CS 88 Course Outline as of Fall 2024

CATALOG INFORMATION

Dept and Nbr: CS 88 Title: FOUND DATA SCI
Full Title: Foundations of Data Science
Last Reviewed: 2/12/2024

Units		Course Hours per Week		Nbr of Weeks	Course Hours Total	
Maximum	4.00	Lecture Scheduled	3.00	17.5	Lecture Scheduled	52.50
Minimum	4.00	Lab Scheduled	3.00	8	Lab Scheduled	52.50
		Contact DHR	0		Contact DHR	0
		Contact Total	6.00		Contact Total	105.00
		Non-contact DHR	0		Non-contact DHR	0

Total Out of Class Hours: 105.00 Total Student Learning Hours: 210.00

Title 5 Category:	AA Degree Applicable
Grading:	Grade or P/NP
Repeatability:	00 - Two Repeats if Grade was D, F, NC, or NP
Also Listed As:	
Formerly:	

Catalog Description:
In this course, students will study the Foundations of Data Science from three perspectives: inferential thinking, computational thinking, and real-world relevance. Given data arising from some real-world phenomenon, how does one analyze that data so as to understand that phenomenon? The course teaches critical concepts and skills in computer programming and statistical inference, in conjunction with hands-on analysis of real-world datasets, including economic data, document collections, geographical data, and social networks. It delves into social issues surrounding data analysis such as privacy and design.

Prerequisites/Corequisites:

Recommended Preparation:
Course Completion of CS 81.41A and one of the following MATH courses (MATH 15, MATH 1A, MATH 4) or equivalent

Limits on Enrollment:

Schedule of Classes Information:
Description: In this course, students will study the Foundations of Data Science from three

perspectives: inferential thinking, computational thinking, and real-world relevance. Given data arising from some real-world phenomenon, how does one analyze that data so as to understand that phenomenon? The course teaches critical concepts and skills in computer programming and statistical inference, in conjunction with hands-on analysis of real-world datasets, including economic data, document collections, geographical data, and social networks. It delves into social issues surrounding data analysis such as privacy and design. (Grade or P/NP)
Prerequisites/Corequisites:
Recommended: Course Completion of CS 81.41A and one of the following MATH courses (MATH 15, MATH 1A, MATH 4) or equivalent
Limits on Enrollment:
Transfer Credit: CSU;UC.
Repeatability: Two Repeats if Grade was D, F, NC, or NP

ARTICULATION, MAJOR, and CERTIFICATION INFORMATION:

AS Degree:	**Area**			Effective:	Inactive:
	L2	Mathematical Concepts & Quantitative Reasoning		Fall 2025	
CSU GE:	**Transfer Area**			Effective:	Inactive:
IGETC:	**Transfer Area**			Effective:	Inactive:
CSU Transfer:	Transferable	Effective:	Fall 2024	Inactive:	
UC Transfer:	Transferable	Effective:	Fall 2025	Inactive:	

CID:

Certificate/Major Applicable:
Both Certificate and Major Applicable

COURSE CONTENT

Student Learning Outcomes:
At the conclusion of this course, the student should be able to:
1. Employ foundational programming concepts to explore and analyze datasets.
2. Apply foundational data science to explore and analyze datasets.
3. Analyze real-world data sets using a modern programming language, problem decomposition, and code design strategies.
4. Identify limitations and issues surrounding data analysis in terms of bias, ethics, establishing causality, and privacy.

Objectives:
At the conclusion of this course, the student should be able to:
1. Employ foundational programming concepts such as data types, basic data structures such as lists and tables, functions, looping, decision making, and input/output commands to explore and analyze datasets.
2. Apply foundational data science concepts including extracting data from tables based on specific criteria, computing summary statistics, creating data visualizations, simulating experiments, and inferential statistics.
3. Analyze real-world data sets using Python, problem decomposition methods, and code design

strategies.
4. Use computer simulations to explore concepts in probability and statistical inference including machine learning techniques.
5. Recognize limitations and issues surrounding data analysis in terms of bias, causality, ethics, and privacy.

Topics and Scope:

I. Causality and Experiments
- A. Establishing causality
- B. Randomization

II. Programming Skills for Use in Applications
- A. Relevant programming libraries and utilities
- B. Expressions
- C. Variables
- D. Data types
- E. Tables and arrays
- F. Operators
- G. Errors
- H. Functions and methods
- I. Iteration

III. Statistical Concepts through Computer Simulations
- A. Computer-generated descriptive statistics
- B. Data visualizations
- C. Randomness and probability
- D. Sampling and empirical distributions
  1. Sampling from a population
  2. Empirical distribution of a statistic
  3. Normal distributions
  4. Central Limit Theorem
- E. Estimation
  1. Bootstrapping
  2. Confidence intervals
- F. Hypothesis testing
  1. Test statistics
  2. P-value
  3. A/B testing
  4. Decision errors

IV. Machine Learning Techniques for Use in Applications
- A. Linear regression
  1. Correlation coefficient
  2. Linear regression equation
  3. Least-squares
  4. Predictions
  5. Residuals and residual plots
- B. Classification
  1. Training and testing
  2. Accuracy
  3. Proximity algorithms
  4. Multiple linear regression

V. Ethical Concerns in Data Science
- A. Data privacy

B. Machine learning and bias

All sections are covered in the lecture and lab portions of the course

Assignment:

Lecture- and Lab-related Assignments:
1. Read approximately 0-50 pages per week from the course Jupyter Notebook on topics such as forming and testing hypotheses, interpreting graphical and numerical summaries of data sets, identifying features in data to use for machine learning techniques such as classification.
2. Weekly discussions.
3. Data analysis and interpretation.
4. Completion of four projects: each shall include use of relevant tools, analysis of data sets, report of findings, and statistical inference addressing potential bias with machine learning algorithms.
5. Written reports demonstrating students' ability to make inferences about populations based on random sample data.
6. Weekly assignments using Jupyter Notebook and Python programming language.
7. Exam(s) (0-8) and final exam.

Methods of Evaluation/Basis of Grade:

Writing: Assessment tools that demonstrate writing skills and/or require students to select, organize and explain ideas in writing.

Written reports	Writing 10 - 20%

Problem Solving: Assessment tools, other than exams, that demonstrate competence in computational or non-computational problem solving skills.

Weekly assignments. Projects. Data analysis and interpretation	Problem solving 10 - 50%

Skill Demonstrations: All skill-based and physical demonstrations used for assessment purposes including skill performance exams.

None	Skill Demonstrations 0 - 0%

Exams: All forms of formal testing, other than skill performance exams.

Exam(s) and final exam	Exams 20 - 60%

Other: Includes any assessment tools that do not logically fit into the above categories.

Participation and discussions	Other Category 5 - 20%

Representative Textbooks and Materials:
Computational and Inferential Thinking: The Foundations of Data Science. 2nd ed. Adhikari, Ani and DeNero, John and Wagner, David. UC Berkeley. 2021.