Data Science Survival Skills

Introduction

CORONA RULES

- We will check **at each lecture** if you (>= 10%) are either
  - Vaccinated
  - Recovered
  - Tested
- Wear your mask throughout the whole class
- If you don’t feel well, stay at home!
- Register at darfichrein.de
- More information: https://www.fau.de/corona

Agenda

- Who we are
- What to expect
- Administration stuff
- Exam, exercises etc

- What is data science?
- What skills do you need to survive?
- Being a problem solver.

Who we are

Andreas Kist

René Groh

Hernan Aguilera

What to expect

Lectures: We explain how things work

∞

Exercises: You experience how things work

Administration stuff

- Please subscribe to the **StudOn** course!

- Register for the exam on **meincampus**!

- Attendance is not mandatory, but strongly encouraged.
  ➜ You get access to the slides, but I won't guarantee that this is everything you should know.

- Exercises are not mandatory, but strongly encouraged.
  ➜ you get access to the solutions, but if you don't understand them, you should have asked in the exercise!

- Each **successfully** submitted exercise gives up to 1 bonus point 

Lectures + Exercises

Lectures are Monday 10-12

Exercises are Fridays 12-14

All in this seminar room!

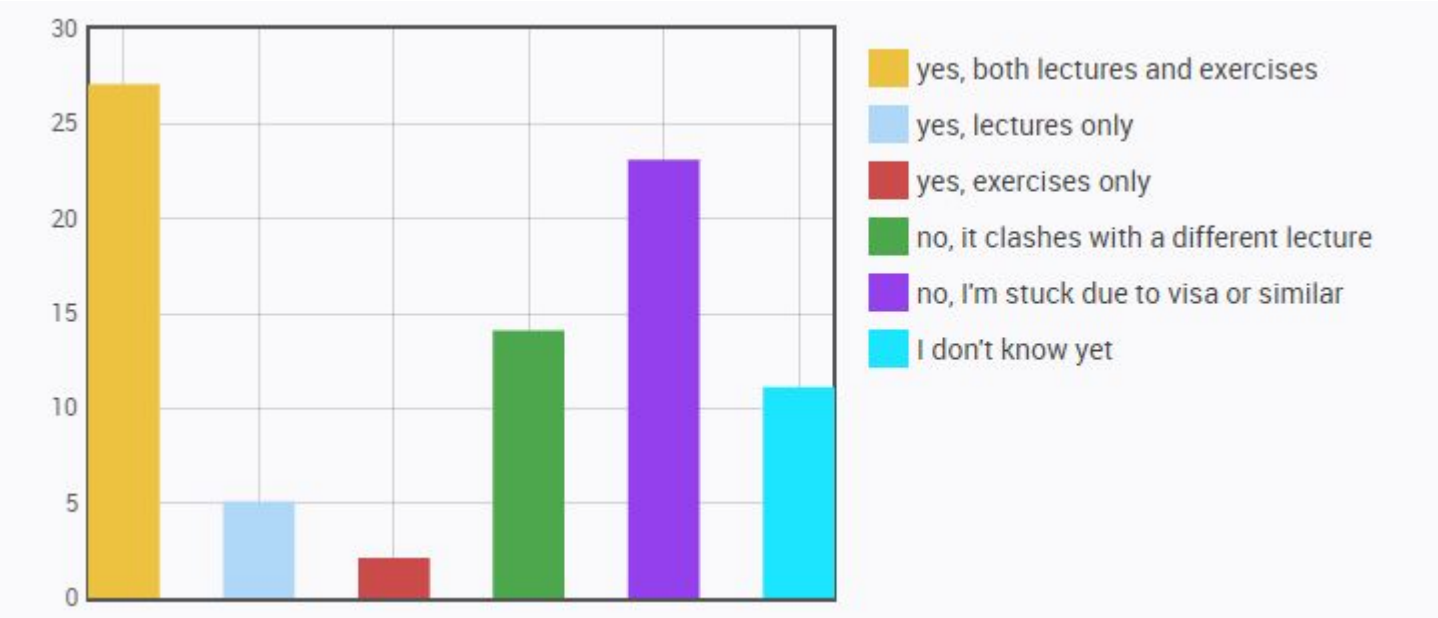
Exercises:
Task + Colab environment
On Fridays, we discuss the exercise and solve parts of it together.

