

DSCB 230 - Aufgabenblatt 4

Die Aufgabenblätter werden in dieser Veranstaltung in Jupyter Notebooks veröffentlicht und bearbeitet. Diese finden Sie in der Github Organisation für Data Science 2 unter dem Repository *dscb230-tutorial* (<https://github.com/hka-mmvmv/dscb230-tutorial>). Die Musterlösung wird ebenfalls in Form eines Jupyter Notebook in Github hochgeladen.

Aufgabenteil 1: List Comprehension

Das Listenverständnis, das oft als Teil der funktionalen Programmierung in Python angesehen wird, ermöglicht es, mit weniger Code Listen für eine Schleife zu erstellen. Kurz gesagt, es ist eine wahrhaft pythonische Art des Programmierens. Weniger Code - mehr Effektivität.

Versuchen Sie also Listen zu befüllen, mit möglichst wenig Zeilen Code.

1. Füllen Sie eine Liste, klassisch mit einer For-Schleife. Es sollen alle Zahlen von 0 bis 9 enthalten sein.
2. Füllen Sie eine Liste, nach der Regel der List Comprehension. Es sollen alle Zahlen von 0 bis 9 enthalten sein.
3. Formen Sie Teilaufgabe 2 so um, dass nun nur noch alle geraden Zahlen enthalten sind.
4. Formen Sie Teilaufgabe 3 so um, dass nun nur noch alle geraden Zahlen enthalten sind, und diese mit 2 multipliziert werden.
5. Gegeben seien folgende Listen, erarbeiten Sie mithilfe der List Comprehension folgende Ausgabe:

```
nums = [1, 2, 3, 4, 5]
letters = ['A', 'B', 'C', 'D', 'E']
```

```
output:
[[1, 'A'], [1, 'B'], [1, 'C'], [1, 'D'], [1, 'E'], [2, 'A'], [2, 'B'], [2, .....]
```

6. Schreiben Sie ein Programm, welches mittels der List Comprehension, einen Beispieltext `dscb230` in jeden einzelnen Char aufteilt und in einer Liste speichert, also:

```
output:
['d', 's', 'c', 'b', '2', '3', '0']
```

Aufgabenteil 2: Tesla-Aktienkurs

In der CSV-Datei "TSLA.csv" liegen Aktienkurse von Tesla seit April 2017 vor. Ihre Aufgabe ist es, den Tagesdurchschnitt des Kurses der letzten drei Jahre zu plotten.

- 1) Benutzen Sie pandas, um die ersten **vier** Spalten der CSV-Datei zu laden. Die Spaltennamen sind leider verloren gegangen, beschriften Sie die Spalten in folgender Reihenfolge: "Date", "Open", "High", „Low“

Hinweis: Lesen Sie die Docs (https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read_csv.html)

- 2) Lassen Sie sich die Datentypen des Dataframes anzeigen. Wie Sie sehen, stimmt der Datentyp des Datums noch nicht. Ändern Sie diesen so, dass er stimmt.
- 3) Ergänzen Sie das Dataframe um eine weitere Spalte "Avg" in der Sie den Durchschnitt aus Low und High des Tages berechnen.
- 4) Erzeugen Sie nun ein neues Dataframe (auf Grundlage des bisherigen), welches nur Daten ab dem 22.04.2019 beinhaltet.
- 5) Benutzen Sie Matplotlib um den Tagesdurchschnitt (Avg) des Kurses der letzten 3 Jahre zu plotten. Die Linie soll die Farbe rot haben und in der Legende als "TSLA" bezeichnet werden.

Hinweis: Nutzen Sie "plt.gcf().autofmt_xdate()", um die X-Achse lesbar zu formatieren.

Aufgabenteil 3: Try-Except

1. Schreiben Sie ein Programm, welches für 2 vom Benutzer eingegebene Zahlen den Quotienten bildet. Fangen Sie hierbei zum Einen den ValueError an, falls der Nutzer keine Zahl eingibt, und den ZeroDivisionError ab, wenn versucht wird, durch 0 zu dividieren, sodass das Programm dabei nicht abstürzt, sondern den Fehler „Sie dürfen nur Zahlen eingeben“ bzw. „Eine Division durch 0 ist nicht möglich, geben Sie bitte andere Werte ein“ ausgibt.
2. Im Folgenden ist ein Programmcode, welcher aus einem langen Satz zählt, wie oft ein Buchstabe vorkommt. Ohne den Programmcode auszuführen, welcher Fehler könnte hierbei auftreten und wie kann man diesen mithilfe der Ausnahmebehandlung lösen?

```
sample_sentence = """Data Science (von englisch data „Daten“ und science „Wissenschaft“, im Deutschen auch Datenwissenschaft) bezeichnet generell die Extraktion von Wissen aus Daten. Data Science ist ein interdisziplinäres Wissenschaftsfeld, welches wissenschaftlich fundierte Methoden, Prozesse, Algorithmen und Systeme zur Extraktion von Erkenntnissen, Mustern und Schlüssen sowohl aus strukturierten als auch unstrukturierten Daten ermöglicht. Der Studiengang Data Science verwendet Techniken und Theorien aus den Fächern Mathematik, Statistik und Informationstechnologie, einschließlich der Signalverarbeitung, verwendet Wahrscheinlichkeitsmodelle des maschinellen Lernens, des statistischen Lernens, der Programmierung, der Datentechnik, der Mustererkennung, der Prognostik, der Modellierung von Unsicherheiten und der Datenlagerung. Personen, die im Bereich Data Science arbeiten, werden als Data Scientist bzw. Datenwissenschaftler bezeichnet, wobei meist speziellere oder Spezialisierungen anderer, übergeordneter Berufsbezeichnungen üblich sind (z. B. Statistiker, Informatiker)"""

char_counter = {}
for char in sample_sentence:
    char_counter[char] += 1

print(sorted(char_counter.items(), key=lambda x: x[1], reverse=True))
```