DIS08 – Data Modeling

07 – Basic Data Processing with Python

Philipp Schaer, Technische Hochschule Köln, Cologne, Germany

Version: WS 2021

Technology
Arts Sciences
TH Köln

Agenda

Last weeks

- The shell
- GIT / GitHub
- Regular expressions
- Open data
- Data cleaning in a nutshell

This week

- We will start with Python!
- Our final goal is to **program our own web scraper** to gather data from the web for further analyses.

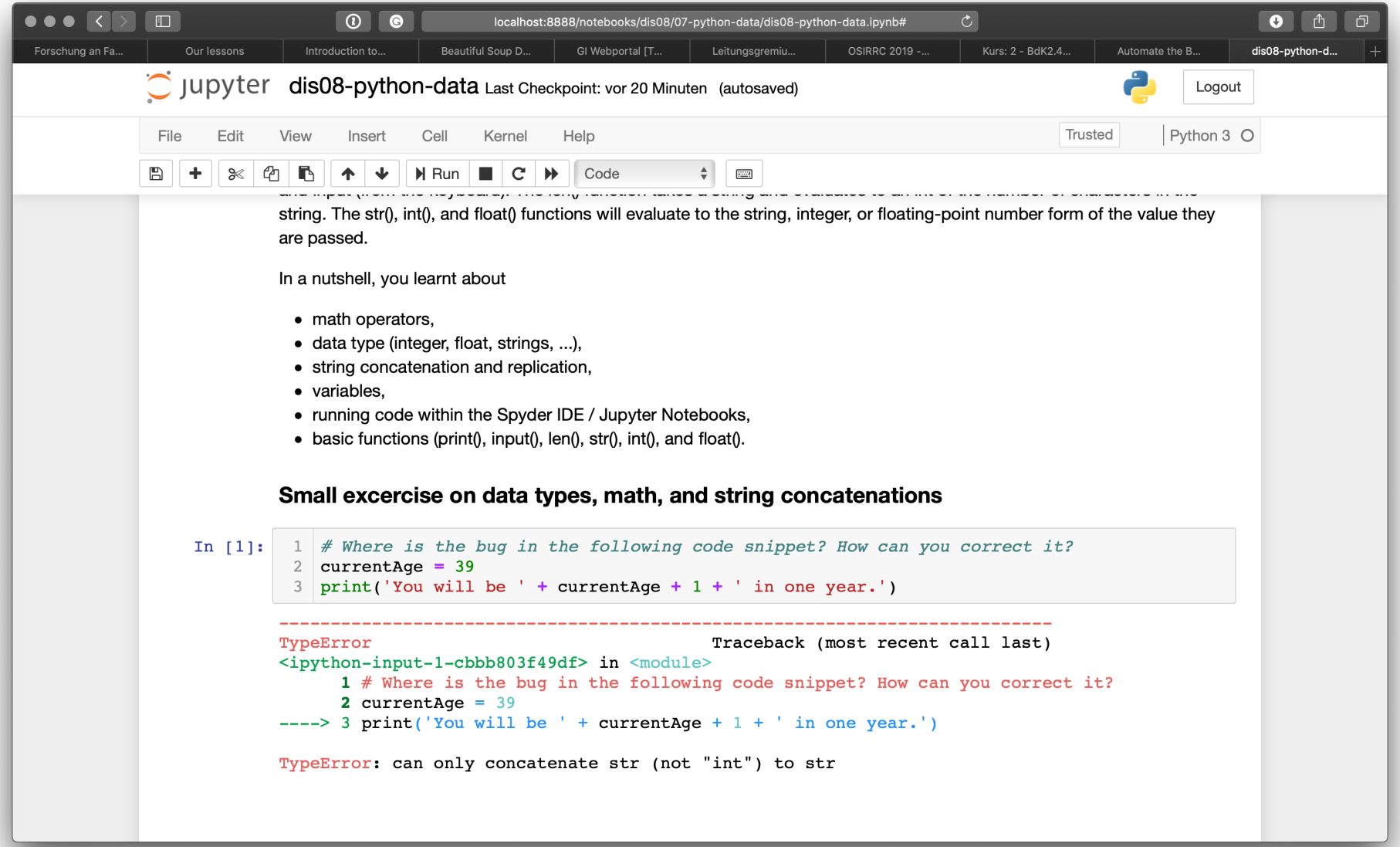
Zahlenchaos Darum war der Inzidenzwert in Köln nicht verlässlich

Screenshot taken from https://www.ksta.de/koeln/zahlenchaos-darum-war-der-inzidenzwert-in-koeln-nicht-verlaesslich-39161766

Schedule for lectures WS 2021/22

15.10.21	Introduction, Markdown		first week (no tutorial)
22.10.21	Unix Shell		Markdown and style guides
29.10.21	Versioning, Git, GitHub		Shell tutorial
05.11.21	Regular Expressions		Hands-on Git, GitHub
12.11.21	CSV, JSON, XML		Regex tutorial
19.11.21	OpenData, Tidy Data Principles		open.cologne open data
26.11.21	Project week (no lecture)		**Submit assigment 1** (no tutorial)
03.12.21	Python: Data structures		Hands-on Python
10.12.21	Python: Files, folders and more		Hands-on Python
17.12.21	Python: Pandas		Hands-on Python
24.12.21	Christmas (no lecture)		Christmas (no tutorial)
31.12.21	Christmas (no lecture)		Christmas (no tutorial)
07.01.22	Python: Web Scraping		Hands-on Python
14.01.22	Q&A - Summary		Hands-on Python
04.02.22			**Submit assigment 2**

We will work with Jupyter Notebooks...



For today we will focus on...

Repeating some basic Python stuff that you already heard from Prof. Strahringer or Dr. Schaible/Dr. Carevic
(very quick and dirty repetition)

- Python basics
- Flow Control
- Functions
- Lists
- Dictionaries and structuring data
- Manipulation Strings
- Regex in Python
- Reading and writing files