	Dates		
	VL (Monday)	UE (Friday)	Topic
1	18/10/2021	x	Welcome, Overview, DSSS
2	25/10/2021	29/10/2021	Technical Equipment - CPU, GPU, TPU, Jetson, Firefly-DL
3	01/11/2021	05/11/2021	Data handling, ZIP, HDF5, JPG, PNG, MP4, AVI, XML, JSON
4	08/11/2021	12/11/2021	Version control, Python package management, Cookie cutter
5	15/11/2021	19/11/2021	Machine Learning, Deep Learning, Linear/Logistic regression, Correlation
6	22/11/2021	26/11/2021	Webscraping, Beautiful Soap, REST-API, Node.JS
7	29/11/2021	03/12/2021	Staying up-to-date: WhatsApp, Telegram, Notify
8	06/12/2021	10/12/2021	Graphical User Interface, Qt5, PyQt5, pyqtgraph, napari, kivi
9	13/12/2021	17/12/2021	Deploying code, PyInstaller, fbs, NSIS, flask, dask, spark
10	20/12/2021	x	Christmas VL - only fun^^
11	x	x	
12	x	x	
13	10/01/2022	14/01/2022	Multiprocessing and multithreading
14	17/01/2022	21/1/2022	C performance with Python: Cython and numba, Julia
15	24/01/2022	28/01/2022	Visualization I, Plots, Barplots, ... matplotlib, plotly, bokeh
16	31/01/2022	4/2/2022	Visualization II, Images, Colors, Colorspace
17	07/02/2022	x	Conclusion, Q&A, Exam questions

Please submit exercise until Friday noon to get potentially the bonus point! 

Students

- We planned with ~ 20
- We have a room for ~ 50
- We have 135 registered students

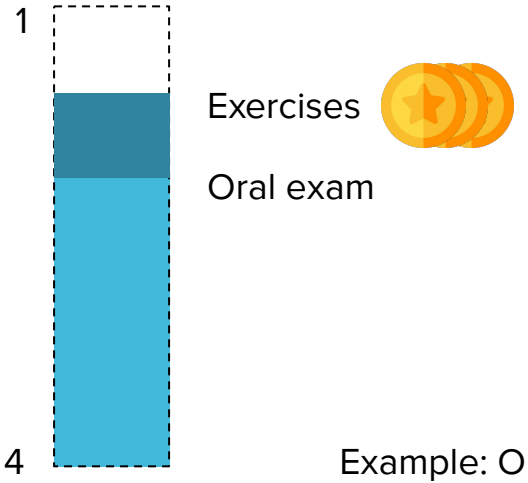


Exam

- Oral exam, max. 30 min **WE RE-EVALUATE**
- Content: Lectures + Exercises
- I am aiming for CONCEPTS and LOGICAL THINKING, not for F***ING DETAILS :-) (SAME FOR WRITTEN EXAM)

0-4 bonus points:	-0.0
5-8 bonus points:	-0.3
9+ bonus point:	-0.7

Grade:



Example: Oral exam 2,3 + 10 bonus points ➜ 1,7

Expectations

Student expectations

Please get in touch with your fellow students and ask yourself the following questions:

- What do I want from the course?
- How can I achieve this?
- How can I actively contribute to the course?
- What do expect from lecturers?

