



Hochschule
Bielefeld
University of
Applied Sciences
and Arts

Pratikum Blatt 3 (4 Punkte)

mit Dateien „countries_population.csv“ und „bundeslaender.txt“

Datenanalyse und Einführung in Maschinelles Lernen
WS 2025/26

Datenanalyse mit Python
Abgabe am 19.November/20.November 2025

Dozentin: Grit Behrens
mailto: grit.behrens@hsbi.de

A decorative graphic on the right side of the slide, consisting of a series of overlapping, slanted rectangular blocks in various shades of blue and teal, creating a stepped, architectural effect.

Studiengang Informatik Fachbereich Campus Minden

Aufgabe 1 (1 Punkt)

Lesen Sie die Datei `countries_population.csv` ein (liegt mit auf ILIAS). Sie enthält die Bevölkerungszahlen von allen Ländern (Stand Juli 2014) . Finde heraus, mit welchen Parametern die folgenden Probleme gelöst werden können: Als Trennzeichen fungiert ein Leerzeichen, Kommas trennen die Tausenderpositionen. Geben Sie die ersten fünf Zeilen des DataFrames aus. Erzeugen Sie die folgende Ausgabe:

Country	Population
China	1355692576
India	1236344631
European Union	511434812
United States	318892103
Indonesia	253609643

Aufgabe 2 (2 Punkte)

Lesen Sie die CSV-Datei **bundeslaender.txt** ein.

- a) Erzeugen Sie eine neue Datei mit den Spaltennamen „land“, „area“, „female“, „male“, „population“ und „density“. Die density ist die Anzahl der Einwohner pro Quadratkilometer. (1 Punkt)

	land	area	female	male	population	density
0	Baden-Württemberg	35751.65	5465	5271	10736	300.0
1	Bayern	70551.57	6366	6103	12469	177.0
2	Berlin	891.85	1736	1660	3396	3808.0
3	Brandenburg	29478.61	1293	1267	2560	87.0
4	Bremen	404.28	342	321	663	1640.0
5	Hamburg	755.16	894	849	1743	2308.0
6	Hessen	21114.79	3109	2983	6092	289.0
7	Mecklenburg-Vorpommern	23180.14	861	846	1707	74.0
8	Niedersachsen	47624.20	4076	3918	7994	168.0
9	Nordrhein-Westfalen	34085.29	9261	8797	18058	530.0
10	Rheinland-Pfalz	19853.36	2069	1990	4059	204.0
11	Saarland	2568.70	540	510	1050	409.0
12	Sachsen	18415.51	2191	2083	4274	232.0
13	Sachsen-Anhalt	20446.31	1264	1206	2470	121.0
14	Schleswig-Holstein	15799.38	1448	1385	2833	179.0
15	Thüringen	16172.10	1185	1150	2335	144.0

Aufgabe 2 b und c

b) Geben Sie alle Zeilen aus mit einer größeren weiblichen Bevölkerungszahl als die der Männer. Wie viele sind es? **0,5 Punkte**

c) Geben Sie alle Zeilen aus, deren Bevölkerungsdichte größer als 1000 Einwohner pro Quadratkilometer ist. **0,5 Punkte**

	land	area	female	male	population	density
2	Berlin	891.85	1736	1660	3396	3808.0
4	Bremen	404.28	342	321	663	1640.0
5	Hamburg	755.16	894	849	1743	2308.0

Aufgabe 3 (1 Punkt)

- a) Schreiben Sie ein Script und finden Sie heraus, wieviele Arbeitstage/reguläre Wochentage es zwischen dem 24. 12. 2025 und dem 6.01.2026 gibt – generieren Sie die unten angegebene Ausgabe. (0,5 Punkte)

```
DatetimeIndex(['2025-12-24', '2025-12-25', '2025-12-26', '2025-12-29',  
              '2025-12-30', '2025-12-31', '2026-01-01', '2026-01-02',  
              '2026-01-05', '2026-01-06'],  
              dtype='datetime64[ns]', freq='B')  
Es gibt 10 normale Wochentage/reguläre Arbeitstage in den Weihnachtsferien 2025/26.
```

- a) Wieviele Sonntage gibt es zwischen dem 24.12.2025 und dem 6.01.2027, die gleichzeitig auf den 1. des Monats fallen? – generieren Sie die unten angegebene Ausgabe. (0,5 Punkte)

```
DatetimeIndex(['2026-02-01', '2026-03-01', '2026-11-01'], dtype='datetime64[ns]', freq=None)  
Vom 24.12.2025 bis zum 6.01.2027 gibt es 3 Sonntage, die auf den 1. des Monats fallen.
```