Course Title:	**Intro to Social Media Analytics using Python (aka "Big Data & Media Analytics")**
Course Number:	MEJO 490.1
Instructor:	Prof. Deen Freelon, Ph.D.
Time:	MW 1230pm – 145pm
Room:	Carroll 011
Office hours:	M 2pm – 5pm and by appt
My office:	Carroll 380
My email:	freelon@email.unc.edu
Course website:	https://sakai.unc.edu/portal/site/mejo490.001.sp19

Course introduction

Millions of people use social media every day. Making sense of all that content is one of the great challenges of the early 21st century. The skills needed to conduct such research are prized not only by social science researchers like me, but also by companies who want to know how customers are responding to their products and nonprofits interested in measuring the efficacy of their cause campaigns. Students in this course will learn computer programming skills (in the Python language) and apply them to an individual research project that analyzes, visualizes, and draws meaningful conclusions about social media data. By the end of the course they will be prepared to conduct basic programming-based social media research as well as enroll in advanced programming classes that build on these skills.

By the end of this course, you should be able to:

- Preprocess and standardize Twitter data using the Python programming language
- Conduct basic descriptive social media analysis (see below)
- Visualize research results
- Explain clearly how research findings relate to practical and/or theoretical concerns

Summary of course requirements

- Attend and participate in all class meetings – 10%[1]
- Slack – 5 %
- Four programming homework assignments – 40% (10% each)
- Final research project and presentation - 45%

Detailed course plan

This course will be split into two units whose exact lengths will depend on how quickly we cover the material in the first unit. We will begin with an intensive introduction to Python, the

---

[1] For class participation, I give all students a starting grade of 90 and adjust it upward for volunteering to speak in class and downward for unexcused absences and failing to complete in-class assignments.

programming language we will use to analyze our social media data. The idea is to help you learn the Python skills that are most important to social media analysis as quickly as possible. You will spend the second unit learning intermediate social media analytic techniques and working on your research project. Some of the second unit's class time will be devoted to independent lab work. That means you will work on your project in class and I will be available to assist you if you need help.

You will conduct your research project on a Twitter dataset which I will provide. Midway through the semester, I will introduce a group of datasets from which you will be allowed to choose on a first-come-first-serve basis. If you care about what you'll be analyzing I suggest you move quickly, as once someone chooses a dataset, it will not be available to anyone else.

The overall goal of your final research project will be to quantitatively describe and explain your dataset. The project will include three distinct quantitative analyses and/or visualizations. A series of possibilities is listed below; however, you may design your own analyses as long as you clear them with me in advance. Here is what you will turn in when you are finished:

- All the code you wrote to generate the analyses
- Any visualizations you created
- A brief (400-600 word) summary of your three analyses **in Microsoft Word format.** This summary should focus primarily on what you found, why it matters, and why you made the programming choices you did.

You will also give an 6-8 minute oral presentation of your work during our finals period during which you will use the projector to show the visuals you created.

Here are a few examples of analyses that you could conduct for the final project. I will send out more specific instructions later in the semester explaining exactly what will count for this.

- A bar chart of the top 10 most-retweeted users in your dataset showing the number of times they were retweeted
- A table of the top 10 most-mentioned users in your dataset showing the number of times they were mentioned
- A line chart showing the number of retweets, non-retweets, and total tweets per day.
- A table containing the top 5 hashtags overall AND a line chart showing the number of times each hashtag appeared over time
- A list of the top 10 most-tweeted links
- A line chart showing the number of times the top five most-linked-to web domains appeared overall and over time.

Grade key

95-100%	A
90-94%	A-

87-89%	B+
84-86%	B
80-83%	B-
77-79%	C+
74-76%	C
70-73%	C-
66-69%	D
65% and below	F

Required materials

You'll need to bring a laptop running Windows, Mac OS, or Linux to class every day (tablets, phones, and some Chromebooks won't work). We will install the Python interactive programming environment, or "shell" (which is free) on the first day, and we'll use it every class day.

