

DATA SCIENCE

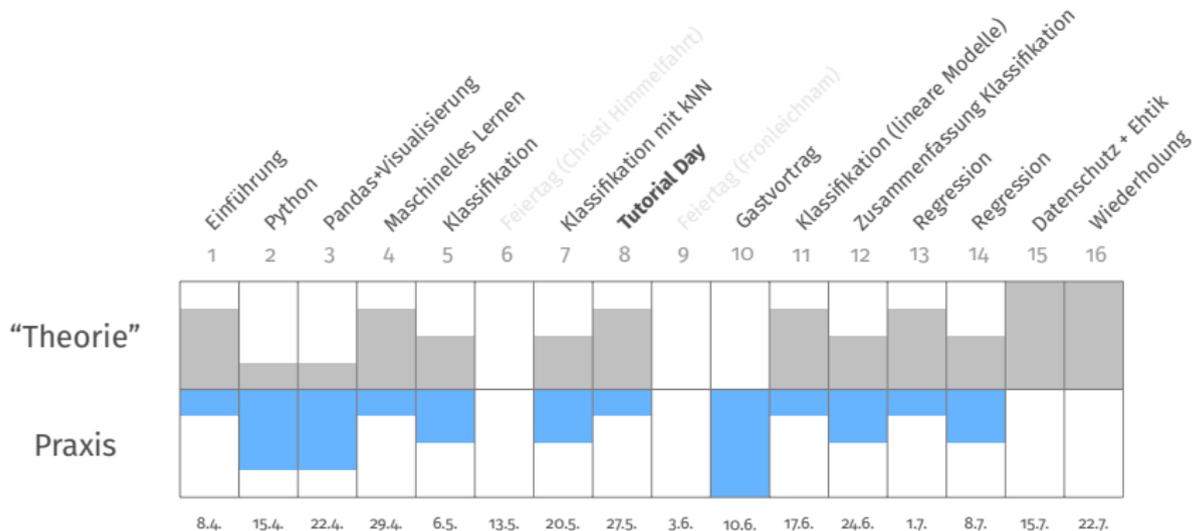
VORLESUNG 3 - INTRO

PROF. DR. CHRISTIAN BOCKERMANN

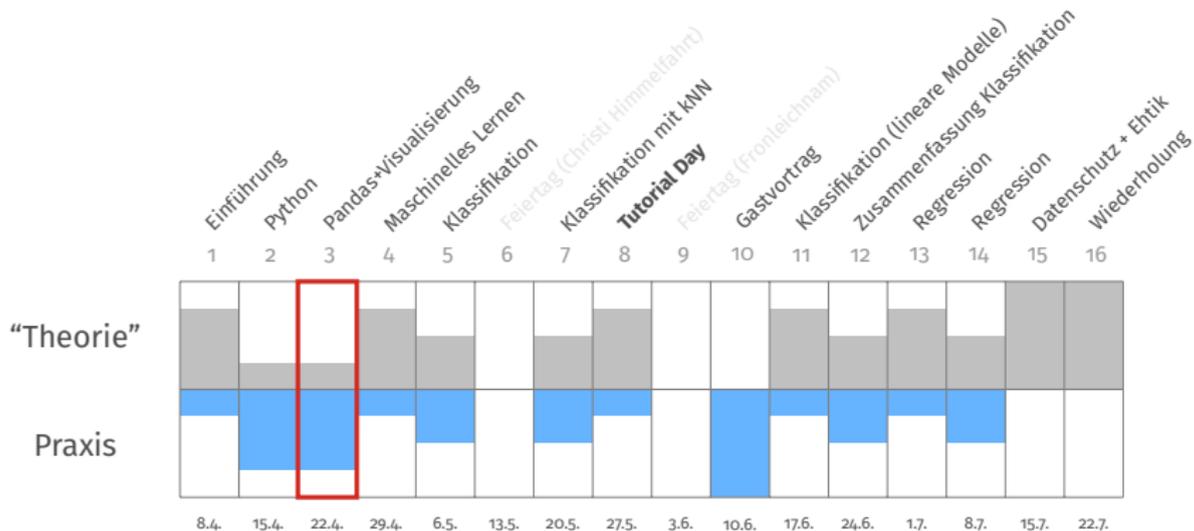
HOCHSCHULE BOCHUM

SOMMERSEMESTER 2021

Wo sind wir?



