# YAP 101: Introduction to Data Science
## 2024 Spring

**Instructor:** Muhammed YILMAZ
**Office:** 145/A
**Office Phone:** 4655
**Email:** muhyilmaz@etu.edu.tr
**Office Hours:** TBD

**TA:** TBD
**Email:** TBD

**Class Hours:** 10:30 - 12:20, Tuesday at ST03.
　　　　　　　12:30 - 14:20, Wednesday at Amfi 2.

**Prerequisites**: BİL 113 (You need to have knowledge of basic programming.)

## Course Description

The aim of this course is to explain the basics of data science and prepare you for advanced topics by teaching how to use tools for acquiring, cleaning, analyzing, exploring, and visualizing data; making data-driven inferences and decisions; and effectively communicating results. In the course, first the concepts of causality and experimental research will be explained. Afterwards, it aims to provide basic programming skills on data analysis with Python programming, DataFrame library and visualization. Then, randomness, sampling, testing hypotheses, comparing two samples, prediction and the importance of the mean will be discussed. At the end of the course, a very superficial introduction to machine learning will be made and topics such as prediction, classification and Bayes Theorem will be discussed.

## Course Outline

- Causality and Experiments
- Programming in Python
- Randomness
- Sampling and Empirical Distributions
- Testing Hypotheses
- Comparing Two Samples
- Why the Mean Matters
- Prediction
- Inference for Regression
- Classification
- Updating Predictions

**Course Material**

All course material such as the syllabus, lecture notes, homework assignments, and any announcements will be posted at the Piazza site for this course.

**Textbook:** *Computational and Inferential Thinking: The Foundations of Data Science* by Ani Adhikari, John DeNero, David Wagner, https://applieddatascience.cmp.uea.ac.uk/intro.html
**Supplementary Book 1:** *Introduction to Probability*, 2nd Edition, by Dimitri P. Bertsekas and John N. Tsitsiklis, Athena Scientific, 2008.
**Supplementary Book 2:** *Learning to Program with Python* by Richard L. Halterman

**Homeworks:** There will be total of 7 homework assignments. Assignments will be in Python. In order to be accepted, your code should run without any errors. Late submissions will not be accepted.

**Grading**
- 20% Homeworks
- 35% Midterm
- 45% Final
- In order to pass the course, students must collect at least 40 points without counting homeworks and must collect total of at least 50 points.

**Attendance:** 70% attendance is mandatory.

**Additional Notes**
- Exams will be written, closed book and in English.

**Academic Honesty**

All course work you submit must be done on your own. Note that academic dishonesty includes not only cheating, fabrication, and plagiarism, but also includes helping other students commit acts of academic dishonesty by allowing them to obtain copies of your work. You are allowed to discuss homework problems and general solution strategies with your classmates, however when it comes to formulating/writing/programming solutions, you should work independently. You may use free and publicly available sources, such as books, journal and conference publications, and web pages, as research material for your answers. (You will not lose points for that.) You may not use any service that involves payment, and you must clearly and explicitly cite all outside sources and materials that you made use of. Failure to comply with these rules constitutes academic dishonesty and will be dealt with harshly. Each academic dishonesty case will be referred to the university administration.