



University of North Texas  
College of Information  
Data Science  
Stats for Data Science  
Spring 2025

---

## Course Information

In this course, we will explore the role of statistics in data science. We will not only introduce a broad range of statistical terms but also apply them to typical data science projects. Statistics involves the collection, organization, analysis, interpretation, and presentation of data. Through hands-on tools, we will practice these skills and expand our data science toolkit.

### Instructor

**Name:** Dr. Yang Zhang

**Office:** Discovery Park E296J

**University of North Texas**

**Email:** [Yang.zhang@unt.edu](mailto:Yang.zhang@unt.edu)

### TA Contact

**Prabu, Alwin** [alwinprabu@my.unt.edu](mailto:alwinprabu@my.unt.edu)

## Course Description

This course introduces students to both theories and applications of statistical methods. Students will learn the core concepts of statistical learning for data analysis while working hands-on with real data using Python and interactive notebooks. The topics taken from a data science lifecycle include data ingestion and manipulation, basic descriptive statistics, standard parametric statistical models, linear regression, classification, and tree-based models.

## Course Objectives

Upon completion of this course, students should be able to:

- Talk comfortably about data science and model-building
- Understand the core concepts of statistical learning related to regression, classification, and tree-based models
- Evaluate various data models, their use, their strengths, and their weaknesses
- Design solutions and use software to build, train, and test models
- Present data analysis results

## Course Prerequisites

This class is an entry-level class. There is no formal course prerequisite for this class.

## Class Location & Time:

This class is scheduled for **Tu 6:30PM - 9:20PM at NTDP D215**. Supporting material will be offered in Canvas at <https://unt.instructure.com>

## Office Hours

Students are welcome to make an appointment with the instructor at any time to discuss course-related questions by available means of communication (Mail, Zoom, etc.).

## Teaching Philosophy

Data Science, while inherently fascinating, demands a significant investment of time and experience to achieve true mastery. This course endeavors to maintain an approachable atmosphere, where concepts and techniques are laid out through practical, step-by-step instructions. While the final project invites your creativity and independence, for most of the class, it is advisable to follow along closely. The intent here is not to turn you into an expert overnight but rather to immerse you in the foundational ideas and methodologies, sparking curiosity rather than immediate proficiency.

### Textbook

An Introduction to Statistical Learning 2<sup>nd</sup> Edition by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani  
The book is available for download at <https://statlearning.com/>

### Software/Hardware Requirements

We will use the following applications:

- Python 3+
- Jupyter Notebook
- Anaconda

### Course Modules

This course consists of **16 Weeks**. This is a hands-on course, and we will be working on most of each assignment in class. There are assignments, Quizzes, in-class activities, Group Projects and one exam.

We will be using Anaconda and the Python programming language for analysis and statistical methods related to data science. This is an entry-level course, so no prior experience is required. Additionally, please note that this course provides an overview of statistical learning and introduces many topics. Mastery is not expected; the goal is for you to enjoy learning something exciting.

### Tentative Schedule (Subject to change):

- Week 1: Introduction
- Week 2: Understanding the Problem and Getting the Data
- Week 4: Data Prep and Exploratory Data Analysis
- Week 5: Probability
- Week 6: Distributions
- Week 7: Feature Engineering
- Week 8: Feature Selection
- Week 9: Hypothesis Testing, Test of Means, and ANOVA
- Week 10: Linear Regression
- Week 11: Classification
- Week 12: Trees and Random Forests
- Week 13: Unsupervised Learning
- Week 14: AB Testing
- Week 16: Final Presentations

Turn in your assignments by submitting them to the submission tools in Canvas by the date specified in the **UNT canvas portal**. If an emergency arises which prevents you from submitting your assignments, you should contact the instructor as soon as possible before the due date.

## Projects

The projects consist of real-world data science problems and include Exploratory Data Analysis, Feature Selection, Model Comparison, Cross-Validation, Grid Search, and Visualizations spread out over the project. You will share your projects with your peers.

## Grading

Grades are determined by a simple points system, with a total of *at least* 100 pts given though more than 100 points are likely. The expected distribution of points is given below, with the exact scale determined by point values given for each assignment, project, or exam - this is subject to minor modification based on actual points given. Note, due to the nature of the course, exams are a significant means of establishing your final grade, so please complete the assignments in a timely way and study appropriately prior to each exam.

