
Übungsblatt 1

Computergestützte Statistik

HWR Berlin, Wintersemester 2024

Prof. Dr.-Ing. Sebastian Schlesinger

Aufgabe 1 (Erste Explorative Datenanalyse in Google Colab)

(5 Punkte)

Macht Euch mit Kaggle Datensätzen vertraut. Ihr findet eine Übersicht unter <https://www.kaggle.com/datasets>.

Ladet einen Datensatz Eurer Wahl in Google Colab und lasst Euch für eine Spalte Eurer Wahl (die quantitative Werte enthält) die deskriptive Statistik ausgeben.

Lasst Euch für die Spalte auch den Boxplot ausgeben. Wie interpretiert ihr den?

Aufgabe 2 (Lokale Ausführung Datenanalyse mit Jupyter Notebook)

(5 Punkte)

Installiert Python (wenn nicht vorhanden) und Jupyter Lab auf Eurem Rechner, startet Jupyter Lab und führt die gleiche Analyse wie in der vorherigen Aufgabe durch.

Aufgabe 3 (Fortgeschritten: Erste Infos mit Pandas extrahieren)

(5 Punkte)

Ladet Euch den Datensatz über die meist gestreamten Songs in Spotify herunter und analysiert ihn. Ihr findet ihn unter <https://www.kaggle.com/datasets/abdulszz/spotify-most-streamed-songs>.

Ladet als erstes den Datensatz in Google Colab hoch und versucht, folgende Dinge herauszufinden:

- a) Wie viele Zeilen und Spalten hat der Datensatz?
- b) Welche Spalten sind vorhanden?
- c) Welche Wertetypen sind den jeweiligen Spalten enthalten?
- d) Wie viele Streams hat der Song Shape of You?
- e) Wie viele Streams haben die Songs von Ed Sheeran insgesamt?

Aufgabe 4 (Fortgeschritten: Python)

(5 Punkte)

Installiert Python (wenn nicht vorhanden) und schreibt ein Programm, das folgendes Problem löst: Eine Funktion

`two_sums`

erhält eine Liste von Zahlen (Integer) und ein Zahl (Integer) als Zielwert. Die Funktion soll durch die Liste gehen und die Indizes von zwei Zahlen der Liste zurückgeben, die die Summe des Zielwerts ergeben. *Hinweis:* Die Indizes in Listen starten bei 0.

Beispiel:

```
two_sums([2,7,11,15],9)=[0,1],  
two_sums([3,2,4],6)=[1,2]
```