Professor Nicholas Schettini
nicholas.schettini@sps.cuny.edu
linkedin.com/in/nschettini

DATA 602: Advanced Programming Techniques

Python for Data Science

Course Description

This 15-week course will provide you with advanced programming techniques in Python in the context of data manipulation, data analysis, basic machine learning and visualization of data.

We will rely on Python 3.8.x to study these techniques.

Prior knowledge of Python is not required

Course Objectives

1. Understand core Python concepts, data structures and programming techniques.
2. Learn data science-specific Python packages like NumPy, pandas, matplotlib and sci-kit learn.
3. Perform data analysis using Python libraries as to mimic that of a work-place environment
4. Understand basic machine learning usage with Python using regression and KNN
5. Gain exposure to real world tools and techniques like SQL and form connections to the python environment.

Grading

Topic	**Weight**
Pre/post course assessment	5%
Syllabus Quiz	5%
Participation	10%
Homework	20%
Weekly Assignments	30%
Data Project	20%
Final Exam	10%

Topics by Week

Week	Topic
1	Getting Started
2	Python Basics - variables, types, conversion
3	Python Basics - Data Structures
4	Python Basics - Conditionals, Loops, Functions
5	Python Basics - Classes
6	Data Manipulation with NumPy
7	Data Manipulation with Pandas - Gathering and loading data
8	Data Manipulation with Pandas - Data Manipulation
9	Data Manipulation with Pandas - Data Analysis
10	Data Visualization - visualization best practices, matplotlib
11	Data Visualization - seaborn
12	Data Visualization - plotly
13	Intro to Machine Learning - Scikiit learn
14	Intro to Machine Learning - Scikiit learn
15	Python and SQL

Pre-Course Assessment

A pre-course assessment must be completed by the end of Week 1. The assessment questions are not graded, only whether the full assessment form was completed or not is. The link to the assessment is on BlackBoard. The assessment can be attempted only once and is not timed.

Syllabus Quiz

A syllabus quiz must be completed by the end of Week 2. The quiz tests your knowledge of the contents of this document. The link to the quiz can be found on the course website or in the Week 1 BlackBoard Course Materials folder. The exam is

aggressively timed and randomized from a pool of questions, which means you should read the syllabus before you take the quiz. The quiz can be attempted only once and must be completed in one sitting.

Participation

Participation in this course will be evaluated from the following: blackboard discussion board weekly questions and our slack channel. The most valuable form of participation is where you are helping to solve other people's problems and providing interesting content. Casual conversations and general discussions do count in favor of your participation grade. Class participation will be calculated by the volume AND quality of contribution and the professor's subjective measure of that contribution.

The link to the semester **slack** channel will be posted as a link in BlackBoard at the beginning of the semester.

Homework

Homework will consist of chapters from various datacamp courses. The specific courses will be listed in the corresponding week's content folder in blackboard. These courses must be completed using your CUNY email address. Each datacamp chapter is aligned to the current week's material, but can be turned in by the end of week 16 for full credit. However, it is in your benefit to complete them weekly. A partial competition of a chapter will result in a missed grade. Most weeks will have homework assignments to be completed.

Weekly Assignments

Generally for each topic in the course, an assignment will be assigned to reinforce the skills covered. **Most** weeks will have assignments to be completed, although some will take more than a week. You may work in small groups to complete these assignments. Time will be given to work on these assignments during class meetups. Breakout rooms will be made available.

Data Project

The purpose of the data project is for you to conduct an analysis with a dataset of your choosing. You will use skills learned from the course to assist in your analysis. There are three components to the project, the proposal, which will be graded on a pass/fail basis, the final project (your analysis) and a video recording presentation. You may choose to work in a group of 2-3 people.

Project proposal can be an informal document using bullet points if you'd like and include any python code or output. It must include at the minimum:

- Research Question
- Justification - why is this relevant to you or industry?
- Data Sources - did you find this data online or collect yourself? Provide links.
- Libraries potentially being used.
- EDA and summary statistics.

The **Final Project** should include:

- Abstract (no more than 300 words) - summarizes analysis, provides conclusions and implications
- Introduction - research question clearly stated, can be answered by the data
- EDA - well-labeled graphs/plots that are appropriate to your analysis. Each visualization includes a description of what is being shown.
- Data Wrangling - tidying of your data, manipulating missing values, structuring of your data, etc.
- Data Analysis and Machine Learning
- Conclusions

More information will be released in BlackBoard in the upcoming weeks.

Final Exam

The final exam will be a timed test of 25-30 multiple choice questions out of a different pool of questions, and will be administered via Blackboard. The exam is open book and individual only. A twelve day window will be established starting from day 1 of Week 15 for when you can start the final exam. After the window, access to the exam will close and any students not completing the exam during that window will receive a zero for the exam. The exam will last 60 minutes and must be completed within one sitting, there will be no ability to pause and resume the exam. The exam will contain a mix of easy, intermediate and difficult questions and will feel like it is aggressively timed for those who are not reasonably prepared. The exam is forward only, meaning that once a question has been answered you cannot return to it to change the answer.

