Data Science - Exercises
Institut für Informatik, Christian-Albrechts-Universität zu Kiel
WiSe 2023/24
Prof. Dr. Matthias Renz, Christian B. Niko P., Yannick W.

Data Science Assignment 1 - Data Acquisition and Integration

0 Info

- Please hand in one pdf that covers your solution for task 1 and exports of the jupyter notebooks for tasks 2 and 3.
- Only one upload per Team, please.
- Both English and German (and mixtures thereof) are accepted.
- In order to "pass" this assignment you need to hand in sensible solution attempts for all the tasks on this assignment sheet.
- The course element for uploading your solution will be enabled on Monday November 20th.
- The pdf export for the notebooks must include the output of the code cells
- Deadline: 23:59 November 26, 2023

1 Data Acquisition

1. Visit https://pangaea.de/ and select any dataset you deem interesting.
2. From the paper *Gebru et al.: Datasheets for Datasets, 2020* (https://arxiv.org/abs/1803.09010) select at least one question in each of the subcategories 3.1-3.7. Answer the questions with regard to the dataset you chose in (1.). If you can't answer a question, explain why the missing information would be relevant.
3. Chose one of the questions you could answer from (2.) and briefly explain why it is relevant for using the dataset. If applicable give example uses of the data, where the information is particularly relevant.

2 Pandas tutorial

In the Moodle course you find the folder "Homework_01_Material". The archive named "1_2_Pandas_workshop.zip" contains a Jupyter notebook titled "workbook.ipynb," sourced from https://github.com/stefmolin/pandas-workshop. While we suggest engaging with the workshop notebooks, you have the liberty to tackle the exercises in the provided workbook by any means.

1. Solve the tasks in "workbook.ipynb".

3 Fairness

In the Moodle course you find the folder "Homework_01_Material". The archive named "1_3_Fairness.zip" encompasses a Jupyter notebook titled "Data Bias - Fairness Gerrymandering.ipynb".

1. Solve the tasks in the provided notebook.