Program Learning Outcomes
School of Information and Computer Sciences
Data Science B.S.

The broad educational goals of the program are to produce students with foundational training in data analysis, with a dual emphasis on the principles of statistics and computing, and the ability to apply these principles to a range of data-driven problems. More specifically, the educational goals of the proposed major are to:

- Provide students with a foundation in mathematical and statistical aspects of data analysis;
- Provide students with a foundation in the general principles of computer science;
- Teach students how to utilize their knowledge of statistical and computing principles to develop algorithms, and software for solving real-world data analysis problems;
- Provide students with practical experience in applying their knowledge of theories, methods, and tools, to a variety of data analysis problems;
- Teach students how to communicate effectively using data.

The following is a list of learning outcomes (LOs) for graduating seniors from this major:

LO1: Knowledge of foundational mathematical concepts relevant to data analysis, including
- calculus (integration and differentiation), both univariate and multivariate
- applied linear algebra (matrix operations, linear transformations, projections in Euclidean space)
- concepts in probability including univariate and multivariate distributions for both discrete and continuous densities, random variables, independence and conditional independence, expectation, limit theorems

LO2: Knowledge of basic principles in computer science, including
- basic concepts in Boolean algebra and logic
- basic concepts in computer hardware
- basic concepts in operating systems and programming languages
- basic concepts in data management
- basic concepts in algorithms and data structures
- basic concepts in computational complexity

LO3: Knowledge of foundational statistical concepts, including
- foundational statistical theory including parameter estimation, hypothesis testing, decision theory, maximum likelihood, and Bayesian methods
- statistical modeling principles, including linear and logistic regression models, non-parametric methods, diagnostic techniques

- design of statistical studies, including sampling methodologies, random assignment, data collection, efficiency, and issues of bias, causality, confounding, and coincidence; sample size and power calculations
- exploratory data analysis methods including visualization, projection methods, clustering, and density estimation

LO4: Knowledge of basic principles in statistical computing including
- ability to understand and program randomization and simulation techniques such as bootstrap, Monte Carlo and cross-validation methods
- ability to understand and program optimization methods such as gradient descent and solving linear systems of equations
- ability to understand and implement basic concepts and algorithms in machine learning, such as predictive modeling for classification and regression
- basic knowledge of issues in handling large-scale data, such as scalability, indexing, data management, and distributed computing

LO5: Ability to take a real-world data analysis problem, formulate a conceptual approach to the problem, match aspects of the problem to previously learned theoretical and methodological tools, break down the solution into a step-by-step approach, and implement a working solution in a modern software language, including
- knowledge of basic concepts in developing and testing software programs
- knowledge of basic concepts in software engineering skills and experience in developing software in at least one modern programming language such as Python
- basic skills in data management, including relational database systems and SQL
- basic skills in at least one statistical programming environment such as R
- basic concepts in data collection such as sampling methods, handling missing data, data provenance, ethics in data collection
- basic concepts in data privacy, e.g., legal aspects, institutional review boards, algorithmic techniques for data anonymization

LO6: Ability to communicate effectively in data analysis projects, including
- effective technical writing and presentations
- teamwork and collaboration
- communicate results to both data analysis specialists and non-specialists accurately, concisely, and effectively