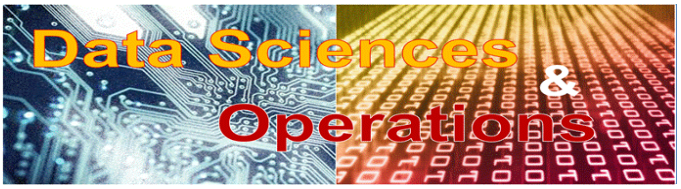
# DSO 545: STATISTICAL COMPUTING AND DATA VISUALIZATION

USC Marshall School of Business




| | |
|---|---|
| Instructor: | Francis Pereira, Ph.D. |
| Office: | BRI 401 M |
| Office Hours: | Tuesdays 12 pm – 1 pm |
| | By Zoom (generally Fridays) |
| Email: | pereira@marshall.usc.edu |
| Teaching Assistant: | TBA |

## COURSE DESCRIPTION

This course aims to prepare the students to learn about and execute the data science process in a business setting. The data science process involves multiple stages that starts by asking the right question, exploring, and cleaning the data, building the model, interpreting the results and finally communicating the results. In this class, we will go through the whole data science process and focus on the exploratory data analysis (EDA) and communications of your results via interactive web dashboards.

The course starts by learning how to frame, refine, and translate a business question into a data science problem. It will be followed by learning different computational techniques on data cleaning, feature engineering, and data visualizations using Python as computational tool. Students will learn how to utilize Python to answer data science problems in marketing, finance, operations, and human resources. Utilizing Python as a tool in our data science process, we will:

- Read different types of data (structured and unstructured)
- Clean, manipulate, and aggregate the data into a useful format
- Explore the data using numerical summaries
- Explore the data using data visualization
- Explore the data using clustering techniques
- Run A/B tests
- Run regression analysis
- Use and create Python data structures, control statements, functions, and classes
- Scrape data from the web
- Create web interactive dashboards to present your results to different business stakeholders

## COURSE OBJECTIVES

By the end of this course, students will be able to:

Understand

- Interpret business problems using exploratory data analysis (EDA)
- Interpret and communicate the outcomes of the EDA process

Apply

- Clean and prepare datasets for analysis
- Produce graphs that follow the grammar of graphics to support the EDA process
- Produce interactive dashboards to support the EDA process
- Access data through different sources (databases, cloud, and web scrapping)

Analyze

- Analyze business case studies using data science tools to make decisions in marketing, finance, human resources, and operations

Evaluate

- Recommend business strategies based on evidence coming from data

Create

- Prepare a data project management plan for real-world data science problems
- Prepare data science solutions for real-world business problems

## COURSE MATERIALS

All class material will be found in blackboard. There will be no specific book for this class. All resources will be listed in each module.

Books *(soft copy available for USC students https://libraries.usc.edu)*

1. Python for Marketing Research and Analytics. Springer 2020. (by Jason Schwarz, Chris Chapman, and Elea McDonnell Feit)

Software *(open source)*

2. Anaconda (https://www.anaconda.com/products/individual)
3. Plotly Python Open-Source Graphing Library (https://plotly.com/python/)

## GRADING

Grades are determined by your homework average, and your scores on the quizzes, midterm and final exam. Letter grades will be determined at the end of the course according to a curve centered on B+. Letter grades will not be given for individual assignments.

| Assessment | % of Grade |
|---|---|
| Homework | 20% |
| Team Project | 30% |
| Midterm | 25% |
| Final Exam | 25% |

HOMEWORK

There will be several homework assignments. In general, the homework assignments are due one week after they are assigned (by the end of the day at 11:59pm).

Failing to submit the homework assignment on time will result in penalties as follows:

| | |
|---|---|
| 1 day late (1 min late is considered 1 day late) | 15% off |
| 2 days late | 30% off |
| 3 days late | 50% off |
| 4 days or more late | 100% off |

TEAM PROJECT

For the group presentation, teams are expected to find their own datasets, pose a question, explore and visualize the dataset using the tools we've developed throughout the course. Finally, teams will write a report and present the results in-class. More info will be shared during the semester.

## Unclaimed Graded Paperwork

Returned paperwork, unclaimed by a student, will be discarded 4 weeks after the end of term, and hence will not be available should a grade appeal be pursued by a student following receipt of his/her course grade.

## Add/Drop Process:

Please note that the last day to register and add classes is September 12. The last day to drop a class without a mark of "W" is also September 12 and receive a tuition refund. The last day to drop a class without a mark of "W" is October 6. For more information, visit https://classes.usc.edu/term-20233/calendar/

Open Expression and Respect for All

An important goal of the educational experience at USC Marshall is to be exposed to and discuss diverse, thought-provoking, and sometimes controversial ideas that challenge one's beliefs. In this course we will support the values articulated in the USC Marshall "Open Expression Statement" (https://www.marshall.usc.edu/about/open-expression-statement).

