DTSC 3010 – Introduction to Data Science (Fall 2025)

University of North Texas – College of Information, Department of Information Science
Course Number: DTSC 3010
Course Title: Introduction to Data Science
Credit Hours: 3.0
Semester: Fall 2025

Course Information

Class Meetings: Wednesdays 2:00 PM – 5:00 PM (August 20 – December 12, 2025)

Location: UNT at Frisco (Hall Park) – Room FRLD 458, 12995 Preston Rd, Frisco, TX 75033

Instructor: Jiyuan Li (Principal Lecturer)

Contact Email: jiyuanli@my.unt.edu

Office Hours: By appointment (via Zoom or in-person at UNT Frisco) – please email to schedule.

Prerequisites: None

Course communication will be through UNT Canvas and official UNT email. Students should check Canvas regularly for announcements, course materials, and updates.

Course Description

This course introduces students to the concepts, principles, tools, and professional practices of data science. Students will explore different types of data and learn how data can be acquired, stored, organized, analyzed, and visualized to support decision-making and the creation of data-driven products. Through hands-on assignments and a term project, students will tackle real-world data challenges, learning to use data to answer questions and make informed decisions. Key topics include the data science process and lifecycle, basic programming for data analysis, data visualization, statistical inference, and introductory machine learning techniques. The course will also survey applications of data science in areas such as natural language processing, databases, financial modeling, social network analysis, and more. Ethical issues in the data science process (privacy, bias, etc.) are discussed to foster responsible data practices. *No prerequisites.*

Course Objectives

Upon successful completion of this course, students will be able to:

- **Understand the data science lifecycle:** Identify the stages of data collection, processing, analysis, inference, and product delivery, and explain how they interrelate in typical data science projects.
- **Manage and integrate data from multiple sources:** Acquire data in various formats (structured and unstructured) and use tools to store, clean, and organize data for analysis.
- **Use computational tools for data analysis:** Develop proficiency in Python programming for data science, including use of Jupyter/Google Colab notebooks to manipulate data tables, compute summary statistics, and create data visualizations.
- **Apply statistical and machine learning techniques:** Perform basic exploratory data analysis, apply inferential statistics (e.g. hypothesis tests, confidence intervals) to draw conclusions from data, and implement simple predictive models (such as regression and classification) to make predictions.
- **Think critically about data and results:** Interpret analytical results correctly; assess the validity and limitations of conclusions (e.g. understanding correlation vs. causation, statistical significance vs. practical importance); and consider alternative explanations.
- **Address ethical and societal considerations:** Recognize ethical issues in data science (privacy, fairness, transparency) and incorporate responsible practices when handling and analyzing data.
- **Communicate findings effectively:** Document data analysis workflows and report results in a clear, concise manner suitable for both technical and non-technical audiences, including using appropriate visualizations and presenting project outcomes.

Required Textbook and Materials

- **Textbook**: Computational and Inferential Thinking: The Foundations of Data Science (2nd Edition) by Ani Adhikari, John DeNero, and David Wagner. This textbook is available for free online in an interactive format. Weekly readings will be assigned from this book (see course schedule below).
    - Citation: Adhikari, A., DeNero, J., & Wagner, D. (2021). Computational and Inferential Thinking: The Foundations of Data Science (2nd ed.). (Creative Commons BY-NC-ND licensed online textbook)
- **Computing**: We will use Python as the primary programming language. All in-class coding will be done using Google Colab (a cloud-based Python environment), which requires a Google account. Students should bring a laptop to each class for hands-on exercises. (If you need to borrow a device, contact the instructor in advance.) No prior programming experience is required.
- **Canvas**: Course materials, announcements, assignments, and grades will be managed through UNT Canvas. It is the student's responsibility to ensure access to Canvas and to check it regularly for updates.

Course Requirements and Grading

Students are expected to attend the weekly class sessions (3 hours each, roughly 50% lecture/discussion and 50% hands-on lab activities), complete the required readings before each class, and actively participate in exercises. Outside of class, students should budget time to complete readings and assignments (approximately 6–9 hours per week of independent study is recommended).

Grading Breakdown: The final course grade is composed of the following components:

- **Assignments (6 total):** 60% (6 assignments at 10% each)
- **Mid-Term Exam:** 10%
- **Final Exam:** 10%
- **Final Project:** 20%

Final letter grades will follow the standard UNT scale: **A** = 90–100%, **B** = 80–89%, **C** = 70–79%, **D** = 60–69%, **F** = below 60%. (Grades may be rounded to the nearest whole percent; no additional curve or extra credit is planned beyond what may be offered to the entire class.)

