## **SPAC ugeopgave 1 - Python intro**

## Det er vigtigt at kunne udføre grundlæggende dataanalyse, fordi det gør det muligt at træffe informerede beslutninger baseret på data. I en verden, hvor der produceres store mængder data dagligt, kan evnen til at analysere og forstå data give afgørende indsigt i trends, mønstre og sammenhænge, som ellers ville være skjulte.

I denne opgave skal du skrive et program, der sorterer en liste af navne først efter længde, derefter alfabetisk. Brug derefter et dictionary til at tælle forekomsten af hvert bogstav i alle navne.

Under hver opgave finder du ressourcer der kan bruges til at løse den. Gennemlæs gerne hele dokumentet før du begynder.

**Tilgang til Opgaven**

* **Sortering af Navne**: Start med at oprette en liste af navne. Brug **Python's indbyggede sortering** med en brugerdefineret nøgle for at sortere navnene først efter længde og derefter alfabetisk.
* **Optælling af Bogstaver**: Gå igennem hvert navn i den sorterede liste og tæl forekomsten af hvert bogstav. Gem disse tællinger i et **dictionary**, hvor nøglen er bogstavet og værdien er antallet af forekomster.

**Hints**

* Brug **sorted()** funktionen med key parameteren for at sortere listen som ønsket.
* For at tælle bogstaver, kan du bruge en for løkke til at iterere igennem hvert navn og en indre for løkke til at iterere igennem hvert bogstav i navnet.
* Overvej at bruge **defaultdict**fra collections modulet for at forenkle optællingen af bogstaver.

**Ressourcer**

* Real Python - Python Data Structures Tutorial: <https://realpython.com/python-data-structures>
* Pluralsight: Core Python: Getting Started

**Avanceret Dataanalyse og Visualisering**

Hvis du er hurtigt færdig med denne opgave så kan du få yderligere udfordringer, enten med følgende udvidelse af opgaven eller ved at fortsætte med ugeopgave 2.

**Opgavebeskrivelse**  
Avanceret dataanalyse og visualisering hjælper med at forstå komplekse datasæt på en intuitiv og dybere måde. Avanceret analyse kan identificere skjulte mønstre, tendenser og sammenhænge, som grundlæggende metoder ikke afslører. Ved at visualisere disse data kan man let kommunikere komplekse indsigter, så de bliver forståelige for beslutningstagere, der måske ikke har teknisk baggrund.

Byg videre på den oprindelige opgave ved at udføre en detaljeret analyse af navnedataene. Efter at have talt forekomsten af hvert bogstav i alle navne, udfør følgende:

* **Frekvensanalyse**: Beregn og visualiser frekvensen af hvert bogstav anvendt i navnelisten. Overvej at bruge biblioteket **matplotlib**eller **seaborn**til at lave et histogram- eller søjlediagram, der viser de mest almindelige bogstaver
* **Ordsky**: Generer en ordsky (word cloud) baseret på hyppigheden af hvert bogstav. Jo oftere et bogstav forekommer, desto større skal det vises i ordskyen. Dette kan gøres med **wordcloud biblioteket i Python**.
* **Navnelængde Analyse**: Analyser fordelingen af navnelængder i dit dataset. Beregn **gennemsnitlig** og **median**for navnelængderne, og visualiser dataen med passende plots.

**Yderligere Overvejelser**

Overvej at filtrere navnelisten for duplikater før analysen, og gentag kørslen og sammenlign resultaterne.

**Ressourcer for Udvidet Opgave**

* Matplotlib Dokumentation: For grundlæggende til avancerede plots: [https://matplotlib.org](https://matplotlib.org/)
* Seaborn Dokumentation: For statistiske data visualiseringer: [https://seaborn.pydata.org](https://matplotlib.org/)
* WordCloud Dokumentation: For at generere ordsky billeder:   
  <https://github.com/amueller/word_cloud>