There are no predefined out-of-class readings for this course. Instead, the readings will serve as reference materials for your in-class and homework assignments (first unit) and your final project (second unit). These readings will consist primarily of online materials written by me, but here are a couple other resources you might find helpful. They are both freely available online:

The Non-Programmer's Tutorial for Python 3: http://en.wikibooks.org/wiki/Non-Programmer%27s_Tutorial_for_Python_3 – this is a good general reference text written for beginners. I would advise you to take a look at it if/when you get stuck.

Pandas Tutorials: https://pandas.pydata.org/pandas-docs/stable/tutorials.html: This will help when we learn about Pandas, a Python data analysis framework we will use in the second half of the semester.

Stack Overflow: http://stackoverflow.com/ -- This is an excellent question-and-answer site that almost certainly already has the answer to whatever programming questions you may have. I strongly suggest searching for the answer rather than asking a question yourself—you might get a rather brusque response if you ask a question that's already been asked several times.

Class Slack channel

In this class everyone (including me) will participate in a Slack channel to help you understand the material. If you've never used Slack, it's a lot like an iMessage group chat, but more geared toward work and academic collaboration. We'll practice using Slack on the first day, but you'll be able to use it at any time (including class time) to ask and answer questions about the material.

Use of the class Slack channel will account for 5% of your final grade. To get full credit, you must post something relevant to the course content (e.g. a question, an answer to someone else's

question, or other helpful comment) on five different class days. None of your contributions for credit can be code—that's too easy. Also, to discourage plagiarism, you can only post one line of code at a time (not for credit, but maybe to help someone else). The easiest way to get full credit is to use the channel at least once a week, which will hopefully also help you learn the material better.

You can access the class Slack channel online at https://prof-freelon.slack.com or through the Slack app for iPhone or Android. (You'll need to create an account first.)

My classroom expectations

In this class, I expect that you will:

- Come to class prepared to engage with the day's material
- Come to class on time
- Complete all assignments on time
- Silence your mobile phone during class
- Not waste class time on electronic or online services unrelated to class.
- Speak up regularly and relevantly
- Let me know if and when you're having trouble understanding anything (feel free to do so publicly or privately)
- Not insult or belittle me or your fellow classmates
- Refrain from plagiarism and other violations of UNC's Honor Code (see below)

Additionally, given recent events, I feel it is important to clarify the bounds of class conduct and discussion in advance to reduce confusion about what is permitted and what is not. I undertake this task in the spirit of one of every university's main purposes: to distinguish between valid and invalid knowledge and judgments. Therefore, over and above UNC's official diversity statement (reproduced below), I hereby establish the following bounds of classroom conduct. All students in this class will:

- Refrain from judging individuals according to the collective groups of which they are members (e.g. race, gender, class, sexual orientation, disability status, etc.);
- Assess intellectual ideas and arguments strictly according to the evidence supporting them, and not based on the identities of the individual who created them;
- Acknowledge that due to historical and contemporary systems of oppression, allegations of racism, sexism, homophobia, ableism, etc. are not symmetrical between social groups. This means such claims can only be valid when advanced by members of a less powerful group against a more powerful group. Allegations in the opposite direction (e.g. of "reverse racism") will not be tolerated. Such notions have been definitively debunked by many strong arguments for which I am happy to provide references upon request.

By the same token, you can expect that I will:

- Come to class prepared and enthused to engage with the day's material

- Treat your personal views with respect
- Carefully explain any concepts that don't make sense
- Cultivate a civil and welcoming class environment
- Return your graded assignments within about a week
- Reward good-faith efforts to engage with course material
- Refer plagiarism and other violations of UNC's Honor Code to the proper authorities (see below)

My policies

- *Lateness and absences:* Please arrive promptly for class; lateness is disruptive and inconsiderate. Chronic lateness will count against your grade.
- *Late assignments:* Turning in your assignments on time will be absolutely critical in this class. Otherwise you will fall behind, which will jeopardize your ability to complete the final assignment. So please keep current with these.
- *Mobile phones:* These should not be used during class under any circumstances, and your ringer should be set to silent.
- *Bathroom:* Feel free to use the bathroom whenever you need to; just leave and re-enter as quietly as possible.