Exams

There will be two exams in this course: one mid-term and one final. Both will be administered via UNT Canvas using the quiz tool.

Mid-Term Exam (10%): Scheduled for Week 9 (Oct 13-19). This exam will consist of signle-choice, multiple-choice, and true/false questions covering material from Weeks 1–8. The mid-term will be taken on Canvas during our class meeting. Please ensure you have reviewed all relevant topics and have a stable internet connection for the quiz.

Final Exam (10%): Scheduled for Week 16 (Dec 1-7). The final exam will be delivered as a Canvas quiz with single-choice, multiple-choice, and true/false questions. It is cumulative, covering material from the entire course. The exam will take place during the scheduled class period on Dec 3.

Make-up exams will only be offered in exceptional circumstances (e.g. serious illness or emergency) and require advance notification and appropriate documentation. Both exams are open-notes/open-book within Canvas but must be completed individually without collaboration.

Assignments

There will be **6 assignments** and **1 final project** in this course, as described below. All assignments are to be completed **individually** (no collaboration unless explicitly allowed). Submit all assignment deliverables via Canvas by the specified deadlines.

Assignment 1 (10%): Data Science Essentials and Python Basics. This introductory assignment ensures you have set up the programming environment and understand basic Python and Jupyter/Colab operations. You will perform simple calculations, create basic visualizations, and answer questions about a small dataset to demonstrate understanding of foundational concepts (e.g. types of variables, cause vs. correlation). **Due:** Tuesday, Sep 2, 2025, by 11:59 PM.

Assignment 2 (10%): Data Manipulation with Tables. In this assignment, you will work with tabular data. Tasks include reading in a dataset (CSV file), using Python table libraries to filter and sort data, computing summary statistics, and interpreting results. You will practice transforming datasets and answering questions by writing code (using operations like selection, filtering, grouping, etc.). **Due:** Tuesday, Sep 16, 2025, by 11:59 PM.

Assignment 3 (10%): Data Visualization and Exploratory Analysis. This assignment focuses on exploring data through visualization. Using a provided dataset, you will create plots such as histograms, bar charts, or scatter plots to reveal patterns or relationships. You will be asked to describe insights from the visuals and to experiment with different types of chart or categories. The goal is to demonstrate best practices in data visualization and interpretation of descriptive statistics. **Due:** Tuesday, Sep 30, 2025, by 11:59 PM.

Assignment 4 (10%): Probability and Simulation. This assignment introduces randomness and simulation in data science. You will write code to simulate random processes (for example, drawing samples, shuffling data, or running Monte Carlo trials) to solve probability problems. You will compare simulation results to theoretical expectations and gain intuition about concepts like randomness, distributions, and the Central Limit Theorem. **Due:** Tuesday, Oct 14, 2025, by 11:59 PM.

Assignment 5 (10%): Hypothesis Testing (A/B Testing). Conduct an analysis to test a hypothesis using real or synthetic data. Formulate null and alternative hypotheses for a comparison between two groups, then use a statistical test (e.g. a permutation test) to determine if an observed difference is statistically significant or could be due to chance. Compute p-values and/or confidence intervals, interpret the results, and discuss any assumptions or potential biases. **Due:** Tue, Oct 28, 2025, by 11:59 PM.

Assignment 6 (10%): Regression Analysis. Perform a simple linear regression analysis on a dataset (for example, predicting housing prices). Calculate the correlation between variables, fit a regression line to the data, interpret the slope and intercept of the line, and use the model to make predictions. **Due**: Tue, Nov 11, 2025, by 11:59 PM.

Final Project (20%) – Comprehensive Data Science Project: The final project synthesizes all course learning outcomes in an applied context. You will select a dataset or problem domain of interest (subject to instructor approval) and carry out an end-to-end data science project. This includes formulating a clear question or hypothesis, obtaining or collecting relevant data, performing data cleaning/preprocessing, conducting exploratory analysis and visualizations, and applying at least one inferential or predictive method from the course (e.g. a hypothesis test, regression, or classification). You will interpret the results to draw conclusions and discuss

insights. The deliverables are a written project report (including narrative, code, figures, and analysis) and a short in-class presentation (10–15 minutes) to share your findings. The project will be evaluated on technical correctness, depth of analysis, creativity, and clarity of communication. Due: Wednesday, Nov 19, 2025 – written report and code submitted on Canvas by class time; in-class presentations will occur on Nov 19.

