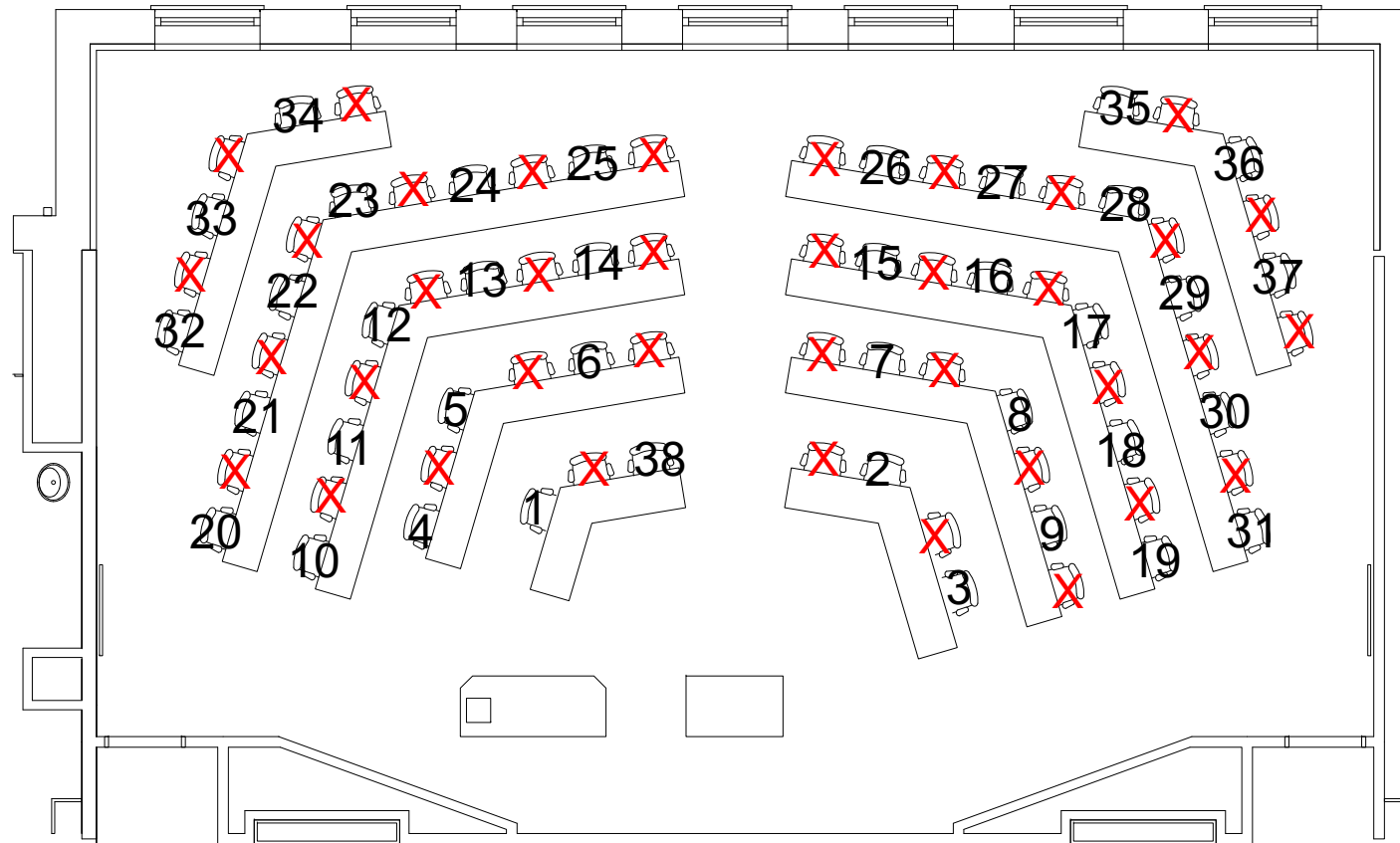# CSC 405/605 Seating Assignments

X: DO NOT SIT

CSC 405/605 Seating Assignments

X: DO NOT SIT

Office of Space Utilization and Planning

800 Oakland Ave.
Greensboro, NC
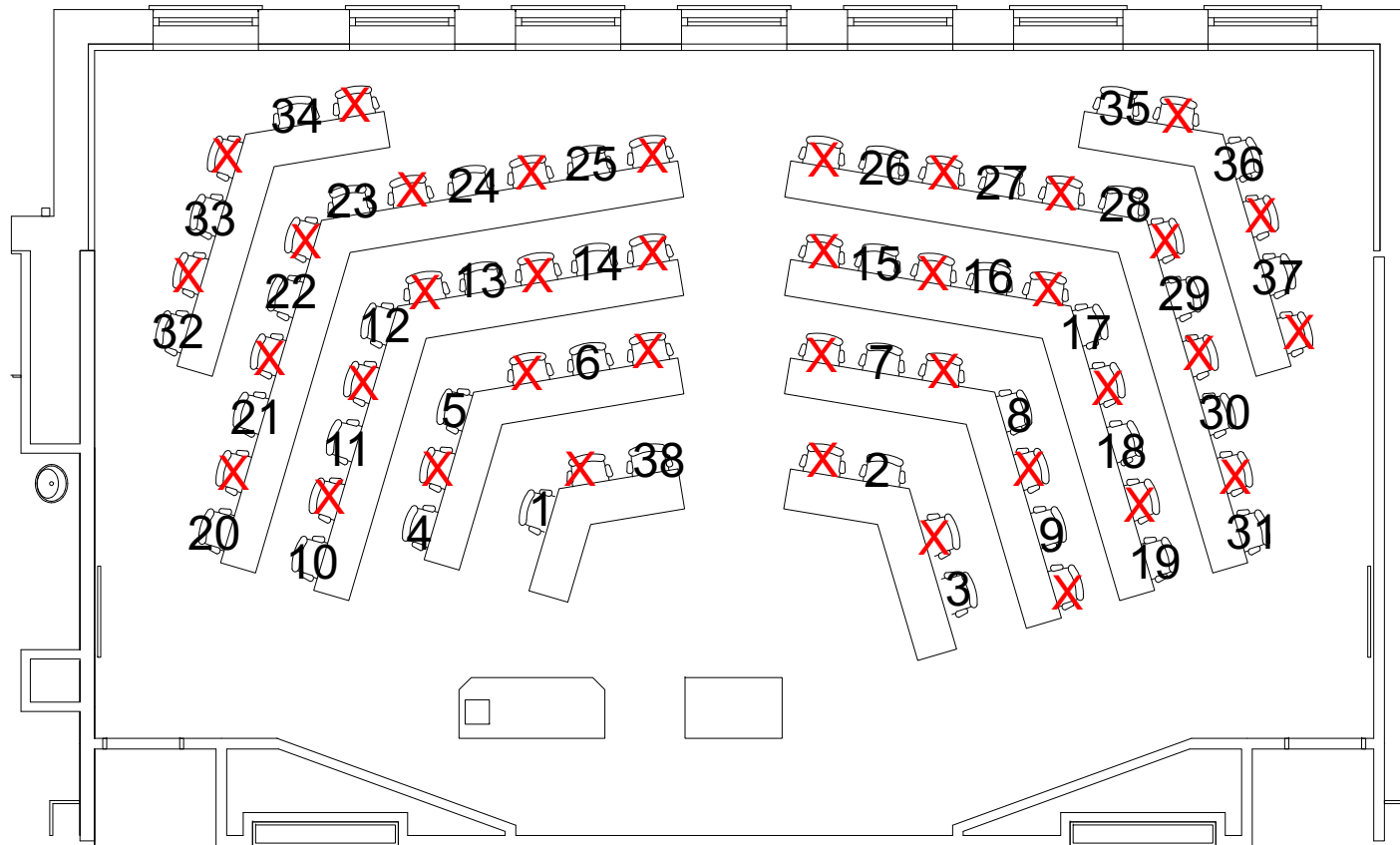27402-6170

BUILDING

PETTY BUILDING

CLASSROOM

219

MUST MAINTAIN AT LEAST 3' CLEARANCES IN ALL WALKWAYS FOR ADA ACCESSIBILITY.

https://forms.gle/RvZsNGjFS25Ac7zz9

Data Science

CSC 405/605

Course Info

- Course: Data Science
  - CSC 405/605
  - Monday & Wednesday 5:30PM-6:45PM
  - Office Hours: Thursday 5:00PM – 6:00 PM via Zoom only (Email for appointment and zoom link)
  - Location: Petty Building 219
  - Class Discussions: https://discord.gg/N2rhnw7h3J

