

Individual report – Rebekka Gorge- 59055

For the dataset project I worked in a group together with Aron Gaden and Annie Witschas. We performed the group project on the dataset “Beijing Multi-Site Air-Quality Data set”. We selected this dataset together because we are all interested in this subject. From the beginning of the project we all worked closely together, and we tried together to get an overview over the subjects for the project. Then we tried to split the tasks in a senseful way.

In the beginning Aron and me worked corporately on the linear regression task, while Annie already started to work on the PCA. Later Aron switched to the clustering part of the project. Even if we split the tasks up, we frequently had a video conference to discuss and recheck the different results.

For the regression analysis Aron and me firstly tried to understand and implement the given tutorial about linear regression. Then we transferred the results to our dataset. To find good features for the linear regression we plotted all features together and considered which features could be interesting and have a useful plot. Both of us implemented the tasks and then we compared our results. In the end Aron started to work on the clustering and I tried to improve some of our results, finished the last tasks and tried to analyse our results.

For the analysis and writing part of our project, Annie worked again on the PCA, Aron on clustering and I analysed and wrote the part about linear regression. If one of us was insecure about a question or had a problem with a task, we tried to find together a solution. It was a good teamwork.

The project was useful to understand the dataset in a better way, and while implementing the parts of the project, it was easier to understand the sense of the single theoretical subjects. As I started to work on linear regression, I did not really understand why we are doing these steps, but now in the end I recognized that it helped a lot to understand the features and to make predictions about their relationship and behaviour. When I started to analyse the results, I also understood why we need to calculate the single steps. At some points it was a bit difficult to find the right equations and to understand completely what we should do in the single tasks. For the linear regression part, it was a bit easier because the tutorial was a big help. I also learned a lot about python in this project, because I did not really use it before.