Assignment Submission: All assignments must be submitted via Canvas by the stated due date and time (usually 11:59 PM on the due date). Follow the submission instructions for each assignment (e.g. upload a ".ipynb" notebook or PDF report as specified). Ensure that your code is well-documented and that your answers/explanations are clearly provided.

Late Work Policy: Students are expected to submit assignments and projects on time. The due dates are specified in each assignment. If an extenuating circumstance such as a medically diagnosed illness or family emergency arises, which prevents you from submitting your assignments, you should contact the instructor and the TA as soon as possible before the due date. Late work without the permission of the instructor will not be accepted. A student who is having trouble with assignments is strongly encouraged to contact the instructor and the TA as early as possible for personal advising.

Course Schedule

The schedule below outlines the topics, assignment due dates, and readings for each week of the semester. This schedule is **tentative** and subject to adjustment; any changes will be announced on Canvas. All readings refer to chapters in the course textbook *Computational and Inferential Thinking (2nd Edition)* by Adhikari et al. Students are expected to complete the specified reading **before** each class meeting to be prepared for the lecture and activities.

Week	Date	Lecture	Assignment	Reading
1	Aug 18-24	*Course Introduction.* What is Data Science? Overview of the data science process; introduction to Python & Google Colab; basic Python syntax (data types, variables, simple calculations).		**Chapter 1:** What is Data Science?
2	Aug 25-31	*Data Generation and Causality.* Cause vs. effect, observational studies vs. experiments (John Snow cholera case); **Data Ethics & Privacy** introduction; Hands-on Python review (expressions, using notebooks).	**Assignment 1**– Due Tue, Sep 2 @ 11:59 PM (Python basics & DS essentials)	**Chapter 2:** Causality and Experiments
3	Sep 1-7	*Data Tables and Data Wrangling.* Tables in Python; importing datasets (CSV files); selecting columns, filtering rows, sorting; hands-on work with data tables (summary statistics).		**Chapter 3 & Chapter 4:** Introduction Python programming I

Week	Date	Lecture	Assignment	Reading
4	Sep 8-14	*Groups and Aggregation.* Grouping data and computing summary tables; aggregation (group by, summarize); joining tables; basics of data cleaning (handling missing values); hands-on grouping and pivoting exercises.	**Assignment 2**– Due Tue, Sep 16 @ 11:59 PM (Data table manipulations)	**Chapter 5 & Chapter 6:** Introduction Python programming II
5	Sep 15-21	*Data Visualization.* Principles of graphical data representation; basic chart types (bar charts, line plots, histograms); interpreting distributions (center, spread, shape); hands-on building and customizing plots.		**Chapter 7:** Introduction Python programming III
6	Sep 22-28	*Functions for Data Analysis.* Writing reusable code (defining functions, lambda); applying functions to data tables; transformation pipelines (combining operations); hands-on feature engineering with functions.	**Assignment 3**– Due Tue, Sep 30 @ 11:59 PM (Data visualization & exploration)	**Chapter 8:** Introduction Python programming IV
7	Sep 29-Oct 5	*Randomness and Simulation.* Introduction to probability concepts (random trials, distributions, law of large numbers); using code to simulate random processes (coin flips, etc.); hands-on Monte Carlo simulations to estimate probabilities.		**Chapter 9:** Randomness
8	Oct 6-12	*Statistical Inference – Hypothesis Testing.* Hypothesis testing framework (null vs. alternative hypotheses, test statistics, p-values, significance levels); A/B testing example; hands-on permutation test for practice.	**Assignment 4**– Due Tue, Oct 14 @ 11:59 PM (Probability & simulation)	**Chapter 11:** Testing Hypotheses; **Chapter 12:** Comparing Two Samples
9	Oct 13-19	*Estimation and Confidence Intervals.* Estimation vs. hypothesis testing; building confidence intervals using the bootstrap method; interpreting confidence levels (e.g. 95% CI); hands-on bootstrap resampling demonstration.		**Chapter 13:** Estimation
10	Oct 20-26	*Correlation and Regression.* Examining relationships between variables; scatter plots and the correlation coefficient; fitting a simple linear regression (least-squares line); interpreting slope,	**Assignment 5**– Due Tue, Oct 28 @ 11:59 PM (Hypothesis testing)	**Chapter 15:** Prediction

