

Data Science for Business (M-IBM-2.4)

Syllabus

Prof. Dr. Stephan Huber

2024-01-01

Regardless of position and discipline, evidence-based decision-making is a key competence in business, especially in an increasingly data-driven environment. Managers need to be able to situate practical problems in scientific contexts, generate reliable arguments from scientific discourse, critically evaluate research findings, and eventually interpret and present data they or the company have gathered themselves. This module furthers qualitative and quantitative research by enabling them to apply said methods in a data-driven business environment.

Learning Outcomes

After successful completion of the module, students are able to:

- demonstrate their proficiency in using computer programs to solve business problems,
- identify, process, visualize, and analyze data sets using the programming language R,
- explain various methods of data science, such as data mining, regression analysis, or clustering,
- assess and apply data science methods to solve problems in practical and professional contexts as well as in theoretical and scientific contexts,
- visualize solutions and communicate the challenges and outcomes of a data analysis process in written and oral form.

Lecturer

Prof. Dr. Stephan Huber

- *Email:* Stephan.Huber@hs-fresenius.de
- *Office:* Building 4b OG1 Room 1
- *Office hours:* Thursday 1-2pm and [upon request](#)

Lecture Times

13 meetings scheduled for Thursdays from 13:05 to 16:15, to be held in either HS 4c OG2 or SR4, spanning calendar weeks 11 to 17 and 20 to 25.

Module Content

Scope and Nature of Data Science

- What defines Data Science (Artificial Intelligence, Machine Learning, Big Data, Data Mining, Data Analytics, Statistics, Information Technology, Computer Science)
- The Importance of Data Science in Businesses
- Data Science Process
- Automated Decision-Making in Businesses
- Overview of Data Science Software Tools

The Programming Language R

- Installation (R, RStudio)
- Data Management and Writing Scripts
- Rmarkdown
- Git and Github

Data Visualization

- Descriptive Statistics
- Data Mining
- Empirical Models (Regression, Clustering, Difference-in-Difference)

Case Studies: Application of Data Science Methods

- Data Mining
- Regression Analysis
- Clustering and Similarity Analysis

Weekly Learning Plan

Meeting Content		Readings
1	Introduction, project description, R installation, swirl, PC basics	Huber (2023a)
2	R basics, R scripts, assignment operator, data import, swirl	Huber (2023a); Wickham & Grolemund (2023, ch. 3)
3	Data visualizations, descriptive statistics	Huber (2023a); Wickham & Grolemund (2023, ch. 2)
4	Data management, tidyverse, dplyr	Huber (2023a); Wickham & Grolemund (2023) [ch. 4], Neth (2023, ch. 3)
5	Quantitative analysis (causal inference)	Huber (2023b, ch. 1-2)
6	Quantitative analysis (data mining)	Huber (2023b, ch. 1-2)
7	Exercises (data management)	Huber (2023a); Neth (2023, ch. 4)
8	Exercises (exploratory data analysis)	Huber (2023a); Neth (2023, ch. 4)
9	R Markdown	Huber (2023a)
10	Regression Analysis	Huber (2023b, ch. 3)

Meeting Content		Readings
11	Git, GitHub	
12	Student presentations	
13	Student presentations	
14	Review	

Workload

125h Overall

56h physical synchronous contact hours

69h private study hours

Learning Material:

Can be found on ILIAS.

Assessment Methods and Criteria

Students complete this module with a project work. The project work includes a project report (15-20 pages) and a project presentation (20-30 minutes).

The project is explained in a separate document.

Course Policy

- Punctuality is expected. Please arrive on time for class.
- Refrain from taking photos or recording during the class.
- Always take your laptop with you to class.
- Feel free to ask questions during the class or reach out via email (stephan.huber@hs-fresenius.de) after the class. Your inquiries are welcome and encouraged.

Syllabus Disclaimer

Please note that the syllabus is subject to change. While I will make every effort to inform you in advance of any modifications, there may be instances where adjustments are necessary to enhance the learning experience or address unforeseen circumstances. It is your responsibility to stay updated on any changes, so I recommend regularly checking the course announcements and your email for notifications regarding syllabus updates. Your flexibility and understanding in such situations are greatly appreciated.

Literature

- Huber, S. (2023a). *How to use R for data science: Lecture notes*. Accessed September 11, 2023. <https://hubchev.github.io/ds/>
- Huber, S. (2023b). *Quantitative methods: Lecture notes*. Accessed September 11, 2023. <https://hubchev.github.io/qm/>
- Neth, H. (2023). *ds4psy: Data science for psychologists*. Social Psychology; Decision Sciences, University of Konstanz. <https://doi.org/10.5281/zenodo.7229812>
- Wickham, H., & Grolemund, G. (2023). *R for data science (2e)*. Accessed November 30, 2023. <https://r4ds.hadley.nz/>