

Hacettepe University
Department of Industrial Engineering

EMÜ 430- Data Analytics

2023-2024 Fall

Friday 13:40 – 16:30

Faculty Member Information

| Instructor | E-mail | Office hours |
|-------------------|----------------------------|---|
| Erdi Dasdemir | edasdemir@hacettepe.edu.tr | Office 319, by appointment Slack, 7/24 |

General Information

Course objectives

- provide basics and usefulness of data analytics
- develop hands-on skills for applying tools and techniques of data analytics
- collecting and reshaping raw data having different formats
- implementing advanced visualizations to extract usable conclusions
- using productivity tools to organize data analytic projects and generate reproducible reports.

Course Content

- fundamentals of data analytics
- programming basics for data analytics (basic syntax, data types, vectors, vector arithmetic, indexing, conditional expressions, loops, iteration, functions)
- principles and techniques of data visualization
- data wrangling: import, tidy, and process data (importing spreadsheets, web scraping, reshaping data, combining tables, string processing)
- productivity tools for organized and reproducible data analytics projects (basic Unix, Git and GitHub, markup languages for reproducible reports)

Programming Language

- this course uses R as the programming language

Web Page and Communication Server

- Learning Management Platform: www.hadi.hacettepe.edu.tr
- Communication Server: Slack - *EMU Hacettepe Özel Grubu*

References

- Irizarry, Rafael A. (2019). Introduction to data science. CRC Press, FL, US, ISBN-9780367357986
- Peng, R. D., & Matsui, E. (2016). The art of data science. A Guide for Anyone Who Works with Data. Skybrude Consulting, LLC, ISBN-9781365061462
- Freeman, M. & Ross, J. Programming skills for data science, Pearson Education India, ISBN – 9789353942632,
- Zamora Saiz, A., Quesada González, C., Hurtado Gil, L., & Mondéjar Ruiz, D. (2020). An Introduction to Data Analysis in R. Springer, Cham, Switzerland, ISBN- 978-3-030-48996-0
- R for Data Science (2e) <https://r4ds.hadley.nz/>
- MEF BDA530 – Essentials of Data Analytics by Berk Orbay <https://mef-bda503.github.io/>

Assessment

| Activity | Percentage |
|---|------------|
| Assignments | %15 |
| Project: | |
| (Deliverables %20) | |
| (Presentation %5) | %25 |
| 1-2 minutes video teaser for the project (up to +5) | |
| Midterm (in-class exam) | %20 |
| Final ((in-class exam) | %40 |
| Bonus | |
| Datacamp assignments (up to +10) | |
| participation in course content (up to +5) | + |
| participation in invited talks (up to +5) | |

Activity Details

GitHub Repositories

- **GitHub Classroom** will be our assignment submission platform.
- Each of you will have two repositories for code submissions:
 1. **Individual Repository** for the individual assignments
 2. **Group Repository** for the project deliverables
- Invitation links will be provided to your email addresses.

Datacamp

- We will be utilizing Datacamp courses as part of our learning resources. Normally, Datacamp charges monthly fees, but as a student in this course, you won't have to pay anything. Invitation links will be provided soon.
- There will be optional Datacamp assignments during the term, which will be evaluated as a bonus added to your overall grade. The maximum bonus you can accumulate from this is 10 points. The availability of 10 points depends on the number of assignments given to you and your completion rates.

Assignments

- There will be 2 homework assignments throughout the term. These assignments will be done individually.
- The assignments will be interconnected, with the second building upon the first one. Consequently, all assignments are essential and will be evaluated as a whole at the end.

Project

- You will conduct a comprehensive data analysis using relevant real-life datasets.
- The project will be done by teams of students (4-6 students in a team, except for some special cases).
- There will be at least 3 deliverables during the term. Requirements of each one will be given to you. Your submissions will be through your public GitHub Pages. There may be additional printed reporting requirements.

Participation in the Course Content

- This course is designed to be accessible 24/7 through our communication channels, including the Learning Management System and Slack. You are expected to engage by asking questions, responding to your peers' inquiries, and enhancing the course content with your feedback. Your participation will be monitored throughout the term. The maximum bonus you can accumulate through this activity is 5 points (will be added to your overall course grade).

Invited Talks

- Occasionally, we may have experts from the field of data science as guest speakers in our course. Attending each speech and preparing your questions before the sessions can earn you 2 points (1 point for participation and 1 point for your question). The maximum bonus you can accumulate through this activity is 5 points (will be added to your overall course grade).

Course Outline Weekly

| Week / Date | Subject | Activity (submission) |
|--------------------------------|--|-----------------------|
| [Week 1] October 6, 2023 | Introduction to data analytics and R language | |
| [Week 2] October 13, 2023 | Programming and R basics for data analytics (data types, vectors, vector arithmetic, indexing) | |
| [Week 3] October 20, 2023 | Programming and R basics for data analytics (conditional expressions, loops, iteration, functions) | |
| [Week 4] October 27, 2023 | Productivity tools for data analytics projects (git and GitHub, basic Unix) | |
| [Week 5] November 3, 2023 | Productivity tools for data analytics projects (reproducible reports, Quarto, R markdown) | Assignment 1 |
| [Week 6] November 10, 2023 | Data visualization (plot types, graph components, layers, customization, scales, labels, colors) | |
| [Week 7] November 17, 2023 | Data visualization (grouping, sorting, faceting, transformations, data visualization principles) | Project Deliverable 1 |
| [Week 8] November 24, 2023 | Data visualization: Case study | Assignment 2 |
| [Week 9] December 1, 2023 | Midterm Exam | |
| [Week 10] December 8, 2023 | Data wrangling: data import (importing spreadsheets, web scrapping) | |
| [Week 11] December 15, 2023 | Data wrangling: reshaping and tidying data | Project Deliverable 2 |
| [Week 12] December 22, 2023 | Data wrangling: string processing (splitting, removing, extracting, replacing, regex) | |
| [Week 13] December 29, 2023 | Data wrangling: Case study | Assignment 3 |
| [Week 14] January 5, 2023 | Project presentations and discussions | Project Deliverable 3 |

Invited Speakers

| Week / Date | Invited Speaker | Details |
|--------------------------------|----------------------------|---|
| [Week 3] October 20, 2023 | Cem Vardar (Ph.D.) | Data Scientist, Former data scientist at Intel (2002-2010), Revionics (2010-2015) and Carvana (2015-2021), Founder of Decision Science Lab https://decisionsciencelab.com/cem_with_blogs |
| [Week 7] November 17, 2023 | Mustafa Baydoğan (Ph.D.) | Data Scientist, Associate Professor of IE, Founder of Algopoy, http://www.mustafabaydogan.com/ |
| [Week 10] December 08, 2023 | Baykal Hafizoğlu (Ph.D.) | Data Scientist, Platform Management Director, FICO, Atlanta, USA |
| [Week 12] December 22, 2023 | Sami Serkan Özarık (Ph.D.) | Data Scientist, Operations Research Scientist, Amazon, Luxembourg |