DATA SCIENCE

CSC 405/605

THE UNIVERSITY of NORTH CAROLINA
GREENSBORO

COURSE INFO

- **Course: Data Science**
  - CSC 405/605
  - Tuesday and Thursday 3:30 pm - 4:45 pm
  - Office Hrs: Thursday 2:00 pm - 3:00 pm via Zoom only (email for appointment and Zoom link)
  - Location: 212 Bryan
  - Class Discussions:
    - https://discord.gg/ktdrsUH4Fd (use #csc-405-605 channel for class discussions)
  - Instructor: Dr. Somya Mohanty
    - Email: sdmohant@uncg.edu
  - Course Details:
    - https://github.com/UNCG-CSE/CSC-405-605_Fall_2021

COURSE INFO

What is the course about?

- **Programming your way into Data Science**
- **Theory - Programming**
- **It is not a Statistics or an AI or a Visualization course**
- **The course contains parts of everything**
- **Learn about 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 skills Python**
    - *We will go through Introduction to Python*
      - *You would have to work hard in the early weeks to get comfortable with Python*
  - Linux
  - Terminal, Command-Line

- **Books:**
  - 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)

COURSE INFO

- **Grading**
  - **Class / Homework Assignments (4): 30%**
    - **Individual**
    - 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, and
    - Stage IV Visualization and Dashboard.
    - Deliverables:
      - Report – Canvas & Github
      - Code Jupyter/IPython Notebooks - Github
      - Presentation - Online
- **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: somyamohanty
    - 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 100 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
    - 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	* Assignment 1	09/02/2021
	* Assignment 2	10/05/2021
	* Assignment 3	10/26/2021
	* Assignment 4	11/11/2021
Project	* Stage I	09/21/2021
	* Stage II	10/21/2021
	* Stage III	11/18/2021
	* Stage IV	12/02/2021

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	59%

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***
- **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
    - https://discord.gg/ktdrsUH4Fd (use #csc-405-605 channel for class discussions)
- **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.

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 presentation recording, all members should present. If someone does not, they will not be graded for the stage.

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 - *I will be 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

- **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://uncg.zoom.us/j/3634402596?pwd=N01WaW9QM0c2VlpqUjlPZk0xRjQ5Zz09
    - Email in advance to setup an appointment.
    - Office Hrs: Thursday 2:00 pm - 3:00 pm

QUESTIONS