Introduction to Data Analytics

ITP 249 - 4 Units - 2023 Spring
Section 31832R - Tuesday 3:30-6:50pm

USC Viterbi School of Engineering

Course Description

Data is now an integral part of our lives and to be successful in today's business landscape, we need to be able to leverage data to make critical business decisions. This course will teach students how to use data to make those decisions confidently.

Objectives

After completing this course, students will be able to:

- Use Excel, SQL, NoSQL and leading industry tools
- Collect, clean, and analyze data from multiple sources
- Pose questions, collect relevant data, analyze data, interpret data and provide insights
- Present data-driven insights using data visualization and dashboards
- Use statistical techniques to gain insights

Prerequisites

None

Lectures and Lab

Tuesday 3:30PM to 6:50PM in person in classroom GFS 116.

Zoom: There will be no remote attendance option, nor will there be any recordings of the lecture. Students are expected to attend lectures in-person.

Instructor

Calvin Nguyen <cnt.nguyen@usc.edu>
Office Hours: after class on Tuesday and by appointment.
Contact Instructor for all Lecture-related questions.

Teaching Assistant

Office Hours: see Blackboard/Piazza
Contact Teaching Assistants for all Assignment-related questions.

Course Notes

blackboard.usc.edu
Lecture slides and any supplemental course content will be posted to Blackboard. All announcements for the course will be posted to Blackboard. Information about assignments, due dates, exams and grades will also be posted on Blackboard. Students should check Blackboard regularly for updates.

Discussion Board

piazza.com
Piazza is for students to ask questions and for Instructor & Teaching Assistants to answer questions in written form. Link to Piazza will be posted on Blackboard.

Textbook
No official textbook required.

Optional Books:
Carlos Coronel, Steven Morris and Peter Rob. Database Systems: Design, Implementation, and Management. Boston, MA: Cengage Learning, 2014. (CMR) ISBN-13: 978-1111969608

Additional reference material will be provided in class as needed.

Technological Proficiency and Hardware/Software Required
Most assignments in the class are done using software. Software will be available for download or available through a virtual lab. Students are expected to have access to a computer. ITP has a limited number of laptops that students can request to borrow.

Software
Students can use their own computers or login to the Viterbi Virtual Lab. Students are introduced to several tools most used in the industry:

- Excel and Microsoft Access
- MySQL and MySQL Workbench
- NoSQL, MongoDB and Studio3T
- Neo4J
- Tableau and other statistical software

IT Help
USC IT (ITS): https://itservices.usc.edu/contact
Viterbi IT: https://viterbi.usc.edu/resources/vit/contact-us.htm

Grading
The weight of the graded material during the semester is listed below:

Individual Homework Assignments	35%
Team Project	15%
Exam – Midterm	25%
Exam – Final	25%

The following Grading Scale will be used to determine your letter grade:

A	100-93
A-	92-90
B+	89-87
B	86-83
B-	82-80
C+	79-77
C	76-73
C-	72-70
D+	69-67
D	66-63
D-	62-60
F	59 or below

Half percentage points will be rounded up to the next whole percentage. So for instance, 89.50% is an A-, but 89.49% is a B+

For the Pass/No Pass grading option, you must earn C- or higher to pass for undergraduates students, and B or higher to pass for graduate students.

Individual Homework Assignments

This course will make use of Blackboard for assignments. All assignments will be posted to Blackboard under the “Assignments” section. Each assignment will include instructions, a due date, and a link for electronic submission. Assignments must be submitted using this link.

Team Project

For a hands-on experience, there will be a final team project. The goal of the project is to solve real world problems using data analytics. This is a team project with each consisting of 2 to 3 members. Each team will work together to identify a problem to solve, collect the necessary data, prepare, clean and format the data, analyze the data, create visualizations, dashboards and models to analyze and understand the problem and to use insights to develop solutions.

The team project grading breakdown is listed below:

Project Statement, Methodology and Accuracy	25%
Final Report	50%
Peer Evaluation	25%

Policies & Late Submissions

Students are expected to attend and participate in lecture discussions, in-class exercises and team meetings. Attendance will be taken during lecture sessions electronically. Do not share the code with students that are not in the room; doing so is an academic integrity violation. If they would like to be considered for an excused absence, email the instructor and include name, week (1-15), day, date, reason, and documentation.

Students are responsible for completing individual assignments and their fair share of team assignments by stated deadlines. **Assignments turned in late will have 25% of the total points deducted from the graded score for each late day**. Projects do not have a late policy.

Students can submit an assignment multiple times. The latest submission will be graded.

No make-up exams (except for documented medical or family emergencies) will be offered. If they will not be able to attend an exam due to an athletic game or other valid reason, then they must coordinate with the instructor before the exam is given. They may arrange to take the exam before they leave, with an approved university personnel during the time they are gone, or within the week the exam is given. If students do not take an exam, then they will receive a 0 for the exam.

Disability Services and Programs

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to your course instructor (or TA) as early in the semester as possible. If you need accommodations for an exam, the form needs to be given to the instructor at least two weeks before the exam.

