Data Exploration and Visualization

Welcome to the Data Exploration and Visualization (DEVI) course. This course is mainly designed for M.Sc. Information Systems students interested in the initial phases of data science, that is, in exploratory data analysis.

Through a series of (mostly) practical lectures and exercises, you will build a strong foundation on what one does when performing data exploration and data visualization activities, mainly with Python, Pandas, and Matplotlib.

Learning Goals

The course is designed to equip students with the skills required for in-depth data analysis and the creation of effective visualizations. Focusing on Python, the course introduces essential libraries such as Pandas and Matplotlib for data manipulation, exploration, and visualization.

Students will engage in exploratory data analysis techniques to uncover patterns, trends, and correlations within datasets, enhancing their understanding and interpretation of data.

The module also incorporates principles of design to ensure visualizations are not only informative but also accessible and aesthetically appealing.

Through practical projects, participants will learn to craft visualizations that accurately communicate their findings, preparing them for a career in data science by emphasizing analytical rigor and clarity in data presentation.

Course Structure

This course is structured to provide both theoretical knowledge and practical experience.

It consists **formally** of:

- 12 lectures
- 5 labs

But the execution will be in a hybrid format:

- Some content is in person, in the classical frontal lecture format.
- Some content is in person, as exercises in class.
- Some content will be in the form of a video lecture and a related Jupyter Notebook.
- Each lecture as a video has a corresponding Q/A session in class, which may also contain additional material and/or exercises.

The lectures cover the following topics:

1 - Fundamentals of data analysis (frontal lecture)
2 - Python refresher (frontal lecture)
3 - NumPy (frontal lecture)
4 - Pandas data structures (video)
5 - Data wrangling with Pandas I (video)
6 - Data wrangling with Pandas II (video)
7 - Aggregating Pandas DataFrames I (video)
8 - Aggregating Pandas DataFrames II (video)
9 - Visualizing Data with Pandas and Matplotlib I (video)
10 - Visualizing Data with Pandas and Matplotlib II (video)
11 - Formatting plots with Matplotlib (video)
12 - RSS Best practices for data visualization (frontal lecture)

Make sure to have a look at ILIAS to see what is being offered and when. **You need to watch the video(s) before attending the corresponding in-class event**.

Prerequisites

The course is as self-contained as possible, and it does not have any formal prerequisites.

There is recommended knowledge you should have:

- Basics of Python programming (acquired during the 1st semester of the MSc IS)
- Basics of statistics

In this course, you will program extensively, using Jupyter Notebooks. You do not need to know what a Jupyter Notebook is before starting the course.

You need a laptop for most of the sessions.

Course Materials

The course is 100% self-contained. No book purchase is needed. Attending lectures and labs, watching the videos, and studying from the slides (and notebooks) covers all that is needed.

While no textbook purchase is required, the major sources for the course content are:

- Molin, Stefanie. Hands-On Data Analysis with Pandas: A Python Data Science Handbook for Data Collection, Wrangling, Analysis, and Visualization. Packt Publishing Ltd, 2021.
- Please note that the theory comes from this book, but the examples all deviate from the book and are custom-tailored for the course.
- Nicolas P. Rougier. Scientific Visualization: Python + Matplotlib. Nicolas P. Rougier, 2021, 978–2–

9579901–0–8. hal–03427242
- Online documentation for NumPy: https://numpy.org/doc/stable
- Online documentation for Pandas: https://pandas.pydata.org/docs
- Online documentation for Matplotlib: https://matplotlib.org/stable/index.htm

Course Evaluation

The evaluation for this course is based on a final written exam.

Contact Information

The course is held by:

- Prof. Dr. Daniel Graziotin, graziotin |AT| uni-hohenheim |DOT| de, lecturer, exerciser, examiner, and responsible for the module.