  - Instructor: Dr. Kelechi Ikegwu
    - Email: kmikegwu@uncg.edu
  - Course Details:
    - https://github.com/ikegwukc/CSC-405-605_Spring_2022
    - Course Assignments, Project Details, and Lecture Notes are posted here

UNC
GREENSBORO

About Me

- B.S. In Information Technology at NC A&T - 2016

- Ph.D. In Informatics at UIUC – 2021
    - Machine Learning
    - Information Theory
    - Network Analysis
    - Financial Accounting

- Working as a Data Scientist in Industry

Course Info

- What is the course about?
- Applied Data Science
- Not an AI, Statistics, or Visualization Course
- Course contains a bit of everything
- Learn a lot of tools and how to use them in innovative ways
- We will work with real world data
- Hopefully develop some cool projects

Course Info

- Experience in:
  - Programming in Python
    - We will go through an Introduction to Python
    - You will have to work hard during the early weeks to understand it
  - Linux
    - Basic shell commands
- Course Textbook:
  - Nothing is required ☺
  - Recommended
    - Building Machine Learning Systems with Python (Richert and Coelho)
    - Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (Wes McKinney)
  - Lecture notes will be made available
    - along with online resources

Course Info

- Grading
  - **Class / Homework Assignments (4): 30%**
    - Individual Assignments
    - Deliverables
      - Code Jupyter/IPython Notebooks – Github
  - **Final Project: 70%**
    - ***Team and Individual***
    - Stage I: Data/Project Understanding,
    - Stage II: Modeling,
    - Stage III: Basic Machine Learning
    - Stage IV Visualization
    - Deliverables:
      - Report – Canvas & Github
      - Code Jupyter/IPython Notebooks – Github
      - Presentation
  - **No Exams**

Course Info

- Grading
  - Homework Assignments (4):
    - Utilization of tools learned in class
    - Mostly programming and data analysis
    - The submission will be on IPython notebooks
    - Utilize GitHub for assignments (own account)
      - Create your own repository **(Private repository)**
      - Add me as a collaborator to it, my id is:
        - *ikegwukc*
    - Link to the assignment submission via Canvas for submission

Course Info

- Grading

  Final Project:

  - Most of the grade is based on the Final Project
  - Each Stage is 17.5 points.
  - Projects contain tasks – Team and Member tasks.
    - Team is collaborative
    - Member is individual, discussion allowed
  - All stages are weighted equally for the final project points.
  - At end of each stage, provide deliverables (Report, Notebooks on GitHub, and Presentation)
    - Team Repository

Course Info

- Grading

  Final Project:

  - Presentation (At each stage)
    - 2 min presentation by each member
    - Review of the work done and report **interesting outcomes**
    - Upload presentation recording to canvas.
  - Final Presentation on completion (Finals Week)
    - 20 min presentation (Uploaded to canvas)
      - Data, Methods, Visualization
    - Graduate Students Only:
      - Paper (4 pages minimum) – IEEE/ACM Standard

Course Info

Category	Sub-Category	Deadline
Assignment	GitHub Accounts Created	2022-01-19 11:59 PM
	Assignment 1	2022-01-30 11:59 PM
	Assignment 2	2022-02-13 11:59 PM
	Assignment 3	2022-03-06 11:59 PM
	Assignment 4	2022-04-17 11:59 PM
Project	Groups Formed	2022-01-19 11:59 PM
	Stage 1	2022-02-20 11:59 PM
	Stage 2	2022-03-20 11:59 PM
	Stage 3	2022-04-10 11:59 PM
	Stage 4	2022-04-27 11:59 PM

UNC GREENSBORO

Course Info

A	100%	to	94%
A-	<94%	to	90%
B+	<90%	to	87%
B	<87%	to	84%
B-	<84%	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	

Course Info – Timeline (Tentative)

- Introduction to Data Science: (Week 1)
  - Class Syllabus, Grading, Expectations, and Getting to know each other.
  - Introduction to Data Science.

