

Syllabus

DTSC 1301: Data and Society A (section 02)

Semester: Fall 2025

Format: Studio-based (Hands-on, interdisciplinary, project-oriented)

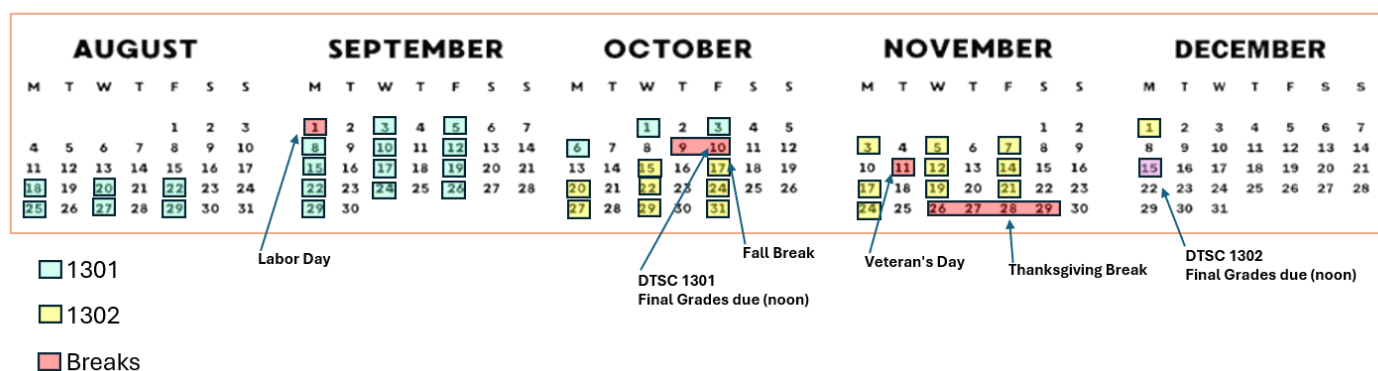
Days & Time: Mondays, Wednesdays, and Fridays, 10:10 am - 12:05 pm

Location: Colvard South, Room 1040

Course Duration: August 18th - October 6th

Final Exam Date: October 1st or October 3rd

Final Grade Due Date: October 10th (noon)



* DTSC 1301 and DTSC 1302 are two separate 3-credit courses part of the Data Science Studio 1 experience. DTSC 1301 = Studio 1A is the first half and DTSC 1302 = Studio 1B is second half.

Instructional Team

INSTRUCTORS

Instructor

Marco Scipioni, PhD

Email: mscipion@charlotte.edu (mailto:mscipion@uncc.edu)

Office Hours:

- Mondays 12:30 - 1:30 pm
- Tuesdays 1:00-2:00 pm

Office Location: 1048E, Colvard Building *South* (School of Data Science)

Zoom Link: <https://charlotte-edu.zoom.us/j/6023604735?omn=95804603877>  (<https://charlotte-edu.zoom.us/j/6023604735?omn=95804603877>)

Zoom Meeting ID: 602 360 4735

Instructor

Marianna B. Ganapini

Email: mbganapini@charlotte.edu (<mailto:mbganapini@charlotte.edu>)

Office Hours: Monday and Wednesday 1-2pm

Office Location: Winningham 109C or on zoom

TEACHING ASSISTANTS

Teaching Assistant (Data Science)

Name: Anirudh Pamidimukkala

Email: apamidim@charlotte.edu (<mailto:apamidim@charlotte.edu>)

Office Hours: TBA

Location: TBA

Teaching Assistant (Data Science)

Name: Emanuel Stanescu Moloman

Email: estanesc@charlotte.edu (<mailto:estanesc@charlotte.edu>)

Office Hours:

- Mondays: 12:30 - 1:30 pm (Colvard Building)
- Wednesdays: 4:00 - 5:30 pm (Atkins Library)
- Fridays: 12:30 - 1:30 pm (Colvard Building)

Location: Zoom or Colvard Building/Atkins Library

Teaching Assistant (Data Ethics)

Name: Johnny Flynn (pronouns: he/him)

Email: jflynn25@charlotte.edu (<mailto:jflynn25@charlotte.edu>)

Office Hours:

- Tuesdays: 10:00 AM - 11:30 AM, or by appointment (and email is always open too)

Location: Garinger 105

Course Description

An introduction to data acquisition, models, and analytic methods for interpreting data and developing hypotheses in the context of the interdisciplinary field of Critical Data Studies. Critical Data Studies seeks to track, analyze, and transform the use of large data sets across a variety of domains, including the health sciences, security and surveillance technologies, social media, marketing and business, government and public policy, and other uses of aggregative and algorithmic data science. With resources from Critical Data Studies, students learn how to use statistical methods/tools and scripting programming languages to explore social problems and the ethical implications of collecting and using tabular data.