Wo sind wir?



Einführung in Python

- Skript-Sprache Python, Programme als Scripte (Text-Datei)
- Grundlegende Datentypen, Schleifen und eigene Funktionen
- Module zur Erweiterung des Funktionsumfangs

```
# Liste ['A', 'B', 'a', 'a',...]
liste = list("ABaaBccbbb")
anzahlBs = 0

# Zaehle die B's:
for x in liste:
    if x == 'B' or x == 'b':
        anzahlBs = anzahlBs + 1
```

Python cont'd

- *List-comprehension* als Schreibweise für Listen

```
liste = list("ABaaBccbbb")  
  
# extrahiere grosse und kleine B's  
#  
bs = [x for x in liste if x == 'b' or x == 'B']  
anzahlBs = len(bs)
```

Python - Module

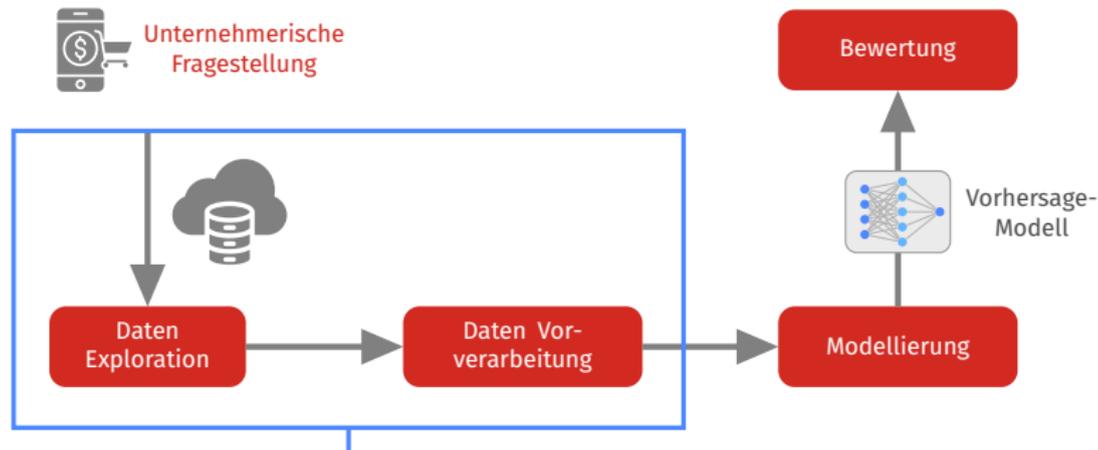
- Module als Bibliotheken mit fertigen Funktionen
- Einbindung von Modulen mit **import**
- Module mit Namensraum

```
# Binde das Modul 'datascience' im Namesraum 'ds' ein  
import datascience as ds  
  
# Einbinden von 'Zufall' in den globalen Namensraum  
from datascience.vorhersage import Zufall  
  
m = Zufall()    # statt ds.Zufall
```

Vorlesung 3

Pandas & Visualisierung

Vorgehen bei der Datenanalyse



Datenvorverarbeitung
hier: Mit Python und Pandas

Vorlesung 3 (heute):

- Vorstellung des Moduls **Pandas**
- Selektieren/Filtern von Daten
- Grundlagen zur Arithmetik mit Pandas

Vorlesung 3 (heute):

- Vorstellung des Moduls **Pandas**
- Selektieren/Filtern von Daten
- Grundlagen zur Arithmetik mit Pandas

Hochschule Bochum
Bochum University
of Applied Sciences



Fachbereich Wirtschaft
Prof. Dr. Christian Bockermann

Data Science

Sommersemester 2021

Übungsblatt 3

Dieses Übungsblatt beschäftigt sich mit dem Einlesen, dem Filtern und der Exploration von Daten. Die Daten liegen als CSV-Dateien in Ihrem Verzeichnis

Vorlesung 4 (nächste Woche):

- Grundlagen zum Maschinellen Lernen
- Klassifikation von Pflanzen (Modell)
- Bewertung von Vorhersagemodellen, Fehlermaße
- Datenvorverarbeitung/Cleaning mit Pandas

