

4 units

**Fall 2023 – Section #20881**

**Monday 6:30-9:20**

**Location:** ASC 331

**Instructor: Joshua Clark**

**Office:** TBD

**Office Hours:** Friday 1:30-4:30pm

**Contact Info:** joshuaac@usc.edu

### Course Description

This course is meant to prepare students for working in a modern data and analytics team and communicate with peers and stakeholders. The goal is to allow students to build a project and portfolio, which they can show to prospective employers in order to demonstrate the skills that they have acquired thus far and communicate their value to a company. This course requires a basic familiarity with data science techniques and focuses on how to explain and communicate complex statistical concepts to both collaborators and a non-technical audience.

### Student Learning Outcomes

Throughout this course students will learn how to.

- Collect a unique dataset distinctive to their interests.
- Manipulate this data to deal with errors, missing values and other issues.
- Collaborate like a data science team, using common co-working methodologies employed in industry.
- Learn to document their own code.
- Undergo a data peer review process and integrate changes into a codebase.
- Create a final portfolio of code, data and visualizations which can be paired with a resume.

**Prerequisite(s):** There are no course requirements for this class, but students are expected to have a familiarity with Python or R and basic data science methods.

**Co-Requisite(s):** None

**Concurrent Enrollment:** None

**Recommended Preparation:** Experience with version control systems and SQL are recommended but not required.

### Course Notes

In mimicking a data organization, this course will use a private Github repository for the provision and submission of assignments. Students will be expected to have their own Github account and will leave the course with a repository that can be used as a portfolio. All lecture materials will also be made available via the course git repository.

### Technological Proficiency and Hardware/Software Required

This course requires access to a computer running Windows, MacOS or Linux. Laptops are available through the [USC Computing Center Laptop Loaner Program](#). Additionally, we will be using a number of free and open source tools including but not limited to Python.

## Laptop Policy

All undergraduate and graduate Annenberg majors and minors are required to have a PC or Apple laptop that can be used in Annenberg classes. Please refer to the [Annenberg Digital Lounge](#) for more information. To connect to USC's Secure Wireless network, please visit USC's [Information Technology Services](#) website.

## Required Readings and Supplementary Materials

There is no required textbook for this class. All readings are linked in the course list below.

## Optional Readings and Supplementary Materials

As a supplemented reference, I would recommend getting An Intro to Statistical Learning, available [here](#).

## Description and Assessment of Assignments

The core element of this class is a portfolio project. Students are expected to find a data source, work with their peers to identify a transformation or data product derived from this data source that creates value and then execute this plan.

**Pitch Deck/Summary** - A brief presentation describing the data source and problem supporting a student's project. Given early in the semester, the pitch deck should be a short, high-level presentation meant to convince the audience that this project is a good use of time and resources. It should be accompanied by a one-page cover sheet summarizing the pitch.

**Weekly Stand Up** - Following practice of data science teams, students are expected to give a short verbal summary of their progress, challenges and achievements every week. If unable to make a class, a written summary can serve as a replacement.

**Peer Review** - As students make progress on their project, they will commit code to the class's common repository. Each student will have a designated reviewer who is required to review and approve commits to the repository. Informative and timely reviews will earn credit for the reviewer.

**Final Project Assessment** - Students' final project will be assessed on two fronts:

- Quality of data product - Points will be assigned based on the scope and quality of the submitted code and data product. Work will be assessed on ease of use, readability of code, usefulness of final project and overall presentation/documentation.
- Final Presentation - Students will be expected to give a 15-20 minute presentation describing their work, the challenges and the value that their data product delivers. Presentations should assume a non-technical audience with little familiarity with the project and demonstrate the value created by the work in a clear, concise and non-confusing way. Presentations will be spread over two weeks to allow time for questions and dialogue, students presenting on the second week will have a higher expectation of polish in their work to account for the extra time.
- Final code review, the final portfolio is open to an entire class code review.

## Participation

Participation will be assessed through the stand-up meetings. Students can make up missed standups by providing written summaries of their reports, however, attendance is strongly encouraged to benefit from instruction and peer review within the classroom setting.

## Grading Breakdown

Description of assessments and corresponding points and percentage of grade.

