CUNY School of Professional Studies

PROFESSOR JAMIEL SHEIKH

http://jamiel.io

http://linkedin.com/in/jamiel

jamiel.sheikh@sps.cuny.edu

ADVANCED PROGRAMING TECHNIQUES

PYTHON AND MACHINE LEARNING

FALL 2020

PLEASE READ THIS SYLLABUS CAREFULLY AND ENTIRELY.

COURSE WEBSITE

http://DATA602.COM

Version 2020.27

Keccak-256 Hash: f53cfb17b40987843439e5d3b9ff826936373bc137bb4ca86841775c11ccbfd3

Jamiel Sheikh

Digitally signed by Jamiel Sheikh
Date: 2020.08.22 05:32:38 -04'00'

SYLLABUS TABLE OF CONTENTS

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, Flask, matplotlib and sci-kit learn.
3. Understand basic machine learning usage with Python using regression and KNN algorithms.
4. Apply Python to concepts like blockchain and simple web applications that leverage data analysis.
5. Gain exposure to real world tools and techniques like SQL and MariaDB.

TOPICS COVERED BY WEEK

WEEK	TOPIC
1	Getting Started
2	Python - Basics
3	Python - Data Structures
4	Python - Functions, Conditional Statements
5	Python - Classes
6	Python - More Classes
7	Blockchain & Crypto
8	Matplotlib, Plotly
9	Flask / Web Apps
10	NumPy
11	pandas
12	Advanced pandas
13	Machine Learning w/ Scikit-learn
14	Machine Learning w/ Scikit-learn
15	Databases & SQL

DELIVERABLES

ALL REQUIRED MATERIALS IN THIS COURSE ARE AVAILABLE FOR FREE AND DO NOT REQUIRE A PURCHASE OF ANY KIND.

In this syllabus or in Blackboard, any reading or video content marked REQUIRED is subject to appear on the midterm or final. Content marked RECOMMENDED deeply enriches your understanding of the required material and better prepares you for the midterm and final exams. OPTIONAL content will not appear on the midterm or final except as extra credit questions.

These are the REQUIRED deliverables in this course:

1. **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 the course website. The assessment can be attempted only once and is not timed. The password to the assessment form will be available and pinned on Slack. The **password will be required** to access the assessment.
   Effort level estimate: 10 minutes.

2. **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.
   Effort level estimate: 30 minutes to read syllabus, 15 minutes for quiz.

3. **DATACAMP COURSES. Six DataCamp courses will be required**. Completion of all six DataCamp videos is required by the *end of* ***Week 16*** midnight EST. There

will be no extensions granted beyond this. Completion will be tracked and evaluated and must be done using your assigned CUNY email address to log into DataCamp.

A partial completion of any individual DataCamp course is equivalent to no completion for that DataCamp course. **Although completion of all DataCamp must occur by the end of week 16 for grading purposes only, completion of the videos to prepare for the midterm and final are aligned week by week in The Tracker. The onus is on you to manage your time accordingly**.

Any deadline information specified inside of the DataCamp.com website **should be ignored and is unrelated** to the official deadline which is the end of week 16 midnight EST.

Any additional DataCamp courses assigned to you beyond the six required are entirely for your benefit and are optional.

An invitation link to join the DataCamp group for this course will be posted and pinned on the Slack #data602 channel. You must use an `sps.cuny.edu` or `cuny.edu` email account to get access.
Effort level estimate: 6-12 hours in total, depending on your experience.

4. **COURSE VIDEOS. Four course videos on YouTube will be required.** Completion of the course videos is not tracked. **The videos to complete in order to prepare for the midterm and final are aligned in The Tracker.** The videos are accessible via the course website or Blackboard.
   Effort level estimate: 3-6 hours in total, depending on your experience.

5. **TEXTBOOK AND READINGS.** The popular *Think Python* book is required reading and central to this course. The link to download the book is available on the course website.

   In addition, select articles will be included in the weekly course materials. See section 13 for a list of those readings.
   Effort level estimate: 6-16 hours in total, depending on your experience.

6. **TEXTBOOK: MASTERING CORDA.** A second textbook, Mastering Corda, has **only one required chapter**. The link to the book and chapter are on the course website. **The PDF of the chapter should not be shared outside with individuals of the class.**
   Effort level estimate: 1-2 hours in total, depending on your experience.

