numerical Analysis I
ISMG 4400 Web Application Development (Programming Fundamentals with Python)

Probability and statistics In order to ensure that students can accurately quantify the likelihood of various outcomes and quantify uncertainty related to estimation and prediction, students should take a course that covers basic probability and statistics. The list of currently approved courses includes:

Math 2830 Introductory Statistics (or equivalent coursework with Undergraduate Committee approval)
Math 3382 Statistical Theory
Math 3800 Probability and Statistics for Engineers

Data manipulation and visualization In order to ensure that students are able to comfortably work with and visualize data, students should take a course developing skills related to obtaining, manipulating, and visualizing data. The list of currently approved courses includes:

Math 3376 Data Wrangling & Visualization

Data modeling In order to ensure that students are able to build reasonably complex models for explaining or identifying patterns in data, students should take a course that largely focuses on describing the behavior of data (whether synthetic or observed) via tools like simulation, direct model building, association, or a complementary approach. The list of currently approved courses includes:

Math 3301 Introduction to Optimization in Operations Research
Math 4387 Applied Regression Analysis
Math 4830 Applied Statistics

Summary of requirements:

Skill	Course
Programming	Choose one: Math 1376 Programming for Data Science Math 4650 Numerical Analysis I ISMG 4400 Web Application Development
Probability and statistics	Choose one: Math 2830 Introductory Statistics (or equivalent) Math 3382 Statistical Theory Math 3800 Probability and Statistics for Engineers
Data manipulation and visualization	Math 3376 Data Wrangling & Visualization
Data modeling	Choose one: Math 3301 Introduction to Optimization in Operations Research Math 4830 Applied Statistics Math 4387 Applied Regression Analysis

At least 9 of the 12 credits used for the certificate must be taken in residency at CU Denver.