Assessment Tool (assignments)	Points	% of Grade
Pitch Deck	10	10%
Stand Up	10	10%
Peer Review	10	10%
Final Code Review	30	30%
Final Presentation	40	40%
<b>TOTAL</b>		<b>100%</b>

## Course Grading Scale

Letter grades and corresponding point value ranges.

Letter grade and corresponding numerical point range		
94% to 100%: A	80% to 83%: B-	67% to 69%: D+
90% to 93%: A-	77% to 79%: C+	64% to 66%: D
87% to 89%: B+	74% to 76%: C	60% to 63%: D-
84% to 86%: B	70% to 73%: C-	0% to 59%: F

## Course Specific Policies

You are welcome/expected to use generative AI tools (e.g. ChatGPT, Dall-e, etc.) in this class as doing so aligns with the course learning goal of developing code for a portfolio in a way that mimics a member of a data team.. You are responsible for the information you submit based on an AI query (for instance, that it does not violate intellectual property laws, or contain misinformation or unethical content). Your use of AI tools must be properly documented and cited in order to stay within university policies on academic integrity.

Copying of code from Chat-GPT, Stackoverflow and other sources is allowed. Any code blocks taken from this sources should have a comment or link to the source to help with reproducibility.

## Classroom Policies

This class will be held to the same standards of a data science team in an industry setting. Students are expected to be punctual, professional and have a high level of respect for their colleagues while still providing constructive feedback.

## Grading Standards

Letter Grade	Description
A	<b>Solid contributions, elevates their peers and produces good quality code. Can communicate complex data processes to technical and stakeholder audiences.</b>

Letter Grade	Description
B	Good work, produces effective code and can communicate simple data processes. Work may lack scope, or there may be room for improvement in tackling more complex topics.
C	Satisfactory/Fair; reasonable knowledge and understanding of subject matter. Demonstrates basic understanding of data science principles but has not produced or communicated result effectively.
D	Marginal; minimal knowledge and understanding of subject matter; more than one significant shortcoming. Either serious issues with the final portfolio or systemic problems with the communication of data topics.
F	Failing; unacceptably low level of knowledge and understanding of subject matter; deficiencies indicate lack of understanding.

### Grading Timeline

All grades will be submitted within a week of receiving the assignment. Feedback will be provided through email and comments on the github repo.

### Assignment Submission Policy

Assignments are due by the start of class. Grade for the stand-up portion of the class will be determined by engagement within the classroom or submission of a written summary of your progress if you have to miss a class.

### Add/Drop Dates for Session 001

**(15 weeks: 8/21/2023 – 12/1/2023; Final Exam Period: 12/6-13/2023)**

Link: <https://classes.usc.edu/term-20233/calendar/>

**Last day to add:** Friday, September 8, 2023

**Last day to drop without a mark of "W" and receive a refund:** Friday, September 8, 2023

**Last day to change enrollment option to Pass/No Pass or Audit:** Friday, September 8, 2023 [All major and minor courses must be taken for a letter grade.]

**Last day to add/drop a Monday-only class without a mark of "W" and receive a refund or change to Audit:** Tuesday, September 12, 2023

**Last day to withdraw without a "W" on transcript or change pass/no pass to letter grade:** Friday, October 6, 2023 [Mark of "W" will still appear on student record and STARS report and tuition charges still apply.]

\*Please drop any course by the end of week three for session 001 (or the 20 percent mark of the session in which the course is offered) to avoid tuition charges.]

**Last day to drop with a mark of "W":** Friday, November 10, 2023

### Course Schedule: A Weekly Breakdown

A weekly schedule of the topics, readings, and deliverables for the course.

**Important note to students:** Be advised that this syllabus is subject to change - and probably will change - based on the progress of the class, news events, and/or guest speaker availability.