7. **MIDTERM EXAM.** A midterm exam is required. It will be a timed test of 25-30 multiple choice questions out of a pool of 300 questions, and will be administered via Blackboard. The exam is open book and individual only. A **ten day window** will be established starting from day 1 of Week 8 for when you can *start* the midterm exam. After the window, access to the exam will close 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.

8. **FINAL EXAM.** A final exam is required. It will be a timed test of 25-30 multiple choice questions out of a different pool of 300 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 midterm exam. After the window, access to the exam will close 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 MIDTERM AND 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.

9. **PARTICIPATION VIA QUESTION OF THE WEEK.** A response to the Question of the Week (QOTW), a weekly question on Python and data science related current events and themes, will be required. All responses must be made in Blackboard's Discussion board. A QOTW will be released every week. Completion of all QOTWs must be done by the end of the 15th week but a student's initial response occurring beyond the week of the question's release receive significantly reduced

grade value. In other words, the later the response after the week the question was posted the lower the partial credit.

Effort level estimate: 10-20 minutes per QOTW

10. **PARTICIPATION VIA SLACK.** Student discussion and participation on Slack will be monitored, quantified and evaluated on a regular basis using automated and manual tools. Interacting with other students, posting thoughtful articles, tips and techniques, useful content, helping students solve assignment problems and exchanging ideas are all types of contributions that receive the highest points that translate to higher grades.

    Effort level estimate: 6-8 hours of contribution time over the course of the semester

11. **POST-COURSE ASSESSMENT.** A post-course assessment is required. The assessment questions are not graded, only whether the full assessment form was completed or not is. The link to the assessment is on the course website. The assessment can be attempted only once and is not timed. The password to the assessment form will be available and pinned on Slack. The **password will be required** to access the assessment and will only accept responses starting Week 15, closing end of Week 16.

    Effort level estimate: 10 minutes

RECOMMENDED OR OPTIONAL DELIVERABLES

12. **OPTIONAL READING**. Two texts, *Practical Data Analysis* and *Building Machine Learning Systems with Python* are optional. Content from these books may appear in optional **extra credit questions** on the midterm or final.

13. **SUPPLEMENTAL MATERIALS**. Additional materials are provided to further enhance or enrich your understanding of the core required materials. Links to the material can be found on the course website or Blackboard weekly Course Materials folders.

Material	Medium	Requirement level
Slack 101	Read	Optional
Google Python Style Guide	Read	Optional
Python Docs Section 9.1 - 9.7	Read	Recommended
Learn Python by Building A Blockchain	Watch	Recommended
Consensus Algorithms	Watch	Optional
Matplotlib Tutorial	Read	Recommended
Plotly Getting Started Guide	Read	Optional
Pyplot Tutorial	Read	Recommended
Flask Quickstart	**Read**	**Required**
Learn HTML	Read	Optional
SciPy	Read	Optional
10 Minutes to pandas	Read	Recommended

Data Analysis in Pandas	Watch	Optional
MariaDB & SQL Primer	Read	Optional

14. **RECOMMENDED UNGRADED ASSIGNMENTS.** Eight *non-graded* coding assignments will be released throughout the semester. These assignments prepare you for the midterm and final and is work you do on your own. **Assignments do not need to be submitted and will not be graded.** Discussion on Slack around the assignments is highly recommended, but **sharing of solution code is not allowed**. Instead feel free to share code hints, links to StackOverflow or code analogies. Content from the assignments will appear on the midterm or final exam.

    Each assignment comes with a bare minimum set of test cases that help provide a basic validation of your solution. How to use the test cases will be discussed at the first office hours.

    Effort level estimate: 15-30 mins (earlier assignments) to 1-3 hours (later assignments) per assignment, depending on the assignment and your experience level

