Lincoln University
LEARN. LIBERATE. LEAD.

Lincoln University
Department of Mathematical Sciences
Master Syllabus

COURSE TITLE:	Foundations of Data Science	**COURSE NUMBER:**	MAT-115
CREDIT HOURS:	3	**PREREQUISITE (S):**	MAT-110 and CSC-151
TERM:		**CO-REQUISITE (S):**	
COURSE MODE / METHOD:	Lecture + Lab	**MEETING DAY AND TIME:**	
INSTRUCTOR:		**CLASSROOM/LAB/STUDIO LOCATION:**	
OFFICE LOCATION:		**E-MAIL:**	
OFFICE HOURS:			

I. COURSE DESCRIPTION:

This course will teach the foundations of data science and data-driven research. It is designed to serve as an optional elective course for mathematics majors, minors, and other STEM, business and social science majors who are interested in pursuing data-driven careers or graduate study programs.

Students will acquire basic computational skills, basic knowledge of statistical analysis, error analysis, and the basics of Machine Learning. Students will also be familiarized with good practices for handling small and big data. After this class, students should be able to formulate a question, identify appropriate data to answer the question, prepare and analyze the data, extract knowledge and insights, make decisions and identify the confidence level of the decisions. The course will be organized in a modular fashion, with labs and projects assigned to students for group work.
Prerequisites MAT-110, CSC-151.

II. READING MATERIALS:

Required Text:

Jake VanderPlas; Python Data Science Handbook: Essential Tools for Working with Data, 1st Edition, O'Reilly Media, 2016; ISBN-13: 978-1491912058; ISBN-10: 1491912057.

Suggested Reading:

Ani Adhikari, John DeNero, David Wagner; Computational and Inferential Thinking: The Foundations of Data Science; 2nd Edition, UC Berkeley, 2021. (https://inferentialthinking.com/chapters/intro.html)

III. REQUIRED MATERIALS (Software/Technology Requirements nn Case Of Remote Classes):

- To access the course in case of remote learning, students must have a laptop or desktop computer with reliable internet connection, webcam, and working microphone and speakers.
- Minimum computer/programing literacy and web navigation skills.
- Access to Canvas and Zoom.

IV. COMMUNICATION MODE:

Subsequent communication will be conducted in-person during class, through email, live Zoom sessions in cases of remote learning and Canvas discussion forums. Students can also track the completion progress of their assignments and real time grade in the course through Canvas.

V. UNIVERSITY ACADEMIC, INSTITUTIONAL and TECHNOLOGY SUPPORT SERVICES:

This section shall outline where a student can access assistance for academic, institutional and technology support. Links to offices to be added here.

Academic Support: https://www.lincoln.edu/departments/academic-support

Institutional Support: https://www.lincoln.edu/departments/institutional-effectiveness-research-and-planning

Technology Support: https://www.lincoln.edu/departments/information-technology

CETL (Center for Excellence in Teaching & Learning): https://www.lincoln.edu/departments/moodle-resources

VI. ASSESSMENT CRITERIA & ALIGNMENT:

Course SLO	PSLOs (indicate #only)	ILOs (indicate #only)	Direct & Indirect Assessment Methods
CSLO#1	PSLO#2	ILO#7, ILO#2	
CSLO#2	PSLO#3	ILO#7	Embedded questions in tests and final exam, assignments, and student projects.
CSLO#3	PSLO#3	ILO#7, ILO#2	
CSLO#4	PSLO#2	ILO#7	

Course Student Learning Outcomes (CSLO):

Upon successful completion of this course, the student will:

CSLO#1: Identify, acquire, and prepare data from various file types (CSV, TSV, ASCII, basic SQL, AP, etc.) and organize the data into charts.

CSLO#2: Test a hypothesis or model against data using p-values, chi-square and z-tests, prior and posterior likelihood ratios, and correlations between variables.

CSLO#3: Perform Exploratory Data Analysis (EDA) to extract important variables, understand the relationship(s), or lack thereof, between variables, identify missing data, errors in data, confounders, and detect underlying structures of the data.

CSLO#4: Correctly code several foundational approaches in machine learning, including clustering, decision trees, and tree ensemble methods.

Program Student Learning Outcomes (PSLO):

PSLO#2: Computational Mathematics - Apply computational techniques and technology to support mathematical reasoning and problem solving.

PSLO#3: Statistics & Applied Mathematics - Apply mathematical concepts of probability, statistics, linear algebra and related applications.

Institutional Learning Outcomes (ILO):

ILO#2: Technology & Information Literacy

ILO#7: Scientific & Quantitative Reasoning

VII. CALCULATION OF FINAL GRADES:

- Assignments: **20%**
- Projects: **30%**
- Tests: **20%**
- Final Exam: **30%**

GRADING SCALE: (Should follow Department and/or College Template)

Grade	A	A-	B+	B	B-	C+	C	C-	D+	D	F
GPA	4.0	3.7	3.3	3.0	2.7	2.3	2.0	1.7	1.3	1.0	0.0
%	100-93	92.9-90	89.9-88	87.9-82	81.9-80	79.9-78	77.9-72	71.9-70	69.9-67	66.9-60.1	60 and under

Late Work Policy:

Submit assignments on time. Partial assignment submissions will result in a penalty. Students who miss the submission due dates will receive a "0" grade for the missed assignment.

