

Leistungsnachweis Pandas

Aufgabenstellung

In diesem Leistungsnachweises sollen die bisher erworbenen Kenntnisse von Python, und insbesondere von Pandas, angewendet werden. Aufgabe ist es, entsprechend der vier Schritte der Datenanalyse eine eigene Datenanalyse anhand von selbst gewählten Daten zu erstellen. Folgende Anforderungen werden an die 4 Schritte gestellt:

1. Daten beschaffen und laden:
 - Wählen eine Datenquelle und einen Datensatz (nicht das Beispiel Lebenserwartung aus der Vorlesung)
 - Für ein «erfüllt» reicht ein Datensatz aus «Our World in Data»
 - Zusatzpunkte gibt es, wenn weitere Argumente für den read-Befehl benutzt werden müssen.
2. Datenvorbereitung
 - Hier sollen die Daten in das Long Format gebracht werden
 - Sind die Daten bereits im Long Format, muss in diesem Schritt nichts gemacht werden
 - Hier gibt es Zusatzpunkte, wenn die melt-Methode verwendet wird oder wenn mehrere Datensetzen vereint werden.
3. Daten-Exploration
 - Mach eine statistische Analyse jeder Spalte (im Long Format)
4. Daten-Analyse. Wie beim Excel-Leistungsnachweise:
 - Überlege eine für die Daten geeignete Forschungsfrage
 - Beantwortet sie mit einer sinnvollen Visualisierung der Daten
 - Beschreibe in wenigen Sätzen (als Richtlinie: max. 10) das Vorgehen und die Resultate der Visualisierung.
 - Zusatzpunkte: Verwendung von Pivot, mehrere Visualisierungen, Recherche zur Plausibilisierung der Resultate.

Rahmenbedingungen

- Ihr arbeitet in Gruppen von maximal 3 Personen.
- In der übungsstunde vom 1. Dezember könnt ihr Fragen stellen und bekommt Hilfe durch Adrian.
- Abgabezeitpunkt ist der 3. Dezember 2022 um 20h.
- Abzugeben ist eine Jupyter-Datei auf Moodle. Es reicht, wenn eine Person die Datei einreicht.
Dateiname: *Titel_Nachname1_Nachname2_Nachname3.ipynb*
- Der Nachteilsausweis zählt 10% der Gesamtnote