

General

Det er fedt at du er ærlig med din brug af ChatGPT. Det giver os god mulighed for at forstå hvor du har været udfordret og kontekst for koden.

Der er ikke et requirements.txt således at folk kan køre scripts, og du gør ikke brug af opdelingen af data og source. Du har korrekt en mappe med data og en med source, men vælger at lægge data i source mapperne stadig. Ideen er at du i dit script kan tilgå Data mappen hente data derfra

Det er kun i opgave 1, at du faktisk har fil extension på dit script. Alle de andre er af typen `file` da der mangler `.py`. Så det er dejligt der er rigtig gode kommentarer i din kode. Din README filer er rigtig gode. Dog skal `Kørsel af scriptet` sectionen faktisk forklarer hvilken mappe du har haft som hoved mappe, og blot hvilket script man skal køre. Her mener jeg at sige: "Compiler og kød koden fra src mappen, hoved scriptet er main.py", eller noget i den stil, da dette er uafhængig fra om man bruger en IDE, og hvilken IDE det er.

Dine generelle tanker til videre arbejde er skide gode! Jeg vil også foreslå at du kigger på hvordan du kan gøre nogle af dem mere fleksible. I den forstand at de kan bruges nemt til at analysere lignende på tværs af forskellige formater og filer.

Opgave 1

Opgaven er ganske fint løst, der er nogle fine plots etc. Jeg ville dog have foretrykket at du lavede opgaven med det originale datasæt til sidst, for at få en bedre analyse. Jeg kan dog se at du har fjernet de navne der forekommer 2 gange i datasættet, i din analyse. Det fjerner data fra datasættet der kan berige dine fund, en anden gang så jeg gerne at du lavede analysen, fastslog at der er mange bogstaver der går igen, og konkluderer at der f.eks er 10 personer der hedder det samme.

Hvis du går ud på gaden og sampler data vil du heller ikke fjerne duplikater da det faktisk er en vigtig information. Du skriver at du udmiddelbart ville have brugt en uge på denne opgave alene hvis du ikke havde brugt chatgpt. Til dette vil jeg blot gøre dig opmærksom på at det også ville have været helt fint. Pointen er ikke at nå langt i opgaverne, men at få erfaring og lære det der af indhold pr. uge. På akademiet gælder det nogle gange om at 2 fantastiske opgaver istedet for 17 mindre gode.

Opgave 2

Super god ide med at have farve kodet dine error meddelelser. Dog skulle jeg manuelt rette nogle ting for at få det til at virke: Rettet `from colorama import init, Fore, Style` og tilføjet

```
if __name__ == "__main__":  
    init(convert=True)
```

Desuden er der ingen grund til at du i linie 78 og ned resetter style i hver linie, for så at sætte farven igen. Hvis du siger `print(Fore.RED + f"{message}")` bliver farven sat indtil du kalder Fore med en ny farve, eller du re-setter alt med `Style.RESET_ALL`. Det var en god ide, at du har håndteret at der kan være andre log-filer og at de ikke bliver overskrevet ved at du giver dem et tidsstempel. Dog er der allerede tidsstempel inde i filen, så du kunne bare have appendet til filen i stedet. Derudover skriver du at det er omstændigt at skulle forholde sig til så mange linier og det derfor kom bag på dig at der var SUCCESS og INFO etc. Her kunne du have brugt at log typen altid stod det samme sted, og generere en ny log fil til hvergang du stødte på et nyt ord i selve koden, i stedet for at kigge på dem selv.

Opgave 3

Her vil jeg specifikt kigge på de ting du har skrevet om i Udfordringer.

Din filsti er slet ikke dynamisk detsvære da din `base_path` er kun specifik for din computer. Fejlen her er at din sti er statisk, altså hardcoded i selve scriptet.

```
base_path = r"C:\Users\spac-46\Documents\GitHub\Uge 10, 2. undervisningsuge\Opgave  
3 - Kortfejl håndtering"
```

min pc user, heder ikke "spac-46" og jeg har ikke lagt dit script i "Documents" mappen. Hvis du skal have den absolutte sti til documentet kan du bruge `base_path = os.path.dirname(__filename__)`. Dog vil jeg råde dig til bare at bruge en relative path `data_path = os.path.join(".", "data")` og ellers få brugeren til at indskrive det når du kører scriptet. Eventuelt via commandline arguments men også bare mens du kører scriptet `data_path = input("Please input your data source ")`. En sidste bemærkning er at dine fejlmedelenser er gode.

Opgave 4

Lidt det samme filsti som før. Til dine udfordringer og feedback. Det ser generalt rigtig godt ud, hvis du gerne vil have lidt extra tydelighed kunne du eventuelt printe nogle vandrette linier meller overskrift og data, men det er en små ting. Hvis du gerne vil have at det skal se indu federe ud kan du også kigge her https://pandas.pydata.org/docs/user_guide/style.html. Dette er til atter at gøre stilen mere lækker og synsvenlig. Dine plots er også generalt fine, dog overvej at bruge bar plots mere, specielt til den første graf. Hvilken type graf du vælger afhænger i stor grad af hvad du vil informere. En line graph basser bedst til hvis der er et logisk sammenhæng eller vækst. Det er derfor den første tanke en person får når du ser en line graf. Det ville have

været bedre med bar plots. Dog er det fedt at se at du har titel, og akse værdier og at du har tænkt på significant decimal!