Introduction to Data Science
CS 210
Fall 2023 - 2024

Instructor: Onur Varol, PhD
Email: onur.varol@sabanciuniv.edu

Time and location
- Wednesdays 16:40 – 17:30, FENS G077
- Fridays 12:40 – 14:30, FENS G077

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
  - Kerem Aydın
  - Halil İbrahim Ergül
  - Mansur Kiraz
  - Damla Erden
  - Ö zgün Yargı
- LAs
  - Ege Demirci
  - Ahmet Yavuz Lülecioğlu
  - Berk Sezer

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

Main references: No dedicated textbooks. There will be required and suggested reading and online materials on SuCourse platform.

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
# Tentative Course Outline:

<table>
<thead>
<tr>
<th>Week</th>
<th>Dates</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>(4-6 Oct.)</td>
<td>Introduction to data science</td>
</tr>
<tr>
<td>Week 2</td>
<td>(11-13 Oct.)</td>
<td>Traveling for a conference</td>
</tr>
<tr>
<td>Week 3</td>
<td>(18-20 Oct.)</td>
<td>DS project life cycle and Ethics</td>
</tr>
<tr>
<td>Week 4</td>
<td>(25-27 Oct.)</td>
<td>Data collection: APIs &amp; web scraping and parsing</td>
</tr>
<tr>
<td>Week 5</td>
<td>(1-3 Nov.)</td>
<td>Homework #1 due Probability review &amp; Explanatory data analysis</td>
</tr>
<tr>
<td>Week 6</td>
<td>(8-10 Nov.)</td>
<td>Data visualization</td>
</tr>
<tr>
<td>Week 7</td>
<td>(15-17 Nov.)</td>
<td>Hypothesis testing</td>
</tr>
<tr>
<td>Week 8</td>
<td>(22-24 Nov.)</td>
<td>Hypothesis testing</td>
</tr>
<tr>
<td>Week 9</td>
<td>(29 Nov. – 1 Dec.)</td>
<td>Working with text data</td>
</tr>
<tr>
<td>Week 10</td>
<td>(6-8 Dec.)</td>
<td>Project proposal due Network analysis</td>
</tr>
<tr>
<td>Week 11</td>
<td>(13-15 Dec.)</td>
<td>Traveling for a project meeting</td>
</tr>
<tr>
<td>Week 12</td>
<td>(20-22 Dec.)</td>
<td>Homework #2 due Machine Learning – Supervised learning techniques</td>
</tr>
<tr>
<td>Week 13</td>
<td>(27-29 Dec.)</td>
<td>Machine Learning – Unsupervised learning techniques</td>
</tr>
<tr>
<td>Week 14</td>
<td>(3-5 Jan.)</td>
<td>Machine Learning – Intro to Deep Learning</td>
</tr>
</tbody>
</table>

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

- **Homework** (2x15=30%): There will be 2 assignments on data collection, explanatory analysis, and machine learning experiment
- **Exam** (40%): 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 will analyze, visualize and communicate a dataset about themselves. Proposal will worth (5%) and the final presentation (20%) and material (5%) will be evaluated by the team. Evaluation criteria will rank students by their performance and their rank will be reflected to the score.
Class Policies and advice:
- Regular attendance is essential and class participation is expected in paper discussions.
- Late assignments. There will be 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 public release of their code and data with a proper documentation.

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