First-year Common Core Course

# ENGR 1330 Computational Thinking with Data Science

**Important Guidelines:**

- If Texas Tech University campus operations are required to change because of health concerns related to the COVID-19 pandemic, it is possible that this course will move to a fully online delivery format. Should that be necessary, students will be advised of technical and/or equipment requirements, including remote proctoring software.
- Policy on absences resulting from illness: We anticipate that some students may have extended absences. To avoid students feeling compelled to attend in-person class periods when having symptoms or feeling unwell, a standard policy is provided that holds students harmless for illness-related absences (see **Section A** below).

## A. Illness-Based Absence Policy

If at any time during this semester you feel ill, in the interest of your own health and safety as well as the health and safety of your instructors and classmates, you are encouraged ***not*** to attend face-to-face class meetings or events. Please review the steps outlined below that you should follow to ensure your absence for illness will be excused. These steps also apply to not participating in synchronous online class meetings if you feel too ill to do so and missing specified assignment due dates in asynchronous online classes because of illness.

1. If you are ill and think the symptoms might be COVID-19-related:
    a. Call Student Health Services at 806.743.2848 or your health care provider. During after-hours and on weekends, contact TTU COVID-19 Helpline at (806) 743-2911.
    b. Self-report as soon as possible using the Dean of Students COVID-19 webpage (https://www.depts.ttu.edu/dos/COVID-19Absence.php). This website has specific directions about how to upload documentation from a medical provider and what will happen if your illness renders you unable to participate in classes for more than one week.
    c. If your illness is determined to be COVID-19-related, all remaining documentation and communication will be handled through the Office of the Dean of Students, including notification of your instructors of the time you may be absent from and may return to classes.
    d. If your illness is determined not to be COVID-19-related, please follow steps 2.a-d below.

2. If you are ill and can attribute your symptoms to something other than COVID-19:
    a. If your illness renders you unable to attend face-to-face classes, participate in synchronous online classes, or miss specified assignment due dates in asynchronous online classes, you are encouraged to contact either Student Health Services at 806.743.2848 or your health care provider. Note that Student Health Services and your own and other health care providers may arrange virtual visits.
    b. During the health provider visit, request a "return to school" note.

c. E-mail the instructor a picture of that note.
d. Return to class by the next class period after the date indicated on your note.

Following the steps outlined above helps to keep your instructors informed about your absences and ensures your absence or missing an assignment due date because of illness will be marked excused. You will still be responsible to complete within a week of returning to class any assignments, quizzes, or exams you miss because of illness.

## Course Instructor and Teaching Assistants:

Instructors : Long Nguyen
Email : long.nguyen@ttu.edu
Office hours : By appointment via email

Teaching assistant: Bipana Thapaliya
Lab sections : D58 and D61
Email : bipana.thapaliya@ttu.edu
Office hours : During tutoring hours

Teaching assistant : Shweta Dabetwar
Lab sections : D59 and D60
Email : shweta.dabetwar@ttu.edu
Office hours : During tutoring hours

## Catalog Course Description:

Introduces Python programming, its relevant modules and libraries, and computational thinking for solving problems in Data Science. Students will learn data science approaches to importing, manipulating, and analyzing data as well as modeling and visualizing real-world data sets in various science and engineering disciplines.

- 3 credit hours comprising of lectures and hands-on lab sessions.
- This course provides a hands-on learning of principles of programming and data science by introducing Python programming, its relevant modules and libraries, and computational thinking for solving problems in data science. Students will learn data science approaches to importing data, manipulating data, and analyzing it as well as modeling and visualizing real-world data sets in various science and engineering disciplines.

**Pre-requisite:**

- No technical/programming background is required.

**Textbook:**

Ani Adhikari and John DeNero, *Computational and Inferential Thinking, The Foundations of Data Science,* Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). **Link:** https://www.inferentialthinking.com/chapters/intro.

**Course Contents:**

- Computational thinking for problem-solving: Logical problem solving, decomposition, pattern recognition, abstraction, representation, algorithm design, and generalization.

- Python Programming: Variables, constants, data types, data structures, strings, math Operators, boolean operators, expressions, program constructs, functions, loop, I/O files, modules, and database.
- Data science fundamentals:
  - ✓ *Experimental setup*: Importing and formatting data sets, displaying data, data pre-processing.
  - ✓ *Introductory statistical analysis with Python:* Elementary statistics, randomness, sampling, probability distribution, confidence intervals, hypothesis testing, and A/B testing .
  - ✓ *Basic data analysis, visualization, and machine learning*: Data pre-processing, basic supervised/unsupervised learning, performance evaluation metrics.

## Learning Outcomes:

On completion of the course, students should –

- Be able to implement basic Python programs using computational thinking concepts.
- Know basic Python programming constructs and libraries relevant to data science.
- Be able to write Python scripts to perform fundamental data analytics and basic visualization.

### ABET Student Outcomes

- **Engineering:**
  - ✓ An ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics.
  - ✓ An ability to acquire and apply new knowledge as needed, using appropriate learning strategies.

- **Computer Science:**
  - ✓ Analyze a complex computing problem and to apply principles of computing and other relevant disciplines to identify solutions.
  - ✓ Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline.

## Resources/Tools

### Platforms for Python Programming

1. **Anaconda platform** (https://www.anaconda.com/)**:** Anaconda distribution is an open-source Data Science Distribution Development Platform. It includes Python 3 with over 1,500 data science packages making it easy to manage libraries and dependencies. Available in Linux, Windows, and Mac OS X.

**2. Jupyter** (https://jupyter.org/)**:** JupyterLab is a web-based interactive development environment for Jupyter notebooks, code, and data. JupyterLab is flexible: Configure and arrange the user interface to support a wide range of workflows in data science, scientific computing, and machine learning.

**Modules for Python Programming**

**3. Math module** (https://docs.python.org/3/library/math.html): Gives access to the mathematical functions defined by the C standard e.g. factorial, gcd, exponential, logarithm.

**4. Operator module** (https://docs.python.org/3/library/operator.html): Helps in exporting a set of efficient functions corresponding to the intrinsic operators of Python. For example, the operator add(x, y) is equivalent to the expression x + y.

**Python Modules for Data Science**

**5. Scipy module** (https://www.scipy.org/): A Python-based ecosystem of open-source software for mathematics, science, and engineering. Some of the core packages are:
• **Numpy:** Provides n-dimensional array package
• **Scipy:** Fundamental for scientific computing (e.g. linear algorithm, optimization)
• **Matplotlib**: Visualizations/2D plotting
• **IPython**: Enhanced interactive console
• **Pandas:** Data structures and data analysis

**6. Scikit-learn module** (https://scikit-learn.org/stable/): A library for machine learning in Python. It is a simple and efficient tool for predictive data analysis. It is built on NumPy, SciPy, and matplotlib modules.

**Course Schedule (Tentative – Lectures may be recorded):**

A few hours at the back end of the semester are left open to adjust for teaching speed and final project discussion.

| Week | Lecture |
|---|---|
| 1 | **Introduction to Computational Thinking with Data Science:**<br>- Computational thinking<br>- Programming principles<br>- Data science |
| | **Programming Principles:**<br>- Data types (int, float, string, bool)<br>- Variables, operators, expressions, basic I/O<br>- String operations |

| | |
|---|---|
| | **Programming Principles:**<br>- Data structures: Array, list, tuple, set, dictionary<br>- Conditional statements |
| 2 | **Programming Principles:**<br>- Loops |
| | **Programming Principles:**<br>- Functions<br>- Variable scope |
| | **Programming Principles:**<br>- Class and objects<br>- File handling |
| 3<br>Holiday on 09/07 | Quiz 1<br>Review – Python programming principles |
| | **Data Representation and Operations:**<br>Python library: NumPy<br>- Data representation: Arrays, vectors, matrices<br>- Data operations: Indexing, math functions |
| 4 | **Data Query and Manipulation:**<br>Python Library: Pandas<br>- Data frame: Create, index, read/write to file, summarize statistics, and fill and drop values |
| | **Data Display:**<br>Python Libraries: Matplotlib<br>- Data Display for line charts, bar charts, box plot, scatter plot, and histograms |
| | **Review – NumPy, Pandas, Matplotlib** |
| 5 | **Midterm-1** |
| | **Data Modeling: Statistical Approach:**<br>- Establishing causality<br>- Randomness: Iteration, simulation |
| | Randomness: Probabilities |
| 6 | Sampling and empirical distributions |
| | Hypothesis testing: General concept and examples of assessing models. |

