

DS 250: Introduction to Data Science

Overview

In this course, we are going to do a thorough overview of 'data science', which is a brand new, exciting and extremely relevant (to you!) discipline. We will cover data sources, collection, storage, analysis, and interpretation (including visualization); we will also look at intersections with related areas such as Big Data, Data Mining and Machine Learning, and examine a variety of real-world applications. 'Data' is the new 'oil', and this course will explain 'why' and 'how'.

Pre-requisites

As stated in the university catalog, you don't need a lot of prep. Note that this course involves data analysis - so there will be a bit of math, and a small amount of coding (in R, Python) - but, this is not a programming course, or a math-heavy theory course. That said, note that ITP 115 is a co-req.

Topics

Here is what we'll cover:

- Introduction(s)
- Data Science - an overview
- Data sources, representation
- Data storage, formats
- EDA [Exploratory Data Analysis]
- Big Data: fundamentals
- Big Data: processing
- Data Mining
- ML [machine learning] ML APIs, tools, platforms...
- genAI [generative AI]
- Spatial data
- Data visualization
- Applications_1: Business
- Applications_2: Medical/health, Societal
- Applications_3: Environmental, Agricultural, Manufacturing
- Applications_4: Communications
- Applications_5: Knowledge, Entertainment, Commerce
- Governance
- What's next?

Textbook - optional: 'The Data Science Design MANUAL', by Steven S. Skiena.

Supplementary material: As we progress through the course, there will be a wealth of additional content (PDFs, links to sites, etc) that will be put up, related to current industry practices, news items, etc.

Course structure

Your letter grade will be determined by your scores on:

- four homeworks, worth 15%, 15%, 20%, 20%
- two exams (midterm and final), each worth 15%; samples will be provided before the exams

Questions?

Do feel free to ask: saty@usc.edu.