<b>Activities</b>	<b>Points Possible</b>	<b>Percentage of Final Grade</b>
<b>Assignments</b>	50 points	50%
<b>Group Project</b> <ul style="list-style-type: none"> <li>• <b>Proposal</b></li> <li>• <b>Presentation &amp; Report</b></li> <li>• <b>Attendance</b></li> </ul>	15 points 20 points 5 points	35% 5%
<b>Exams</b> <ul style="list-style-type: none"> <li>• <b>1 exams @ 10 points</b></li> </ul>	10 points	10%
<b>Total Points</b>	100 points	100%

**Grading Scale:** A>=90, B=80-89.9, C=70-79.9, D=60-69.9, F=0-59.9 pts. No exceptions. If class grades are low (e.g. I expect the vast majority of students will end with A's and B's), extra points will be added to all the students in the class (we do not assign extra work or extra points for individuals).

**Note:** All questions about grades must be brought up within a week after the grades are released. No grade change can be made afterward.

## Policy About Incomplete Grades:

Per UNT policy, a grade of Incomplete can only be awarded to a student who is 1) passing the course and 2) has a justifiable and documented reason, beyond the control of the student, for not completing the course work on schedule. Notification and submission of documentation must be provided to the instructor at the time of the emergency. Please see <http://essc.unt.edu/registrar/academic-record-incomplete.html> (Links to an external site.) for information.

## Course Evaluation

UNT will make available evaluations near the end of the semester.

## Course Policies Assignment Policy

Please become familiar with the due dates. Sometimes technology doesn't cooperate. Communication is key to a successful online experience. Please be patient if something unavoidable happens and we will work something out. The University is committed to providing a reliable online course system to all users. However, in the event of any unexpected server outage or any unusual technical difficulty that prevents students from completing a time-sensitive assessment activity, the instructor will extend the time windows and provide an appropriate accommodation based on the situation. Students should immediately report any problems to the instructor and contact the UNT Student Help Desk: [helpdesk@unt.edu](mailto:helpdesk@unt.edu) or 940.565.2324. The instructor and the UNT Student Help Desk will work with the student to resolve any issues at the earliest possible time.

### Examination Policy

All quizzes are closed-book exams. Web searching and the use of AI are strictly prohibited. Students are expected to complete the exam individually, with no discussions allowed during the exam. Failure to follow these rules will result in an 'F' grade for the course.

### Instructor Responsibilities and Feedback

I will do everything I can to help you succeed in this class. Communication is important. Contact me if you have any concerns.

### Late Work

Late work is not accepted

### Attendance Policy

Students are encouraged to attend every class. Attendance is mandatory. You can miss 3 out of 16 classes.

The University of North Texas' Attendance Policy may be found at: <http://policy.unt.edu/policy/15-2-5>.

### Class Participation

In-class activities are essential for succeeding in this class, as they provide practical exposure to the concepts.

### Syllabus Change Policy

There may be occasions when adjustments to assignments, grading criteria, or due dates listed in this syllabus are necessary. If any changes occur, you will be notified immediately.

### Academic Dishonesty – Plagiarism (taken from the UNT Student Code of Conduct)

The term "plagiarism" includes, but is not limited to (a) the knowing or negligent use by paraphrase or direct quotation of the published or unpublished work of another person without full and clear acknowledgment and (b) the knowing or negligent unacknowledged use of materials prepared by another person or by an agency engaged in the selling of term papers or other academic materials.

Plagiarism is copying by retyping, cutting and pasting, or paraphrasing. In this course, beware of the following:

- Do not quote or paraphrase published sources without explicit reference to the original work. Information used or quoted from other sources must contain a citation, whether the source is a print or electronic source.
- APA Style: <http://www.apastyle.org/learn/tutorials/basics--tutorial.aspx>
- Citation Machine: <http://citationmachine.net/index2.php?reqstyleid=1>
- EasyBib: <http://www.easybib.com/>
- Owl Purdue: <http://owl.english.purdue.edu/owl/resource/560/01/>
- Do not insert parts of another student's work into your work. That student trusts you to respect his/her intellectual product.
- Do not copy and paste parts of the course material into your work.

### Penalties for Plagiarism

Plagiarism is illegal, unethical, and unacceptable. Any instances of plagiarism in student work will result in the following penalties: First offense: Grade of zero for the assignment. Second offense: Final course grade reduced by one complete grade. Third offense: Assignment of F (Fail) for final course grade.

**Important Notice for F-1 Students taking Distance Education Courses Federal Regulation**

To read detailed Immigration and Customs Enforcement regulations for F-1 students taking online courses, please go to the Electronic Code of Federal Regulations website: <http://ecfr.gpoaccess.gov>. The specific portion concerning distance education courses is located at "Title 8 CFR 214.2 Paragraph (f)(6)(i)(G)" and can be found buried within this document:

<http://www.gpo.gov/fdsys/pkg/CFR---2013---title8---vol1/xml/CFR---2013--title8---vol1---sec214---2.xml>

The paragraph reads:

(G) For F-1 students enrolled in classes for credit or classroom hours, no more than the equivalent of one class or three credits per session, term, semester, trimester, or quarter may be counted toward the full course of study requirement if the class is taken online or through distance education and does not require the student's physical attendance for classes, examination or other purposes integral to completion of the class. An online or distance education course is a course that is offered principally through the use of television, audio, or computer transmission including open broadcast, closed circuit, cable, microwave, or satellite, audio conferencing, or computer conferencing. If the F-1 student's course of study is in a language study program, no online or distance education classes may be considered to count toward a student's full course of study requirement.