	Topics/Daily Activities	Readings and Homework	Deliverable/Due Dates
<b>Week 1</b> Dates: 8/21	Introduction to class, discussion of project deliverables, tech setup and bug hunting.		
<b>Week 2</b> Dates: 8/28	Understanding the modern data team. Where and how to find data.	<a href="#">Harvard Business Review: Data Scientist.</a>  <a href="#">The Ambiguity of Data Science Team Roles and the Need for a Data Science Workforce Framework</a>  <a href="#">List of open data sources</a>  <a href="#">Los Angeles Open Data</a>	Special Stand Up - Describe your career goals within the data space.
<b>Week 3</b> Dates: 9/4			[Labor Day: Monday, September 4]
<b>Week 4</b> Dates: 9/11	What is a data product? Finding data, work scoping, timelines and deliverables	<a href="#">Tidy Data</a>  <a href="#">Map Reduce Simplified Data Processing on Large Clusters</a>  <a href="#">State of Data Science 2022</a>  <a href="#">W3 SQL Reference (if you need it)</a>	Stand Up - data exploration
<b>Week 5</b> Dates: 9/18	Communication within data teams. Peer review and project management.	<a href="#">The Agile Manifesto</a>  <a href="#">Agile Data Science</a>  <a href="#">Hello World - Github</a>  <a href="#">Git and Github for beginners.</a>	Pitch Deck presentations
<b>Week 6</b> Dates: 9/25	Data visualization, dashboarding and avoiding common graphing mistakes.	Student reading assignment, find and share favorite data visualization from media or culture.	Stand Up - present visualization as part of stand up.

	Topics/Daily Activities	Readings and Homework	Deliverable/Due Dates
		<a href="#">The Grammar of Graphics</a>	
<b>Week 7</b> <b>Dates: 10/2</b>	Experiments and communication. AB Testing and data driven product development	<a href="#">Avoiding the Pitfalls of AB Testing</a>  <a href="#">Here's How Cornell Scientist Brian Wansink Turned Shoddy Data Into Viral Studies About How We Eat</a>  <a href="#">Bayesian Methods for Hackers, Chp 1 &amp; 2, PYMC Version (Optional)</a>	Stand Up
<b>Week 8</b> <b>Dates: 10/9</b>	Delivering supervised learning results. Classification, regression and understanding models.	<a href="#">SHAP Scoring Repo and Documentation</a>  <a href="#">"Why Should I Trust You?" Explaining the Predictions of Any Classifier</a>  <a href="#">Interpretable ML. Chapters 1-3</a>	Stand Up - midterm course evaluation survey.
<b>Week 9</b> <b>Dates: 10/16</b>	Clustering and unsupervised models	<a href="#">Clustering in machine learning</a>  <a href="#">Comprehensive Survey of Clustering</a>  <a href="#">Clustering. That's Cute.</a>	Stand Up - code should have initial peer review by this point at the latest.
<b>Week 10</b> <b>Dates: 10/23</b>	Machine learning, neural networks, LLMs and black box modeling.	<a href="#">Statistical Learning, the two Cultures.</a>  <a href="#">Stop explaining black box machine learning models for high stakes decisions</a>  <a href="#">Sparks of AGI (skim)</a>	Stand Up
<b>Week 11</b> <b>Dates: 10/30</b>	Working with qualitative and mixed methods researchers.	<a href="#">When User Experience Designers Partner with Data Scientists</a>	Stand Up

	Topics/Daily Activities	Readings and Homework	Deliverable/Due Dates
		<a href="#">Investigating How Experienced UX Designers Effectively Work with Machine Learning</a>	
<b>Week 12</b> <b>Dates: 11/6</b>	Homestretch hack-a-thon week. Prep for stakeholder presentations.		
<b>Week 13</b> <b>Dates: 11/13</b>	Resume and CV Workshop	<a href="#">Intro to LaTeX</a> <a href="#">Overleaf - LaTeX in 30</a> <a href="#">Github pages.</a>	Bring your current CV/Resume.
<b>Week 14</b> <b>Dates: 11/20</b>	Stakeholder Presentations pt1	Reading summaries, code and decks from other student's projects for feedback and questions.  Final Code review, entire class comments on code.	Presentation decks and errata due for early presenters.
<b>Week 15</b> <b>Dates: 11/27</b>	Stakeholder Presentations pt2	Reading summaries, code and decks from other student's projects for feedback and questions.  Final Code review, entire class comments on code.	Presentation decks and errata due. Last chance of peer review of code
<b>STUDY DAYS</b> <b>Dates: 12/2-12/5</b>			
<b>FINAL EXAM PERIOD</b> <b>Dates: 12/11</b>	Portfolio due		Complete portfolio (resume, code, deck and errata) need to be committed to github by 9pm at the latest.

