

# Introduction to Data Science

## CS 210

Spring 2023

<b>Instructor:</b> Onur Varol, PhD	<b>Email:</b> <a href="mailto:onur.varol@sabanciuniv.edu">onur.varol@sabanciuniv.edu</a>				
<b>Time and location</b> <ul style="list-style-type: none"><li>- Wednesdays 11:40 – 12:30, FENS G077</li><li>- Fridays 12:40 – 14:30, FENS G077</li></ul> Link for Zoom: <a href="https://sabanciuniv.zoom.us/j/98524232845">https://sabanciuniv.zoom.us/j/98524232845</a> <p><i>We may have to revise the course plan according to the countrywide reassessment to be made regarding higher education. This is expected to happen at the beginning of April. The content to be delivered is certain but the method of course delivery, the number and dates of exams, and some other details are subject to change.</i></p>					
<b>CS210 Team:</b> <table><tr><td><b>TAs</b></td><td><b>LAs</b></td></tr><tr><td><ul style="list-style-type: none"><li>○ Baris Alparslan</li><li>○ Kerem Aydin</li><li>○ Ahmet Yasin Aytar</li><li>○ Damla Erden</li><li>○ Ali Najafi</li><li>○ Vahid Khalili Param</li><li>○ Yasser Zouzou</li></ul></td><td><ul style="list-style-type: none"><li>○ Ege Demirci</li><li>○ Furkan Eris</li><li>○ Ali Mirik</li></ul></td></tr></table>		<b>TAs</b>	<b>LAs</b>	<ul style="list-style-type: none"><li>○ Baris Alparslan</li><li>○ Kerem Aydin</li><li>○ Ahmet Yasin Aytar</li><li>○ Damla Erden</li><li>○ Ali Najafi</li><li>○ Vahid Khalili Param</li><li>○ Yasser Zouzou</li></ul>	<ul style="list-style-type: none"><li>○ Ege Demirci</li><li>○ Furkan Eris</li><li>○ Ali Mirik</li></ul>
<b>TAs</b>	<b>LAs</b>				
<ul style="list-style-type: none"><li>○ Baris Alparslan</li><li>○ Kerem Aydin</li><li>○ Ahmet Yasin Aytar</li><li>○ Damla Erden</li><li>○ Ali Najafi</li><li>○ Vahid Khalili Param</li><li>○ Yasser Zouzou</li></ul>	<ul style="list-style-type: none"><li>○ Ege Demirci</li><li>○ Furkan Eris</li><li>○ Ali Mirik</li></ul>				
<b>Office hours and contact information off the TAs and LAs shared on SUCourse</b>					
All email communication will be done through special course email address and one of the CS210 team member will respond to you. <i>We won't respond emails sent to other addresses: <a href="mailto:cs210.fens@sabanciuniv.edu">cs210.fens@sabanciuniv.edu</a></i>					

**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:

Week 1		Introduction to data science
Week 2		DS project life cycle and Ethics
Week 3		Data collection: APIs & web scraping and parsing
Week 4		Probability review
Week 5	Homework #1 due	Explanatory data analysis
Week 6		Data visualization
Week 7		Hypothesis testing
Week 8		Hypothesis testing
Week 9	Homework #2 due	Working with text data
Week 10		Network analysis
Week 11		Machine Learning – Basic concepts
Week 12		Machine Learning – Supervised learning techniques
Week 13	Homework #3 due	Machine Learning – Unsupervised learning techniques
Week 14		Machine Learning – Intro to Deep Learning

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

- **Homework** (3x10=30%): There will be 3 assignments for 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%): Team of students (at most 6) will work on a data science problem related to theme, "Data Science for Social Good in Türkiye", and they will collect, analyze, and present their predictive and exploratory analysis. Project proposal will worth (10%) and the final presentation (10%) and material (10%) will be evaluated by the team.

### 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.
- 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>