Credit Hours: (3)

Co-requisites. **DTSC 1302** (https://catalog.charlotte.edu/content.php?filter%5B27%5D=-1&filter%5B29%5D=&filter%5Bkeyword%5D=DTSC+1301&filter%5B32%5D=1&filter%5Bcpag%5D=1&cur_cat_oid=27&expand=&navoid=2571&search_database=Filter#tt8448); **STAT 1220**

(https://catalog.charlotte.edu/content.php?filter%5B27%5D=-1&filter%5B29%5D=&filter%5Bkeyword%5D=DTSC+1301&filter%5B32%5D=1&filter%5Bcpag%5D=1&cur_cat_oid=27&expand=&navoid=2571&search_database=Filter#tt2426), **STAT 1221** (https://catalog.charlotte.edu/content.php?filter%5B27%5D=-1&filter%5B29%5D=&filter%5Bkeyword%5D=DTSC+1301&filter%5B32%5D=1&filter%5Bcpag%5D=1&cur_cat_oid=27&expand=&navoid=2571&search_database=Filter#tt945), **STAT 1222** (https://catalog.charlotte.edu/content.php?filter%5B27%5D=-1&filter%5B29%5D=&filter%5Bkeyword%5D=DTSC+1301&filter%5B32%5D=1&filter%5Bcpag%5D=1&cur_cat_oid=27&expand=&navoid=2571&search_database=Filter#tt5631).

Course Materials

- Students must have access to a laptop or desktop (Linux, Windows, or macOS operating system preferred).
- All software and learning materials required for the course are free and open source.
- **Reference Open-source Textbook:** *Python Data Science Handbook* by Jake VanderPlas is available at <https://jakevdp.github.io/PythonDataScienceHandbook/> 
(<https://jakevdp.github.io/PythonDataScienceHandbook/>)

Notebooks for the book: <https://github.com/jakevdp/PythonDataScienceHandbook> 
(<https://github.com/jakevdp/PythonDataScienceHandbook>)

Course Objectives

Upon successful completion of DTSC 1301 and DTSC 1302, students will develop technical proficiency in data science, build awareness of the ethical implications and challenges in the practice of data science, and foster critical thinking in data analysis.

Data Science Research Skills

- Develop hypotheses that anticipate associations in data based on a thorough literature
- Evaluate the unique position of citizens in social phenomena that occur in the US and the world
- Understand and linking social, economic, and political concepts to measurement, data collection, and analysis

The Ethics of Data Science

- Evaluate ethical and policy-based debates within data
- Apply ethical principles/values/frameworks within data-driven organizational
- Understand and present the positions on multiple sides of an ethical issue, including the reasons, principles, and values offered in ethical arguments.
- Understand and assess the ethical stakes of conducting research on or about human subjects
- Critique arguments that involve ethical issues in data

Data, Computing, Statistics, and Probability

- Collect and clean data obtained from various datasets
- Appropriately deal with missing values
- Use a statistical programming language to perform data analysis
- Create statistical models using the concepts of correlation, linear regression, least squares, residuals
- Test hypotheses and identify statistically significant results

Course Topics

The table below outlines the weekly schedule for the course. Each week builds on the previous one, progressing from foundational computing and Python programming toward data handling, modeling, and deployment. Ethics in data science is integrated throughout to foster responsible and thoughtful application of technical skills. In DTSC 1301, Students will be working on different synthetic datasets (provided by the instructors) with only minor missing data.

<u>Week 1</u>	Technical Data Science
---------------	------------------------

a) General Computing Concepts

- Files, folders, absolute vs. relative paths
- Environment variables (especially PATH)
- Terminal basics and command-line navigation (cd, ls, python, script.py)

b) Python & Environment Setup

- Installing Python and IDEs (VS Code, Jupyter Notebooks)

c) Data Types in Python

- Variables and data types (int, float, str, bool)
- Expressions, type conversion using int(), float(), str()

d) Data Structures

- Lists (indexing, slicing, mutability)
- Tuples and sets
- Dictionaries (key-value pairs, access and mutation)

Data, Ethics, and Society**I. General ethical concepts**

- What is ethics?
- Subjective vs objective morality
- Making an 'ethical' choice
- Ethical lenses

II. Ethics of Tech and Data

- What is ethics of tech and data?