Week	Date	Lecture	Assignment	Reading
		intercept, and residuals; hands-on linear regression exercise.		
11	Oct 27-Nov 2	*Regression Evaluation.* Using regression for prediction; goodness-of-fit metrics (R^2, RMSE); prediction vs. causation; introduction to multiple regression concepts; caution about overfitting; hands-on evaluating model predictions and residuals.		**Chapter 15:** Prediction **Chapter 16***:* Inference for Regression *(optional)*
12	Nov 3-9	*Classification and k-NN.* Introduction to classification problems and basic machine learning; k-Nearest Neighbors classification method; training vs. testing data; evaluating classification accuracy; hands-on building a simple k-NN classifier.	**Assignment 6**– Due Tue, Nov 11 @ 11:59 PM (Regression analysis)	**Chapter 17:** Classification
13	Nov 10-16	*Case Studies in Data Science & Ethics.* Special topics and real-world applications (e.g. basics of NLP, recommender systems, social network analysis); revisiting ethical considerations (bias in algorithms, privacy issues); hands-on case study analysis; **Final Project Q&A** (last-minute project questions and consultation).		*(No new textbook reading – focus on project work)*
14	Nov 17-23	**Final Project Presentations & Course Wrap-Up.** In-class presentations of final projects (each team presents problems, dataset, methods, findings); celebrate learning outcomes; course wrap-up and discussion of next steps in data science.		
15	Nov 24-30	**Thanksgiving Break – No Class**		
16	Dec 1-7	*Final Project Presentations (continued, if needed); Course Conclusion.* Conclude any remaining project presentations. No new content – course concluded.	**Final Project due** – Written report & code due; in-class project presentations.	

Course Policies

Attendance and Participation

Attendance is expected at all class meetings. This course involves in-class discussion and hands-on work, so your presence and participation are critical. If you must miss a class due to illness or emergency, notify the instructor as soon as possible (preferably **before** the class). You are responsible for catching up on any material covered in your absence. More than two unexcused absences may result in a lower participation grade. Consistent tardiness or leaving early without a valid reason will also count against attendance. Participation means actively engaging during class – asking questions, contributing to discussions, and putting effort into lab exercises. Simply attending without engagement may only earn partial participation credit. Students are expected to contribute to a positive, collegial classroom environment.

Late Submission Policy

All students are expected to submit assignments and project deliverables on time. Deadlines are set to keep you on track and allow timely feedback. As noted above, late work will incur a penalty of **–10 points per day** (each 24-hour period past the deadline). An assignment submitted at 12:05 AM (just after the deadline) is considered one day late. If an extenuating circumstance (such as a serious illness or family emergency) prevents timely submission, you **must** contact the instructor **before** the due date to request an extension and provide appropriate documentation if requested. Extensions are not guaranteed and are granted at the instructor's discretion. Work submitted late without prior approval will be subject to the standard penalty or may not be accepted for credit if excessively late. Plan ahead and start assignments early to avoid last-minute problems.

Incomplete

An Incomplete Grade ("I") is a non-punitive grade given only during the last one-fourth of a term/semester and only if a student (1) is passing the course and (2) has a justifiable and documented reason, beyond the control of the student (such as serious illness or military service), for not completing the work on schedule. The student must arrange with the instructor to finish the course later by completing specific requirements. Please refer to http://essc.unt.edu/registrar/academic-record-incomplete.htmlfor more information. You are responsible for contacting the instructor to request an incomplete and discuss requirements for completing the course. If you do not remove the incomplete within one calendar year, you will receive a grade of F.

