

# Data Science

Wintersemester 2022 / 2023

## Übungsblatt 2

### Aufgabe 1

Gegeben sei das folgende Python Skript:

```
a = 4.3
b = 2
c = a + b + 0.7
d = [a, 2, 7, c]
e = list("Hello, World!")
f = e[2:4]
g = e[4] == "l" or d[2] < 8
```

1. Geben Sie für jede der Variablen den Typ an, den Python ermitteln würde.
2. Welchen Wert hat **f** ?
3. Welchen Wert hat **g** ?
4. Was ergibt **len(e)** nach der obigen Zuweisung?

### Aufgabe 2 (Listen)

Gegeben sei eine Liste **xs** mit den Elementen 1, 2, 4, 5, 6 und 7.

1. Definieren Sie die Variable **xs** mit den genannten Elementen als Liste in einem Python Notebook.
2. Ergänzen Sie eine Code-Zelle, die die ersten beiden Elemente als Teil-Liste extrahiert.
3. Benutzen Sie *slicing* in einer weiteren Code-Zelle und den **+** Operator, um das Zahl 3 an die dritte Stelle in die Liste einzufügen.

### Aufgabe 3 (Funktionen)

Gegeben sei eine Liste mit den Zahlen

1.4, 2.3, 8.3, 2.3, 7.4, 5.6, 6.9, 7.2

1. Definieren Sie eine Variable mit dieser Liste in ihr Notebook ein.
2. Definieren Sie in Ihrem Notebook eine Funktion **avg**, die eine Liste von Zahlen bekommt und den *arithmetischen Mittelwert* der Zahlen zurückliefert.
3. Erweitern Sie ihr Skript so, dass das arithmetische Mittel der obigen Liste mit der Funktion **avg** berechnet und ausgegeben wird.

## Aufgabe 4 \*

Geben Sie die folgende Liste mit Wörtern (Zeichenketten, Typ `str`):

```
ProductId  
ProductName  
ProductCategory  
Price:EUR  
Price:USD  
Price:DKK  
Price:PLN
```

1. Definieren Sie diese Liste von Wörtern in einem Notebook als Liste in der Variablen `alle_spalten`.
2. Schreiben Sie eine Python Funktion, die aus einer Liste die Elemente extrahiert, die entweder dem String `ProductId` entsprechen, oder den Teilstring `Price:` enthalten.
3. Schauen Sie sich die Hilfe zu der String-Funktion `str.replace` an! Dazu können Sie in einem Notebook einfach die Zeile `help(str.replace)` ausführen. Erzeugen Sie nun eine Liste `spalten`, die die ausgewählten Spalten `ProductId` und die `Price:-`Spalten enthält, wobei bei den `Price:-`Spalten der Präfix `Price:` entfernt werden soll.

Das Ergebnis in `spalten` soll also folgendermaßen aussehen:

```
['ProductId', 'EUR', 'USD', 'DKK', 'PLN']
```

4. Jetzt soll das letzte Element aus der Liste `spalten` entfernt werden. Geben Sie dafür ein Stück Python-Code an, das dies erledigt.

## Aufgabe 5 \*

Ein Unternehmen verkauft Waren in mehreren Ländern. Die folgenden Listen enthalten jeweils das Länderkürzel, sowie den letztjährigen Umsatz. Dabei steht an der  $i$ -ten Stelle der ersten Liste das Länderkürzel und an der  $i$ -ten Stelle der zweiten Liste der Umsatz.

```
countries = ['DE', 'NL', 'AT', 'CH', 'PL']  
sales = [ 1000000, 500000, 750000, 400000, 175000 ]
```

1. Erstellen Sie eine Liste mit 2-Tupeln, die jeweils das Länderkürzel und den zugehörigen Umsatz enthalten. (Die Funktion `zip` ist hier sicherlich hilfreich.)
2. Erstellen Sie aus der Liste mit 2-Tupeln nun eine Liste mit Tripeln (= 3-Tupel), die als weitere Komponente den relativen Anteil des jeweiligen Landes am Gesamtumsatz enthält.

Das Ergebnis sollte in etwa so aussehen:

```
[('DE', 1000000, 0.35398...),  
 ('NL', 500000, 0.1769...),  
 ...]
```