THE FINAL EXAMS IS TO BE TAKEN INDIVIDUALLY ONLY. ACCESS TO BOOKS AND ONLINE MATERIAL IS ALLOWED, BUT COMMUNICATION TO ANY INDIVIDUAL DURING TEST TAKING IS UNCONDITIONALLY DISALLOWED.

SHARING EXAM INFORMATION, INCLUDING QUESTIONS OR SOLUTIONS, ANYTIME AFTER THE EXAM IS ALSO DISALLOWED.

ANY VIOLATION OF THESE RULES IS GROUNDS FOR AN IMMEDIATE ZERO FOR THE COURSE AND A REPORT TO THE DEAN. PLEASE MAKE SURE TO USE A RELIABLE INTERNET CONNECTION WHEN TAKING THE EXAM AND TO NOT CLOSE THE BROWSER ON WHICH THE EXAM IS RUNNING

Software

To study the required texts and work on the assignments, access to Python is required. There are a number of ways of getting Python on your computer or using it in the cloud. Please install version 3.8.x, preferably the 64-bit edition.

- http://python.org The official Python website where it can be downloaded from. Use the 64-bit version.
- https://colab.research.google.com/ Requires no installation and allows you to code in python in code blocks similar to jupyter notebook.
- http://repl.it Requires no installation and allows you to code in Python and in other languages. An editor is included.
- https://anaconda.com/ A full suite that installs Python in addition other packages
- https://code.visualstudio.com/ Free editor from Microsoft
  https://www.jetbrains.com/pycharm/ Python-specific editor from JetBrains. Is not free, but CUNY SPS students may be able to download for free license.
- http://atom.io Free editor

Textbook and Readings

The popular Think Python book is required reading. The link to download the book is available on blackboard via the course dropbox menu option. In addition, select articles will be included in the weekly course materials.

Live Sessions & Office Hours (they're different)

Live Sessions

Live sessions will be conducted and announced as needed, conducted **roughly** every 2-3 weeks and will be held on Thursday evenings from 6-7 PM EST. Since we have an active Slack group, live sessions may not occur on a regular basis.

Zoom link to be posted on the blackboard.

Tentative Format of live meetups:

Activity	Estimated Time
Announcements and/or Important Information	6:00-6:15
Student Group Collaboration	6:15:-7:00

The structure of the meetup allows for students to collaborate with one another, share ideas, and learn from each other. This structure follows a more student-to-student environment - where you will collaborate with your classmates for the majority of the session. During these live sessions, you will join a breakout room and can spend this time collaborating with classmates on the week's assignments, homework, and/or projects.

Being enrolled in an asynchronous class where the vast majority of your work is completed "off-line", can be an overwhelming experience at times. The live-sessions are here for you to connect and work with fellow classmates.

Students are expected to read and watch the week's content to further support the student-to-student collaboration.

As busy professionals, sometimes it is not possible to attend a live meetup. All meetups will be recorded and are ***not required to attend.*** However, missing a live meetup will mean missing time collaborating with classmates to work on assignments.

The main purpose of meetups will be to collaborate with classmates.

Office Hours

7 PM-8 PM EST Thursdays via Slack

While I am generally available all throughout the week via slack, office hours are a dedicated time where you can chat with me. Office hours occur every **Thursday from 7-8 PM EST** via the chat program **SLACK** - not zoom.

Office hours are where you can ask questions related to assignments, career advice, grades, and so on. If a question can't be solved via slack, a scheduled zoom call can be made available.

However, as mentioned above, I am almost always available via slack. Please DO NOT wait for office hours to ask a question. Ask your question in slack at any time of the day - myself or classmates can help answer your question - there is no need to wait for the specific time of office hours.

Accessibility and Accommodations

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University. Early planning is essential for many of the resources and accommodations provided. Please see: http://sps.cuny.edu/student_services/disabilityservices.html

Online Etiquette and Anti-Harassment Policy

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University's policies. Please see: http://catalog.sps.cuny.edu/content.php?catoid=2&navoid=205

Academic Integrity

Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see: https://sps.cuny.edu/about/dean/policies/academic-and-student-policies/academic-integrity

Student Support Services

If you need any additional help, please visit Student Support Services: http://sps.cuny.edu/student_resources/

Recordings

Students who participate in this class with their camera on or use a profile image are agreeing to have their video or image recorded solely for the purpose of creating a record for students enrolled in the class to refer to, including those enrolled students

who are unable to attend live. If you are unwilling to consent to have your profile or video image recorded, be sure to keep your camera off and do not use a profile image. Likewise, students who un-mute during class and participate orally are agreeing to have their voices recorded. If you are not willing to consent to have your voice recorded during class, you will need to keep your mute button activated and communicate 26 exclusively using the "chat" feature, which allows students to type questions and comments live