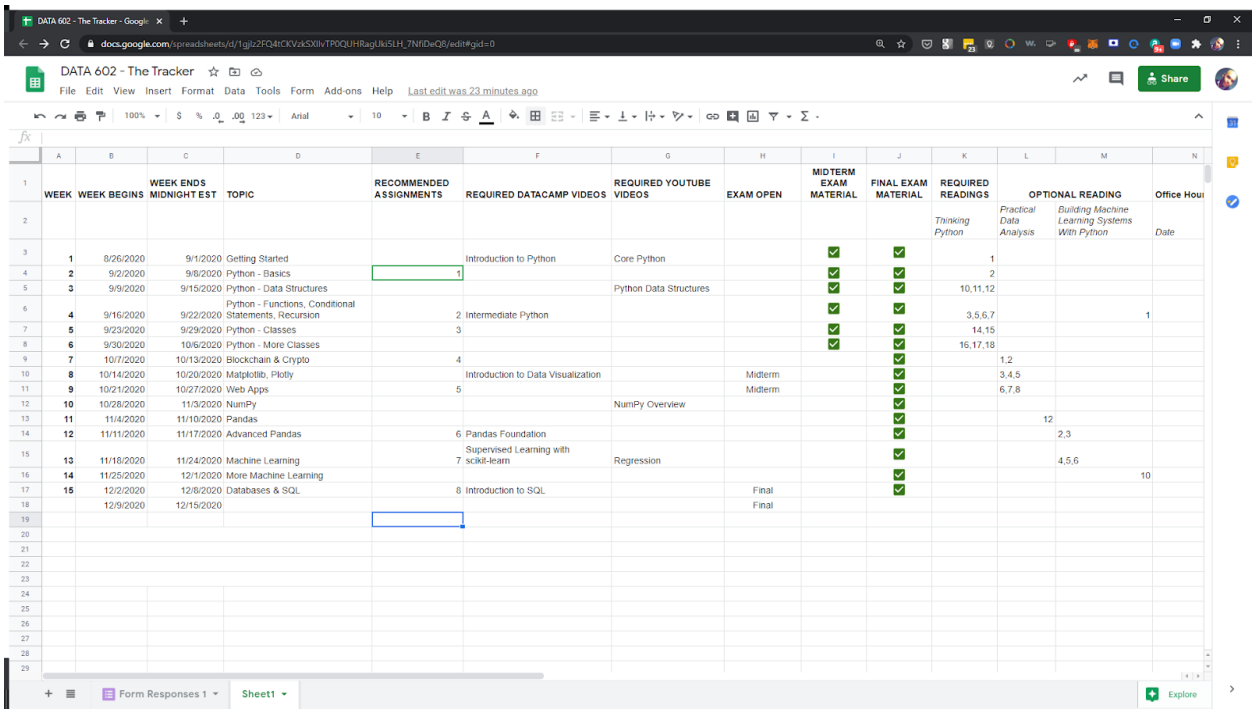
GRADE WEIGHTS

WEIGHT	DELIVERABLE	GRADE
2.5%	Pre-Course Assessment	0 = Incomplete 100 = Complete
5.0%	Syllabus Quiz	Partial Credit Min = 0, Max 100
15.0%	Completion of 6 DataCamp courses by end of week 16	*Per DataCamp Course* 0 = Incomplete, 100 = Complete
10.0%	QOTW via Blackboard	Partial Credit Min = 0, Max 100
25.0%	Participation via Slack	Partial Credit Min = 0, Max 100
20.0%	Midterm Exam	Partial Credit Min = 0, Max 110, Capped At 100
20.0%	Final Exam	Partial Credit Min = 0, Max 120, Capped At 100
2.5%	Post-Course Assessment	0 = Incomplete 100 = Complete

THE TRACKER

The Tracker is a dashboard Google spreadsheet that aligns what needs to be done week by week, giving you a bird's eye view and roadmap of the course. The Tracker should allow you to think about how you need to pace yourself and correlate all of a week's deliverables and readings, **tying together other deliverables to what you can expect on the midterm and final** exams.

The link to the The Tracker can be found on the course website.



WEEK	WEEK BEGINS	WEEK ENDS MIDNIGHT EST	TOPIC	RECOMMENDED ASSIGNMENTS	REQUIRED DATACAMP VIDEOS	REQUIRED YOUTUBE VIDEOS	EXAM OPEN	MIDTERM EXAM MATERIAL	FINAL EXAM MATERIAL	REQUIRED READINGS	OPTIONAL READING		Office Hou
										Thinking Python	*Practical Data Analysis*	*Building Machine Learning Systems With Python*	*Date*
1	8/26/2020	9/1/2020	Getting Started		Introduction to Python	Core Python		☑	☑	1			
2	9/2/2020	9/8/2020	Python - Basics	1				☑	☑	2			
3	9/9/2020	9/15/2020	Python - Data Structures			Python Data Structures		☑	☑	10,11,12			
4	9/16/2020	9/22/2020	Python - Functions, Conditional Statements, Recursion	2	Intermediate Python			☑	☑	3,5,6,7		1	
5	9/23/2020	9/29/2020	Python - Classes	3				☑	☑	14,15			
6	9/30/2020	10/6/2020	Python - More Classes					☑	☑	16,17,18			
7	10/7/2020	10/13/2020	Blockchain & Crypto	4					☑		1,2		
8	10/14/2020	10/20/2020	Matplotlib, Plotly		Introduction to Data Visualization		Midterm		☑		3,4,5		
9	10/21/2020	10/27/2020	Web Apps	5			Midterm		☑		6,7,8		
10	10/28/2020	11/3/2020	NumPy			NumPy Overview			☑				
11	11/4/2020	11/10/2020	Pandas						☑		12		
12	11/11/2020	11/17/2020	Advanced Pandas	6	Pandas Foundation				☑			2,3	
13	11/18/2020	11/24/2020	Machine Learning	7	Supervised Learning with scikit-learn	Regression			☑			4,5,6	
14	11/25/2020	12/1/2020	More Machine Learning						☑				10
15	12/2/2020	12/8/2020	Databases & SQL	8	Introduction to SQL		Final		☑				
	12/9/2020	12/15/2020					Final						

