|  |  |
| --- | --- |
| **Module description** | Data Science 1  This module introduces students to the fundamental concepts and techniques for extracting useful knowledge from data representations thereof. These concepts are of three types. First, it will discuss data-analytic thinking, and introduce data science standards that are commonly used in the industry. Second, the module introduces students to concepts in statistics and probability theory that form a basis for modern data science. Lastly, the module will zoom in on data analysis, visualization and reporting. |
| Code: | DS1 |
| Title: | Introduction to Data Science: Statistics & Probability |
| ECTS: | 4 |
| Intended Learning Outcome: | **ILO 4. Researching and Analysis:**  1) The student demonstrates a full understanding of different data types, descriptive data, graphical representations of data, statistical inference, basic probability theory, correlation and simple linear regression.  2) The student can apply the CRISP-DM, sigma-notations and linear transformations.  **ILO 5. Conceptualizing**  3) The student can transform a business requirement into a data science problem and propose an effective solution.  4) The student is able to formulate an insightful data-driven research question, quantify appropriate real-world phenomena and objects into data and covert this data into meaningful graphical representations thereof. |
| Pre-requirements: | None |
| Approach: | 3 days a week where every 8 hour course day is structured as follow:   * Video-lecture for presenting and discussing the main concepts, using practical examples: 30 – 1.5 hours * Interactive Workshop, at home or during Data Lab with lecturer support: 4 - 6 * Interactive Mock Assessment: create building-blocks, at home or in Data Lab: 1 – 3 Hours |
| Topics by week: | |  |  |  | | --- | --- | --- | | **Week** | **Day** | **Lecture-Workshop Hybrid** | | 3 | 1 | Introduction to Data Science(codifying the world in data, attributes) | | 3 | 2 | Variables (data frames, continuous, nominal, ordinal etc.) | | 3 | 3 | Descriptive analyses (mean, sd, range, IQR) & visualisation (boxplots) | | 4 | 4 | Introduction to Probability (random variable, distributions) | | 4 | 5 | Introduction to statistical inference (sample, population, hypothesis testing) | | 4 | 6 | Analysing Relationships between variables (e.g., compute correlation by hand) | | 5 | 7 | Reporting & visualising (Academic Skills) | | 5 | 8 | Introduction to simple regression & machine learning (supervised, unsupervised, the CRISP DM model) | | 5 | 9 | Conference Poster Creation | |
| Assessment: | Create a Conference Poster based on case studies and examples introduced online and in-class; the focus is on all stages of the data science process (formulating the problem; selecting the data and exploratory data analyses technique(s); interpretation of descriptive data; reporting and visualizing the data) from an elementary level with specific emphasis on data visualisation as a means for understanding the data . Propose next steps: an analysis to run on the data. |
| Literature/Resources: | * OnlineStatBook: [Online Statistics Education: A Free Resource for Introductory Statistics (onlinestatbook.com)](https://onlinestatbook.com/2/index.html) * Learning Statistics with R, by D. Navarro (2018): <https://learningstatisticswithr.com/> * Discovering Statistics using R, A. Fields et al (2012): <https://uk.sagepub.com/en-gb/eur/discovering-statistics-using-r/book236067> * Swirl package for Interactive Programming Courses in R: <https://swirlstats.com/> |