

# DATA SCIENCE 1

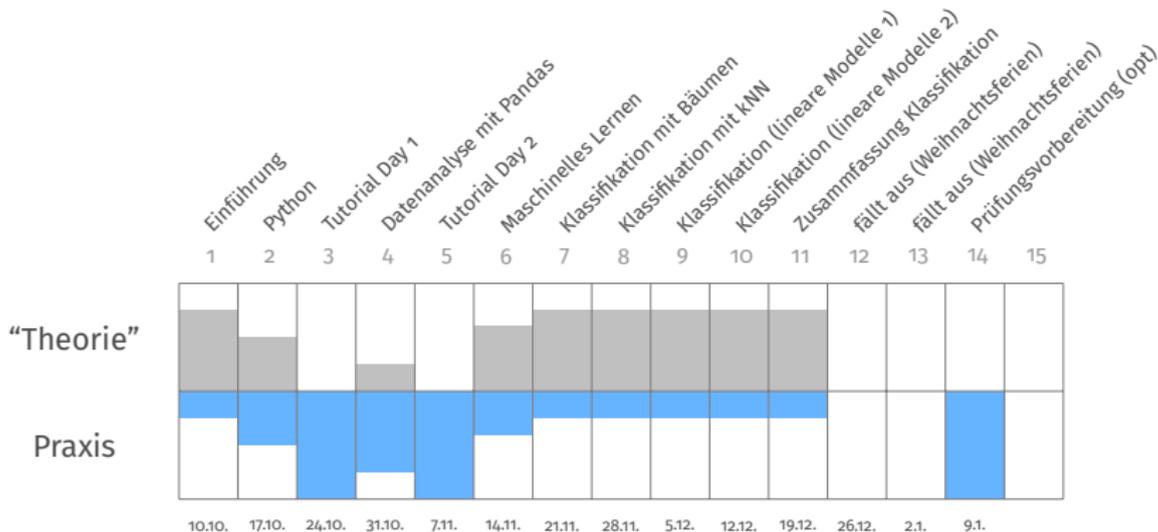
## VORLESUNG 3 - INTRO

PROF. DR. CHRISTIAN BOCKERMANN

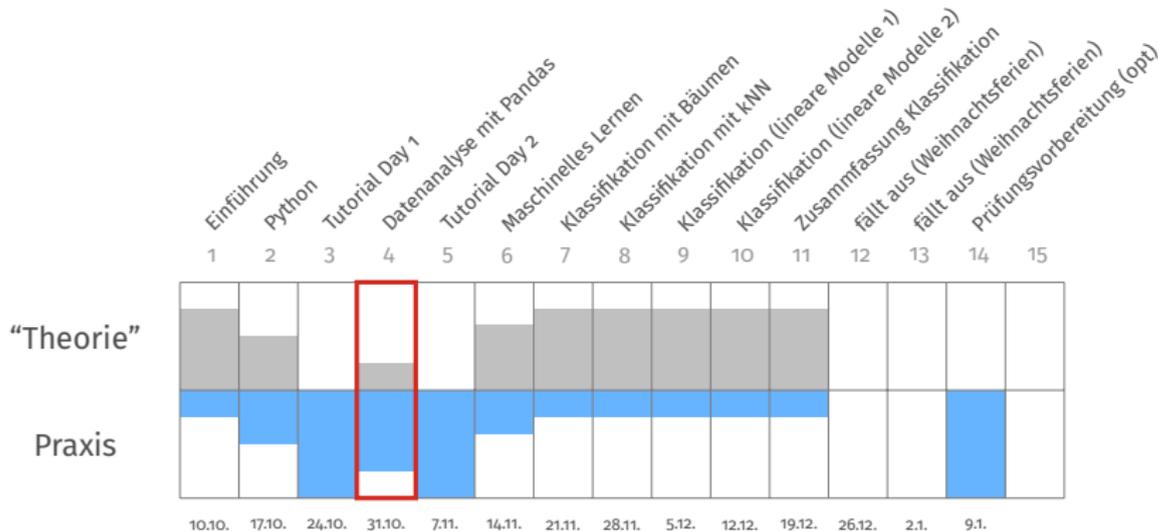
HOCHSCHULE BOCHUM

WINTERSEMESTER 2023/2024

## 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
```

## Listen mit Strings

- Strings sind wie Listen von Buchstaben
- mit **in** kann geprüft werden, ob etwas in einer Liste enthalten ist:

