DSO 585: Data-Driven Consulting

Topic: Subject to Students' Proposed Project

(I can also provide a project related to Quantitative Trading with Reinforcement Learning)

Summer 2024 - 1.5 Units

Instructor:	Austin Pollok	**Time:**	M 5 - 8pm
E-mail:	Please use Slack (if needed - pollok@usc.edu)	**Room:**	TBA

Course Description and Policies

Course Description:

Data-driven analytics continues to reshape all major areas within an organization and across business domains such as financial products and services, manufacturing, retail, transportation, health care, entertainment, education, and many more. Data-driven consulting is the hands-on application of descriptive, predictive, and prescriptive analytics to a unique business problem. Such business problems can range from minimum viable product (MVP) development to technical product management, as well as improving existing components of a business on the operational side or on the product and services development side. Wherever there is data, there is a place for data-driven consulting!

This course will allow students, either individually or with a team, to apply descriptive, predictive, and prescriptive analytics tools to a problem of their choosing (should be related to business problems faced by companies they aspire to work for). From the outset, we will discuss the relevant business intelligence specific to their projects. We will then move into properly scoping the project by setting objectives and key results (OKRs), as well as key performance indicators (KPIs). Finally, we will use Python programming, and any other necessary tools, to implement an analytics-based solution for the OKRs. This course can be challenging due to the hands-on nature of the course as well as the business problems and practical constraints faced by real-world data and companies. These features of the course will benefit graduate students interested in real-world applications of business analytics.

Course Learning Objectives:

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

1) **Identify** opportunities to apply descriptive, predictive, and prescriptive analytics as a solution for a particular business problem.

2) **Translate** a high-level business objective into a clear and precise data-driven objective.

3) **Build** novel data sets by gathering, cleaning, and exploring data sources and vendors.

4) **Apply** statistical analysis, machine learning, and optimization to help advance a data-driven solution for a business-oriented project.

5) **Communicate** effectively with both technical and non-technical business stakeholders on blockers, OKRs, KPIs, and recommendations.

Prerequisites:

- DSO 530, Applied Modern Statistical Learning Methods.

- DSO 545, Statistical Computing and Data Visualization.

- DSO 570, The Analytics Edge: Data, Models, and Effective Decisions.

- Students are also expected to have Python programming training and experience.

Textbook Requirements:

There will be no official textbook for this course. We will be using course notes from Jupyter notebooks, distributed through Blackboard, as well as selected readings as needed.

There are two recommended (not necessary to buy) books for analytics and data wrangling, respectively:

1) Kuhn, M. and Johnson, K. (2018). *Applied Predictive Modeling.* Springer.
(You can download a PDF copy through USC's SpringerLINK:
https://libraries.usc.edu/databases/springerlink)

2) Banachewicz, K. and Massaron, L. (2022). *The Kaggle Book.* Packt Publishing.

Technology Requirements:

Software: Download and install Anaconda (a popular Python distribution platform). Throughout the class, we will mainly rely on Jupyter Notebooks (a way to run Python in your web-browser). For instructions on how to install Anaconda, follow these instructions. We will be using Python, including packages such as Pandas, Scikit-Learn, Keras and TensorFlow, and CVXPY as needed throughout the course to implement the data-driven analytics.

Laptop: Students must have access to a laptop computer on which Anaconda is installed. Students who do not own a laptop that can be brought to class may consider the USC Laptop Loaner Program: https://itservices.usc.edu/spaces/laptoploaner/.

Large Datasets: Students may be required to work with large datasets. Students are fully responsible for ensuring their laptops can comply with the datasets for any coursework. If your laptop cannot work with large datasets, please consider the following references to options for borrowing laptops from Marshall and from USC libraries:

From USC Libraries: https://itservices.usc.edu/spaces/laptoploaner/

- Mac or PC

- Rent on Hourly Basis for free (4-hour max, can renew in person for more hours.)

- Can come any time the library is open.

- If requested from Professor, Libraries may allow advance reservation, Professor would email ITS Learning environments at spaces@usc.edu for this special request.

Office hours (subject to change):

- TBA on Zoom at https://usc.zoom.us/j/98024219791,

- TBA in-person at BRI 307J

AI Policy:

I expect you to use AI (e.g., ChatGPT and image generation tools) in this class. Learning to use AI is an emerging skill, and I welcome the opportunity to meet with you to provide guidance with these tools during office hours or after class. Keep in mind the following:

- AI tools are permitted to help you brainstorm topics or revise work you have already written.

- If you provide minimum-effort prompts, you will get low-quality results. You will need to refine your prompts to get good outcomes. This will take work.

- Proceed with caution when using AI tools and do not assume the information provided is accurate or trustworthy. If it gives you a number or fact, assume it is incorrect unless you either know the correct answer or can verify its accuracy with another source. You will be responsible for any errors or omissions provided by the tool. It works best for topics you understand.

Course Communication Policy:

Please communicate! I would like to unblock students as early as possible, meaning I want to clear any questions, confusions, or concerns. We can meet in office hours, over Zoom, on Slack, or in-person, as well as grabbing a coffee for a quick chat.

I am very interested in having discussions with all students, whether it be regarding course material, extensions of course materials, programming, or general career discussions in the field of analytics. I would also appreciate hearing any feedback regarding the course, as it is in a state of continual evolution.

Please Slack me to set up any times for discussion, I will typically respond to all Slacks within three business days, though commonly quicker. Any content specific questions should also be posted on the course Slack channel where I will respond for any students who may be having similar trouble.

Course Grading Policy:

This course will be offered as **Credit / No Credit**. To receive **Credit**, all presentations and the final report must be fully completed, and the final total percentage points must be above the threshold of 85%.

Assignment	Weight
Participation	10%
Presentation 1	15%
Presentation 2	15%
Presentation 3	15%
Final Report	45%

The group presentations and individual final report will be evaluated according to the following criteria:

- Relevance: should be addressing the business problem posed by the company using a data-driven solution.

- Usefulness: should add value towards an implementable solution.

- Soundness: choose data sets with enough sample size; conduct sound numerical experiments (split the data into training/validation/test sets); make comparative result tables

using validation or cross-validation; use the test set only for final assessment; add graphs and other good ways of visualization.

- Clarity/presentation: good slide deck or report organization, good bibliography (if any), enough graphs and visual support, length should not exceed 6 pages.

- Novelty: novelty is not strictly required but strongly encouraged.

The final report is to be submitted by **TBA**, which is the University's scheduled date for the final exam.

You will receive a grade of zero for a final report that is submitted after the deadline unless you have a written excuse from your doctor.

Assignments and Grading Policies

Student Deliverables:

Students are expected to produce 4 deliverables by the end of the course. See Blackboard TurnItIn Assignments for due dates.

1) Presentation 1: First stakeholder update. .... ??
2) Presentation 2: Second stakeholder update. .... ??
3) Presentation 3: Final stakeholder presentation. .... ??
4) Final Report: Summarize your findings in an executive report. .... ??

Assignment Descriptions and Criterion:

1. **Participation:**

   Class participation is an extremely important part of the learning experience in this course as the richness of the learning experience will be largely dependent upon the degree of preparation by all students prior to each class session. My expectation and that of your classmates are that you are prepared for all classes and will actively participate in and meaningfully contribute to class discussions. You are also expected to ask questions in class presentations.

   In-class participation grading will be based on students' demonstrated willingness to participate and the quality of the comments expressed, rather than quantity. While some students are far more comfortable than others with class participation, all students should make an effort to contribute meaningfully. The direction and quality of a discussion is the collective responsibility of the class. To underscore the importance of participation, 10 percent of the course grade will be allocated to class participation.

   Students will offer their opinions in group settings many times in their careers; thus, class participation serves to prepare students for this business experience. The evaluating of in-class participation is based on the following:

   - Relevance – Does the comment or question meaningfully bear on the subject at hand? Irrelevant or inappropriate comments can detract from the learning experience.

   - Responsiveness – Does the comment or question connect to what someone else has said?

   - Analysis – Is the reasoning employed consistent and logical? Has data from course materials, personal experience, or general knowledge been employed to support the assertions/findings?

- Value – Does the contribution further the understanding of the issues at hand?

- Clarity – Is the comment concise and understandable?

Class Participation—Behavioral Anchor Rating Scale:

- Excellent Performance

    - Initiates information relative to topics discussed

    - Accurately exhibits knowledge of assignment content

    - Clarifies points that others may not understand

    - Shares personal experiences or opinions related to topic

    - Offers relevant / succinct input to class

    - Actively participates in class exercises

    - Demonstrates ability to apply, analyze, evaluate synthesize course material.

    - Demonstrates willingness to attempt to answer unpopular questions

    - Builds on other students' contributions

- Average Performance

    - Participates in group discussions when asked

    - Demonstrates knowledge of course material

    - Offers clear, concise, "good" information on class assignments

    - Offers input, but tends to reiterate the intuitive

    - Attends class regularly

- Unacceptable Performance

    - Fails to participate even when directly asked

    - Gives no input to discussions

    - Does not demonstrate knowledge of the readings

    - Shows up to class: does nothing

    - Distracts group / class

    - Irrelevant discussion

2. **Group Presentations:**
   There will be three presentations delivered to business stakeholders throughout the semester. Each presentation will contain updates on the progress the class has made on their solution to the business problem. Each presentation is intended to summarize different milestones of the data-driven solution starting with a novel data set construction, results from descriptive analytics, and results from predictive and prescriptive analytics. These

presentations are also intended to give students opportunities to ask questions to the business stakeholders.

The presentations will be completed in self-selected groups of at most five students. Groups can be different from one project to the other. Peer evaluations will be taken into account when accessing individual scores for the team projects. See Appendix II for the peer evaluation form. The in-class presentations will be evaluated by the individual effort and contribution of each member (as measured by the peer evaluation form), quality of slides, presentation, and discussions.

