SABANCI UNIVERSITY

Faculty of Eng. & Natural Sci.

CS-210

Introduction to Data Science

Instructor

Selim Balcısoy balcisoy@sabanciuniv.edu FENS 1096

Head TA: Ekberjan Derman derman@sabanciuniv.edu

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, web programming and visualization. The assignments and term project will involve Python.

Links to Reference Material

<https://probability4datascience.com/index.html>

<https://www.oreilly.com/library/view/data-science-from/9781492041122/>

https://www.kaggle.com

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, web 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.
Make-up only for medical records from the Sabancı University Health Center or similar official documents (national sport competitions).

Grading

Midterm I %20

Midterm II %20

Quizzes %10

Project Part I %25 (Exploratory Analysis and Hypothesis Testing)

Project Part II %25 (Machine Learning and Presentation)

Project Policies

Project team size will be four. If you have no team, we will assign you to a group. Each group will be assigned to a TA and LA independent from your recitation section.

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. A detailed project description will a published in Week 3.

The projects will have four tasks which will be submitted and graded in two parts.

Quiz Policies

There will be random online quizzes.

Outline

W1 Intro March 1

W2 Visualization March 4, 8 and 11

W3 Manipulation of Data March 15 and 18

W4 Exploration of Data March 22 and 25

W5 Statistics March 29 and April 1 (Midterm I)

W6 Statistics April 5 and 8

W7 Statistics April 12 and 15

W8 Feature Engineering April 19 and 22

W9 Machine Learning April 26 and 29

W10 Machine Learning May 10 and 13

W11 Machine Learning May 17

W13 Machine Learning May 24 and 27

W14 Machine Learning May 31 and June 3 (~~Midterm II)~~

W15 Deep Learning June 7 (Midterm II) and 10