CSC 405/605/705: Data Science

Time: Tuesday / Thursday 12:30 pm – 13:45 pm
Date range: Aug 19, 2025 - Dec 03, 2025
Instructor: Dr. Qianqian Tong
Location: Petty Building 224
Office Hours: Tuesday/Thursday 3:15PM – 4:30PM at Petty 157
Email: q_tong@uncg.edu

Course Description

In today's world, where data generation is prolific, both from human and machine sources, the realm of computer science has shifted from focusing primarily on computation-intensive methods to embracing data-intensive strategies. This change underscores the potential to solve real-world problems by analyzing diverse, complex, and unstructured datasets through Data Science methodologies.

This course offers a highly interactive environment where students will delve into the theories, techniques, and tools essential for extracting meaningful insights from large datasets. Adopting a problem-based learning approach, participants will utilize these technologies to create data-driven solutions capable of processing and analyzing real-world data for addressing various challenges in scientific, social, and environmental domains.

The core topics addressed by the course will be:

- Programming for Data Science
- Data Mining, Munging, Wrangling
- Statistics, Analytics, Representation, Visualization
- Introduction to Applied Machine-Learning

Prerequisites

A grade of C+ or better in CSC 330 and (STA 271 or STA 290), or permission of instructor (prior programming and statistics experience is required).

Textbooks

There is no required text for the course. Class slides will be available for download. Suggested textbooks are: 1) Building Machine Learning Systems with Python (Richert and Coelho), 2) Data Science from Scratch (Joel Grus)

Course Overview

In this course, participants will delve into the fundamentals of Data Science, exploring a wide array of tools, techniques, and foundational concepts. Throughout the course, students will engage in a detailed examination of the entire data science workflow. This includes the processes of data collection, cleansing, exploration, and feature engineering. Additionally, learners will gain hands-on experience in building models, validating results, and interpreting data insights. By the conclusion of the course, students will have acquired a thorough understanding of how to apply data science principles effectively in practical scenarios.

Course Topics and Schedule (Tentative)

1. **Introduction to Data Science: (Week 1-3)**
    - o Data Science Introduction
    - o Class Project discussion and team formed
    - o Programming prepare
        - 1). Re/Introduction to Python
        - 2). IPython, IPython-Notebook
    - o Data Science Reproducibility
        - 1). Setting up your Repository – Data, Code, and Documentation
        - 2). Using Version Control with Git
    - o Project Review - Stage I

2. **Data Munging, Wrangling, Cleaning (Week 4-5)**

    - o Data Structures
    - o Data Manipulation
        - 1). Selection - Indexing
        - 2). Handling Missing Data
        - 3). Aggregation
        - 4). Descriptive Statistics
        - 5). Merging / Join
        - 6). Working with Date-Time
    - o Assignment 1 due
    - o In class quiz 1

3. **Data and Statistics (Week 6-9)**
    - o Distributions
    - o Estimates
    - o Statistical Hypothesis Testing
    - o Correlation
    - o Distribution Estimators: MoM, MLE, KDE
    - o Project Review - Stage II
    - o In class quiz 2

4. **Introduction to Applied Data Modeling: (Weeks 10-12)**
    - o Applied Machine Learning
    - o Regression and Feature Selection
    - o Bias versus Variance
    - o Clustering and Dimensionality Reduction
    - o Validation and Model Performance
    - o Mathematical optimization (if time allowed)
    - o Stochastic thinking (if time allowed)
    - o Invited talk (if time allowed)
    - o Assignment 2 due

5. **Data Visualization (Week 13-14)**
    - o Graph Generation
        1). Types of Graphs
        2). Customizing Plots
        3). Visualizing Errors
        4). Interactive / Dynamic Graphs
    - o Visualization Best Practices
    - o Project Review - Stage III
    - o In class quiz 3

6**. Project Presentations: (Week 15–16)**
    - o Assignment 3 due
    - o Project Review - Stage IV-Final presentation
    - o Graduate Students report submission

Grading Policy (No curve in the final grade)

Grade	Max%	to	Min%
A	100%	to	92%
A-	< 92%	to	89%
B+	< 89%	to	86%
B	< 86%	to	83%
B-	< 83%	to	80%
C+	< 80%	to	77%
C	< 77%	to	74%
C-	< 74%	to	70%
D+	< 70%	to	67%
D	< 67%	to	64%
D-	< 64%	to	60%
F	< 60%	to	55%

1. Class Participation: 10%

Attendance is mandatory for all class meetings. If a student is unable to attend an in-person class, they must inform the instructor in advance by providing a valid reason for their absence. This communication should be done through email and must be sent before the class session begins. Failure to notify the instructor prior to the start of class will result in the student losing credit for that absence. It should be noted that attendance records may be taken either at the beginning or the end of the class. Students are advised to ensure their presence throughout the session to avoid any discrepancies in the attendance record.

2. In class quizzes (3): 40%

Throughout the course, students will be assessed through **three in-class quizzes**, each designed to evaluate their understanding of the lecture material and class discussions. These quizzes will collectively contribute **40 points** toward the final course grade, distributed as follows:

- **Quiz 1** – 10 points
- **Quiz 2** – 10 points
- **Quiz 3** – 20 points

Each quiz is scored out of 100 points and scaled to its respective weight in the overall grade. All quizzes must be completed **in class**, and students are expected to answer all questions thoroughly. To receive full credit, quizzes must be **submitted within the allotted timeframe**. Late submissions may result in partial or no credit, depending on the circumstances.

3. Assignments (3): 15%

Three programming-based assignments will be given covering the utilization of the tools learned in class. Each assignment accounts for 5 points. Absolutely no collaboration on assignments. Students must upload (Notebooks) individual assignments to canvas before deadline. Later submission (per day) will have a 50% deduction, late for more than 2 days will directly have zero grade.

4. Project: 35%

The project for this course will involve the complete end-to-end development of an analytical model. It will be organized into the following stages:

- Stage I. Dataset Selection and Project Setup,
- Stage II. Data Analysis, Distributions and Hypothesis Testing,
- Stage III. Machine Learning and Deep Learning Model Development,
- Stage IV. Visualization & Final presentation

This will be a team-based effort, where in the first week of the course the students split into teams of 3-4 students. After completing each stage, the teams will have to give a short presentation (10 mins) and a report of their progress with the project. The projects will be open-source, and the teams will have to use GitHub as their code repository. Upon completion of the project the teams will present their software along with the results in form of a presentation (15-20 minutes).

Each Stage of the Final Project has 100 points. They will be equally weighted for the project final score. Each stage consists of: 1). Report; 2). Code Jupyter/IPython Notebooks; 3). Presentation. To get the full points in each stage you need to finish all the deliverables.

Graduate Students Only: In addition to the final presentation, graduate students are required to submit a project report in IEEE format. The report should be at least 3 pages for a single author, 5 pages for two authors or more, inclusive of figures and references. Please ensure that the report is submitted before the final week of the course.

Academic Honesty Policy

The instructor will deal strictly with any violations of academic honesty and integrity in this course. ***Absolutely no discussion, collaboration, copying, and sharing on assignments. This includes coping from the internet. Any student who violates this policy will receive "F" directly in the course. The instructor will report the case to the university.***

Special Needs and/or Disabilities

Students with disabilities should have documentation from the Office of Accessibility Resources & Services. This documentation should be provided to the instructor for review. In the case of major provisions such as separate testing environment or test-readers, the student must make arrangements with Office of Accessibility Resources & Services so that suitable accommodations can be provided.

Midterm Grades

The midterm grade for Fall 2025 dues on September 26, 2025. During this time, I will assign all undergraduates a midterm grade for this course, which you can access in UNCGenie. Your midterm grade in this course is a snapshot of how you are currently performing academically based on the assignments we have had to date. It will let you know if you are on the right track or if you need to take action to do something differently to improve your grade. If you have a D

or an F at the midterm, we should definitely talk further about strategies and options for continuing in the class. You can find more information about midterm grades here: https://spartancentral.uncg.edu/student-records/grades/ Once midterm grades are assigned, reach out to me if you have questions. You should also talk with your academic advisor if you are considering withdrawal from this class.

Health and Wellness

Health and well-being have a big impact on your learning and academic success. Throughout your time at UNCG, you may experience a range of concerns that impact your personal and academic success. These might include illnesses, strained relationships, anxiety, high levels of stress, alcohol or drug concerns, crime victimization, feeling down, loss of motivation, or death of a loved one. It is OK TO ASK FOR HELP!

· Student Health Services (SHS) (336-334-5340): For preventative and acute healthcare, SHS offers a primary medical clinic, full pharmacy, and over-the-counter medications.
· Counseling & Psychological Services (336-334-5874): free confidential mental health services
· Spartan Well-Being
· Campus Violence Response Center (336-334-9839)
· Spartan Recovery offers recovery support services (SRP@uncg.edu)