3. **Final Report:**

Each student will individually write a comprehensive report summarizing the business problem and the data-driven solution implemented throughout the semester. The report should include:

- all objectives and key results,

- any relevant business recommendations,

- details about the data set construction,

- results from descriptive analytics with clear and concise recommendations,

- results from predictive or prescriptive analytics with clear and concise recommendations,

- and finally, any personal take-aways from the project and the process of creating and implementing a data-driven solution to a business problem.

Assignment Policies:

Regrading:

If you believe that an error has occurred in the grading of any assignment, you may submit a regrading request within one week (7 calendar days) of the date that the assignment is returned to the class. Please attach a hard-copy memo (a post-it note is fine) to your hard-copy assignment submission (if it is code, print out the relevant parts). The memo should describe in detail which part of the assignment should be regraded, and why. Please be mindful that regrading may result in an upward adjustment, downward adjustment, or no adjustment of the grade.

Plagiarism:

Students are permitted and encouraged to discuss with other students in the class about ideas for completing assignments. However, once a student begins writing code, everything must be individual and independent work. Students may not seek help from anyone outside

the class, including but not limited to: former students of this class, friends and family, tutors, and online forums. Students may reference course materials and web resources, but may not post anything related to the assignments online. Failure to abide by the above guidelines may constitute a case of suspected plagiarism or cheating, which will be reported and investigated by USC. Please see the "Academic Integrity and Conduct" section below for further details.

Course Schedule

Tentative Course Outline: **WORK IN PROGRESS - Subject to Change**

This outline is tentative and subject to change. It will be updated as the course progresses.

Interesting links: There is a number of great websites and podcasts that provide additional interesting information. See for instance MIT Machine Learning News, MIT Big Data News, Hitchhiker's Guide to Python, Towards Data Science, FiveThirtyEight, and Real Python. Other interesting links related to analytics being applied in various companies include (these are really good): Netflix Tech Blog, Microsoft Experimentation Platform, Uber Engineering Blog, Airbnb Data Science Platform Blog, Etsy Tech Blog, Unofficial Google Data Science Blog, DoorDash Data Science Blog, LinkedIn Engineering Blog, Stitch Fix Tech Blog, Lyft Data Science Blog, Booking.com AI Blog, Spotify Data Science Blog, Yelp Engineering Blog

---

Chosen by individual students or teams of students.

I can also give a project related to: Quantitative Trading with Reinforcement Learning

---

Week 1: Discussion of how to structure an unstructured project and frame a business problem.

Discussion of OKRs, KPIs, and project scoping.
Introduction to data sources and vendors.
Creating a novel data set to help achieve OKRs.

Week 2: Descriptive analytics techniques

Exploratory data analysis (EDA).
Feature engineering your curated data.

Week 3: Predictive and prescriptive modeling techniques I:

Presentation 1 on data sources and data set construction.

Week 4: Predictive and prescriptive modeling techniques II

Presentation 2 on descriptive analytics results and initial recommendations.

Week 5: Predictive and prescriptive modeling techniques III

Cover modern deep-learning based methods.

Week 6: Predictive and prescriptive modeling techniques IV

Continue covering modern deep-learning based methods.

Week 7: Spillover

We will use this extra time for anything that has spilled over from the previous weeks.

Week 8: Final Presentations

Presentation 3 on predictive and prescriptive analytics results and final recommendations. Students will submit their final reports.

University Scheduled Final Exam Day TBA

USC 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."

USC Statement on Academic Conduct and Support Systems

Academic Conduct:

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". Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on Research and Scholarship Misconduct.

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

eeotix.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.

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 for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

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

osas.usc.edu

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) 821-4710

campussupport.usc.edu

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

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.

Office of the Ombuds - *(213) 821-9556 (UPC) / (323-442-0382 (HSC)*
ombuds.usc.edu
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) 442-3340 or otfp@med.usc.edu*
chan.usc.edu/otfp
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.

SAMPLE PEER EVALUATION FORM

Please identify your team and team members for the ___ Project(s) that you worked on. Then rate all your team members, *including yourself,* based on the **contributions** of each team member for the selected assignment according to the criteria listed below. On a scale of 0 to 2, with 0 indicating does not meet expectations, 1 meets expectations and 2 exceeds expectations, rate each person on each of the five criteria. Lastly, add up the points for each person with the maximum number of points for each person being 10. In the box below, describe the exact contributions of each team member, including yourself.

Team Members/ Assessment Criteria of Team Contributions	Team Member 1	Team Member 2	Team Member 3	Yourself
1. Conceptualization of the project (coming up with the idea, the scope, and/or the application)				
2. Managing the assignment (making a work plan, keeping the group on track)				
3. Execution (locating and cleaning data, coding)				
4. Written exposition (writing, preparing figures)				
5. Presentation prep (creating presentation slides, prepping group members for questions)				
Total				

Contribution details: