Scripting Project

In the project work, you will combine the theoretical and practical learnings from previous lectures and exercises. The overarching goal is to demonstrate scripting literacy in a convincing, results-oriented project. You will eventually produce a data processing application that runs live on the Internet.

The application needs to fulfil a number of criteria to pass the project:

- It needs to load data from at least one, better two different sources over the network. (Use the well-known public open data portals for this matter.)
- The data needs to be cached intelligently with a timeout of one minute. (It is recommended to use higher timeouts during the development.)
- The data needs to be cleaned up. Data cleansing, preprocessing, reordering etc. are essential task to keep the actual data analysis lean.
- A correlation of features needs to be performed. Data also needs to be aggregated and/or resampled. Aggregation and correlation can be combined but can also be done separately.
- Either a statistical analysis or a machine learning process needs to be demonstrated.
- Afterwards, the resulting data needs to be visualised in properly labelled plots.
- Finally, the solution needs to run on a server as Jupyter notebook. Alternatively, you can also opt in for running as web application on a VM or cloud function with web action functionality (resources provided by the lecturer).

You will need to show some flexibility in choosing the dataset(s) because each set must only be used by one course participant. The course assistant will maintain a registration list to guarantee uniqueness.