DSP is located in STU 301 and is open from 8:30am to 5:00pm, Monday through Friday. Contact info: 213-740-0776 (Phone), 213-740-6948 (TDD only), 213-740-8216 (FAX), ability@usc.edu, http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html.

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" https://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, http://policy.usc.edu/scientific-misconduct.

Support Systems

Student Counseling Services (SCS) - (213) 740-7711 – 24/7 on call
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention. https://engemannshc.usc.edu/counseling/

National Suicide Prevention Lifeline - 1-800-273-8255
Provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week. http://www.suicidepreventionlifeline.org

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 - 24/7 on call
Free and confidential therapy services, workshops, and training for situations related to gender-based harm. https://engemannshc.usc.edu/rsvp/

Sexual Assault Resource Center
For more information about how to get help or help a survivor, rights, reporting options, and additional resources, visit the website: http://sarc.usc.edu/

Office of Equity and Diversity (OED)/Title IX Compliance – (213) 740-5086
Works with faculty, staff, visitors, applicants, and students around issues of protected class. https://equity.usc.edu/

Bias Assessment Response and Support
Incidents of bias, hate crimes and microaggressions need to be reported allowing for appropriate investigation and response. https://studentaffairs.usc.edu/bias-assessment-response-support/

The Office of Disability Services and Programs
Provides certification for students with disabilities and helps arrange relevant accommodations. http://dsp.usc.edu

Student Support and Advocacy – (213) 821-4710
Assists students and families in resolving complex issues adversely affecting their success as a student EX: personal, financial, and academic. https://studentaffairs.usc.edu/ssa/

Diversity at USC
Information on events, programs and training, the Diversity Task Force (including representatives for each school), chronology, participation, and various resources for students. https://diversity.usc.edu/

USC Emergency Information
Provides safety and other updates, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible, http://emergency.usc.edu

USC Department of Public Safety – 213-740-4321 (UPC) and 323-442-1000 (HSC) for 24-hour emergency assistance or to report a crime.
Provides overall safety to USC community. http://dps.usc.edu

Course Outline

	Topics	Reading	Homework
Week 1 **Jan 10**	**The Value of Data** ● Explanation of course objectives and tools ● Syllabus Review ● Discussion of the value and impact of data-driven decision making ● Discussion of visual analytics and common presentation strategies ● Excel Analytics ● Brief history of databases and their role in information systems ● Different types of databases and their organizational context ● Survey of DBMS ● Introduction to Database ● Keys: primary, foreign, candidate, surrogate, super	Check Blackboard	HW1 Model
Week 2 **Jan 17**	**Foundations of Databases and SQL Data Modeling** ● Data models ● Business rules ● Relational and entity-relationship modeling ● Entities, attributes, relationships ● Minimum and maximum cardinality ● Designing Entity Relationship Diagram		HW2 ERD
Week 3 **Jan 24**	**Normalization** ● Anomalies and the need for normalization ● Normal forms ● First, second, third normal forms ● Denormalization ● Dependency Diagrams		HW3 Normalization
Week 4 **Jan 31**	**Introduction to SQL** • Database structures • Introduction to SQL's SELECT statement with WHERE clauses • Query command tools: GROUP BY, HAVING, DISTINCT, COUNT, AND, and OR • Conditional operators: =, !=, >, <, IN, NOT IN, and BETWEEN • Aggregation functions: MIN, MAX, SUM, AVG, and COUNT		HW4 SQL
Week 5 **Feb 7**	**Combining Data in SQL** • Appending similar data together • Combining data from different tables together • Commands for combining data: JOIN and UNION **Cleaning Data and Creating Multiple Joins** • Creating relationships between tables: INNER, RIGHT, FULL OUTER, EXCEPTION and CROSS JOINs • Optimizing queries: WHERE, LIMIT and COALESCE		HW5 SQL Joins
Week 6 **Feb 14**	**Subqueries** • Asking multiple questions in a single query • Nesting queries • Multi-step aggregation or filtering		HW6 SQL Subqueries
Week 7 **Feb 21**	**Data Visualization with Tableau**		HW7 Tableau
Week 8 **Feb 28**	**Midterm Review**		
Week 9 **Mar 7**	**Midterm Exam**		
Mar 12–19	**Spring Break**		

Week 10 **Mar 21**	**Fundamentals of MongoDB** ● Data storage ● Data Retrieval ● Queries		HW8 MongoDB
Week 11 **Mar 28**	**MongoDB Aggregation Framework**		HW9 MongoDB Aggregation
Week 12 **Apr 4**	**Graph DBs** ● Why graph databases? ● Introduction to Neo4J		Project Comprehensive
Week 13 **Apr 11**	**Working with graph databases** ● Querying ● Graph Visualization		
Week 14 **Apr 18**	**TBD**		
Week 15 **Apr 25**	**Final Review** **Final Project Due**		
Week of Finals	**Study Days** **Sat-Tue Apr 29 - May 2** **Final Exam** **Tuesday, May 9 2-4 pm**		