LIVE SESSIONS (AKA OFFICE HOURS)

Class office hours will be conducted and announced as needed, conducted *roughly* every other 2-3 weeks and will be held on Thursday evenings at 7PM EST. Since we will have an active Slack group, office hours sessions may not occur on a regular basis. We will be using Blackboard's Collaborate tool to conduct office hours. All sessions will be recording and will be available for playback. **Material covered in these sessions may appear on the midterm or final and is fair game.** Although attendance is not required, attendance is tracked and used to enhance, potentially significantly, grades in the participation component of your final grade. This is especially important if you are lacking in participation in Slack.

To be useful to all participants, office hours should be used to discuss issues that cannot be resolved via Slack and should address questions and issues that are broad instead of specific to one person (i.e. *"What is OOP?"* is a valid question instead of *"Why doesn't Python run on MY computer?"*). If you have an installation problem or are getting a Python error you can't resolve then that might not be the best topic to raise on a call with the rest of the class. However, if you have questions on the requirements for the assignments that all participants can benefit from then please do bring it up.

The first office hours will be held Thursday night of Week 1 at 7PM EST via Blackboard Collaborate, unless announced otherwise on Slack. Please make sure you attend.

OFFICE HOURS RECORDINGS

All office hours will be audio recorded using Blackboard Collaborate. Recordings are usually available within a few hours after the event in Collaborate.

ANNOUNCEMENTS

All announcements will be made and pinned on the Slack #data602 channel. Blackboard's announcements page **will not** be used in this course. It is your responsibility to check Slack announcements regularly. For CRITICAL announcements, for example a change in the syllabus, an email will be sent to your CUNY account.

PARTICIPATION

Participation in this class is a big part of your grade and has at least three vectors:

1. Slack
2. QOTW
3. Office hours

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.

PARTICIPATION RUBRIC

QOTW	Qualitative: Provide useful, insightful, current, relevant and discussion provoking responses to the question of the week Quantitative: Number of posts made in the QOTW discussion areas
Slack	Help resolve issues raised by other students Discuss current events in the context of the course Solve pop trivia

