جامعة نيويورك أبوظبي

NYU | ABU DHABI

Course Title	**SOCSC-UH 2214: Applied Data Science (4 credits)**
Professor	**Bedoor AlShebli**
Term	**Fall 2023**
Lecture Days	**Tuesday and Thursday**
Lecture Times	**Sec 1: 12:45 pm - 2:00 pm, Sec 2: 2:10 pm - 3:25 pm**
Lecture Locations	**Sec 1: A2-006, Sec 2: A2-006**
Recitation Days	**Wednesday**
Recitation Times	**Sec 1: 12:45 pm - 2:00 pm, Sec 2: 2:10 pm - 3:25 pm**
Recitation Locations	**Sec 1: C2-W006, Sec 2: A2-006**
Prerequisites:	**SOCSC-UH 1010Q: Statistics for the Social and Behavioral Sciences**
Cross-lists:	**Economics (Data Science and Econometrics track), Political Science (Methods elective) and SRPP (Methods elective)**

Contact Details:

	Professor	**Teaching Assistant**
Name	Bedoor AlShebli	Sara Saboor
Email	bedoor@nyu.edu	ss17229@nyu.edu
Workspace	A5 - 1119	A5 - 1172A
Office Hours	Thurs 3:30-4:30 pm	Tues 3:30-4:30 pm

جامعة نيويورك أبوظبي

NYU | ABU DHABI

Course Description:

This will be an applied course that will introduce students to the python programming environment. It is intended for students who want to apply statistical analysis, information visualization, machine learning, text analysis, and data collection techniques through popular python toolkits such as scipy, numpy, pandas, matplotlib, seaborn, beautiful soup, scikit-learn, nltk, and more to gain insight into any data.

By the end of this course, students will be able to: (1) take any tabular data, clean it, manipulate it, and run inferential statistical analyses, (2) identify best practices in data visualizations, (3) identify the difference between supervised (classification) and unsupervised (clustering) techniques, and identify which technique they need to apply for a particular dataset and need, as well as, engineer features to meet that need, (4) be able to perform basic text mining and text manipulation, and (5) be able collect large amounts of data via some form of automation.

This course requires basic competency in statistics, and preferably some very basic programming knowledge (in any language). The basic programming knowledge expected involves knowing how to print statements, assign values to variables, write an if-statement, and write for and while loops. Students with no prior programming experience will be provided with a short (ungraded) assignment at the beginning of the course covering all the basic knowledge they expect to have in order to continue with the course. Students are also not expected to have any prior knowledge in machine learning algorithms either.

Course Learning Outcomes and link to Program Learning Outcomes:

Course Learning Outcome	**Link to Major PLOs**	**Level of Contribution to PLO**
1. Develop an understanding of Python fundamentals	Econ 6, Pol Sci 5	Medium, High
2. Gain practical Python skills and apply them to data analysis	Econ 6, Pol Sci 5	High, High
3. Obtain, clean/process, and transform data	Econ 6, Pol Sci 5	High, High
4. Communicate data insights effectively through data visualizations	Econ 2, Econ 6, Pol Sci 5, Pol Sci 6, SRPP 8	High, High, High, High, High
5. Interpret data findings effectively orally, visually, and in written formats	Econ 1, Econ 2, Econ 3, Pol Sci 5, Pol Sci 6, SRPP 8	High, High, High, High, High, High

جامعة نيويورك أبوظبي

NYU | ABU DHABI

6. Effectively execute projects demonstrating an understanding of data science techniques and tools using Python.	Econ 1, Econ 2, Econ 3, Pol Sci 5, Pol Sci 6, SRPP 8	High, High, High, High, High, High

Teaching and Learning Methodologies:

This is a lecture-based course, however, students are expected to bring their laptops to perform hands-on exercises during the lecture. There will also be a recitation / lab once a week that will provide material complementary to that covered in the lectures, as well as exercises that will help solidify the student's understanding of concepts taught in the lectures.

