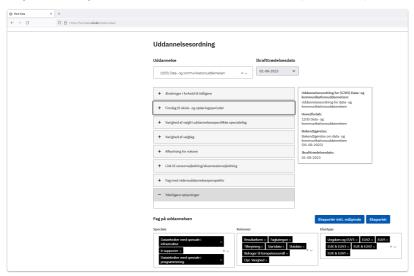
Data cleaning and filtering

Baggrund

På Industriens Uddannelsers hjemmeside https://iu.dk/, kan man hente et regnark med alle fag og målepinde for Datatekniker med speciale i programmerings uddannelsen.

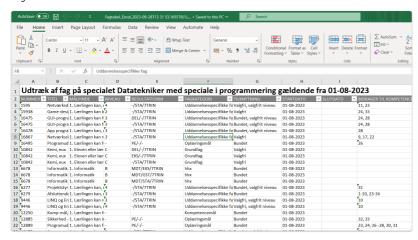
Gå til https://hentdata.stil.dk/uddannelser

Vælg (1205) Data- og kommunikationsuddannelsen + alle kolonner + "Eksporter inkl målepinde"



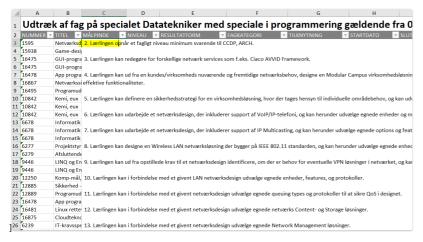
(eller hent regnearket i Moodle)

Desværre er formatet ikke helt lige til. Denne opgave går ud på at indlæse regnearket i en Python data struktur, og derefter filtre data, så vi kan finde præcist det som vi søger.



Kolonne nr. tre: Målpinde indeholder en liste af fagets målpinde i et sjovt format med linie-skift.

Vi vil gerne have den som en liste vi kan slå op i.



Udfordringer

- Regnarket indholder 4 faneblade: ét til fælles fag, ét til programmering, ét til infrastruktur og ét til IT-supporter
- Arket indeholder (af én eller anden grund) fag med kategorier som Grundfag og htx. Disse vil vi gerne filtrere væk.
- Nogle fag findes flere gange. Bla. én gang for hvert niveau (2-4). Vi ønsker blot at kigge på højeste niveau, som vi finde for hvert fag
- Nogle fag findes endda flere gang med samme niveau, men med forskellige Resultatformer. Vi ønsker at finde dag med resultatformen "-/STA/7TRIN"

• Målepindene står i ét felt, formateret med linie skift.

Se næste side

Opgaven

- Indlæs regnearket i en Pandas datastruktur
- Læs en Python funktion som tager et fag-nummer som input og returnerer en (Python) liste

```
1. Lærlingen kan redegøre for, hvordan en maskine kan lære.
2. Lærlingen kan anvende maskinlæring til at løse praktiske opgaver.
3. Lærlingen kan samle og forberede data til maskinlæring.
4. Lærlingen kan redegøre for forskellige maskinlæringsmetoder.
5. Lærlingen kan træne og teste sin model.
6. Lærlingen kan anvende klassificering og regression til at løse praktiske opgaver.
7. Lærlingen kan anvende et maskinlærings-framework til at løse en praktisk opgave.
1. Lærlingen kan, ud fra en case-opgave, omhandlende krav til en nyudviklet applikation, selvstændigt planlægge, gennemføre og dokumentere en softwaretest
med fokus på funktionalitet, performance og sikkerhed, og kan herigennem demonstrere viden, færdigheder og kompetencer ud over de i faget beskrevne mål og
målniveauer
2. Lærlingen har en generel viden om metoder og løsninger til at fejl- og funktionsteste en udviklet softwareløsning, samt om løsninger, der kan forbedre
sikkerheden på en web-enabled applikation.
3. Lærlingen kan tilrettelægge en testplan, som indeholder test til funktionelle krav.
4. Lærlingen kan anvende forskellige relevante testmetodikker.
5. Lærlingen kan vurdere, hvornår en software-release er tilstrækkeligt testet.
6. Lærlingen kan gennemføre en scriptet test på en app- eller webplatform
\textbf{7. L\"{a}rlingen kan integrere automatiserede tests i software udviklingscyklussen.}
8. Lærlingen kan gennemføre en tryktest, hvor performance og krav til hardware medtages.
9. Lærlingen kan aflæse performancetabeller og identificere flaskehalse i et kompliceret system.
10. Lærlingen kan redegøre for UX-tests (User Experience), som evaluerer kundeoplevelsen.
11. Lærlingen har en generel viden om data-kryptering, og kan anvende krypteringsformer som f.eks. SSL i forbindelse med udvikling af webbaserede
applikationer.
12. Lærlingen kan indhente viden omkring softwaresikkerhedsstandarder og kendte sikkerhedshuller i eksisterende software.
13. Lærlingen har en generel viden om authentication og kan udvikle programmer, der anvender authentication.
14. Lærlingen har en generel viden om Hashing-kodning, og kan anvende kodningsformer som f.eks. MD5 i forbindelse med udvikling af webbaserede
applikationer.
15. Lærlingen har en generel viden om forskellige former for hacking angreb, som f.eks. SQL injection og XSS (Cross-Site Scripting), og kan anvende denne
viden til at sikre egne udviklede webbaserede applikationer.
16. Lærlingen kan anvende forskellige website security tools til at teste sikkerheden på udviklet software.
```

Hints

Python pakken Pandas kan bruges til at indlæs CSV filer og Excel filer.

```
excel_file = 'Fagtabel_Excel_2023-09-28T13 31 53.165576.xlsx'

df = pd.read_excel(excel_file, sheet_name=1)
```

Brug Panda tutorials til at lære at organisere og filtre data, så vi kan lave den ønskede funktion!