USC Marshall School of Business

DSO 545: STATISTICAL COMPUTING AND DATA VISUALIZATION
FALL 2021
3 units, Wednesdays, 6:30pm-9:20pm

Instructor:	*Abbass Sharif*
Office:	https://usc.zoom.us/my/alsharifabbass
Office Hours:	Fridays 1pm – 2pm
Email:	*asharif@marshall.usc.edu*
Teaching Assistant:	- *Name: **Grant Smith*** - *Contact: Use the class's Slack channel to reach out for Bowen*

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 ask, 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. The 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, you 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)
2. The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios. Wiley 2017. (by Steve Wexler, Jeffrey Shaffer, and Andy Cotgreave)

Software *(open source)*

3. Anaconda (https://www.anaconda.com/products/individual)
4. 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

Homework will be assigned on a weekly basis. In general, the homework assignment is due one week after it was 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

More info will be shared during the semester.

TENTATIVE COURSE OUTLINE

Session	Date	Topic	Due
1	Aug-26	• **Intro to class** • **Overview of Python** • **Data structures:** boolean, numeric, strings, dictionaries • **Control statements:** conditional if statement, for loops	
2	Sep-2	• **Data structures:** sets • **Control statements:** while loops • **Data wrangling**: numpy, pandas, filtering, selecting, mutating, and grouping • **Exploratory data analysis (EDA):** describing data, functions to summarize a variable, summarizing dataframes, single variable visualization • **Loading and saving data**	HW 1
3	Sep-9	• **Functions and classes** • **Data wrangling:** joining data tables • **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 2
4	Sep-16	• **Data wrangling:** tidy data, wide data, long data • **Exploratory data analysis (EDA):** comparing groups using tables, and data visualizations • **Hypothesis Testing**: statistical tests, testing group frequencies, testing observed proportions, testing group means, testing multiple group means, A/B testing	HW 3
5	Sep-23	• **Data Wrangling:** working with dates and time data objects • **Clustering techniques:** market segmentation, unsupervised clustering methods for exploring subpopulations**,** hierarchical clustering, k-means clustering	HW 4
6	Sep-30	• **Linear regression:** simple and multiple linear regression, comparing models, using a model to make predictions, using factors as predictors, interaction terms	HW 5
7	Oct-7	• **Simulating data** • **Retrieving data from SQL databases into Python environment**	

8	Week of Oct-11	Midterm (TBD)	
9	Oct-21	• **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-28	• **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	HW6
11	Nov-4	• **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-11	• **Business Dashboards using Plotly Dash** (part 1)	
13	Nov-18	• **Business Dashboards using Plotly Dash** (part 2)	
14	**Week of Nov 22 (online)**	• **Web data scraping and text data manipulation** (extracting data from websites with or with no APIs)	
15	Dec 2	• **Project presentations** • **Course evaluation**	Project

STATEMENT ON ACADEMIC CONDUCT AND SUPPORT SYSTEMS

Academic Conduct:

Students are expected to make themselves aware of and abide by the University community's standards of behavior as articulated in the Student Conduct Code. Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Part B, Section 11, "Behavior Violating University Standards" policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct at http://policy.usc.edu/scientific-misconduct.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
studenthealth.usc.edu/counseling
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call
suicidepreventionlifeline.org
Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention and Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call
studenthealth.usc.edu/sexual-assault
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED)- (213) 740-5086 | Title IX – (213) 821-8298
equity.usc.edu, titleix.usc.edu
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. The university prohibits discrimination or harassment based on the following *protected characteristics*: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in applicable laws and governmental regulations. The university also prohibits sexual assault, non-consensual sexual contact, sexual misconduct, intimate partner violence, stalking, malicious dissuasion, retaliation, and violation of interim measures.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
usc-advocate.symplicity.com/care_report
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office of Equity and Diversity |Title IX for appropriate investigation, supportive measures, and response.

The Office of Disability Services and Programs - (213) 740-0776
dsp.usc.edu
Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Support and Advocacy - (213) 821-4710
uscsa.usc.edu
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101
diversity.usc.edu
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
dps.usc.edu, emergency.usc.edu
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) 442-120 – 24/7 on call
dps.usc.edu
Non-emergency assistance or information.