Course Materials:

- Python for Data Analysis by Wes McKinney (O'Reilly). Copyright 2017 Wes McKinney, ISBN: 978-1-491-95766-0.

Recommended Books:

- Learning the Pandas Library, by Matt Harrison, ISBN: 978-1-533-59824-0.
- The Functional Art: An Introduction to Information Graphics and Visualizations, by Alberto Cairo, ISBN: 978-0-321-83473-7.
- The Visual Display of Quantitative Information, by Edward Tufte. ISBN: 978-0-961-39214-7.
- Introduction to Machine Learning with Python: A Guide for Data Scientists By Andreas C. Müller and Sarah Guido. ISBN: 978-1-449-36941-5

Graded Activities:

Activity	Grade Percentage
Class Participation	5 %
Recitation Participation	5 %
Problem Sets	66 %
Final Project	24%

Assignments:

Class Attendance and Participation: Full attendance in lectures is mandatory. Every lecture absence needs to be cleared with the professor prior to class. Every unexcused lecture absence (not pre-cleared with instructor, or not with medical documentation provided to instructor) will automatically result in a 0.5% reduction from the 5% final class participation grade.

Class participation will be allocated based on attendance as well as in-class contributions, which includes asking and answering questions, but mainly involvement with the in-class exercises. The

students will receive feedback on their class participation and performance near the midpoint of the semester for information purposes but can request an update on their participation grade (or any other grading component) any time during office hours.

Recitation Attendance and Participation: Full attendance in recitations is mandatory. Every recitation absence needs to be cleared with the professor prior to class. Every unexcused absence (not pre-cleared with instructor, or not with medical documentation provided to instructor) will automatically result in a 0.5% reduction from the 5% final recitation / lab grade. Recitation final grade will be based on attendance, as well as the correctness of the submission of the in-lab exercises.

Problem Sets: Throughout the course, students will be asked to submit 11 problem sets covering various topics in applied data science. Each problem set will be worth 6% of the final grade. Problem sets are expected to be completed **individually** by each student. Each problem set is given at least one week to solve. All timings related to problem sets' release and submission dates are given in advance in the class schedule below. No extensions will be granted without a strong reason. Students are expected to plan their schedule in advance to accommodate the submissions.

Final Project: There will be no exams in this course, however 24% of the final grade will go towards a class project. The project will be done in groups of 3-4 students. There will be 4 milestone submissions throughout the semester, with the aim of completing a data science project from start to finish by the end of the semester. The four milestones will be as follows:

Milestone	Grade Breakdown	Released	Due
1- Project Hypothesis and Data Description	4%	end of week 6	end of week 8
2- Exploratory Data Visualizations	5%	end of week 8	end of week 10
3- Preliminary Results and Machine Learning	5%	end of week 10	end of week 13
4- Final Project Report and Presentations	10%	end of week 13	end of week 14

Grade Distribution: Grades are generally not curved. Course percentages will be translated into letter grades based on these intervals:

A	**A-**	**B+**	**B**	**B-**	**C+**	**C**	**C-**	**D+**	**D**	**F**
[95,100]	[90,94]	[87,89]	[84,86]	[80,83]	[77,79]	[74,76]	[70,73]	[65,69]	[60,64]	[0,59]

جامعة نيويورك أبوظبي

NYU | ABU DHABI

Distribution of Class Materials: Problem sets and Recitations are intellectual property of NYUAD, and we request the students to **not distribute** them or their solutions to other students who have not signed up for this class, and/or intend to sign up in the future. We also request you don't post these problem sets and recitations online or on any public platforms.

Academic Integrity:

Compliance with and respect for academic integrity as defined by NYU and NYUAD policy is expected and enforced. Note NYUAD's community commitment to integrity:

> At NYU Abu Dhabi, a commitment to excellence, fairness, honesty, and respect within and outside the classroom is essential to maintaining the integrity of our community. By accepting membership in this community, students, faculty, and staff take responsibility for demonstrating these values in their own conduct and for recognizing and supporting these values in others. In turn, these values create a campus climate that encourages the free exchange of ideas, promotes scholarly excellence through active and creative thought, and allows community members to achieve and be recognized for achieving their highest potential. As part of the NYU global network, NYUAD students are also subject to NYU's all-school policy on Academic Integrity for Students at NYU.

Alleged integrity violations are resolved using NYUAD's Academic Integrity Procedure: https://students.nyuad.nyu.edu/campus-life/student-policies/community-standards-policies/academic-integrity/

Moses Center for Student Accessibility (CSA):

New York University is committed to providing equal educational opportunity and participation for students with disabilities. CSA works with NYU students to determine appropriate and reasonable accommodations that support equal access to a world-class education. Confidentiality is of the utmost importance. Disability-related information is never disclosed without student permission.

If you have any questions or would like to have further information about this, please visit the following link:

https://www.nyu.edu/students/communities-and-groups/students-with-disabilities.html

Contact: mosescsd@nyu.edu

Mental Health Resources:

As a University student, you may experience a range of issues that can interfere with your ability to perform academically or impact your daily functioning, such as: heightened stress; anxiety; difficulty concentrating; sleep disturbance; strained relationships; grief and loss; personal struggles. If you have any well-being or mental health concerns please visit the Counseling Center on the ground floor of the campus center from 9am-5pm Monday – Friday, or schedule an appointment to meet with

جامعة نيويورك أبوظبي

NYU | ABU DHABI

a counselor by calling: 02-628-8100, or emailing: nyuad.healthcenter@nyu.edu. If you require mental health support outside of these hours call NYU's Wellness Exchange hotline at 02-628-5555, which is available 24 hours a day, 7 days a week. You can also utilize the Wellness Exchange mobile chat feature, details of which you can find on the student portal. If you need help connecting to these supports please contact me directly.

جامعة نيويورك أبوظبي

NYU | ABU DHABI

Course Schedule:

Lecture	Topic	Optional Readings and References	Submissions
Tues Aug 29	**Lecture 1:** Course Introduction + Python Functions	Donoho (2017)	**PS0 released:** Warm up (optional)
Wed Aug 30	**Recitation 1 (optional):** Basics in Python (PS0)		
Thurs Aug 31	**Lecture 2:** Python Types and Sequences	Python for Data Analysis: Appendix A	
Tues Sept 5	**Lecture 3:** Python Special Functions	Python for Data Analysis: Appendix A	
Wed Sept 6	**Recitation 2:** Python Functions, Types, and Sequences		**PS1 released:** Python Functions, Types, and Sequences
Thurs Sept 7	**Lecture 4:** Numerical Python Library (NumPy) I	Python for Data Analysis: Ch 4	
Tues Sept 12	**Lecture 5:** Numerical Python Library (NumPy) II	Python for Data Analysis: Ch 4	
Wed Sept 13	**Recitation 3:** Numerical Python Library (NumPy)		**PS1 submission** **PS2 released:** NumPy
Thurs Sept 14	**Lecture 6:** Text Manipulation using Regular Expressions I	Regular expressions	
Tues Sept 19	**Lecture 7:** Text Manipulation using Regular Expressions II	Regular expressions	
Wed Sept 20	**Recitation 4:** Regular Expressions		**PS2 submission** **PS3 released:** Regex
Thurs Sept 21	**Lecture 8:** Pandas Library: Series	Python for Data Analysis: Ch 5	
Tues Sept 26	**Lecture 9:** Pandas Library: DataFrames	Python for Data Analysis: Ch 5	

