Course Code: AIN2002

Course Name: Introduction to Data Science

SYLLABUS[1]

Instructor

Name	: Mustafa Ümit ÖNER
E-Mail	: mustafaumit.oner@bau.edu.tr
Office	: D531
Office Hours:	: Wednesday 16.30 – 17.20
CV (link)	: https://onermustafaumit.github.io

Course Information

Period	: Spring
Time	: Tuesday 14.30 – 18.20
Course Credit / ECTS	: 3 / 6
Classroom	: D405/Over Teams
Mode of Delivery:	: Online
Course type	: Must
Course ECTS Page Link	: https://akts.bau.edu.tr/bilgipaketi/index/ders/ders_id/46995/program_kodu/04142101/s/4/st/M/ln/en

Course Objectives

Data science is an interdisciplinary field that involves using statistical and computational methods to extract knowledge and insights from structured and unstructured data. It is an important field in today's world because of the increasing amount of data being generated and the need to make data-driven decisions across various industries. Data science is used to solve complex problems in areas such as healthcare, finance, marketing, and social sciences.

By analyzing large datasets, data scientists can uncover patterns, trends, and correlations that are not immediately apparent to humans. They can also develop predictive models that can be used to make informed decisions and optimize business processes. The insights gained from data science can lead to improved efficiency, increased profitability, and better decision-making. Besides, by understanding how data can be used to train and optimize machine learning algorithms, data scientists can contribute to the development of intelligent systems that can make decisions and predictions based on data.

This course aims to help students achieve the following course learning outcomes.

Course Learning Outcomes

The students who have succeeded in this course will be able to:

- Understand the fundamental concepts of data science, including data structures, data analysis, data modeling, and data visualization.
- Acquire knowledge of the basic tools and techniques used in data science, such as Python programming, SQL databases, and statistical analysis.
- Gain experience in collecting, cleaning, and preprocessing data from various sources, including structured and unstructured data.
- Learn how to apply machine learning algorithms, such as regression, classification, and clustering, to analyze and model data.

---

[1] It is essential that the syllabus announced at the beginning of the term is not changed except when necessary. When a requirement occurs, the curriculum can be changed by the lecturer of the course by notifying this situation in writing or verbally beforehand. It is the student's responsibility to follow the current program.

BAU Center for Learning & Teaching

- Explore data visualization techniques and learn how to create clear and effective visualizations to communicate data insights.
- Develop critical thinking skills and learn how to ask the right questions to drive data-driven decision making.
- Understand ethical considerations in data science, such as data privacy, security, and bias, and how they impact the practice of data science.
- Prepare for further studies in data science or related fields and build a foundation for a career in data science or artificial intelligence engineering.

Contribution of the Course to the Program

This course will provide students with valuable skills and abilities that are highly sought after in today's world. These skills include:

- Data analysis: Students will learn how to analyze large datasets and identify patterns and trends, which is a critical skill in many industries such as finance, healthcare, and marketing.
- Programming: Students will learn programming languages such as Python, which is widely used in data science, machine learning, and artificial intelligence. Proficiency in programming is becoming increasingly important in many fields, including finance, engineering, and technology.
- Machine learning: Students will learn how to apply machine learning algorithms to analyze and model data. This is an essential skill in the development of intelligent systems and technologies such as self-driving cars, natural language processing, and robotics.
- Critical thinking: Students will learn how to ask the right questions to drive data-driven decision-making. This skill is essential in today's fast-paced business environment, where companies need to make informed decisions quickly.
- Data visualization: Students will learn how to create clear and effective visualizations to communicate data insights. This skill is critical in fields such as marketing and advertising, where communicating data effectively can make a significant impact.
- Ethical considerations: Students will learn about the ethical considerations in data science, such as data privacy, security, and bias. This is an important skill in today's world where data breaches and privacy violations are becoming more common.

Course Structure

The course consists of two main components: learning foundations of data science and gaining hands-on experience. We will learn the foundations via in-class discussions. Each lecture will have some resources that will provide us with the details of the topic. The resources will be made available before the class. You are strongly urged to go through them before the in-class discussions. We will gain hands-on experience by implementing the concepts and building models using Python as an integral part of the classes.

Course Policies

Communication Channels and Methods:

All the communication (announcements, discussions, submissions, etc.) will go through Itslearning course site.

- You are strongly encouraged to post your questions and comments in the course's discussion forums. If you have any personal issues that you cannot post in the course discussion forum, you can contact your instructor via email.
- You must submit all your work via Itslearning course site. Hard copy submissions and submissions via email will not be accepted.

Usage of Digital Tools:

You are requested **NOT to use** your electronic devices (phone, tablet, laptop etc.) during in-class discussions.

Attendance:

Attendance is not compulsory, yet highly recommended. There is a strong correlation between class attendance and success in the course since it provides a unique experience of interaction and discussion, which promotes learning.

BAU
Bahçeşehir University

BAU Center for
Learning & Teaching

Disabled Student Support:

You can contact me directly regarding the issues that may be an obstacle for you (vision, hearing, etc.). In addition to this, there is a Disabled Student Unit to minimize the difficulties that our disabled students will encounter due to their disabilities and to eliminate the obstacles. You should contact this unit regarding your situation.

Oral and Written Communication Ethics:

You must express yourself respectfully in your communication with everyone in the class. In addition, you are responsible for maintaining this respect in discussions and correspondence on the online platform.

