

L4_Exercises

September 24, 2020

1 ØVELSE: Hvordan fordeler glæde sig på tværs af køn?

I skal i denne øvelse få et overblik over, hvordan fordelingen i glæde (variablen `happy`) er forskellige på tværs af køn.

1. Grupper data efter `gndr` med `.groupby()`
2. Brug `value_counts()` til at optælle fordelingen i variablen `happy`.
3. Brug `unstack()` og `plot.bar()` til at visualisere fordelingen.

BONUS ØVELSE Kategorierne i `happy` står i forkert rækkefølge. Løs dette ved at rekode værdierne sådan, at værdierne står i rækkefølge.

[]:

2 ØVELSE: Flere grupperinger

1. Dan en kategorisk variabel, der inddeler respondenter i 3 lige store grupper efter vægt (brug `pd.cut()`)
2. Dan en visualisering, der viser fordeling af svar i `alcfreq` splittet ud på køn og vægtkategori (brug enten `seaborn` eller `.plot` på grupperet data (`.groupby()`)

[]:

3 ØVELSE: Håndtering af missing

I skal lave jeres visualisering fra sidste øvelse igen (en visualisering, der viser fordeling af svar i `alcfreq` inden for hver gruppering af køn og vægtkategori), men denne gang vil vi gerne have nogen af missingværdierne i `weight` med.

1. Dan en `weight_imputed` variabel, hvori missingværdier for `weight` erstattes med midelvægten.
2. Dan vægtkategorien igen med `weight_imputed` variablen
3. Dan visualiseringen igen

BONUSØVELSE

Sammenlign fordelingerne i `alcfreq` indenfor hver køn og vægtekategori før og efter imputering/erstatning af missingværdi - Er der forskel?

[]:

4 ØVELSE: Krydstabel

Dan en krydstabel mellem `alcfreq` og vægtekategori-variablen fra sidste øvelse.

Kombiner evt. med `seaborn`s heatmap-funktion.

[]:

5 ØVELSE: Merge/Join med Pandas i Python

Denne øvelse samler op på det meste, som er blevet gennemgået i dag.

Vi vil gerne undersøge fordelingen i tillid til politiet (`trstplc`) på tværs af partitilhør.

1. Vores data indeholder endnu ikke partitilhør, men det kan vi måske gøre noget ved...

Indlæs datasættet `ESS2014DK_polpartsub.csv` fra `'https://github.com/CALDISS-AAU/workshop_python-data-analysis/raw/master/datasets/ESS2014DK_polpartsub.csv'`

Undersøg datasættet: Er der en nøgle?

Brug `pd.merge()` til at tilføje `ESS2014DK_polpartsub` til det datasæt, I allerede har.

2. Dan en krydstabel med `pd.crosstab()` over partitilhør (`polpartvt`) og tillid til politiet (`trstplc`)
3. Der er værdier i partitilhør-variablen og variablen for tillid til politiet, som ikke kan bruges til noget. Konverter disse til missing med `replace()` (husk at `np.nan` værdien bruges til at ændre til missing).
4. Dan krydstabellen igen

BONUSØVELSE

1. Undersøg, om fordelingen er forskellig på tværs af køn med `groupby`
2. Find en passende måde at vise forskellen; enten ved brug af grafer eller ved at udregne middelværdi for tillid til politiet (dette kræver, at variablen konverteres til numerisk)

[]: