Analytical Computing

Lecture 3: Pandas data manipulation

hogeschool **Windesheim**

Radboud University

Planning

20-4-2021 **Git**

22-4-2021 **Array Handling**

28-4-2021 Pandas data manipulation

30-4-2021 Kick-off opdracht

11-5-2021 Visualization with MatplotLib

12-5-2021 StatsModels for Python statistics

21-5-2021 Sklearn for regression learning

4-6-2021 Eindpresentaties

Pandas

Wat is Pandas?

- Open source Python package
  - Just like NumPy from last week
  - Fun fact: Pandas is actually *built upon* the NumPy package
- Most widely used for data science/data analysis and machine learning tasks
- Data can be easily manipulated using Pandas by making use of **dataframes**
- **Dataframe:** a 2-dimensional labeled data structure with columns of potentially different (variable) types
  - You can see it like a simple SQL or Excel table

Wat kan ik met Pandas

- Data cleaning
- Data filling (e.g. when facing missing values)
- Data normalization
- Merge and join other dataframes
- Data visualization (with e.g. matplotlib)
- Statistical analysis
- Data inspection
- Loading and saving data
- …and much more

Hands-on Pandas

Hands-on Pandas

- No real 'theoretical' lecture today

- Just like with NumPy and Git, practice makes perfect

- If you're stuck, all answers can be found here: https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html

1) Fork the repository containing the assigments into your own GitHub account: https://github.com/PeaceDucko/pandas_exercises

2) Clone **your** version of the assignments using Git and the HTTP link provided by GitHub onto your local computer

3) You don't have to look at chapters 6, 7, 8 and 9 (we will cover this in future lectures)

4) Try to complete at least one exercise from each chapter (try not and peek at the solutions)

5) We will go through one example together

Fork 5.2k

Vragen?

hogeschool
Windesheim

Radboud University