- Startup Tools and Programming (Weeks 2-3)
  - Programming
    - Re/Introduction to Python
    - IPython, IPython-Notebook
  - Data Science Reproducibility
    - Setting up your Repository – Data, Code, and Documentation
    - Using Version Control with Git
  - Final Project Discussions - Goals and Requirements

Course Info – Timeline (Tentative)

- **Data Munging, Wrangling, Cleaning (Week 4-5)**
  - Data Structures for Data Science
  - Data Manipulation
    - Selection-Indexing
  - Handling Missing Data
  - Aggregation
  - Descriptive Statistics
  - Merging/Join
  - Working with Date-Time
  - *Project Review - Stage I*

Course Info – Timeline (Tentative)

- **Data and Statistics (Week 6-9)**
  - Distributions
  - Point Estimates
  - Statistical Hypothesis Testing
  - Correlation
  - Distribution Estimators
    - MoM,MLE,KDE
  - *Project Review - Stage II*

Course Info – Timeline (Tentative)

- **Introduction to Applied Data Modeling: (Weeks 10-12)**
  - Applied Machine Learning
  - Regression and Feature Selection
  - Bias versus Variance
  - Clustering and Dimensionality Reduction
  - Validation and Model Performance
  - *Project Review - Stage III*

Course Info – Timeline (Tentative)

- **Data Visualization (Week 13-14)**
  - Graph Generation
    - Types of Graphs
    - Customizing Plots
    - Visualizing Errors
    - Interactive/Dynamic Graphs
  - Visualization Best Practices
  - *Project Review - Stage IV*
- **Project Presentations: (Week 15 – Final's Week)**

Course Info – Disclaimer

- **Read the syllabus.**
- **Take regular notes.**
- **Class is encouraged to participate and discuss/ask questions – Class Participation Points!**
- **Communication**
  - Discord channel
- **Questions about assignments and projects**
  - Read the syllabus and assignment/project descriptions carefully.
  - If still unclear search the discord channel to see if someone else has asked the question already.
  - If still not found, post on the channel if it is a general question.
  - Each class we will spend the last 10 mins on discussion regarding assignments and projects.
  - Email should be the last step to communicate with me.

UNC GREENSBORO

Course Info – Disclaimer

- **The course is going to be tough, especially for people with limited programming experience**
  - Work hard, be rewarded with a good data science experience
  - Will talk about the benefits later in course intro
- **Do not cheat in the course – Result will be an 'F' grade.**
  - Assignment solutions are unique, differs from student to student. **No collaboration on Assignments and Project Member Tasks whatsoever.**
  - I will run the code through plagiarism detection software - *single incident reporting to honor committee*
  - In team project
    - Do not think that you can get away without contributing - *Monitoring repositories for work done*
    - Any work done should be reported on the repository – *worked locally on my computer will not count.*
- **Utilization of resources found on the Internet is allowed for project accomplishment, with caveats**
  - Any code/library used should be referenced/cited and thoroughly understood
  - If you use code without understanding, that counts as plagiarism

Course Info – Disclaimer

- **On team projects**
- Start early
  - Emailing me questions about assignments and projects 2 days before submission will not get you a response.
- The team creation can be random or self-assigned
  - **Task for today** – Get in touch with class participants and setup groups of 4-5 students. Mix of graduate and undergraduate
  - Email me the group list (student names, emails, Github ids)
  - Use the Discord channel for group formation and discussion.
  - Project presentations, all members should present. If someone does not show up, they will not be graded for the stage.

Course Info – Disclaimer

- **More on team projects**
  - You will get critical comments from me, both on presentation and project progress
  - Its geared towards making your projects awesome!
  - You will be presenting at the end to the department and external attendees.
    - We are trying to achieve a great presentation *made by you for your project*.
  - ***Use office hours – I am here to help you through your project.***
    - ***https://zoom.us/j/6655909472?pwd=RVZibHNKamJjd25QNHd0cjRkbm1wdz09***
    - Email in advance to setup an appointment.
    - Office Hrs: Thursday 5:00 pm - 6:00 pm

Questions?

Action Items

- Sign Up for GitHub: github.com
  - https://education.github.com/pack
- Sign Up for Discord: https://discord.gg/N2rhnw7h3J
  - Go To Announcement channel and fill out the GitHub form
- Peak at JuypterHub
- Peak at Class Projects
- Team Creation on Discord or FTF
  - Need to have your groups formed by Next Wednesday
  - See Slide 19 for other items

UNC
GREENSBORO