## **Policies and Procedures**

### **Additional Policies**

As with any software development process students are expected to maintain security. Do not share access keys, passwords, code which is not yours or other secrets with third parties without approval from the instructor.

### **Communication**

Please do not hesitate to reach out if you cannot attend office hours or need extra help with assignments. The best route of contact is by email. Barring extenuating circumstances, emailed requests will be responded to within 24 hours.

## **Statement on Academic Conduct and Support Systems**

### **Academic Integrity:**

The University of Southern California is a learning community committed to developing successful scholars and researchers dedicated to the pursuit of knowledge and the dissemination of ideas. Academic misconduct, which includes any act of dishonesty in the production or submission of academic work, compromises the integrity of the person who commits the act and can impugn the perceived integrity of the entire university community. It stands in opposition to the university's mission to research, educate, and contribute productively to our community and the world.

All students are expected to submit assignments that represent their own original work, and that have been prepared specifically for the course or section for which they have been submitted. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s).

Other violations of academic integrity include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), collusion, knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university. All incidences of academic misconduct will be reported to the Office of Academic Integrity and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see [the student handbook](#) or the [Office of Academic Integrity's website](#), and university policies on [Research and Scholarship Misconduct](#).

The School of Communication maintains a commitment to the highest standards of ethical conduct and academic excellence. Any student found responsible for plagiarism, fabrication, cheating on examinations, or purchasing papers or other assignments will be reported to the Office of Student Judicial Affairs and Community Standards and may be dismissed from the School of Communication. There are no exceptions to the school's policy.

In addition, it is assumed that the work you submit for this course is work you have produced entirely by yourself and has not been previously produced by you for submission in another course or Learning Lab, without approval of the instructor.

### **Students and Disability Accommodations:**

USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial



appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at [osas.usc.edu](http://osas.usc.edu). You may contact OSAS at (213) 740-0776 or via email at [osasfrontdesk@usc.edu](mailto:osasfrontdesk@usc.edu).

### **Support Systems:**

#### *Counseling and Mental Health* - (213) 740-9355 – 24/7 on call

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

#### *988 Suicide and Crisis Lifeline* - 988 for both calls and text messages – 24/7 on call

The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline is comprised of a national network of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

#### *Relationship and Sexual Violence Prevention Services (RSVP)* - (213) 740-9355(WELL) – 24/7 on call

Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

#### *Office for Equity, Equal Opportunity, and Title IX (EEO-TIX)* - (213) 740-5086

Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

#### *Reporting Incidents of Bias or Harassment* - (213) 740-5086 or (213) 821-8298

Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

#### *The Office of Student Accessibility Services (OSAS)* - (213) 740-0776

OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

#### *USC Campus Support and Intervention* - (213) 740-0411

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

#### *Diversity, Equity and Inclusion* - (213) 740-2101

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

#### *USC Emergency* - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

#### *USC Department of Public Safety* - UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call

Non-emergency assistance or information.

#### *Office of the Ombuds* - (213) 821-9556 (UPC) / (323-442-0382 (HSC)

A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

*Occupational Therapy Faculty Practice* - (323) 442-2850 or [otfp@med.usc.edu](mailto:otfp@med.usc.edu)

Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.

*Annenberg Student Success Fund*

<https://annenberg.usc.edu/current-students/resources/annenberg-scholarships-and-awards>

The Annenberg Student Success Fund is a donor-funded financial aid account available to USC Annenberg undergraduate and graduate students for non-tuition expenses related to extra- and co-curricular programs and opportunities.

*Annenberg Student Emergency Aid Fund*

<https://annenberg.usc.edu/current-students/resources/annenberg-scholarships-and-awards>

Awards are distributed to students experiencing unforeseen circumstances and emergencies impacting their ability to pay tuition or cover everyday living expenses. These awards are not intended to cover full-tuition expenses, but rather serve as bridge funding to guarantee students' continued enrollment at USC until other resources, such as scholarships or loans, become available. Students are encouraged to provide as much information in their application, as well as contact their academic advisor directly with questions about additional resources available to them.

## **About Your Instructor**

Joshua (Josh) Clark (USC PhD 2016) is a Staff Data Scientist at Singularity6 Corp. Josh has seven years experience working at companies such as Riot Games, Meta and Jam City Games working at scale on major intellectual properties.