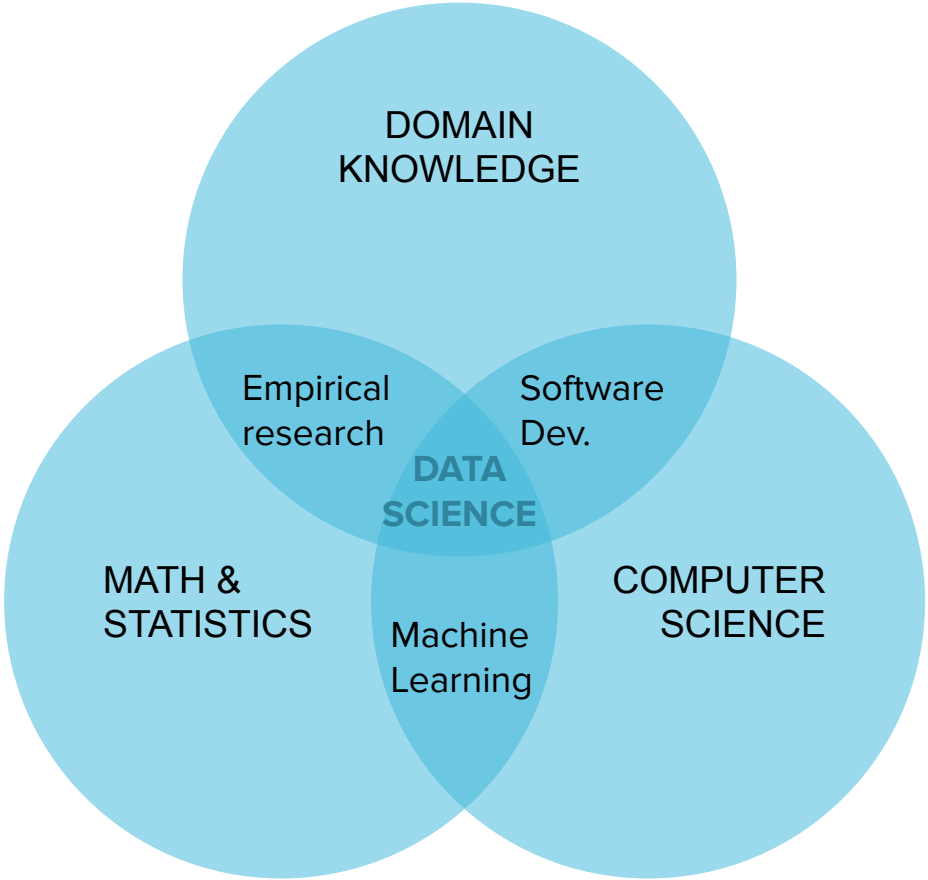
5 minutes

My/our expectations

- Be on time for lectures
- Try to follow and listen, don’t play on your phone
- Do the exercises
- Ask questions
- Use the course forum!

Data Science

What is “Data Science”?

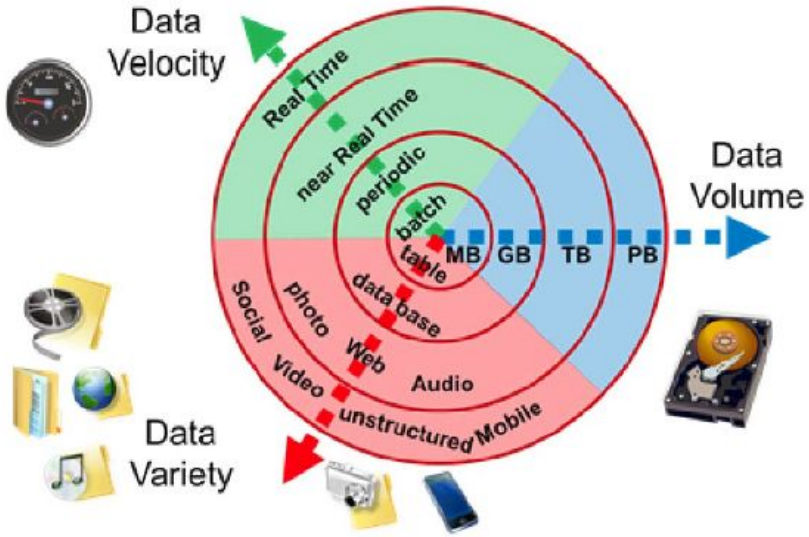


Why do we need it?

BIG DATA

Value	Metric		Value	IEC		Memory	
1000	kB	kilobyte	1024	KiB	kibibyte	KB	kilobyte
1000^2	MB	megabyte	1024^2	MiB	mebibyte	MB	megabyte
1000^3	GB	gigabyte	1024^3	GiB	gibibyte	GB	gigabyte
1000^4	TB	terabyte	1024^4	TiB	tebibyte	TB	terabyte
1000^5	PB	petabyte	1024^5	PiB	pebibyte		–
1000^6	EB	exabyte	1024^6	EiB	exbibyte		–
1000^7	ZB	zettabyte	1024^7	ZiB	zebibyte		–
1000^8	YB	yottabyte	1024^8	YiB	yobibyte		–

