

Zertifikatsklausur 24.11.

- Klausurtyp: Open-Book-Exam
- Bearbeitungszeit: 45 Minuten (+ 15 Minuten für Download & Upload)
- Erreichbare Punkte: 45
- Bestehensgrenze: mit 25 Punkten auf jeden Fall bestanden.

Abgabe bis spätestens 20.00 Uhr per E-Mail an kurse@stads.de.

Der Code muss in Python 3.8.5 mit numpy 1.19.2, pandas 1.1.3, matplotlib 3.3.2, seaborn 0.11.0 bzw. plotly 4.12.0 lauffähig sein. Falls weitere Pakete oder andere Versionen verwendet werden, muss die jeweilige Version angegeben werden.

Aufgabe 1: Grundlagen (18 Punkte)

- Erstellen Sie eine Python-Datei mit dem Namen `<Nachname>_<Vorname>_exam.py` (z.B. `kern_moritz_exam.py`) und bearbeiten in dieser Datei die Aufgabe.

Aufgabe 1a: Get Started (2 Punkte)

- Definieren Sie die beiden Variablen `x` und `y` als `-10` bzw. `20` (1 Punkt)
- Geben Sie das Ergebnis des quadrierten Abstands, d.h. `(x-y)^2` , an. (Tipp: `**` , `print`) (1 Punkt)

Aufgabe 1b: Einfache Funktion (6 Punkte)

- Definieren Sie eine Funktion `abstand_quadriert` , die die zwei Variablen `x` und `y` als Input hat. (1 Punkt)
- Die Funktion soll prüfen, ob die Differenz von `x` und `y` gleich 0 ist, falls dies der Fall ist: `Printe x und y sind gleich.` und gebe `0` zurück. *(2 Punkte)
 - andernfalls, gebe das Ergebnis von `(x-y)^2` zurück. (2 Punkte)
- Werten Sie die Funktion aus für die Inputkombination (x=5, y=5) und (x=4, y=6) (1 Punkt)

Aufgabe 1c: Datentypen (6 Punkte)

- Erstellen Sie eine Variable `x` mit dem Wert 0. Konvertieren Sie `x` explizit zu einem Boolean (True/False) und speichern Sie das Ergebnis als `z` (1 Punkt)
- Definieren Sie die variable `z_is_is_bool` als `TRUE` , falls `z` vom Typ `bool` , andernfalls als `FALSE` . Tipp: Verwenden Sie dazu die Funktion `isinstance` . (1 Punkt)
- Erstellen Sie ein `dictionary` mit dem Namen `semester` mit folgendem Mapping auf jeweils eine Liste. (3 Punkte)
 - `"FSS"` -> [`"Fruehjahrs-Sommer-Semester"`, `"01. Februar"`]
 - `"HWS"` -> [`"Herbst-Winter-Semester"`, `"01. September"`]
- Lassen Sie sich das Element mit dem Key `FSS` ausgeben. (1 Punkt)

Aufgabe 1d: Schleifen (4 Punkte)

- Erstellen Sie mit einer for-Schleife folgende Ausgabe (... ausgeschrieben) (4 Punkte)

```
1
-2
3
-4
5
-6
...
29
-30
```

Aufgabe 2: Wichtige Pakete (27 Punkte)

- Erstellen Sie ein IPython-Notebook mit dem Namen `<Nachname>_<Vorname>_exam.ipynb` (z.B. `kern_moritz_exam.ipynb`) und bearbeiten in dieser Datei die Aufgabe.

Aufgabe 2a: Numpy (6 Punkte)

- Erstellen Sie geschickt einen Vektor `v` , der wie folgt aussieht. (1 Punkt)

```
[ 2000, 2004, 2008, 2012, 2016, 2020, ..., 2092, 2096]
```

- Erstellen Sie eine Matrix `ma_diag` der Dimension (12,12), die auf ihren Diagonalen überall eine 1 hat sonst nur Nullen. (1 Punkt)
- Setzen Sie das letzte Element rechts oben in der Matrix `ma_diag` auf `np.NaN` . (1 Punkt)
- Initialisieren Sie einen Zufallszahlengenerator. (1 Punkt)
- Verwenden Sie den Zufallszahlengenerator, um ein auf dem Intervall [0, 1] gleichverteilte Zufallszahl `z` zu simulieren. (1 Punkt)
- Ersetzen Sie in der Matrix `ma_diag` das NaN-Element durch Ihre erzeugte Zufallszahl `z` .
- Berechnen Sie elementweise die Exponentialfunktion von `ma_diag` und speichern Sie die so erhaltene Matrix als `ma_diag_exp` . (1 Punkte)

Aufgabe 2b: Pandas Basics (9 Punkte)

Für diese Aufgabe benötigen Sie den Datensatz [activity.csv](#).

- Importieren Sie den Datensatz `activity.csv` . (2 Punkte)
- Lassen Sie sich die ersten 10 Zeilen ausgeben. (1 Punkt)
- Aus wie vielen Zeilen und Spalten besteht der Datensatz? (1 Punkt)
- Löschen Sie die Spalte `country_region_code` und wandeln Sie die Spalte `date` ins Datetime-Format um (Tipp: `pd.to_datetime`). Speichern Sie den so erhaltenen Datensatz als `df` ab und verwenden ihn für die folgenden Aufgaben. (2 Punkte)
- Erstellen Sie eine neue Spalte `change_from_baseline` , die sich als Summe der Spalten `retail_and_recreation_percent_change_from_baseline` und `grocery_and_pharmacy_percent_change_from_baseline` ergibt. (1 Punkt)
- Geben Sie die Zeilen mit dem größten Rückgang von der Baseline, also die 5 Zeilen mit den kleinsten Werten in `change_from_baseline` aus. Wie groß war der `change_from_baseline` am 15. November (2020-11-15) in Baden-Württemberg? (2 Punkte)

Aufgabe 2c: Pandas Advanced (7 Punkte)

- Geben Sie die Korrelation zwischen `retail_and_recreation_percent_change_from_baseline` und `grocery_and_pharmacy_percent_change_from_baseline` an. (2 Punkte)
- Gruppieren Sie den Datensatz nach der Spalte `Bundesland` . Filtern ihn nach Beobachtungen seit dem 01.09.2020 . Verwenden Sie diesen für die nächste Fragestellung. (2 Punkte)
- Geben Sie an, in welchem Bundesland der Rückgang bei Einkäufen im Mittel am größten war, d.h. bestimmen Sie das Bundesland, bei dem `grocery_and_pharmacy_percent_change_from_baseline` im Mittel am kleinsten war. (2 Punkte)

Aufgabe 2d: Grafiken (6 Punkte)

- Gruppieren Sie den Datensatz nach der Spalte `Datum` und aggregieren Sie die Spalte `retail_and_recreation_percent_change_from_baseline` mit dem Mittelwert. Verwenden Sie diesen Datensatz für die nächste Aufgabe. (1 Punkte)
- Erstellen Sie einen Scatterplot mit dem Datum auf der x-Achse und den `retail_and_recreation_percent_change_from_baseline` auf der y-Achse. Sie dürfen dabei ein Paket Ihrer Wahl verwenden (z.B. Pandas, Matplotlib, Seaborn, Plotly). (5 Punkte)

Abgabe

Senden Sie die *beiden* von Ihnen erstellten Dateien bis Abgabeschluss (20 Uhr) per E-Mail an kurse@stads.de.