University Policies

The Honor Code

It is my duty to report any and all suspected Honor Code violations to the Student Attorney General. If you are not familiar with the Honor Code, please review it at http://instrument.unc.edu . As stated in the Honor Code, "It shall be the responsibility of every student at the University of North Carolina at Chapel Hill to obey and support the enforcement of the Honor Code, which prohibits lying, cheating, or stealing when these actions involve academic process or University student or academic personnel acting in an official capacity."

A special note about plagiarism: The Instrument of Student Governance at UNC defines plagiarism as "deliberate or reckless representation of another's words, thoughts, or ideas as one's own without attribution in connection with submission of academic work, whether graded or otherwise." Copying-and-pasting from online sources without citing the source from which you obtained the content is clearly an instance of plagiarism. However, it may also be plagiarism if you rely too heavily on the structure and reasoning of another piece (for example, if you rely too much on swapping out synonyms or making only very superficial changes to content that is not yours). This type of extensive paraphrasing is not acceptable in this course, which requires you to demonstrate original thinking and analysis. If you have any questions about whether your use of reference material is appropriate, please see me. If any part of your work is judged by me and an independent faculty member to reflect inappropriate use of reference material, I reserve the right to adjust assignment and course grades downwards, in addition to reporting suspected violations as described in the preceding paragraph.

Students with Disabilities

If you have a diagnosed or suspected disability that you think might affect your performance in this course, you should contact Accessibility Resources & Service to determine whether and to what extent services or accommodations are available. If you think this might apply to you, please contact Accessibility Resources & Service at 962-8300 or visit the department's Website at https://accessibility.unc.edu/ . Please understand that I'm not qualified or permitted under University policies to provide any disability-related accommodations without authorization from ARS.

Diversity

The University of North Carolina at Chapel Hill is committed to equality of educational opportunity. The University does not discriminate in offering access to its educational programs and activities on the basis of age, gender, race, color, national origin, religion, creed, disability, veteran's status, sexual orientation, gender identity, or gender expression. The Dean of Students (Suite 1106, Student Academic Services Building, CB# 5100, 450 Ridge Road, Chapel Hill, NC 27599-5100 or [919] 966-4042) has been designated to handle inquiries regarding the University's nondiscrimination policies.

Accreditation

The School of Media and Journalism's accrediting body outlines a number of values you should be aware of and competencies you should be able to demonstrate by the time you graduate from our program. Learn more about them here:

http://www2.ku.edu/~acejmc/PROGRAM/PRINCIPLES.SHTML#vals&comps

No single course could possibly give you all of these values and competencies; but collectively, our classes are designed to build your abilities in each of these areas. In this class, we will address a number of the values and competencies, with special emphasis on these:

- understand concepts and apply theories in the use and presentation of images and information;
- think critically, creatively and independently;
- conduct research and evaluate information by methods appropriate to the communications professions in which they work;
- apply basic numerical and statistical concepts;
- apply tools and technologies appropriate for the communications professions in which they work.

Course schedule (tentative)

This schedule presents the course content in the rough order I plan on presenting it. I may move faster or slower depending on how quickly students master the concepts.

Date	Unit	Assignment due
1/9	Intro/install Python	
1/14	Basic concepts 1	
1/16	Basic concepts 2	
1/21	(MLK Day; no class)	
1/23	Basic concepts 3	
1/28	String handling 1	Programming homework #1 due
1/30	String handling 2	
2/4	String handling 3	
2/6	String handling 4	
2/11	Comparisons and conditionals 1	
2/13	Comparisons and conditionals 2	Programming homework #2 due
2/18	Comparisons and conditionals 3	
2/20	Working with social media data 1	
2/25	Working with social media data 2	
2/27	Working with social media data 3	
3/4	Advanced string handling 1	
3/6	Advanced string handling 2	Programming homework #3 due
3/11	(no class; Spring Break)	
3/13	(no class; Spring Break)	
3/18	Advanced string handling 3	
3/20	Working with links 1	
3/25	Working with links 2	
3/27	Visualization 1	
4/1	Visualization 2	
4/3	Visualization 3	Programming homework #4 due
4/8	Pulling it all together 1	
4/10	Pulling it all together 2	
4/15	Lab 1	
4/17	Lab 2	
4/22	Lab 3	
4/24	Lab 4	
4/30	Final project presentations (from 12pm - 3pm)	Final project materials due @ 12