Orders of magnitude of data



By Ender005 - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=49888192

Data Science Workflow

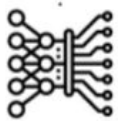
Understand the task

Identify & acquire data

Descriptive and explorative data analysis

Operationalizing and monitoring

Evaluation and acceptance



Modelling and optimization

Data quality

DATADRIVENCOMPANY.DE

DATENQUALITÄT

Der Wert einer hohen Datenqualität

Trust in Data and Analytics

Vertrauen in die Daten und Analysen

Efficient collaboration

Effiziente Kollaboration intern und extern

Stability of Systems and Products

Stabilität von Systemen und Produkten

Save costs, efforts and time

Kosten, Aufwand und Zeit sparen

Compliance Readiness

Zukunftsfähigkeit des Unternehmens

Corporate Sustainability

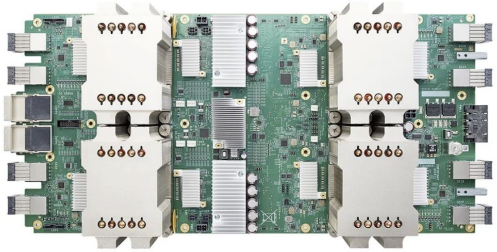
Where is it applicable?

Customer Clustering in Marketing

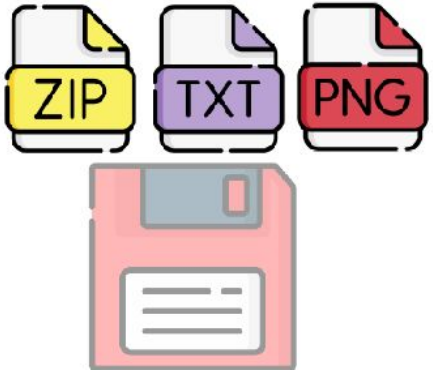
Replacement classification using image recognition in mobile apps