Privacy and Copyright:
In accordance with the Personal Data Protection Law, it is strictly forbidden to register the participants (students and instructors) during the course.

Course Resources
- Adhikari, A., DeNero, J., & Wagner, D. (2022). *Computational and Inferential Thinking: The Foundations of Data Science* (2nd ed.). https://inferentialthinking.com/chapters/intro.html
- Lau, S., Gonzalez, J., & Nolan, D. (2023). *Learning Data Science*. https://learningds.org/intro.html
- Yan, L. and Norouzi, N. *Data 100: Principles and Techniques of Data Science, UC Berkeley, Spring 2023*. https://ds100.org/sp23/
- Gonzalez, J. and Sahai, S. *Data 8: Foundations of Data Science, UC Berkeley, Spring 2023*. https://www.data8.org/sp23/
- Grimson, E., Guttag, J., and Bell, A. *6.0002 Introduction To Computational Thinking and Data Science, MIT, Fall 2016*. https://ocw.mit.edu/courses/6-0002-introduction-to-computational-thinking-and-data-science-fall-2016/

Additional resources will be announced/made available via Itslearning course site.

Course Calendar
Resources specific to each topic will be announced/made available via Itslearning course site.

Week	Course Topic	Assignments
W1	Introduction	Kaggle certificate-I HackerRank certificate-I
W2	Pandas	Kaggle certificate-II
W3	Exploratory Data Analysis and Data Cleaning	Kaggle certificate-III
W4	Text Wrangling and Regex	Kaggle certificate-IV
W5	Visualization	Kaggle certificate-V
W6	SQL	Kaggle certificate-VI HackerRank certificate-II
W7	Confidence Intervals, Sampling, and Standard Error	
W8	Dataset preparation and Modeling-I	Kaggle certificate-VII
W9	Modeling-II	Kaggle certificate-VIII
W10	Feature Engineering	Kaggle certificate-IX
W11	Clustering	HackerRank certificate-III
W12	Classification	
W13	Project presentations	
W14	Project presentations	

Grading and Evaluation

Assignment	Description	Weight (%)
Pop-up Quizzes	• There will be 10 **pop-up quizzes in the class**, weighted equally. • Each quiz may cover topics of previous weeks and the current week.	10
Programming Certificates	• Nine Kaggle certificates, each with one point • Three HackerRank certificates, each with two points • Please see the weekly schedule for the required certificates, details of which will be announced during the semester	15

Project	• The project will be announced around week-3.
	• The project will be conducted as a group of three people. You are required to form your group **by the end of week-3**.
	• **Final code and report (15 points)**
	  o **Submission deadline: 28.05.2023 @23.59**
	  o The report is strictly limited to **four pages including references**.
	    ▪ Should include: Introduction, Related works, Materials and methods, Results, Discussion, References
	    ▪ The report must be in PDF format following NeurIPS paper style and typeset in LaTeX.
	    ▪ https://neurips.cc/Conferences/2022/PaperInformation/StyleFiles
	    ▪ Online editors such as Overleaf may be helpful in writing reports: https://www.overleaf.com.
	  o Code must be deposited at the GitHub and have a step-by-step explanation in README to reproduce the results.
	    ▪ Tips for publishing research code and example README file: https://github.com/paperswithcode/releasing-research-code
	• **Final presentation (10 points)**
	  o **During last two weeks of the semester in the class**
	    ▪ The order of presentations will be based on group registration time: **last-come-first-serve**.
	  o It is strictly limited to **5 minutes**
	    ▪ Should include: Introduction, Related works, Materials and methods, Results, Discussion, References
		25

Midterm Exam	There will be one midterm exam covering all subjects in the first 8 weeks. **It will be given during Week-9.**	10
Final Exam	There will be one final exam covering all subjects. **It will be given during final exams.**	40
TOTAL		100

Matters Needing Attention

• Read all weekly course materials before coming to the class and get prepared for in-class discussions.
• Attend the lectures actively every week.
• Participate in and contribute to in-class activities and discussions.
• Please be careful about requirements to be eligible to take the final exam.

BAU
Bahçeşehir University

BAU Center for
Learning & Teaching

Academic Integrity, Cheating and Plagiarism

Academic integrity is a serious matter in this course. Any violation will be reported to the university's highest levels and maximum punishment will be argued for.

You are encouraged to form study groups to discuss about course assignments. However, you must write down your own solutions without referring to the notes taken during discussions. It is a violation of academic integrity to copy, refer to or look at written solutions and code from another student or any other sources. It is also a violation of academic integrity to post your solutions and code online. Besides, you should write down the names of your collaborators in your submissions.

You are expected to comply with the University Policy on Academic Integrity and Plagiarism. Violations of the university policy can result in severe penalties, including failing this course and possible expulsion from Bahçeşehir University. If you have any questions about this policy and any work you are doing in the course, please feel free to contact your instructor for help.

ARTICLE 25 – (1) In case it is doubled that a student cheats or attempts to cheat, commits plagiarism or similar violations defined in the applicable disciplinary regulation in any exam, assignment or other assessment activities, a disciplinary proceeding is brought against the student. Such activity is not assessed during the proceedings. A student who is found guilty is assigned zero point in addition to the disciplinary punishment. If the student is found innocent because of disciplinary proceeding, the exam taken by the student shall be assessment or a make-up exam or activity is provided.

You can access Bahçeşehir University and Higher Education Institution Regulations by clicking this sentence.

Prepared by: Mustafa Ümit ÖNER
Date of Preparation: 07.03.2023