Project Worksheet Day 2

August 24, 2021

Yesterday you selected a data set to work on and made an inventory of the analysis methods that you could use to answer the research questions in your project. Today your task is identify Python packages (and possibly also other pieces of useful software) that provide essential functions for the implementation of these methods. Two examples of packages that are often used in data science applications (and that will also be discussed in the next days) are the pandas data analysis library (https://pandas.pydata.org/) and the matplotlib 2D plotting library (https://matplotlib.org/). However, many more are available and Google and the Python Package Index (https://pypi.org/) will help you to find the right ones for your project.

Summarize your findings in a table that lists the identified analysis methods from yesterday, and for each of them the identified packages and functions along with a short description of the functions and their main inputs (parameters, but also the object on which a method might be called) and outputs (return values or other kinds of results). If you find more than one package/function for a particular method, you can list the different alternatives, and decide later which one to use in your project.

Don't worry if you don't know yet how to actually use those functions in Python. You will get to that point soon, but it is not required knowledge for this assignment.

For the "ICT kennis en vaardigheden" running example this could look as follows:

Analysis method	***Python Package***	***Function***	***Function Description***	***Inputs***	***Outputs***
(Clustered) bar charts	*matplotlib*	*bar*	*Make a bar plot.*	*A sequence of scalars*	*A BarContainer*
(Clustered) bar charts	*seaborn*	*barplot*	*Show point estimates and confidence intervals as rectangular bars.*	*A variety of possible input formats, pandas objects are preferable*	*A matplotlib Axes object*
Line graphs	*matplotlib*	*plot*	*Plot y versus x as lines and/or markers.*	*Array-like or scalar*	*A list of Line2D objects*
Sort data	*pandas*	*sort*	*Sort DataFrame either by labels or by values*	*A pandas DataFrame*	*A pandas DataFrame*
...	*...*	*...*	*...*	*...*	*...*