Week 2**Technical Data Science****a) Control Flow in Python Programming**

- Conditional logic (if, elif, else)
- Boolean logic (and, or, not)

	- Loops: for, while, control flow (break, continue, pass) b) Functions & Modules - Defining functions, parameters, return values - Importing standard libraries (math, random, os, datetime) - Installing external packages with pip - Create and activate virtual environments using venv or virtualenv for project isolation c) Working with Data in Python (Pandas) - Using pandas for tabular data - Reading and writing CSV/Excel files - String and datetime operations - Filtering, slicing, and inspecting DataFrames - Grouping and aggregation - Joining and merging datasets - String and datetime operations Data, Ethics, and Society I. <i>Stakeholder analysis</i> - Stakeholders' rights and interests - Understanding stakeholders' competing rights - Understanding harm and the utilitarian principle
– <u>Week 3</u>	Technical Data Science a) Understanding and Exploring Variables - Conceptualization and operationalization (abstract ideas to measurable variables) - Types of variables (categorical, continuous, ordinal, binary) - Visualizations (histograms, box plots, scatter plots) - Summary stats (mean, median, mode, std) - Effect size, power analysis b) Handling Data Issues - Introduction to missing values and outliers

	▪ Encoding categorical variables ▪ Feature scaling (standardization and normalization) c) Causal Thinking & Confounding ▪ Confounders, mediators, multicollinearity (introductory)
<u>Week 4</u>	Technical Data Science <i>Variable Selection and Feature Engineering</i> ▪ Domain knowledge and statistical feature selection ▪ Correlation analysis ▪ P-values and t-tests, ANOVA, A/B testing (in samples' analysis and regression) ▪ Stepwise selection, Lasso, Ridge ▪ VIF (Variance Inflation Factor) for multicollinearity ▪ Correlation analysis for bivariate analysis ▪ Automatic feature selection techniques ▪ Feature engineering Data, Ethics, and Society I. How can we lose our data privacy? ◦ Data Leaks, Selling Data, Data Aggregators ◦ Scraping data and the problem of consent ◦ Informed consent, de-identification, re-identification risk ◦ Reasonable expectations II. <i>Privacy and Data</i> ◦ Whose data is private? ◦ Is Privacy a right? ◦ Privacy techniques: data minimization, k-anonymity/l-diversity, intro to differential privacy. III. <i>How data is handled</i>

IV. *Case studies of privacy violation in big data***Week 5****Technical Data Science*****a) Modeling Fundamentals and Machine Learning Basics***

- What is a statistical model?
- Introduction to machine learning; types of ML tasks
- Training vs. testing datasets
- Overfitting and underfitting

b) Regression and Linear Regression

- Coefficients as relationships
- Simple vs. multiple regression
- Assumptions (linearity, normality of residuals, homoscedasticity, no multicollinearity)

c) Regression Metrics

- R^2 , adjusted R^2 , RMSE, MAE

Data, Ethics, and Society***I. Ethical Principles***

- Fairness and Justice
- Rights
- Special Categories Data

II. Data and Bias

- Unjust biases and discrimination in data aggregation
- How can data be biased?
 - Historical, Representational and Measurement bias
 - Benchmarks for data quality

III. Data and Bias in ML applications

- Bias and policing
- Bias and healthcare data

IV. Solving the problem of bias in ML

<u>Week 6</u>	Technical Data Science <i>a) Classification and Logistic Regression</i> ▪ Binary and multiclass classification ▪ Logistic regression, sigmoid function, probability thresholding ▪ Odds and log-odds interpretation ▪ Assumptions <i>b) Classification Metrics</i> ▪ Confusion matrix ▪ Accuracy, precision, sensitivity, log-loss, recall, F1-score ▪ ROC curve and AUC ▪ Pseudo R^2 (McFadden's, Cox, Snell R^2) Data, Ethics, and Society <i>I. Accuracy and robustness of ML</i> ◦ Model drift ◦ Deployment errors <i>II. Residual risks and moral dilemmas</i> ◦ Accuracy vs fairness ◦ Individual fairness vs group fairness ◦ Short-Term Equity vs Long-Term Social Impact
---	--

Course Strategy

Teaching methods are grounded in an interdisciplinary studio style approach to learning in which students are presented with cross-disciplinary data science challenges that guide the mini lectures, in-class activities, and group data-driven projects. Students are expected to participate in discussions, problem solving, critical thinking exercises, analysis of case studies, collaboration, peer-teaching, and design and critique sessions. All students are expected to contribute in a meaningful way to team efforts.

