MSDS-631
Intro to Data Science – Spring 2019

Class Time and Location: W 6:30 – 9:20PM in Lo Schiavo Science 303
Instructor: Jason Shu   **E-Mail**: jwshu2@usfca.edu

Office Hours: Tu 6:30 – 8:30PM in Harney 122E

Course Overview: This course intends to provide the foundational knowledge needed to perform data modeling and analysis for real-world application. In one semester, we will only be able to scratch the surface of Data Science, so my objectives are three-fold: 1) to give you the tools that will allow you to bring Data Science principles into your chosen area of profession or research, 2) to up-level your technical skills in preparation for future job searches, and 3) to help you understand the **when and why** to do things rather than just memorizing the **how and what**.

Each of you come from a unique background where the necessary tools will differ. Social Scientists use a great deal of linear modeling to generate interpretable analyses of what variables affect outcomes. Econometricians use a heavy dose of time series modeling. As such, part of the semester will be about researching the tools most relevant to your respective fields of study (or future career).

By the end of this semester, you will have a comfortable understanding of how to use the Python programming language to visualize, wrangle, manipulate, and explore data of many types and sizes.

Course Website: https://github.com/jasonshu8/MSDS_631

Required Textbook: None! Since this course will cover so many topics, we will pull from the many resources available on the web. We will discuss a few books that are useful for each subject, but I will not require you to purchase any of them. That being said, some of you may want to have some books to reference, so here are a few texts that I've found useful:
- "Python for Data Analysis" by Wes McKinney
- "Introduction to Statistical Learning" by Witten, James, Tibshirani, and Hastie
- "The Visual Display of Quantitative Information" by Edward Tufte

What you are required to bring to class: Please bring a laptop and your phone to every class. This will be a highly interactive course, and having these at your disposal will allow you to participate.

Grades will be assigned according to the following assessment.

- Participation and Professionalism (15%): As just mentioned, this will be a highly interactives class. Students are expected to be punctual and active in class. Activity may come in the form of questions, answers, ideas, demonstrations, or any other communication that brings value to the discussion.
- Assignments (50%): Assignments will be made available each week after class. You are required to upload your homework files to the Canvas site, and unless otherwise noted in the assignment's instructions, files must be electronically submitted **before class begins** the following week. Assignments with timestamps after class begins will not be accepted – no exceptions. Additional grading guidelines can be found below.

- Quizzes (25%): Quizzes will be given periodically throughout the semester to reinforce your learning of the most critical topics covered.
- Final Project (10%): The final project will be a group project that allows you to research Data Science techniques that are relevant to your future professional or academic pursuits.

Participation and Professionalism: Classroom lectures will be a fundamental element of you learning. We will collectively discuss new material each week and partake in various lab exercises, including the sharing many of our homework solutions. All students are expected to participate in all discussions to enrich the conversation. Missing class means missing out on being part of the interaction. Showing up late or having private conversations with classmates interferes with the common discussion. Please be courteous to your fellow classmates and be present, both physically and mentally.

Assignments: Assignments are designed to give you practice so that you can gradually build your Data Science skills. Part of the assignments will be applications of the material discussed during the previous lecture. Additional problems will be assigned in lieu of case studies and will test your ability to identify the correct methods to use to solve real problems. These problems may require using techniques learned at any point during the semester.

Assignments are also designed to force you to apply problem-solving processes to completion. Errors in logic are expected. Errors in logic give us opportunities to identify misunderstandings. Errors in logic will be discussed in class. Errors in syntax do not give us the ability to learn, nor does throwing up your hands and saying "I don't know how to do this." As such, your code **must run from beginning to end and answers must reflect a complete thought**. You will have the opportunity to re-submit your assignments with fixes to logic errors. You will not be allowed to re-submit syntax errors or problems that you did not attempt to completion.

