

DATA SCIENCE 1

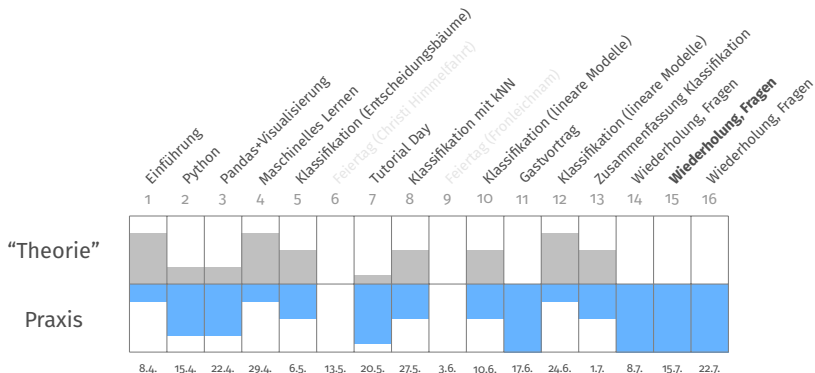
VORLESUNG 11 – WIEDERHOLUNG, FRAGERUNDE

PROF. DR. CHRISTIAN BOCKERMANN

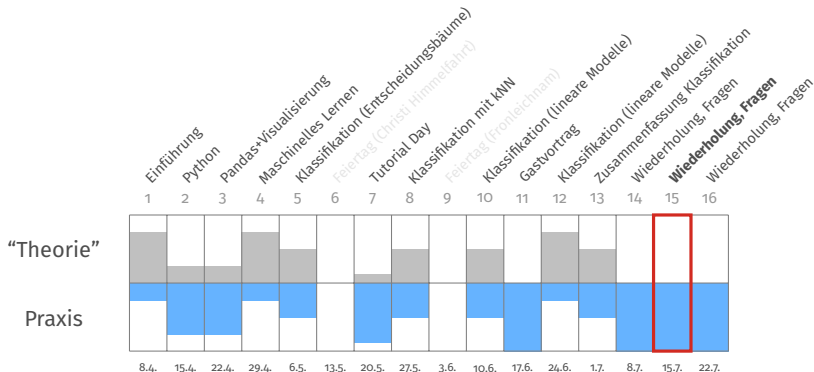
HOCHSCHULE BOCHUM

SOMMERSEMESTER 2021

Wo sind wir heute (Vorlesung 11) ?



Wo sind wir heute (Vorlesung 11) ?



Hausarbeit

- Themenausgabe: 30.7.2021 um 9 Uhr
- Bearbeitungszeit: 30.7. bis 6.8. 23:59 Uhr
- Abgabe als PDF via Moodle

Themenausgabe

Freitag, den 30.7.2021 um 9:00 Uhr
im BBB-Vorlesungsraum

<https://moodle.hs-bochum.de/mod/bigbluebuttonbn/view.php?id=126534>

Wiederholung, Fragen

Wie denkt ein Informatiker?

Produkte als Tupel gegeben:

`(nr, warengruppe, preis)`

Wie denkt ein Informatiker?

Produkte als Tupel gegeben:

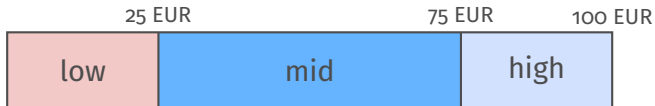
```
(nr, warengruppe, preis)
```

Als Liste in Python z.B.:

```
[(1, 'Sport', 48.99), (2, 'Mode', 49.95), ...]
```

`[(1, 'Sport', 48.99), (2, 'Mode', 49.95), ...]`

Sei der Preis der Produkte aus dem Intervall $(0, 100]$ EUR:



`[(1, 'Sport', 48.99), (2, 'Mode', 49.95), ...]`

Sei der Preis der Produkte aus dem Intervall $(0, 100]$ EUR:



Aufgabe:

Berechnen Sie die **Anzahl** der low/mid/high Produkte!