Course Grading

DTSC 1301 is a PASS/FAIL course which means that students, at the completion of the course, will receive either a passing grade (P) or a failing grade (F) based on the percentage of the final grade that is worth 100%. Taking the final exam is required to pass the class.

Earning 60% or higher = PASSING

Earning 59 % or less = FAILING

Weighted Grading Scheme

Attendance, data science assignments, ethics assignments, and a final exam are assigned % weights out of a total of 100% (see table below).

Task	% of Final Grade
Attendance	10%
In-class activities	10%
Quizzes	25%
Assignments	25%
Final Exam	30%

A student must pass DTSC 1301 to continue on to DTSC 1302.

Students will be assigned a final letter grade for DTSC 1302 at the end of fall semester (mid-December).

To progress to follow-on courses, a student must earn a D or better in DTSC 1302.

Course Policies

Course policies have been provided in a separate document available on our course Canvas page or at the following link

<https://legal.charlotte.edu/policies/up-407> (<https://legal.charlotte.edu/policies/up-407>)

Policy on the Use of Generative Artificial Intelligence (AI)

The use of AI is not mandatory in this course but you are encouraged to consider utilizing AI as a tool to enhance your learning experience:

1. Use ChatGPT as a tool for generating ideas and insights, but verify information through authoritative sources before incorporating it into your work.

2. Utilize ChatGPT to gain alternative perspectives or enhance your creativity, but critically evaluate the output to ensure its relevance and accuracy.
3. Refine your prompts and optimize ChatGPT's responses, but always ensure that the final work reflects your own understanding and analysis.
4. When using ChatGPT to assist with research, complement it with thorough literature review and proper citation practices to maintain academic integrity (see below for guidance on how to cite generative AI tools, and why it is important to cite sources).
5. Leverage ChatGPT's assistance for complex problem-solving or decision-making, but remain cautious of any biases or limitations that may arise from the AI model.

Guidelines to help you make the most out of AI (ChatGPT) and Ensuring Ethical Usage

1. **Understand the limitations of ChatGPT:** Recognize that ChatGPT has its constraints and may not always provide accurate information or answers. It requires thoughtful and well-crafted prompts to yield quality results.
2. **Verify information independently:** If ChatGPT presents a number or fact, assume it is incorrect unless you have reliable knowledge or can cross-reference it with other credible sources. Ultimately, you are responsible for the accuracy of information used in your work. Similar to Wikipedia, ChatGPT can be a good sounding board or beginning place for information, but frequently provides inaccurate information mixed in with true content.
3. **Treat AI as a tool:** When employing AI in your assignments, explicitly acknowledge its use. Include a paragraph at the end of any AI-assisted work, explaining the purpose of using AI, the specific prompts used, and how it contributed to your results. Failure to do so would be a violation of academic honesty policies.
4. **Exercise ethical judgment:** Be discerning about when it is appropriate to use AI as a tool. Consider the context and requirements of each task, ensuring that AI aligns with the assignment's objectives and guidelines.
5. **Information Responsibility:** You are responsible for any errors or omissions provided by the tool. It works best for topics you understand.

Suggested Template for Students' Documentation of AI Use

I acknowledge the use of AI in completing this assignment and would like to provide a brief explanation of how I utilized AI, specifically ChatGPT, as a tool to support my work. For this assignment, I employed ChatGPT to [describe the specific purpose or task]. To do so I crafted the following prompts:

[List prompts used]"

How to cite generative AI in:

- **APA Style** → <https://apastyle.apa.org/blog/how-to-cite-chatgpt>
- **MLA Style** → <https://style.mla.org/citing-generative-ai/>

- [Chicago Style](https://www.chicagomanualofstyle.org/qanda/data/faq/topics/Documentation/faq0422.html) 
(<https://www.chicagomanualofstyle.org/qanda/data/faq/topics/Documentation/faq0422.html>)
- [ChatGPT & AI – Citation Style Guide](https://guides.nyu.edu/c.php?g=276562&p=9693224)  (<https://guides.nyu.edu/c.php?g=276562&p=9693224>)

When and why to cite sources:

- [Citation Style Guide – Research Guides at New York University](https://guides.nyu.edu/citations) 
(<https://guides.nyu.edu/citations>)
- [When and Why to Cite Sources | University Libraries](https://library.albany.edu/infolit/resource/cite-sources) 
(<https://library.albany.edu/infolit/resource/cite-sources>)