Automated heating prediction in flats

Working with data



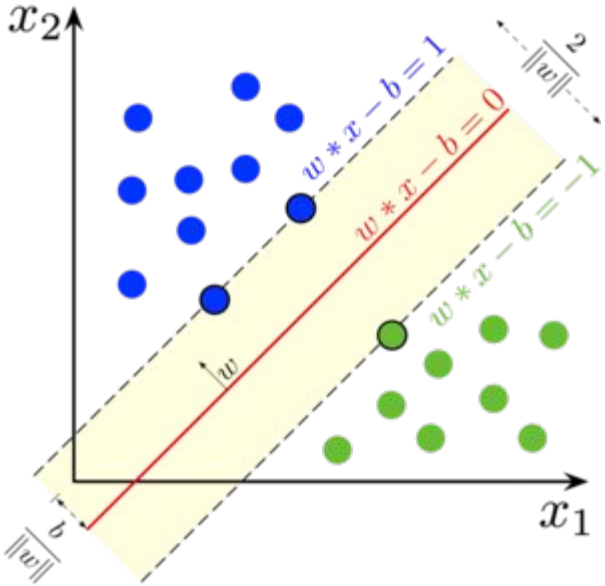
Next Lecture



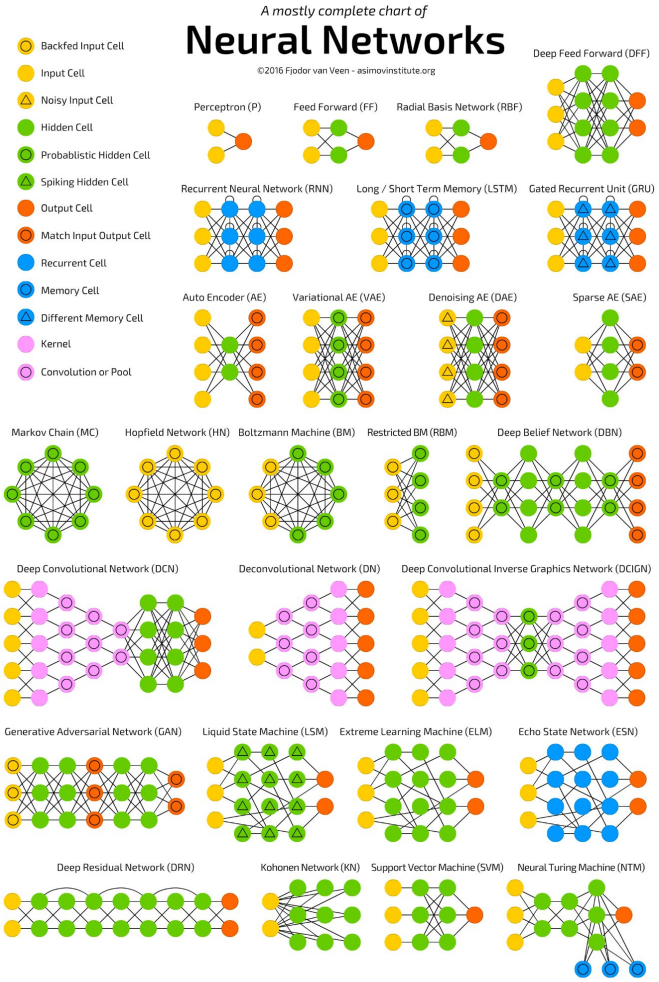
3rd Lecture

Good programming practices, version control, python packaging... (4th lecture)

