MATH 118 - Foundations in Data Science

DIVISION/DEPARTMENT:	Laney - Division of Math & Science / L - Mathematics		
ORIGINATOR:	Oh, Kyla	**BOARD OF TRUSTEES APPROVAL DATE:**	01/24/2023
STATE CONTROL #:	CCC000636044	**STATE APPROVAL DATE:**	01/27/2023
		CURRICULUM COMMITTEE APPROVAL DATE:	11/18/2022
		REQUISITE VALIDATION:	11/18/2022
		CURRENT EFFECTIVE DATE:	08/01/2023

1. REQUESTED CREDIT CLASSIFICATION:
COURSE TYPE: D - Credit - Degree Applicable

CB08 BASIC SKILL STATUS (PBS STATUS): N - Not Basic Skills

CB24-PROGRAM COURSE STATUS: 1 - Program Applicable

2. DEPT/COURSE NO:
MATH 118

3. COURSE TITLE:
Foundations in Data Science

4. COURSE:
Laney New Course

TOP NO.: 1701.00 - Mathematics, General

5. UNITS:
Variable No

Units (Min) 4.000	**Min Total**
Lecture Hours (Min) 3.000	52.5
Lab/Studio/Activity Hours (Min) 3.000	52.5
Units (Max)	**Max Total**
Lecture Hours (Max)	0
Lab/Studio/Activity Hours (Max)	0

6. SELECTED TOPIC:
NO. OF TIMES OFFERED AS SELECTED TOPIC: **AVERAGE ENROLLMENT:**

7. JUSTIFICATION FOR COURSE:
The course is intended to satisfy the Laney College Area 4b, CSU GE Area B4, and IGETC Area 2. The new course will be added to the list of A.S. Math Degree Requirements. Specifically, the degree requirements will be modified to indicate students can take one of: Math 13, Math 11 or Math 118/CIS 118. Additionally, this course will serve as an entry point for an in-development Data Analytics certificate at Laney College.

8. COURSE/CATALOG DESCRIPTION:
Foundations of Data Science: Collecting data, sampling; observational and experimental studies; tables and graphs; histograms; measures of center, spread, and correlation; probability; random variables; normal distributions; confidence intervals; hypothesis testing; Regression; ethical issues; algorithm design, flow charting, and debugging; elements of good programming style. Not open for credit to students who have completed or are currently enrolled in CIS 118.

9. OTHER CATALOG INFORMATION:
a. Modular: No

If yes, how many modules:

b. Open entry/open exit: No

c. Grading Policy: Letter Grade Only

d. Eligible for credit by Exam: No

e. Repeatable according to state guidelines: No

f. Required for degree/certificate (specify):

g. Meets GE/Transfer requirements (specify):

h. C-ID Number:

Expiration Date:

i. Are there prerequisites/corequisites/recommended preparation for this course? Yes

10. LIST STUDENT PERFORMANCE OBJECTIVES (EXIT SKILLS):
1. Distinguish among different scales of measurement and their implications;
2. Use computer scripts to organize data in tables and display data graphically;
3. Compute measures of central tendency and variation for a data set using a programming language;
4. Write scripts to calculate the mean, standard deviation, and variance of a discrete distribution;
5. Identify standard sampling and data-gathering methods and their respective advantages and disadvantages;
6. Distinguish the difference between sampling distributions and population distributions, and use the Central Limit Theorem for analysis;
7. Construct and interpret confidence intervals;
8. Use appropriate statistical techniques to analyze and interpret applications based on data from disciplines including business, social sciences, psychology, life science, health science, and education.
9. Write scripts to calculate probabilities;
10. Write scripts to determine levels of statistical significance and p-values and and interpret the results;
11. Identify basic concepts of hypothesis testing;
12. Interpret the output of a technology-based statistical analysis;
13. Make predictions using machine learning techniques such as clustering and linear regression.
14. Write programs/scripts utilizing variables and data structures.

15. Write and debug programming code using a programming language in a development environment.
16. Write programming code using alternation and iteration.
17. Write programming code using functions.
18. Write appropriate programming algorithms.
19. Write programs / scripts utilizing inputs and producing output.

11. COURSE CONTENT:
LECTURE CONTENT:

1. **Research design (5%**)
    1. Outcomes
    2. Observational studies
    3. Experiments
        1. Treatment
        2. Control
        3. Simulation
    4. Association
    5. Causation
    6. Confounding variables
2. **Probability and statistics (60%)**
    1. Computer-generated descriptive statistics
        1. Data visualizations
            1. Scatter plots
            2. Line graphs
            3. Bar charts
            4. Histograms
            5. Overlaid graphs
        2. Summary statistics
            1. Max
            2. Min
            3. Count
            4. Mean
            5. Median
            6. Standard deviation
            7. Variance
            8. Distance
                1. Absolute value
                2. Euclidean
    2. Theoretical concepts of probability
        1. Events
        2. Probability definitions
        3. Complement rule
        4. Multiplication rule
        5. Addition rule
        6. Conditional probability and Bayes Theorem
    3. Inferential statistics supported by the use of technology
        1. Parameters and statistics
        2. Random sampling variability
            1. Empirical distributions
            2. Normal distributions
                1. Standardization
                2. Central Limit Theorem
        3. Estimation
            1. Percentiles
            2. Bootstrapping
            3. Confidence intervals
        4. Hypothesis testing
            1. Hypotheses
            2. Test statistics
            3. P-value
            4. Statistical significance
            5. A/B testing
            6. Errors in test results
3. **Programming skills for use in applications (28%)**
    1. Relevant programming libraries and utilities
    2. Expressions
    3. Variables
    4. Data types
    5. Tables and arrays
    6. Operators
    7. Errors
    8. Functions and methods
    9. Iteration
    10. Selection

