Programming for Data Scientists: Project 3

Project Setup:

In this project, you should evaluate the data given in one of the following data sets. You can also include data from additional data sets you can e.g. find on Kaggle (`www.kaggle.com/datasets`), Data World (`https://data.world/`) or other web sites.

Group 1 Manifesto Project: `https://manifestoproject.wzb.eu/`

Group 2 Democracy Barometer:
`https://democracybarometer.org/data-and-documentation/`

Group 3 Movies: `https://www.kaggle.com/datasets/rounakbanik/the-movies-datase`

Group 4 Human Development `WorldIndex:https://www.kaggle.com/datasets/iamsouravbanerjee/human-development-index-dataset`

After choosing a research question you should evaluate the data using **pandas** and visualize your results with the help of **Matplotlib**.

Project description:

You will be informed of the group you are assigned to via Email.

Then, proceed as follows:

(1) Specify the research question(s) you would like to answer.

(2) Collect all data you need to answer the question.

(3) Write the Python code to filter the data and to create adequate visualizations for the results. It might be a good idea to create some intermediate data structures (like dictionaries, lists, etc.) with filtered data which can be used more easily to create your results.

(4) Cross check your results (e.g. with literature and information from the internet).

(5) Prepare a presentation where you describe the steps you made (including the code you wrote) and the results you got. Don't forget to list all data sources you used.

(6) Upload all documents (slides, Python code) to Ilias.

In the end, you will have to give a short presentation of around 30 minutes per group (including discussion). In this presentation, you should describe your research question(s), the code you developed and your results. The group presentations will take place in the Pipool, on 28. November, 10:15 - 12:30h.