VIII. SCHEDULE OF LEARNING TOPICS COVERED

Topics	Weeks
1. *Introduction:* Philosophy and good practices of data science: the flow chart of a data-driven project from idea to divulgation, the concepts of falsifiability, reproducibility, open science, the importance of version control. Lab: GitHub repositories, setting up your environment, Jupyter notebooks (on Colab free platform)	Week1
2. *Data manipulation:* Data types, missing data, censored data, organization of data in tables. Data hygiene. Lab: Acquiring and preparing data (CSV, TSV, downloadable ascii files, basic SQL, API) in Pandas: merging data from different files, reading data collections from CSV files into Data-Frames in Python, selecting columns, selecting rows, merging Data-Frames	Week2

3. *Inference and prediction:* Inference from plots: plotting histograms and scatter plots, data types incl ordinal, continuous, categorical data, visual inspection of correlation between variables Lab: read and clean data, CitiBikes, Pluto, Census	Weeks 3
4. *Hypothesis testing:* p-value, chi-square, z-test. Lab: basic statistics on Pluto, Census, Citibikes data, moment extraction, deviations from Gaussianity/Poissonity, histograms, proper binning, confounders	Weeks 4 & 5
5. *PDF/CDF:* Probability Density Function/ Cumulative Distribution Function data dredging, error analysis, testing models (KS, Anderson darling, KL divergence), goodness of fit. Lab: creating and testing simple distribution models in NumPy	Weeks 6 & 7
6. *Basic Bayesian concepts:* Bayes vs Frequentist statistics, Prior, Likelihood, Posterior	Week 8
7. *Databases:* SQL and manipulation of large data	Week 9
8. *Timeseries techniques:* smoothing, detrending, stationary, nonstationary, homeo- & Heteroscedastic noise	Week 10
9. *Multidimensional data:* Spatial + Temporal data (data)	Week 11
10. *Unstructured data:* Text processing and inference from text	Week 12
11. *Machine Learning methods:* Supervised and Unsupervised learning; Regression: ordinary least square (OLS) and weighted least square (WLS) methods; Clustering, Decision trees and Tree ensemble methods	Week 13
12. *Data Visualization:* Matplotlib	Week 14
Final exam week	Week 15

IX. UNIVERSITY ACADEMIC INTEGRITY STATEMENT:

Students are responsible for proper conduct and integrity in all of their scholastic work. They must follow a professor's instructions when completing tests, homework, and laboratory reports, and must ask for clarification if the instructions are not clear. In general, students should not give or receive aid when taking exams, or exceed the time limitations specified by the professor. In seeking the truth, in learning to think critically, and in preparing for a life of constructive service, honesty is imperative. Honesty in the classroom and in the preparation of papers is therefore expected of all students. Each student has the responsibility to submit work that is uniquely his or her own. All of this work must be done in accordance with established principles of academic integrity.

X. University ATTENDANCE POLICY:

Lincoln University uses the class method of teaching, which assumes that each student has something to contribute and something to gain by attending class. It further assumes that there is much more instruction absorbed in the classroom than can be tested on examinations. Therefore, students are expected to attend all regularly scheduled class meetings and should exhibit good faith in this regard.

More information concerning the Academic Integrity Statement and the Attendance Policy may be found in the current Academic Catalog found on the Registrar's webpage:

https://www.lincoln.edu/departments/registrar

XI. STUDENTS WITH DISABILITIES STATEMENT:

Lincoln University is committed to non-discrimination of students with disabilities and therefore ensures that they have equal access to higher education, programs, activities, and services in order to achieve full participation and integration into the University. In keeping with the philosophies of the mission and vision of the University, the Office of Student Support Services, through the Services for Students with Disabilities (SSD) Program, provides an array of support services and reasonable accommodations for students with special needs and/or disabilities as defined by Section 504 of the Rehabilitation Act of 1973 and the Americans with Disabilities Act of 1990. The Services for Students with Disabilities Program seeks to promote awareness and a campus environment in which accommodating students with special needs and/or disabilities is a natural extension of the University's goal.

Any student with a documented disability should contact the Office of Equity and Inclusion:

http://www.lincoln.edu/departments/office-institutional-equity/ada-accommodation/services-students-disabilites-ssd

XII. TITLE IX STATEMENT:

The following person has been designated to handle inquiries regarding the non-discrimination policies: Gerard Garlic, Title IX Coordinator/Director of the Health and Wellness Center, room 126B, Lincoln University, 1570 Baltimore Pike, Lincoln University, PA 19352 (p) 484-746-0000 or Office of Civil Rights, U.S. Department of Education, The Wanamaker Building, 100 Penn Square East, Suite 515, Philadelphia, PA 19107-3323, phone 215-656-8541, fax 215-656-8605, email: ocr.philadelphia@ed.gov

XIII. "NETIQUETTE" EXPECTATIONS:
1. Be professional and respectful
2. Take note of the instructor's introduction and address the instructor as such.
3. Have an opinion, ask questions, and participate in discussions.
4. Respectfully disagree
5. Follow standard rules for grammar and spelling (unless in an online chat) and be clear and concise
6. Avoid using all capital letters as this is interpreted as raising your voice

XIV. POLICY ON ELECTRONIC DEVICES IN CLASSROOM:

(Faculty must include a statement on their policy.)