Assignment for working student candidates

Dear candidate,

One of your tasks at Navigance will be to include data coming from our customers into our database by writing scripts that parses data files and brings it into a standard format. Furthermore, you will be working on some benchmarking tasks, which – in a nutshell – means that data of different chemical plants should be compared with each other in a meaningful way.

We want to see how you approach such tasks. Please find enclosed five datasets in formats as we typically get them from our customers.

Instructions / hints what to do with the data sets:

1. Write simple parsers / scripts which read the files and read them to a pandas DataFrame. The files themselves should not be changed.
2. Bring the data into a standardized format in your DataFrame. Things you might want to consider are, e.g., column names, time format, unit etc.
3. Try to do a meaningful comparison (aka benchmarking) of some variables of your choice (be creative, there is not just one way of doing this).

This mimics a real customer situation, which includes strange data formats, missing information, etc.. In case you feel you are lacking some information, try to guess the missing parts yourself, or reach out to [Philipp.Michel@navigance.com](mailto:Philipp.Michel@navigance.com).

In the technical interview, we will ask you to show your findings. An appropriate format would, e.g., be a Jupyter Notebook or a Python script. The expected time to work on this assignment is not more than ~2-3 hours.

We are looking forward to discussing with you.

Your data science team at Navigance.