| | |
|---|---|
| | Hypothesis testing: Comparing proportions, type1 & type2 errors, p-value. |
| 7 | Comparing two samples: A/B Testing |
| | Comparing two samples: A/B Testing<br>Quiz 2 |
| | Confidence intervals |
| 8 | Interpreting confidence intervals |
| | Center and spread |
| | Normal distribution |
| 9 | Sample means |
| | **Review – Statistical analysis** |
| | **Midterm-2** |
| 10 | **Data Modeling: Machine Learning Approach:**<br>Correlation; Issue final projects + presentation template |
| | Linear regression |
| | Least squares |
| 11 | Residuals |
| | Regression inference |
| | Quiz 3<br>Evaluation metrics: Accuracy, error |
| 12 | **Classification:**<br>- Supervised learning<br>- Nearest neighbor |
| | **Classification Evaluation and Making Decisions:**<br>- Confusion matrix, precision, recall, accuracy, F-score.<br>- Making decisions |
| | **Review – Machine learning** |
| 13 | **Midterm-3** |

**Course Assessment and Grading Criteria:**

There will be three midterm exams and one comprehensive final project for the course. In addition, lab participation, quizzes, and assignments will also be given credits that will contribute to the final grade. **If the assignments and the final project are submitted late, they will not receive any credit.** Students will be assessed based on the following criteria:

| Assessment methods | Weight (%) |
|---|---|
| Midterm-1 | 14 |
| Midterm-2 | 14 |
| Midterm-3 | 14 |
| Lab participation | 6 |
| Quizzes | 12 |
| Lab assignments | 25 |
| Final project | 15 |
| **Overall total** | **100** |

**At the end of the semester, the points will be tallied and converted to a percentage.** Based on the percentage obtained, the following scale will be used to assign grade:

| | **Guaranteed grade** |
|---|---|
| $\geq 90$ | A |
| $80 - 89$ | B |
| $70 - 79$ | C |
| $55 - 69$ | D |
| $< 55$ | F |

**Classroom Policy:**

The following activities are not allowed in the classroom: Texting or talking on the cellphone or other electronic devices, and reading a newspaper.

**ADA Statement:**

Any student who, because of a disability, may require special arrangements in order to meet the course requirements should contact the instructor as soon as possible to make necessary arrangements. Students must present appropriate verification from Student Disability Services during the instructor's office hours. Please note that instructors are not allowed to provide

classroom accommodation to a student until appropriate verification from Student Disability Services has been provided. For additional information, please contact Student Disability Services office in 335 West Hall or call 806.742.2405.

**Academic Integrity Statement:**

Academic integrity is taking responsibility for one's own class and/or course work, being individually accountable, and demonstrating intellectual honesty and ethical behavior. Academic integrity is a personal choice to abide by the standards of intellectual honesty and responsibility. Because education is a shared effort to achieve learning through the exchange of ideas, students, faculty, and staff have the collective responsibility to build mutual trust and respect. Ethical behavior and independent thought are essential for the highest level of academic achievement, which then must be measured. Academic achievement includes scholarship, teaching, and learning, all of which are shared endeavors. Grades are a device used to quantify the successful accumulation of knowledge through learning. Adhering to the standards of academic integrity ensures grades are earned honestly. Academic integrity is the foundation upon which students, faculty, and staff build their educational and professional careers. [Texas Tech University ("University") Quality Enhancement Plan, Academic Integrity Task Force, 2010].

**Religious Holy Day Statement:**

"Religious holy day" means a holy day observed by a religion whose places of worship are exempt from property taxation under Texas Tax Code §11.20. A student who intends to observe a religious holy day should make that intention known to the instructor prior to the absence. A student who is absent from classes for the observance of a religious holy day shall be allowed to take an examination or complete an assignment scheduled for that day within a reasonable time after the absence. A student who is excused may not be penalized for the absence; however, the instructor may respond appropriately if the student fails to complete the assignment satisfactorily.

**Ethical Conduct Policy:**

Cheating is prohibited, and the representation of the work of another person as your own will be grounds for receiving a failing grade in the course.