Withdrawal

Withdrawal: See UNT Graduate Catalog for policies and UNT semester schedule for deadlines. A grade of withdraw (W) or withdraw-failing (WF) will be given depending on your participation and grades to date. If you simply disappear and do not file a formal UNT withdrawal form, you may receive a grade of F. UNT has a new policy for withdrawing from a course. Please see details at: https://registrar.unt.edu/registration/dropping-class

Academic Integrity

The UNT Students Standards of Academic Integrity (2009) are available at the Provost office website: http://vpaa.unt.edu/academic-integrity.htm. The Student Standards of Academic Integrity UNT Policy Manual (http://policy.unt.edu/sites/default/files/untpolicy/pdf/7-Student_Affairs-Academic_Integrity.pdf) defines six categories of academic dishonesty: cheating, plagiarism, forgery, fabrication, facilitating academic dishonesty, and sabotage. The category **plagiarism** defined as follows: "Use of another's thoughts or words without proper attribution in any academic exercise, regardless of the student's intent, including but not limited to:

1) the knowing or negligent use by paraphrasing or direct quotation of the published or unpublished work of another person without full and clear acknowledgement or citation.
2) the knowing or negligent unacknowledged use of materials prepared by another person or by an agency engaged in selling term papers or other academic materials." (http://policy.unt.edu/sites/default/files/untpolicy/pdf/7-Student_AffairsAcademic_Integrity.pdf)

Enrollment in any INFO course is considered implicit acceptance of all DIS and UNT student policies. It is the responsibility of the student to understand and adhere to these policies. DIS has zero tolerance for academic dishonesty. DIS instructors may choose to submit any student work to Turnitin for verification of originality. Penalties for plagiarism in this course follow the UNT guidelines and for repeated offenses, **Student will receive a failing grade for the course. The instructor will direct the student to schedule an in-person conference with the instructor to discuss the suspected misconduct. The instructor will submit to UNT, per the policy, a report of the violation.**

- **Add/Drop Policy**
  - Please refer to the UNT Registrar's Office website regarding the Add/Drop Policy.

- **Code of Conduct**
  - Please refer to the UNT Dean of Students Office website regarding the Student Code of Conduct Policy.

Americans with Disabilities Act Compliance Statement

The Department of Library and Information Sciences, University of North Texas is committed to full academic access for all qualified students, including those with disabilities. In keeping with this commitment and to facilitate equality of educational access, faculty members in the Department will make reasonable accommodations for qualified students with a disability, such as

appropriate adjustments to the classroom environment and the teaching, testing, or learning methodologies when doing so does not fundamentally alter the course.

If you have a disability, it is your responsibility to obtain verifying information from the Office of Disability Accommodation (ODA) and to inform me of your need for accommodation. Requests for accommodation must be given to me no later than the first week of classes for students registered with the ODA as of the beginning of the current semester. If you register with the ODA after the first week of classes, your accommodation requests will be considered after this deadline.

Grades assigned before accommodation is provided will not be changed. Information about how to obtain academic accommodation can be found in UNT Policy 18.1.14, at www.unt.edu/oda, and by visiting the ODA in Room 321 of the University Union. You also may call the ODA at 940.565.4323.

University Policies and Resources

Add/Drop Policy: It is the student's responsibility to manage their enrollment in the course. If you decide to drop or withdraw, be mindful of the official add/drop deadlines published by the Registrar. Refer to the UNT Registrar's resources for policies and dates regarding schedule changes and withdrawals. If you stop attending class but do not officially drop, you may receive an **F** in the course. Always complete the formal withdrawal process if you intend to drop, to avoid academic and financial penalties.

Communication and Email Etiquette: When emailing the instructor, please use your UNT email account and include "DTSC 3010" in the subject line. Be professional and concise in your communications. Allow up to 24 hours for a response on weekdays (longer on weekends). Before emailing, check Canvas and this syllabus to see if your question has already been answered. All communication should be courteous and use appropriate language.

Academic Resources: Take advantage of UNT's academic support services. The UNT Learning Center offers tutoring, study skills workshops, and other assistance. For writing help, you may consult the UNT Writing Lab. If you encounter technical issues with Canvas or other IT resources, contact the UNT IT Help Desk for support.

Emergency Procedures: In the event of an emergency or severe weather that affects class (e.g., a campus closure), monitor official UNT announcements. We will also communicate via Canvas announcements about any changes to class meetings or assignment deadlines in such situations. Safety procedures for on-campus classes (fire evacuation routes, severe weather shelter areas) will be reviewed in class. Please familiarize yourself with the nearest exits and safety protocols for our classroom location.

By remaining enrolled in this course, you agree to abide by the policies and standards set forth in this syllabus and in official UNT policy documents. The instructor reserves the right to make modifications to this syllabus (e.g. adjusting the schedule or assignments) as needed to better achieve the learning objectives or to accommodate unforeseen circumstances. Students will be notified in class and via Canvas of any such changes.