DSA 210: Introduction to Data Science

Fall 2024 - 2025

Syllabus

Instructors:
Section A: Onur Varol, PhD (onur.varol@sabanciuniv.edu)
Section B: Özgür Asar, PhD (ozgur.asar@sabanciuniv.edu)

Time and location

Section A Lectures

- Tuesday 15:40-16:30, FASS G062
- Friday 12:40-14:30, FASS G062

Section B Lectures

- Thursday 14:40 – 15:30, FASS G062
- Friday 09:40 – 11:30, FASS G062

Recitations

- A: Wednesday 17:40 - 19:30, FENS G035
- B: Tuesday 10:40 - 12:30, FENS G035
- C: Tuesday 12:40 - 14.30, FENS G035
- D: Tuesday 16:40 - 18:30, FENS G035

Recitations: Every week or two to cover basic tools and techniques for hands-on experience. All recitations will be delivered online and links for connecting will be provided on SuCourse

CS210 Team:

- **TAs**
  - Beyza Çokkeçeci
  - Ekin Başar Gökçe
  - Mansur Kiraz
  - Mehrdad Heshmatnajafabad
  - Özgün Yargı
  - Sina Salehnia
- **LAs**
  - Alp Önder Yener
  - Doğukan Tosun
  - Kaan Güray Şirin
  - Osman Kantarcıoğlu

Office hours and contact information off the TAs and LAs shared on SUCourse

All email communication will be done through special course email address and one of the CS210 team member will respond to you.
We won't respond emails sent to other addresses: cs210.fens@sabanciuniv.edu

Sabancı Üniversitesi

Tentative Course Outline:

Week 1 (23 Sept - 27 Sept)		DS project life cycle and ethics
Week 2 (30 Sept - 04 Oct)		DS project life cycle and ethics
Week 3 (07 Oct - 11 Oct)		Data collection, APIs and web scraping
Week 4 (14 Oct - 18 Oct)		Probability review
Week 5 (21 Oct - 25 Oct)		Exploratory data analysis
Week 6 (28 Oct - 01 Nov)		Data Visualization
Week 7 (04 Nov - 08 Nov)	Midterm Exam	Hypothesis testing
Week 8 (11 Nov - 15 Nov)		Hypothesis testing
Week 9 (18 Nov – 22 Nov)		Hypothesis testing
Week 10 (25 Nov - 29 Nov)		Machine Learning - Supervised
Week 11 (02 Dec - 06 Dec)		Machine Learning - Unsupervised
Week 12 (08 Dec - 13 Dec)	OV will be out for a conference. ÖA can teach this week.	Causal Inference
Week 13 (16 Dec - 20 Dec)		Working with textual data
Week 14 (23 Dec - 27 Dec)		Working with other data types and pretrained models

Course summary: Data science topics span a large variety of disciplines and require a collection of skills. This course is intended to cover data science's fundamental principles and techniques, emphasizing data-centric quantitative thinking. We will tour the basic data science techniques from manipulation and summarizing the essential characteristics of a data set, basic statistical modeling, visualization, and prediction

Objectives and learning outcomes: Fundamentals of data analytics pipelines: i) data collection and ethics, ii) basic statistics and hypothesis testing, iii) explanatory data analysis, iv) information extraction from basic data types, and v) building machine learning models.

Prerequisites: IF100 and MATH 203

Grading Policy: These percentages are tentative and subject to change.

- **Midterm** (35%): Exam will be held in person in Week 7 (or following the university guidelines)
- **Final** (35%): Exam will be held in person during the final's exam week (or following the university guidelines)
- **Project** (30%): The project will be done individually by each student, and they are expected to analyze, visualize and communicate a dataset about themselves. Proposal will be worth 5%, the final presentation 20% and peer-evaluation 5%. Evaluation criteria will rank students by their performance and their rank will be reflected to the score.
- **Homework**: There will be a few assignments on data collection, explanatory analysis, and machine learning experiment. The assignments will not be graded, though there will be related questions in the exams.

Class Policies and advice:

- Regular attendance is essential and class participation is expected in paper discussions.
- Late assignments. There will be a 10% late penalty for up to 3 days and 20% penalty for assignments submitted in the next 10 days.
- Maximum score you can receive from the projects and assignments cannot be more than 1.5 of the exam score. *For instance, if your exam score is 60 any HW or project score higher than 90 will be lowered to 90*.
- Students have the responsibility of backing up all their data and code. At the end of the semester, they are expected to prepare a public release of their code and data with proper documentation.

Academic honesty: All students must follow the university guidelines of academic integrity. https://www.sabanciuniv.edu/en/academic-integrity-statement

Main references: There is no dedicated textbook for this course. Suggested textbooks are given below:

- G James, D Witten, T Hastie, R Tibshirani, J Taylor (2023) Introduction to Statistical Learning, with Applications in Python. Springer.
- Peter Bruce, Andrew Bruce & Peter Gedeck Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python. 2nd edition. O'Reilly.
- Joel Grus. Data Science from Scratch. O'Reilly.