We will cover a tremendous range of technical topics in this class. Just like during lectures, I hope that you and your classmates share your collective knowledge outside of class to explain concepts to one another in ways that enable a deeper and more intuitive understanding. That being said, you should work on your assignments without help from your classmates. This will prepare you to pass each quiz. It is easy to find yourself getting stuck and asking your classmates how to solve a problem. Powering through and figuring things out on your own will be critical to achieving true understanding of the concepts covered. Half of learning comes from understanding your mistakes. To alleviate the pressure that grades can have and to encourage you to work on your assignments alone, you will be allowed to resubmit graded assignments with corrected version of your answers.

Resubmittals are only eligible for assignments submitted on-time and with good-faith effort. This means that neither late assignments nor unexecutable code will be eligible for resubmittal. "Good-faith" effort (i.e. completing thoughts/solutions to a problem) is solely at my discretion.

Quizzes: The majority of Data Science skills can be looked up online. Memorizing *everything* is both impossible and unnecessary. However, there are a few things that require a deep understanding that enable you to learn more effectively. Quizzes will test your ability to apply learning. I will provide cheat sheets for items I consider pure memorization.

Final Project: As mentioned earlier, this class will only scratch the surface of Data Science topics. You will be expected to learn tools and techniques that are relevant to your future professional or academic pursuits. The final project will be your opportunity to collaborate with classmates with similar futures to

your own. Teams of up to three will be formed to write a research paper about a relevant topic or technique. Additionally, you will present your learnings so that others can learn from your research.

Possible final project topics include A/B Testing, Time Series Analysis, Data Visualization packages such as Seaborn and Dash, Conjoint Analysis, Monte Carlo Simulations, Principal Component Analysis, Clustering, Linear Programming, Google Analytics

Important Dates:

- Monday, 1/28/19 - Last day to add
- Friday, 2/8/19 - Census date. Last day to withdraw with tuition reversal
- Wednesday, 3/13/19 - Spring Break (no class)
- Monday, 4/8/19 - Last day to withdraw
- Wednesday, 5/8/19 - Last day of class!

Academic Integrity

As a Jesuit institution committed to cura personalis - the care and education of the whole person - USF has an obligation to embody and foster the values of honesty and integrity. USF upholds the standards of honesty and integrity from all members of the academic community. All students are expected to know and adhere to the University's Honor Code. You can find the full text of the code online at www.usfca.edu/academic integrity. The policy covers:

- Plagiarism: intentionally or unintentionally representing the words or ideas of another person as your own; failure to properly cite references; manufacturing references.
- Working with another person when independent work is required.
- Submission of the same paper in more than one course without the specific permission of each instructor.
- Submitting a paper written by another person or obtained from the internet.
- The penalties for violation of the policy may include a failing grade on the assignment, a failing grade in the course, and/or a referral to the Academic Integrity Committee.

Students with Disabilities

If you are a student with a disability or disabling condition, or if you think you may have a disability, please contact USF Student Disability Services (SDS) at 415 422-2613 within the first week of class, or immediately upon onset of disability, to speak with a disability specialist. If you are determined eligible for reasonable accommodations, please meet with your disability specialist so they can arrange to have your accommodation letter sent to me, and we will discuss your needs for this course. For more information, please visit: http://www.usfca.edu/sds or call (415) 422-2613.

Behavioral Expectations

All students are expected to behave in accordance with the Student Conduct Code and other University policies (see http://www.usfca.edu/fogcutter/). Open discussion and disagreement is encouraged when done respectfully and in the spirit of academic discourse. There are also a variety of behaviors that, while not against a specific University policy, may create disruption in this course. Students whose behavior is disruptive or who fail to comply with the instructor may be dismissed from the class for the remainder of the class period and may need to meet with the instructor or Dean prior to returning to

the next class period. If necessary, referrals may also be made to the Student Conduct process for violations of the Student Conduct Code.

Learning & Writing Center
The Learning & Writing Center provides assistance to all USF students in pursuit of academic success. Peer tutors provide regular review and practice of course materials in the subjects of Math, Science, Business, Economics, Nursing and Languages. https://tutortrac.usfca.edu. Students may also take advantage of writing support provided by Rhetoric and Language Department instructors and academic study skills support provided by Learning Center professional staff. For more information about these services contact the Learning & Writing Center at (415) 422- 6713, email: lwc@usfca.edu or stop by our office in Cowell 215. Information can also be found on our website at www.usfca.edu/lwc.