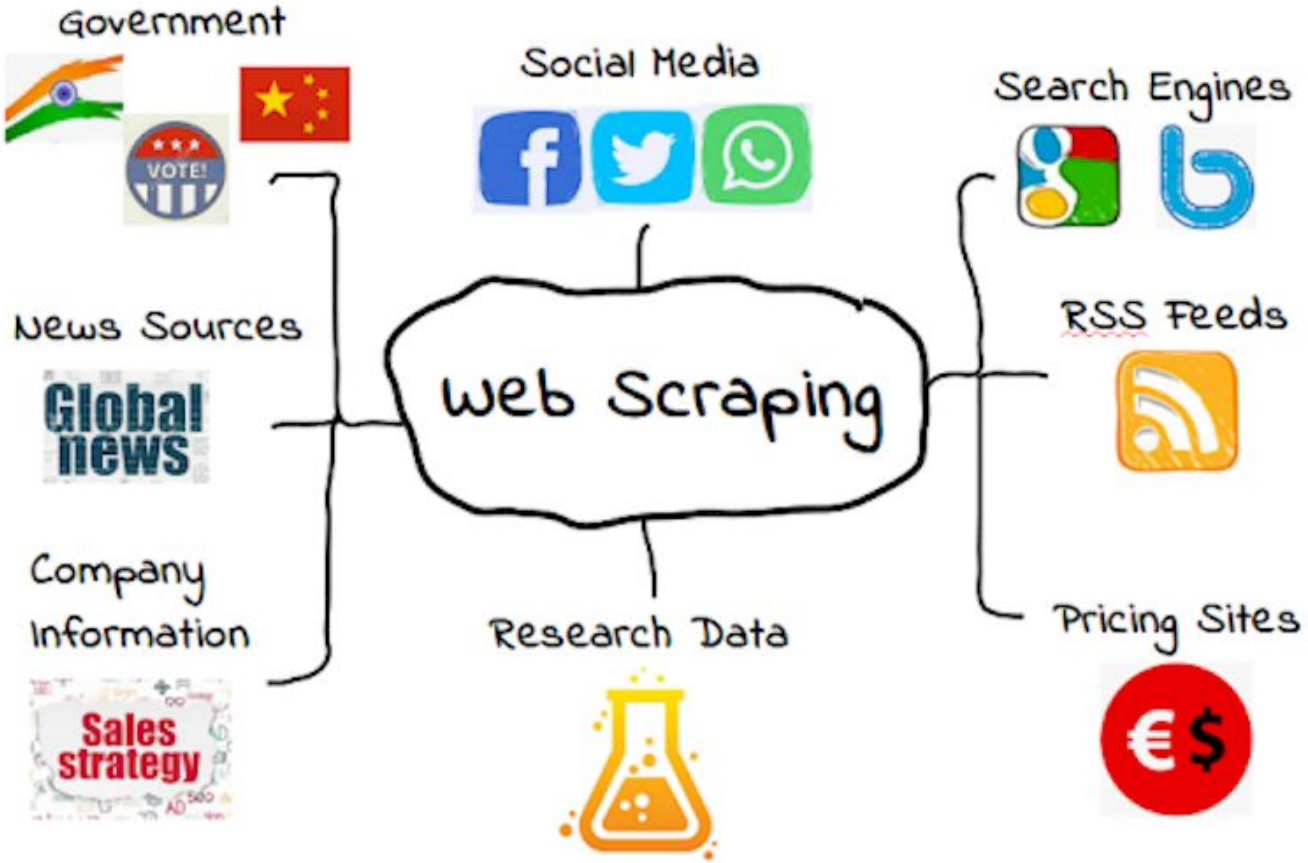
Data analysis



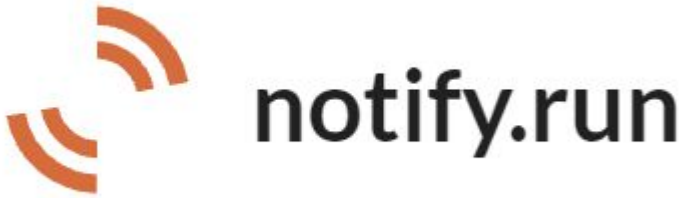
5th lecture



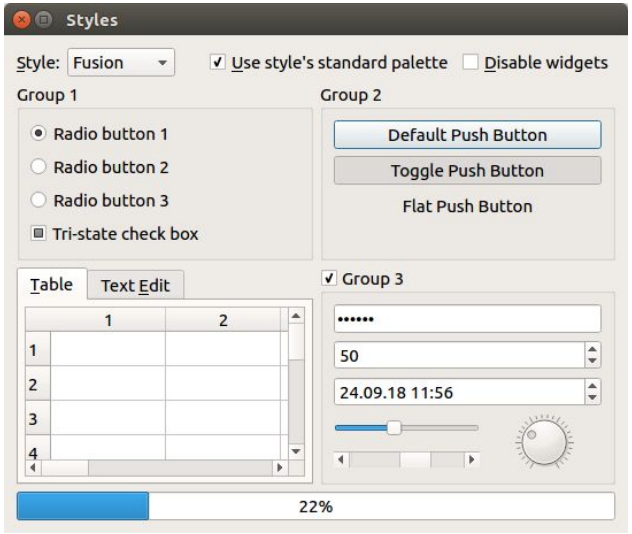
Getting data -> REST-APIs



Staying in contact



GUI Creation



© Michael Herrmann's PyQt5 book.

8th lecture

Deploying

Dev, “you”

“Download and install Anaconda, Python 3, 64 bit, then go to the command prompt and pip install ....”

The end user

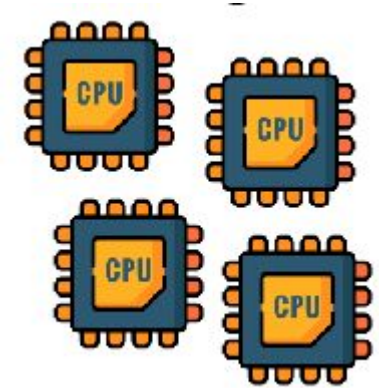
Lecture No. 9

Christmas!

10th lecture will be about very cool stuff!

https://www.history.com/topics/christmas/history-of-christmas-trees

Code optimization



Multithreading
Multiprocessing

11th lecture

Cython

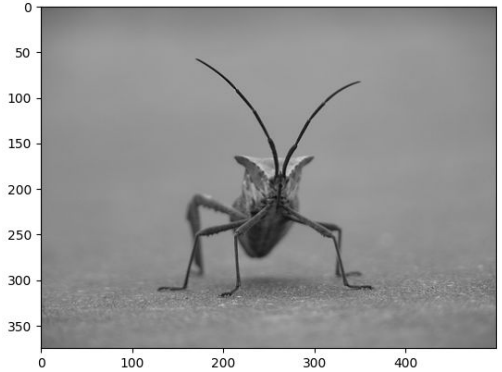
Numba

C-performance with
Cython and numba

12th lecture

Data visualization

Bar plots, etc



Images, scale bars etc

Lectures 13 & 14

The last lecture

- Q&A session - are there any open questions?
- Talking about the exam
  - Sample questions
  - When
  - Where
  - Etc.
- Feedback from you

The last slide

Today we covered

- How the course is setup
- Your and my/our expectations
- The content of the course throughout this semester

Exercise

Description of the exercise

- Now I would explain the task of the exercise.
- After the lecture, the exercises will be released on StudOn
- Please send them in by Friday noon.
    - ➜ Everything will be explained in the exercise w/ René and Hernan