

## 3. Dataanalys (Curie)

19 juli

Programmering är väldigt användbart ifall vi har stora mängder av data som vi vill analysera. Idag kommer vi undersöka hur vi kan använda oss av Python för att analysera en stor mängd data på ett effektivt sätt!

1. Ladda ner csv-filerna från <https://github.com/hairez/mattekollo-2023>.
2. Spara filerna i en directory du gillar (och som du har lätt tillgång till).
3. Starta en Python fil i samma directory.
4. Ta reda på den exakta directoryn du har filerna. Kör följande kod för att hitta directoryn som filen körs från just nu:

```
import os
print(os.path.dirname(__file__))
```

För att kunna arbeta med datan kan du först öppna filen i Python:

```
file=open("/Users/dirc/svenska-stader.csv")

inp=[*file] #Varje rad från filen blir ett element i arrayen.
file.close()

print(inp[0])
print(inp[1])
```

1. (a) Hur många platser (Locality) i Sverige börjar på en vokal?  
(b) Hur många av dessa ligger i Skåne (County)?
2. Vilken "County" i Sverige har flest platser (Locality) i datan svenska-stader.csv?
3. Vilka 100 platser (Locality) ligger mest i norr i Sverige?
4. Vilka 100 platser ligger närmast Nynäshamn, Stockholm?
5. Hur många städer i US har både militär och är "incorporated"? (Svaret är inte 0. Men ifall du får 0 som svar, tar reda på varför!)
6. Hur många ZIP-nummer har alla städer i Texas?
7. (a) Vad är medelvärde på invånarantal för alla städer i US?  
(b) Vilken stad har ett invånarantal som ligger närmast medelvärdet?
8. Finns det fler invånare i Västra-USA, eller Östra-USA?
9. — **Extra.** Placera ut alla städer i Sverige/US på ett koordinatsystem på ett bra sätt med hjälp av matplotlib.pyplot.

Ledtrådar: [https://github.com/hairez/diploma-project/blob/main/code/tsp/data/manual\\_testdata/csvDataRetriever.py](https://github.com/hairez/diploma-project/blob/main/code/tsp/data/manual_testdata/csvDataRetriever.py) och <https://github.com/hairez/diploma-project/blob/main/code/visualizer/plotTestdata.py>

**10. — Extra.** Ändra storleken på alla punkter i koordinatsystemet så att storleken på punkterna representerar invånarantalet.

**11. — Extra.** I `uscities.csv` finns det data av städer i USA som ligger relativt långt borta från alla andra punkter. Lista ut vilken del av USA de tillhör, och filtrera ut dessa ur datan.

**12. — Extra Extra.** Googla upp data på något annat som du vill undersöka. Skapa någon sorts graf eller placera ut datan på ett koordinatsystem.

