

DOPP Workplan - Group 29

Serban ..., Fritz ..., Dennis ..., Tobias Slowiak, 01204691

Introduction Our Group has decided for Question 21, which deals with the topic of nuclear energy. In the following, we draft a work plan for this project. We structured the work plan according to the data science process discussed in the lecture.

Ask interesting Questions We have the following basic questions:

How has the use of nuclear energy evolved over time? How well does the use of nuclear energy correlate with changes in carbon emissions? Are there characteristics of a country that correlate with increases or decreases in the use of nuclear energy?

Additional Questions regarding this topic, that we want to consider are:

- still missing.

We reserve the possibility to adapt these questions in the process.

Concluding task: give a description in the Jupyterhub notebook on why we chose our Questions and why they are good questions.

Get the data First, we will find some datasets until we feel like we have enough data to start the rest of the process.

For a start, we will have a look at the following datasets:

- Eurostat database: energy.
- UN-data.
- statista: Nuclear electricity generation worldwide from 1985 to 2021 (free?)

still find more possible sources. only the first 2 are from the list and they seem to be pretty poor :/.

Additional thing: Is anyone familiar with kaggle kernels/wants to set one up?

In this step we will also clean the data and filter out what we need. This will also include handling missing values and outliers. Should we find inconsistencies in the data in the next step, we will come back to this step and treat them here. We might also find that we need some additional data, that is not directly connected to nuclear energy, then we also come back to this step and collect more data.

Concluding task: Give an overview and description of the datasets in the Jupyterhub notebook. Also describe what was done to the data.

Explore the data This step is intertwined with the previous step. We will Explore the data, make provisional visualizations etc. to get a better feeling for what we are working with. Here we also have to think about if the necessary data is there to answer the questions.

The question *Are there characteristics of a country that correlate with increases or decreases in the use of nuclear energy?* is especially interesting here. We can find hints for what to model later in this step.

Concluding task: Discussion.

Model the data Here, especially the question *How well does the use of nuclear energy correlate with changes in carbon emissions?* will be interesting for modelling. We will have to be quite careful to work out the correlation, since the carbon emissions are determined by a multitude of factors.

Also, the question *Are there characteristics of a country that correlate with increases or decreases in the use of nuclear energy?* can be modelled here, where we will try to work out correlations, that may have shown in the previous step.

Concluding task: Discussion.

Communicate and visualize the results The question *How has the use of nuclear energy evolved over time?* will be answered here by visualizing different aspects of nuclear energy and how they evolved over time.

Concluding task: Answer/Discuss the questions. Talk about how the work was spread between the members.

Also make a 2-page PDF as stated in the exercise statement.

Also pack the data in a zip file. And in case some of the datasets are too large, provide a `install_data.txt` as stated in the exercise statement.