Counseling and Psychological Services

Our diverse staff offers brief individual, couple, and group counseling to student members of our community. CAPS services are confidential and free of charge. Call 415-422-6352 for an initial consultation appointment. Having a crisis at 3 AM? We are still here for you. Telephone consultation through CAPS After Hours is available between the hours of 5:00 PM to 8:30 AM; call the above number and press 2.

Confidentiality, Mandatory Reporting, and Sexual Assault

As an instructor, one of my responsibilities is to help create a safe learning environment on our campus. I also have a mandatory reporting responsibility related to my role as a faculty member. I am required to share information regarding sexual misconduct or information about a crime that may have occurred on USFs campus with the University. Here are other resources: – To report any sexual misconduct, students may visit Anna Bartkowski (UC 5th floor) or see many other options by visiting our website: www.usfca.edu/student life/safer. – Students may speak to someone confidentially, or report a sexual assault confidentially by contacting Counseling and Psychological Services at 415-422-6352. – To find out more about reporting a sexual assault at USF, visit USF's Callisto website at: www.usfca.callistocampus.org. – For an off-campus resource, contact San Francisco Women Against Rape (SFWAR) (415) 647-7273 (www.sfwar.org).

Student Accounts - Last day to withdraw with tuition reversal

Students who wish to have the tuition charges reversed on their student account should withdraw from the course(s) by the end of the business day on the last day to withdraw with tuition credit (census date) for the applicable course(s) in which the student is enrolled. Please note that the last day to withdraw with tuition credit may vary by course. The last day to withdraw with tuition credit (census date) listed in the Academic Calendar is applicable only to courses which meet for the standard 15-week semester. To find what the last day to withdraw with tuition credit is for a specific course, please visit the Online Class Schedule at www.usfca.edu/schedules.

Ability to Change Syllabus

I, Jason W. Shu, will do my best as an instructor to abide by the guidelines set forth in this syllabus throughout the year. I do, however, have the right to change components of this syllabus at my own discretion if I deem such changes as necessary.

Schedule

Week	Lecture Title	Specific Topics Covered	Graded Material
1	Introduction to Python Programming	• Introductions and Course overview • Intro to Github • Intro to iPython / Jupyter Notebooks • Coding workflow • Data types and structures • Variables and operators	• None
2	Introduction to Python Programming (cont.)	• Basic built-in methods • Control flow • Loops	• HW1 Due • Quiz #1
3	Introduction to Python Programming (cont.)	• Functions • Reading / writing data • Numerical methods	• HW2 Due • Quiz #2
4	Introduction to Python Programming (cont.) Final Project	• Numerical methods • Class objects • Overview of key libraries • Final Project Discussion	• HW3 Due • Quiz #3
5	EDA with Pandas	• Rows and columns • Dataframes and series • Basics of vectorizing • Filtering and indexing • Missing Data	• HW4 Due
6	EDA with Pandas	• Vectorizing • Aggregations • Plotting	• HW5 Due • Quiz #4 • Final Project Proposal
7	EDA with Pandas	• Joining • Pivoting • More Plotting	• HW6 Due • Quiz #5
8	Data Visualization	• Fundamentals of visualization • Interpreting data	• HW7 Due • Quiz #6
9	Data Visualization	• Matplotlib	
10	Statistics	• Fundamentals • Distributions	• HW8 Due
11	Applied Data Science	• Case studies	• Quiz #7
12	Linear Regression	• Parameter estimates • Predictive modeling	• HW9 Due
13	Applied Data Science	• Case studies	• Quiz #8
14	Logistic Regression and Intro to Machine Learning	• Classification vs. regression • Parameter estimates • Classification diagnostics	• HW10 Due
15	Machine Learning	• Various models • Hyperparameters	• Project Paper Due
16	Survey of DS Topics	• Present project topics	• Project Presentation Due