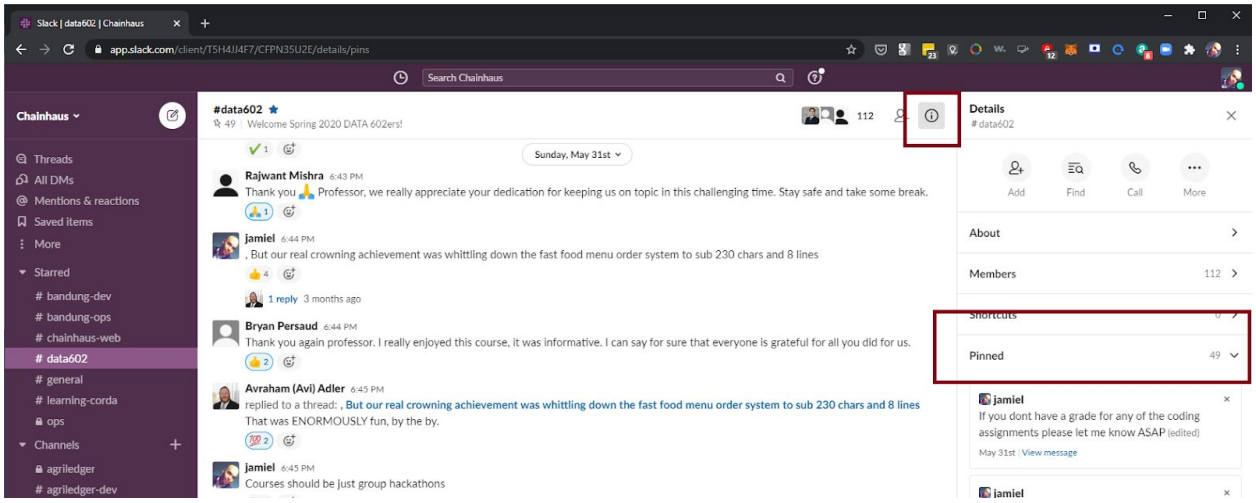
SLACK

The link to join Slack will be on the course website. Registration into Slack will require your CUNY email address. **Please make sure to join the #data602 channel after joining Slack. The default channel will be #general.**

CHANNEL	PURPOSE
#data602	Channel dedicated to this course
#data602-grads	Alum of the DATA 602 course
#general	General discussions related to a number of subjects
#gigs, #jobs	Career-oriented for sharing job openings between classmates

PINNED MESSAGES

Because Slack can involve a lot of chatter and textual content, important information can be buried quickly. To resolve this, any important information will be **pinned to the #data602 channel**. To view pinned information, from #data602 clock on the information icon () at the top right and expand the Pinned section on the right, as depicted below in red rectangles.


Slack | data602 | Chainhaus
app.slack.com/client/T5H4JJ4F7/CFPN35U2E/details/pins
Search Chainhaus
Chainhaus
Threads
All DMs
Mentions & reactions
Saved items
More
Starred
bandung-dev
bandung-ops
chainhaus-web
data602
general
learning-corda
ops
Channels
agriledger
agriledger-dev
#data602
49 | Welcome Spring 2020 DATA 602ers!
112
Sunday, May 31st
Rajwant Mishra 6:43 PM
Thank you Professor, we really appreciate your dedication for keeping us on topic in this challenging time. Stay safe and take some break.
jamiel 6:44 PM
, But our real crowning achievement was whittling down the fast food menu order system to sub 230 chars and 8 lines
1 reply 3 months ago
Bryan Persaud 6:44 PM
Thank you again professor. I really enjoyed this course, it was informative. I can say for sure that everyone is grateful for all you did for us.
Avraham (Avi) Adler 6:45 PM
replied to a thread: , But our real crowning achievement was whittling down the fast food menu order system to sub 230 chars and 8 lines
That was ENORMOUSLY fun, by the by.
jamiel 6:45 PM
Courses should be just group hackathons
Details
data602
Add
Find
Call
More
About
Members 112
Shortcuts
Pinned 49
jamiel
If you dont have a grade for any of the coding assignments please let me know ASAP (edited)
May 31st | View message
jamiel

SOFTWARE

To study the required texts and work on the optional 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.

Link	Notes
http://python.org	The official Python website where it can be downloaded from. Use the 64-bit version.
http://repl.it	**Recommended** - 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

If you wish to use an editor for your Python code, the following editors are some options

Link	Notes
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

DEFINITION OF “SCHOOL WEEK” OR “WEEK *N*”

Deadlines are established by school week and not calendar week. A school week is defined in The Tracker under columns A, B and C. For example, if the first day of class this semester is Tuesday then the end of school week is the following Monday 11:59PM EST. Week of the course is the first day of the semester + 6 days. All deadlines for deliverables in this course are based on school week and not calendar week.

All times are stated in EST.

MIDTERM AND FINAL EXAM QUESTIONS

Questions for the midterm and final exam will be sourced from content in approximate proportions described below. Understanding the concepts and usage of the concepts taught in these sources is crucial. *The proportions listed below are indicative and may not be the exact precise breakdown on the real exam.* Extra credit questions allows you to score above 100% for each exam but each exam is capped at 100%, therefore the extra credit is used to compensate for questions you may have answered incorrectly.

Source	**Midterm**	**Final**
Think Python	30%	20%
6 x DataCamp, 4 x Course Videos, Office Hours	30%	30%
Assignments	30%	40%
All Other Required and Recommended Course Content	5%	10%
Extra Credit (Optional Materials)	10%	20%

DISALLOWED COMMUNICATIONS

Please note, unless for personal, health or emergency reasons, **no direct emails or Slack DMs (private chat messages) to the professor is permitted**. This is to discourage students from obtaining special access to information and to make the course more transparent and fair for all.

Any and all questions should be raised **publicly** on Blackboard or Slack so that all students can benefit from the response, also providing other students an opportunity to gain participation points if they respond.

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/

-- END OF SYLLABUS --