جامعة نيويورك أبوظبي

NYU | ABU DHABI

Wed Sept 27	**Recitation 5:** Pandas I		**PS3 submission** **PS4 released:** Pandas
Tues Oct 3	**Lecture 10:** Pandas Library: Data Manipulation I	Python Data for Analysis: Ch 7	
Wed Oct 4	**Lecture 11:** Pandas Library: Data Manipulation II	Python Data for Analysis: Ch 9	
Thurs Oct 5	**Recitation 6:** Pandas II		**PS4 submission** **PS5 released:** More Pandas
Tues Oct 10	**Lecture 12:** Statistical Testing using SciPy Library + Project Discussion	Science isn't Broken: p-hacking	**Project Milestone 1 released**
Wed Oct 11	**Lecture 13:** Principles of Information Visualization	The Functional Art, by Albert Cairo The Visual Display of Quantitative Information The 5 Graph Algorithms you should know Examples of interesting and interactive visuals	
Thurs Oct 12	**Lecture 14:** Visualizing Data with Matplotlib I	Matplotlib Ten simple rules for better figures	
Tues Oct 17	**Recitation 7:** Statistical Testing and Basic Visualization	Selecting the number of bins in a histogram: a decision theoretic approach	**PS5 submission** **PS6 released:** Visualization
Oct 18 - Oct 25	**Break (No Classes)**		
Thurs Oct 26	**Lecture 15:** Project Milestone 1 Presentations		**Project Milestone 1 submission** **Project Milestone 2 released**

جامعة نيويورك أبوظبي

NYU | ABU DHABI

Sat Oct 28	**Recitation 8 (optional):** Assistance with project		
Tues Oct 31	**Lecture 16:** Visualizing Data with Matplotlib II		
Wed Nov 1	**Lecture 17 :** Visualizing Data with Pandas and Seaborn Libraries	Spurious correlations	
Thurs Nov 2	**Recitation 9:** Visualizations		**PS6 submission** **PS7 released:** More on Visualizations
Tues Nov 7	**Lecture 18:** Fundamentals of Applied Machine Learning: SciKit Learn Library	Introduction to Machine Learning with Python: A Guide for Data Scientists	
Wed Nov 8	**Lecture 19:** Supervised Machine Learning I + Project Milestone 2 Presentations	The foundations of algorithmic bias A few useful things to know about machine learning	**Project Milestone 2 submission** **Project Milestone 3 released**
Thurs Nov 9	**Recitation 10:** Machine Learning I		**PS7 submission** **PS8 released:** Machine Learning
Tues Nov 14	**Lecture 20:** Project Milestone 2 Presentations cntd		
Wed Nov 15	**Lecture 21:** Supervised Machine Learning II	Practical guide to controlled experiments on the web	
Thurs Nov 16	**Lecture 22:** Supervised Machine Learning III		
Tues Nov 21	**Recitation 11**: Machine Learning II		**PS8 submission** **PS9 released**: Machine Learning II

جامعة نيويورك أبوظبي

NYU | ABU DHABI

Wed Nov 22	**Lecture 23:** Supervised Machine Learning IV	Neural networks made easy Deep learning in a nutshell	
Thurs Nov 23	**Recitation 12:** Machine Learning III		
Tues Nov 28	**Lecture 24:** Unsupervised Machine Learning		**PS9 submission** **PS10 released:** Machine Learning III
Tues Dec 5	**Lecture 25:** Project Milestone 3 presentations		**Project Milestone 3 submission** **Project Milestone 4 released**
Wed Dec 6	**Lecture 26:** Web Scraping		
Thurs Dec 7	**Recitation 13:** Web Scraping		**PS10 submission** **PS11 released:** Web Scraping
Tues Dec 12	**Recitation 14 (optional):** Assistance with Project		
Wed Dec 13	**Lecture 27:** Milestone 4: Final Project Presentations I		**Project Milestone 4 presentation submission**
Thurs Dec 14	**Lecture 28:** Milestone 4: Final Project Presentations II		**PS11 submission**
Friday Dec 15	**Project Milestone 4 Final Report submission**		