**University of North Texas Compliance**

To comply with immigration regulations, an F-1 visa holder within the United States may need to engage in an on-campus experiential component for this course. This component (which must be approved in advance by the instructor) can include activities such as taking an on-campus exam, participating in an on-campus lecture or lab activity, or other on-campus experience integral to the completion of this course.

If such an on-campus activity is required, it is the student's responsibility to do the following:

1. Submit a written request to the instructor for an on-campus experiential component within one week of the start of the course.
2. Ensure that the activity on campus takes place and the instructor documents it in writing with a notice sent to the International Student and Scholar Services Office. ISSS has a form available that you may use for this purpose.

Because the decision may have serious immigration consequences, if an F-1 student is unsure about his or her need to participate in an on-campus experiential component for this course, s/he should contact the UNT International Student and Scholar Services Office (phone 940.565.2195 or email [internationaladvising@unt.edu](mailto:internationaladvising@unt.edu)) to get clarification before the one-week deadline.

**UNT Policies**

**Academic Integrity Standards and Consequences.** According to UNT Policy 06.003, Student Academic Integrity, academic dishonesty occurs when students engage in behaviors including, but not limited to cheating, fabrication, facilitating academic dishonesty, forgery, plagiarism, and sabotage. A finding of academic dishonesty may result in a range of academic penalties or sanctions ranging from admonition to expulsion from the University. [Insert specific sanction or academic penalty for specific academic integrity violation.]

**ADA Policy**

UNT makes reasonable academic accommodation for students with disabilities. Students seeking accommodation must first register with the Office of Disability Accommodation (ODA) to verify their eligibility. If a disability is verified, the ODA will provide a student with an accommodation letter to be delivered to faculty to begin a private discussion regarding one's specific course needs. Students may request accommodations at any time, however, ODA notices of accommodation should be provided as early as possible in the semester to avoid any delay in implementation. Note that students must obtain a new letter of accommodation for every semester and must meet with each faculty member prior to implementation in each class. For additional information see the ODA website at [disability.unt.edu](http://disability.unt.edu).

**Emergency Notification & Procedures**

UNT uses a system called Eagle Alert to quickly notify students with critical information in the event of an emergency (i.e., severe weather, campus closing, and health and public safety emergencies like chemical spills, fires, or violence). In the event of a \_\_\_\_\_

university closure, please refer to Blackboard for contingency plans for covering course materials.

### Acceptable Student Behavior

Student behavior that interferes with an instructor's ability to conduct a class or other students' opportunity to learn is unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Dean of Students to consider whether the student's conduct violated the Code of Student Conduct. The University's expectations for student conduct apply to all instructional forums, including University and electronic classroom, labs, discussion groups, field trips, etc. The Code of Student Conduct can be found at [deanofstudents.unt.edu/conduct](http://deanofstudents.unt.edu/conduct).

### Student Evaluation Administration Dates

Student feedback is important and an essential part of participation in this course. The student evaluation of instruction is a requirement for all organized classes at UNT. The survey will be made available during weeks 13, 14 and 15 [insert administration dates] of the long semesters to provide students with an opportunity to evaluate how this course is taught. Students will receive an email from "UNT SPOT Course Evaluations via IASystem Notification" ([no-reply@iasystem.org](mailto:no-reply@iasystem.org)) with the survey link. Students should look for the email in their UNT email inbox. Simply click on the link and complete the survey. Once students complete the survey they will receive a confirmation email that the survey has been submitted. For additional information, please visit the SPOT website at <http://spot.unt.edu/> or email [spot@unt.edu](mailto:spot@unt.edu).

### Services Mental Health

UNT provides mental health resources to students to help ensure there are numerous outlets to turn to that wholeheartedly care for and are there for students in need, regardless of the nature of an issue or its severity. Listed below are several resources on campus that can support your academic success and mental well-being:

- Student Health and Wellness Center (<https://studentaffairs.unt.edu/student-health-andwellness-center>)
- Counseling and Testing Services (<https://studentaffairs.unt.edu/counseling-and-testing-services>) • UNT Care Team (<https://studentaffairs.unt.edu/care>)
- UNT Psychiatric Services (<https://studentaffairs.unt.edu/student-health-and-wellnesscenter/services/psychiatry>)
- Individual Counseling (<https://studentaffairs.unt.edu/counseling-and-testingservices/services/individual-counseling>)