

Extrauppgift | Datatvätt

Bifogat finns två `csv`-filer med data som inte är helt som den ska.

Använd `pandas` för att läsa in datan och träna på att tvätta den.

Datatvätt inkluderar bland annat att sätta rätt datatyper, hantera saknade värden och droppa oönskade kolumner.

Analysera datan, identifiera datatyper och vilka ändringar du måste göra för att kunna sätta rätt datatyper på kolumnerna. Resonera kring saknad data och hur du ska hantera den.

I mappen `cleaned` finns exempel på tvättade versioner av dataseten.

Tip

`messy_IMDB_dataset.csv` använder `;` som kolumnseparator istället för `,`.
`pd.read_csv()` har en parameter `sep` som bestämmer vilket tecken som ska användas för att separera kolumnerna.