
Übungsblatt Nr. 7

Datum: 11.10.2023

Dateien

Hinweis: Bearbeiten Sie die Aufgaben in einer Reihenfolge, die Ihnen sinnvoll erscheint.

Zeigen Sie alle Pflichtaufgaben vor.

Empfehlung: Probieren Sie alle gegebenen Beispiele aus.

[Pflichtaufgabe] Aufgabe 7.1: Wiederholung

Zur Übung der Inhalte der letzten Woche.

Schreiben Sie eine Python-Funktion, welche eine Liste mit Ganz- und Fließkommazahlen entgegen nimmt und die Anzahl der Auftritte der beiden Datentypen darin bestimmt. Die errechneten Werte sind als ein Tupel zurück zu geben.

Lassen Sie diese Funktion von Ihrer main-Methode einsetzen und geben Sie das Ergebnis mit zwei print-Anweisungen (Ganzzahlen und Fließkommazahlen) in einer Zeile aus. Denken Sie an `end` in der `print`-Funktion.

Aufgabe 7.2: NumPy

Erstellen Sie mit NumPy zwei Vektoren der Länge 100 mit Zufallszahlen von 5 bis 25. Schreiben Sie eine Funktion, welche den Abstand jeder Zahl zu jeder anderen berechnet und ein NumPy array (in diesem Fall der Größe 100×100) zurückliefert.

[Pflichtaufgabe] Aufgabe 7.3: Einlesen

Erstellen Sie mit einem Texteditor eine Datei `schueler.txt`, in der die Punktzahl von Schülern in Tests abgelegt ist:

```
1 Max 10 15 20 30 40 32
2 Ada 23 16 19 22 17 2
3 Tom 8 22 17 14 24 9
4 Amy 12 28 21 45 26 10
5 Tim 14 32 25 16 8 21
6 Ana 6 26 30 37 22 16
7 Ben 11 19 27 21 35 22
```

1. Lesen Sie diese Datei ein und legen Sie die Daten in einer Liste ab
2. Plotten Sie mit Hilfe von `matplotlib` die Ergebnisse der Schüler
3. Berechnen Sie den Durchschnitt und speichern Sie eine anonymisierte Datei `anonyme_ergebnisse.txt` ab, in welcher nur die Punktzahl und der Durchschnitt gespeichert sind.

Aufgabe 7.4: Und nochmal Einlesen

Im Moodle finden Sie eine Datei `Namen.csv`. Was sind `.csv` Dateien und wie können diese in Python eingelesen werden? Legen Sie eine Liste mit den einzelnen Tabellenzeilen an und berechnen Sie das Durchschnittsalter der Personen. Schreiben Sie eine Funktion, welche aus einer solchen Liste errechnet, welcher Vorname am häufigsten auftritt.

[Pflichtaufgabe] Aufgabe 7.5: Größere Dateien

Unter dem Link <https://my.pgp-hms.org/profile/huD54F5E> finden Sie eine `.txt` Datei mit Daten von AncestryDNA. Zu Beginn der Datei finden Sie eine Erklärung zum Aufbau der Daten.

1. Laden Sie Diese herunter und versuchen Sie sie in Python einzulesen.
2. Löschen sie den Erklärungs-Kommentar und Kopfzeile
3. Wie viele Zeilen enthält die Datei?
4. Wie oft tritt das 'A' als Allel auf?
5. Welches Allel-Paar finden Sie bei `rs9555101`, welches bei `rs58073564`
6. Nutzen sie die Funktion `readlines()`, um den Inhalt in einer Liste abzulegen.

7. Lassen Sie sich die ersten 5000 Zeilen ausgeben. Nehmen Sie die Zeit (z.B. mit dem Modul time)
8. Zerlegen Sie die Strings in der Liste noch einmal beim Auftreten des Tabs. Nehmen Sie wieder die Zeit.

Aufgabe 7.6: Recherche

Recherchieren Sie: Was sind gängige Datenformate in Ihrer Wissenschaft? Gibt es eine einfache Möglichkeit, diese in Python einzulesen?