NICHT:

```
low = []  
mid = []  
high = []  
  
for p in produkte:  
    if p[2] < 25:  
        low.append(p)  
    else if p[2] < 75:  
        mid.append(p)  
    else:  
        high.append(p)
```

Wir zerlegen das Problem:

1. Preis für Produkt p berechnen
2. Für Produkt p die Kategorie berechnen
3. Kategorien berechnen
4. Für jede Kategorie die Liste der Produkte bestimmen

Wir zerlegen das Problem:

1. Preis für Produkt p berechnen
2. Für Produkt p die Kategorie berechnen
3. Kategorien berechnen
4. Für jede Kategorie die Liste der Produkte bestimmen

1. Aufgabe: Preis für p berechnen

Wir zerlegen das Problem:

1. Preis für Produkt p berechnen
2. Für Produkt p die Kategorie berechnen
3. Kategorien berechnen
4. Für jede Kategorie die Liste der Produkte bestimmen

1. Aufgabe: Preis für p berechnen

```
def preis(p):  
    return p[2]
```

2. Aufgabe: Berechne Kategorie von p

2. Aufgabe: Berechne Kategorie von p

```
def kategorie(p):  
    pr = preis(p)  
    if pr <= 25:  
        return "low"  
    if pr <= 75:  
        return "mid"  
    return "high"
```

3. Aufgabe: Kategorien berechnen

```
def kategorien(produkte):  
    menge = set()  
  
    for p in produkte:  
        menge.add(kategorie(p))  
  
    return menge
```


3. Aufgabe: Kategorien berechnen

```
def kategorien(produkte):  
    menge = set()  
  
    for p in produkte:  
        menge.add(kategorie(p))  
  
    return menge
```

Alternative:

```
def kategorien2(produkte):  
    ks = [kategorie(p) for p in produkte]  
    return set(ks)
```

4. Aufgabe: Listen bauen

```
ergebnis = []  
  
for k in kategorien(produkte):  
    liste = [p for p in produkte if kategorie(p) == k]  
    ergebnis.append( (k, len(liste)) )
```

4. Aufgabe: Listen bauen

```
ergebnis = []  
  
for k in kategorien(produkte):  
    liste = [p for p in produkte if kategorie(p) == k]  
    ergebnis.append( (k, len(liste)) )
```

Ergibt dann z.B.

```
[('low', 41), ('mid', 23), ('high', 5)]
```

4. Aufgabe: Listen bauen

```
ergebnis = []  
  
for k in kategorien(produkte):  
    liste = [p for p in produkte if kategorie(p) == k]  
    ergebnis.append( (k, len(liste)) )
```

Ergibt dann z.B.

```
[('low', 41), ('mid', 23), ('high', 5)]
```

Warum so umständlich?

Lesbarkeit!

```
for p in produkte:  
    summe = summe + p[2]
```

```
for p in produkte:  
    summe = summe + preis(p)
```

Lesbarkeit!

```
for p in produkte:  
    summe = summe + p[2]
```

```
for p in produkte:  
    summe = summe + preis(p)
```

- Sie wollen das ja auch in 2 Woche noch lesen!!

Lesbarkeit!

```
for p in produkte:  
    summe = summe + p[2]
```

```
for p in produkte:  
    summe = summe + preis(p)
```

- Sie wollen das ja auch in 2 Woche noch lesen!!
- Was ist, wenn sich das Tupel ändert? Preis an 5. Stelle?

Wiederverwendbarkeit!

- Nicht verschwenden - wiederverwenden!



Wiederverwendbarkeit!

- Nicht verschwenden - wiederverwenden!
- Gleiche Dinge mit Funktionen lösen - an allen Stellen das gleiche Ergebnis!
- Erhöht die Wartbarkeit + Lesbarkeit!



Aufgabe

Berechnen Sie den Durchschnittspreis je Kategorie!

1. Preis für Produkt berechnen
2. Durchschnittspreis auf Produktliste berechnen
3. Kategorien berechnen
4. Listen bauen und Durchschnittsfunktion (2.) benutzen

1. Preis für Produkt berechnen

```
def preis(produkt):  
    return produkt[2]
```

2. Durchschnittspreis für Produktliste berechnen

```
def durchschnittspreis(produkte):  
    preise = [ preis(p) for p in produkte ]  
    summe = sum(preise)  
    return summe / len(preise)
```

1. Preis für Produkt berechnen



```
def preis(produkt):  
    return produkt[2]
```

2. Durchschnittspreis für Produktliste berechnen

```
def durchschnittspreis(produkte):  
    preise = [ preis(p) for p in produkte ]  
    summe = sum(preise)  
    return summe / len(preise)
```

3. Kategorien berechnen



```
def kategorien(produkte):  
    ks = [kategorie(p) for p in produkte]  
    return set(ks)
```

4. Listen bauen und Durchschnittspreise berechnen

```
ergebnis = []  
  
for k in kategorien(produkte):  
    liste = [p for p in produkte if kategorie(p) == k]  
    ergebnis.append( (k, durchschnittspreis(liste) ) )
```

Übung

Übung

- Pandas: 2. Aufgabe der Test-Hausarbeit2