

Aufgabenstellung:

Führen Sie danach die folgenden Schritte durch (Datenvorbereitung + Visualisierung) - teilen Sie ihr Notebook auch entsprechend in Sektionen ein, dokumentieren und begründen sie ausführlich ihre Schritte:

1. Team
2. Daten:
 1. Link auf Datensatz bzw. Sätze
 2. Laden sie die notwendigen Dateien runter, beschreiben sie die Attribute im Notebook
 3. initiale Std-Analyse mit sample, head, info, describe (und unique Werten wo sinnvoll!)
3. Datenbereinigung
 1. falsche Werte
 2. fehlende Werte
 3. begründete Featurereduktion bzw. Typsetzung
 4. Speichern sie die - falls notwendig aus verschiedenen Dateien zusammengeführten und integrierten - Ergebnissdaten in korr.csv ab
4. Datenvorbereitung
 1. numerische Werte
 2. Speichern sie diese in num.csv ab
 3. normalisierte Werte
 4. Speichern sie diese in norm.csv ab
 5. nominale Werte
 6. Speichern sie diese in nom.csv ab
5. Visualisierung mit Seaborn
 1. Bereiten sie ihren Datensatz mit verschiedenen geeigneten Diagrammen auf (aus allen 4 Datensätzen möglich) inkl. mindestens
 1. eigener Farbpalette
 2. Korrelationen mit Matplotlib.pyplot.matshow() sowie seaborn.heatmap()
 3. Pairplot
 4. Box vs Violinplot inkl. Subplots
 5. Histogramme/KDE
 6. Scatterplots
 7. Linien/Balkendiagramme

2. Besprechen sie jedes Diagramm - Auffälligkeiten?

Geben Sie sowohl die IPYNB Datei sowie eine PDF Datei mit Ergebnissen nach dem Durchlauf ab sowie alle Eingabedateien!

Tip: Für ToC in Jupyter Notebooks wie in colab gibt es mehrere Möglichkeiten:

<https://stackoverflow.com/questions/21151450/how-can-i-add-a-table-of-contents-to-a-jupyter-jupyterlab-notebook>