4. **Machine learning techniques for use in applications (5%)**
    1. Linear regression
        1. Correlation coefficient
        2. Linear regression equation
        3. Least-squares
        4. Predications
        5. Residuals and residual plots
    2. Classification
        1. Training and testing
        2. Accuracy
        3. Proximity algorithms
        4. Multiple linear regression
5. **Ethical concerns in data science (2%)**
    1. Data privacy
    2. Machine learning and bias

LAB CONTENT:

1. Computer Programming Skills using Python (40%)
    a. Expressions
    b.Types and Sequences
    c. Arrays and Tables
2. Statistical Inference Through Simulation (5%)
3. Data distributions (5%)
    a. Histograms and Functions
    b. Sampling
4. Data Analysis using Python (20%)
    a. Tables
    b. Visualization
    c. Estimation
    d. Modeling
    e. Prediction
5. Testing and Verifying Hypothesis (5%)
6. Bootstrap(2%)
7. Correlation (3%)
8. Regression (10%)
9. Decisions (5%)
10. Predictions (5%)

12. METHODS OF INSTRUCTION:

- Lab
- Activity
- Lecture
- Discussion
- Critique
- Individualized Instruction
- Multimedia Content
- Threaded Discussions

13. ASSIGNMENTS:

Out-of-class Assignments (List all assignments, including library assignments. Requires two (2) hours of independent work outside of class for each unit/weekly lecture hour. Outside assignments are not required for lab-only courses, although they can be given.)

Override Outside Class Hours: No

Outside-of-Class Hours (Min): 6.000

Outside-of-Class Hours (Max): 0.000

Override Outside-of-Class Hours (Min):

Override Outside-of-Class Hours (Max):

Out of class Assignment:

1. Readings from the textbook 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. Writing assignments based on readings which include summarizing and analyzing real-world data.

3. Weekly assignments that focus on topics including creating and manipulating tables, operating on arrays, visualizing data, performing A/B tests, using bootstrap methods to create confidence intervals, and classifying data. Each student uses appropriate programming tools such as the Jupyter notebook application and representative programming languages. These assignments mirror instructor led in-class Jupyter notebook assignments.

4. Students complete four projects using relevant tools such as Jupyter notebooks to analyze real data sets and report their findings. Students use iterative programming to conduct statistical inference. The students are provided with a data set, and the project notebook contains a guiding series of questions that reflect the activities performed in the associated in-class and out-of-class assignments. Additionally, students reflect on the unintentional bias introduced when applying machine learning techniques to data sets that contain biased classifications.

14. STUDENT ASSESSMENT:

- ESSAY (Includes "blue book" exams and any written assignment of sufficient length and complexity to require students to select and organize ideas, to explain and support the ideas, and to demonstrate critical thinking skills.)
- COMPUTATION SKILLS
- NON-COMPUTATIONAL PROBLEM SOLVING (Critical thinking should be demonstrated by solving unfamiliar problems via various strategies.)
- SKILL DEMONSTRATION
- MULTIPLE CHOICE
- OTHER (Describe)
- **Other:** Participation: Active participation in class discussions, showing skills inferring information and forming predictions.

15. TEXTS, READINGS, AND MATERIALS:

A. Textbooks:

Adhikari, Ani;DeNero, John; and Wagner, David. *Computational and Inferential Thinking: The Foundations of Data Science*. second edition Creative Commons, 2021.

Website: Computational and Inferential Thinking: The Foundations of Data Science: https://inferentialthinking.com/chapters/intro.html Website: datascience Documentation: https://datascience.readthedocs.io/en/master/?badge=master Website: Project Jupyter: https://jupyter.org/ Website: Google Colaboratory: https://colab.research.google.com/

*Date is required: Transfer institutions require current publication date(s) within 5 years of outline addition/update.

B. Additional Resources:

Library/LRC Materials and Services:

The instructor, in consultation with a librarian, has reviewed the materials and services of the College Library/LRC in the subject areas related to the proposed new/updated course
Print Materials were reviewed? Yes

Non-Print Materials were reviewed? Yes

Online Materials were reviewed? Yes

Services were reviewed? Yes

Specific materials and/or services needed have been identified and discussed. Librarian comments:
Please provide a list of current recommended non-textbook titles to Acquisitions Librarian (asullivan@peralta.edu) to support this curriculum.

C. Readings listed in A and B above are:

Primarily college level

16. DESIGNATE OCCUPATIONAL CODE:
CB09 SAM Code: E - Non-Occupational

17. LEVEL BELOW TRANSFER:
CB21 Levels Below Transfer: Y - Not applicable

18. CALIFORNIA CLASSIFICATION CODE:
CB11 California Classification Codes: Y - Credit Course

19. NON CREDIT COURSE CATEGORY:
Y - Not Applicable, Credit course

20. FUNDING AGENCY CATEGORY:
CB23 Funding Agency Category: Y - Not Applicable (funding not used to develop course)

REQUISITES AND ADVISORIES:
PREREQUISITE:
MATH 203 Intermediate Algebra or Placement through multiple-measures assessment process

RECOMMENDED PREPARATION:
MATH 013 Introduction to Statistics or CIS 006 Introduction to Computer Programming