## Use of AI Platforms

In this course, I encourage you to use artificial intelligence (AI)-powered programs to help you with assignments that indicate the permitted use of AI. You should also be aware that AI text generation tools may present incorrect information, biased responses, and incomplete analyses; thus they are not yet prepared to produce text that meets the standards of this course. To adhere to our university values, you must cite any AI-generated material (e.g., text, images, etc.) included or referenced in your work and provide the prompts used to generate the content. Using an AI tool to generate content without proper attribution will be treated as plagiarism and reported to the Office of Academic Integrity. Please review the instructions in each assignment for more details on how and when to use AI Generators for your submissions.

In this regard, please note that part or all of examinations may involve python writing code in a non-electronic medium.

## Class Notes Policy

Notes or recordings made by students based on a university class or lecture may only be made for purposes of individual or group study, or for other non-commercial purposes that reasonably arise from the student's membership in the class or attendance at the university. This restriction also applies to any information distributed, disseminated or in any way displayed for use in relationship to the class, whether obtained in class, via email or otherwise on the Internet, or via any other medium. Actions in violation of this policy constitute a violation of the Student Conduct Code and may subject an individual or entity to university discipline and/or legal proceedings.

## USC 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, comprises 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.

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

**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. You may contact OSAS at (213) 740-0776 or via email at 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
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.

# TENTATIVE COURSE TOPICS and SCHEDULE

| Week | Week of | Topic | Due |
|---|---|---|---|
| 1 | Aug-21 | • Intro to class<br>• Overview of Python<br>• Data structures: boolean, numeric, strings, dictionaries<br>• Control statements: conditional if statement, for loops | |
| 2 | Aug 28 | • Data structures: sets<br>• Control statements: while loops<br>• Data wrangling: numpy, pandas, filtering, selecting, mutating, and grouping<br>• Exploratory data analysis (EDA): describing data, functions to summarize a variable, summarizing dataframes, single variable visualization<br>• Loading and saving data | |
| 3 | Sep-4 | • Functions and classes<br>• Data wrangling: joining data tables<br>• Exploratory data analysis (EDA): relationship between continuous variables, exploring associations between variables with scatterplots, combining plots, scatterplot matrices, correlation coefficients, exploring associations in survey responses | HW#1 |
| 4 | Sep-11 | • Data wrangling: tidy data, wide data, long data<br>• Exploratory data analysis (EDA): comparing groups using tables, and data visualizations<br>• Hypothesis Testing: statistical tests, testing group frequencies, testing observed proportions, testing group means, testing multiple group means, A/B testing | |
| 5 | Sep-18 | • Data Wrangling: working with dates and time data objects<br>• Clustering techniques: market segmentation, unsupervised clustering methods for exploring subpopulations, hierarchical clustering, k-means clustering | HW#2 |
| 6 | Sep-25 | • Linear regression: simple and multiple linear regression, comparing models, using a model to make predictions, using factors as predictors, interaction terms | |

| | | | |
|---|---|---|---|
| 7 | Oct-2 | • Simulating data<br>• Retrieving data from SQL databases into Python environment | HW#3 |
| 8 | Week of Oct-9 | Midterm (TBD) | |
| 9 | Oct-16 | • Interactive plotting I: plotly express package, scatter line charts, area charts, bar charts, part of whole charts, distributions, images and heatmaps, maps, polar coordinates, 3D coordinates | |
| 10 | Oct-23 | • Interactive plotting II: plotly graphing library basic charts and animation, scatter plots, line charts, bar charts, pie charts, bubble charts, dot charts, filled area plot, horizontal bar charts, gantt charts, sunburst chart, tables, sankey diagrams, treemap charts, categorical axes, icicle charts, animated figures with graph objects, frames, adding control buttons to animations | HW#4 |
| 11 | Oct-30 | • Interactive plotting III: plotly graphing library statistical charts , financial charts, geospatial charts, and subplots, box plots, histograms, distribution plots, scatterplot matrix, faced and trellis plots, parallel categories diagram, marginal distribution plot, heatmaps, dendograms, polar charts, radar charts, wind rose and polar charts, time series and dates, candlestick charts, waterfall charts, funnel chart, gauge charts, bullet charts, choropleth maps, filled area on maps, bubble maps, mixed subplots, table, mixed subplots, map subplots, table and chart subplots, figure factory subplots | Project Proposal |
| 12 | Nov 6 | • Business Dashboards using Plotly Dash (part 1) | |
| 13 | Nov-13 | • Business Dashboards using Plotly Dash (part 2) | HW#5 |
| 14 | Nov 20 | • Web data scraping and text data manipulation (extracting data from websites with or with no APIs) | Time Permitting |
| 15 | Nov 27 | • Project presentations<br>• Course evaluation | Project |