

Data Processing with R and Python

Assignment 3

As an important note, it should not be forgotten that the notebook file that contains all the code is assumed to be placed in the Desktop and the data sets are assumed to be downloaded into the Downloads folder and the paths of files and folders are arranged accordingly.

Data Collections

In this assignment 3 data sources must be chosen, one of which must be a big sized file, and data analysis must be done for them.

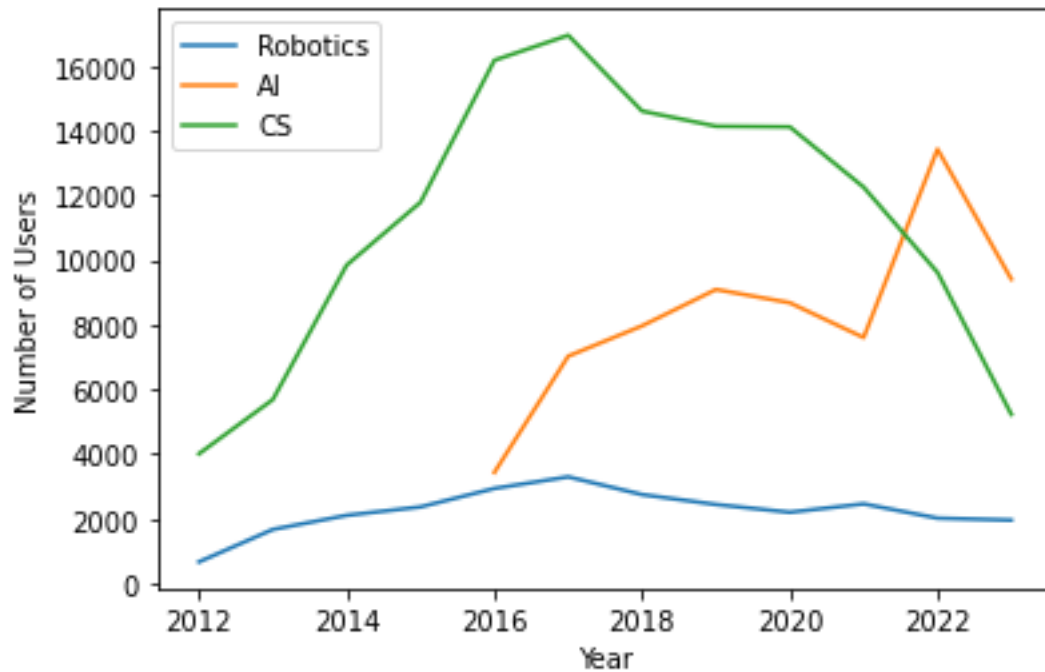
I used data from <https://archive.org/download/stackexchange>

The chosen data sets are:

1. [robotics.stackexchange](#) - 81.3 MB
2. [ai.stackexchange](#) – 30.6 MB
3. [cs.stackexchange](#) – 111 MB

Data Analysis

In the analysis, the query showing how many users created in each year is executed for all the subjects and a line plot is created for comparison. Below the result can be seen:



Interest in AI and CS websites has been significantly decreased while for robotics mostly the interest has steady state. And we can see that AI has gotten attention from the users after 2016 and it has its peak in 2022. It is worth noting that after the rise of interest in AI the interest in CS started to decline and it has declined ever since, which is after year 2017. We can also see that a very little amount of users Robotics has for each year.