Session 2: Problem Solving

Objectives

Our business objective is to create profiles for the companies in Denmark, which host our DAT and SOFT practices.
There are numerous benefits in knowing the profiles, such as:

- we have more precise estimation of the workplaces and the opportunities they provide to our students
- we have a real-life feedback from the companies regarding the knowledge and skills they are searching for in new employees.

Our learning objective is to get broad experience in collecting data from various sources and preparing it for data analysis and AI implementations.

Tasks Related to Achievement of the Objectives

1. Download the MS Excel file provided to you by the instructor. It contains the real *SOFT 2023spring* data.
   Feel encouraged to extend the table with more data available to you.
   In case you add more rows to the table that concern students from DAT, WEB, or another programme, add a new column, which would contain the name of the programme.
2. Anonymise people mentioned in the table by either change their names with fake names or scramble/encrypt them.
3. Use CVRAPI () to find and extract all available data for the companies listed in the source file. Store the extracted data in a file and a *pandas* data frame in your Python application.
4. Use a Python geo-library to extend the available company data with geolocation of companies' branches. Add the new data in new columns of the data frame.
5. Clean the data in the data frame.
6. Explore the set by applying methods and measures of descriptive statistics.
7. Store the data frame into a CSV file
8. When you are ready, send a mail to the instructor programmatically from your application.

Help

Feel free to refer to/use the sample code snippets provided by the instructor, but don't miss to search Internet for more.
Contact the instructor in the classroom any tie you need external assistance.

Note

This problem stays at the basis of the first Mini Project – MP1 that gives study points.
MP1 operates with more data.