```
wort = "Data Science"  
if "en" in wort:  
    print("'en' ist enthalten!")
```

- Listen mit Strings funktionieren natürlich auch:

```
woerter = [ "Data", "Science", "Daten"]  
for wort in woerter:  
    if "at" in wort:  
        print(wort)
```

## Eigene Funktionen

- Erweiterung um eigene Funktionen mit **def**
- **return** beendet die Funktion mit Rückgabewert

```
def avg(zahlenListe)
    # Wenn Liste leer => Ergebnis 0
    if len(zahlenListe) == 0:
        return 0

    sum = 0
    for zahl in zahlenListe:
        sum = sum + zahl

    return sum / len(zahlenListe)
```

## Listen für Fortgeschrittene – *list comprehension*

- Aus der Mathematik – Mengennotation:

$$M = \{x^2 \mid x \in \mathbb{N}, 5 \leq x \leq 10\}$$

## Listen für Fortgeschrittene – *list comprehension*

- Aus der Mathematik – Mengennotation:

$$M = \{x^2 \mid x \in \mathbb{N}, 5 \leq x \leq 10\}$$

- *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

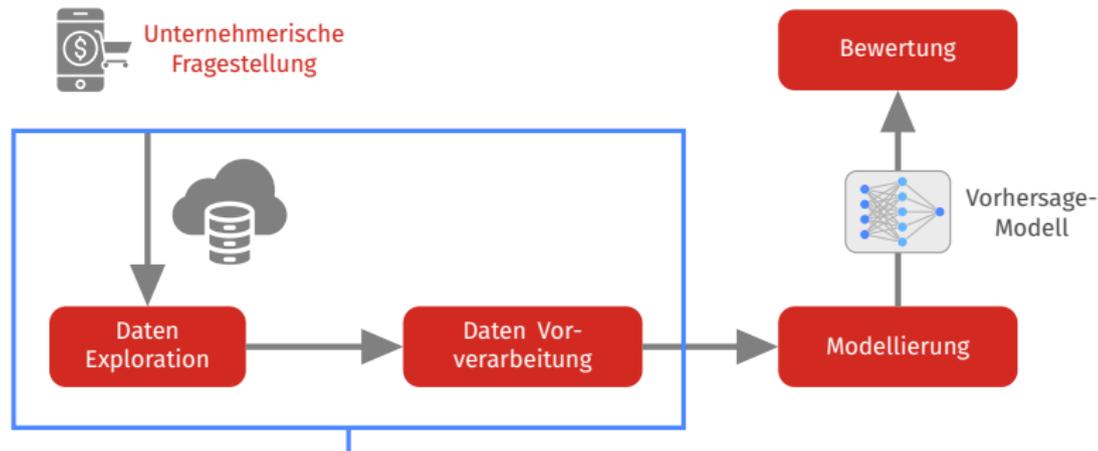
- Module als Bibliotheken mit fertigen Funktionen
- Einbindung von Modulen mit **import**
- Module mit Namensraum

```
# 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 2023

### 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 auf dem Notebook Server.

## Vorlesung 4 (nächste Woche): **Tutorial Day**

- Keine neuen Inhalte (bzgl. Maschinelles Lernen)
- Wiederholung + Vertiefung
- Zentrales Thema: Python + Pandas

### **Ablauf:**

- Vorlesung um 8:30 Uhr in Präsenz
- Es gibt Übungsaufgaben/Folien und Notebooks