## SABANCI UNIVERSITY

# Faculty of Eng. & Natural Sci.

# CS-210

## Introduction to Data Science

Instructor: Selim Balcisoy

Course Content

Data science spans a large variety of disciplines and requires a collection of skills. This course is intended to tour the basic techniques of data science from manipulation and summarizing the important characteristics of a data set, basic statistical modeling, programming and visualization. The assignments and term project will involve Python.

### Objectives

Data science spans a large variety of disciplines and requires a collection of skills. This course is intended to tour the basic techniques of data science from manipulation and summarizing the important characteristics of a data set, basic statistical modeling, programming and visualization.

Learning Outcomes

Learning the fundamentals of data science pipeline.

Learning how to explore and experiment with data.

Learn basic statistics (sampling techniques, mean, variance, outliers, Central Limit theorem, distributions) and machine learning techniques (clustering) that are necessary to analyze data: big and small.

Perform a statistical analysis on sample socio-economic data.

Building an understanding of data analytics techniques (data collection, cleaning, exploratory techniques, modeling and presentation).

Develop competency in the Python programming language within the course project. Design and run experimental tests to evaluate hypotheses about data.

**Course Policies** 

Academic Dishonesty and Plagiarism will not be tolerated and closely monitored. Attendance will be enforced by random assignments. Make-up only for official excuses with written records.

Grading

Midterm I	%15	W5
Midterm II	%15	W10
Quiz	%15	Weekly
Project Part I	%30	Submission (Progress   April 18, Progress    May 2)
Project Part II	%25	Submission May 20 (Final)
Oral Exams	(Midterm Makeup, discussion after May 24 up to %60)	

## **Project Grading**

All deliverables contribute to the project grading, we will be conducting random interviews with some of the project groups. The interview will ensure if you have done the project without any external help. The interviews will be announced and conducted in the week of May 24. The Interview will be graded in base-2 numeral system.

Grading Algorithm: Grade = (Part I + Part II)\*Interview

Outline

22 February W1 Intro
1 March W2 Visualization
8 March W3 Manipulation of Data
15 March W4 Exploration of Data
22 March W5 Statistics
29 March W6 Statistics
5 April W7 Statistics
12 April W8 Feature Engineering
19 April W9 Machine Learning (23April Friday)
26 April W10 Machine Learning
3 May W11 Machine Learning
10 May W13 Machine Learning (Break)
17 May W14 Deep Learning (19 May Wednesday)
24 May W15 Deep Learning