College of San Mateo
Official Course Outline

1. **COURSE ID:** CIS 124 **TITLE:** Foundations of Data Science
   Units: 4.0 units **Hours/Semester:** 48.0-54.0 Lecture hours; 48.0-54.0 Lab hours; 96.0-108.0 Homework hours; 192.0-216.0 Total Student Learning hours
   Method of Grading: Grade Option (Letter Grade or Pass/No Pass)
   Recommended Preparation:
   Eligibility for MATH 200, Eligibility for CIS 254

2. **COURSE DESIGNATION:**
   Degree Credit
   Transfer credit: CSU; UC

3. **COURSE DESCRIPTIONS:**
   Catalog Description:
   Introduces the core concepts of data science, including statistical inference and computational thinking, while working with real data, such as economic data, geographic data and social networks. Designed for entry-level students from any major, teaches basic concepts of computer programming.

4. **STUDENT LEARNING OUTCOME(S) (SLO'S):**
   Upon successful completion of this course, a student will meet the following outcomes:
   1. Write small programs that manipulate and combine data sets and carry out iterative procedures.
   2. Calculate and interpret specified statistics of a given dataset.
   3. Carry out statistical analyses including computing confidence intervals and performing hypothesis tests in a variety of data settings and form correct conclusions.
   4. Articulate the benefits and limits of computing technology for analyzing data and answering questions.
   5. Make accurate predictions using regression and classification techniques.
   6. Assess the accuracy and variability of a prediction.

5. **SPECIFIC INSTRUCTIONAL OBJECTIVES:**
   Upon successful completion of this course, a student will be able to:
   1. Write small programs that manipulate and combine data sets and carry out iterative procedures.
   2. Extend a program with multiple functions to add functionality.
   3. Identify the sources of randomness in an experiment.
   4. Calculate and interpret specified statistics of a given dataset.
   5. Generate and interpret histograms, bar charts, and box plots.
   6. Carry out statistical analyses including computing confidence intervals and performing hypothesis tests in a variety of data sets and form correct conclusions.
   7. Given a question and an analysis, explain whether the analysis addresses the question.
   8. Articulate the benefits and limits of computing technology for analyzing data and answering questions.
   9. Make accurate predictions using regression and classification techniques.
   10. Assess the accuracy and variability of a prediction.

6. **COURSE CONTENT:**
   Lecture Content:
   1. Introduction
   2. Causality and experiments
   3. Programming in Python
      A. Expressions, names and data types
      B. Sequences
      C. Tables
      D. Functions
      E. Visualization
   4. Randomness
   5. Sampling and empirical distributions
   6. Testing hypothesis
   7. Comparing two samples
   8. Estimation

9. Why the mean matters
10. Prediction
11. Inference for regression
12. Classification

Lab Content:

1. Write Python scripts to:
    A. evaluate expressions
    B. group values into collections
    C. sort and organize data in tables
    D. generate data visualization:
        a. scatter plots
        b. bar graphs
        c. histograms
2. Write custom functions for data visualization in Python
3. Write Python scripts for real data analysis

7. **REPRESENTATIVE METHODS OF INSTRUCTION:**

Typical methods of instruction may include:

A. Lecture
B. Lab
C. Activity
D. Discussion

8. **REPRESENTATIVE ASSIGNMENTS**

Representative assignments in this course may include, but are not limited to the following:

Writing Assignments:

A. Weekly lab assignments to practice programming skills needed for data manipulation, organization and visualization
B. Programming assignments to write scripts for data visualization
C. Group assignments to perform data analysis
D. Discussions on statistics and probability concepts
E. End-of-semester project including the following steps:
    a. Proposal submission
    b. Implementation upon approval
    c. Presentation

Reading Assignments:

Weekly readings from textbook.

9. **REPRESENTATIVE METHODS OF EVALUATION**

Representative methods of evaluation may include:

A. Class Participation
B. Exams/Tests
C. Group Projects
D. Homework
E. Lab Activities
F. Oral Presentation
G. Projects
H. Quizzes

10. **REPRESENTATIVE TEXT(S):**

Possible textbooks include:

A. Ani Adhikari, John DeNero, David Wagner. *Computational and Inferential Thinking: The Foundations of Data Science (OER)*, 2nd ed. Berkeley: UC Berkeley, 2021

Possible software includes:

A. Python. www.python.org, 3.0 ed.

Origination Date: December 2021
Curriculum Committee Approval Date: January 2022
Effective Term: Fall 2022
Course Originator: Hellen Pacheco