Python advanced – 2021 spring

**Task 0:**

Create a folder MSPython with the following subfolders: soundcloud, scikitlearn, mypackages. Please use the subfolders for completing the tasks.

**Task1.1** – SoundCloud environment:

Create a virtual environment **soundcloud\_env** to the soundcloud folder. Activate the environment and install *pandas*, *requests* and *bs4* packages. Create requirements.txt file. Add the virtual environment to Jupyter and use that for Task1.2.

**Task1.2** – SoundCloud scraping:

Soundcloud is one of the biggest online audio distribution and music sharing website, and it does not block scraping robots.

(Allowed methods and useful information usually available in robots.txt - <https://soundcloud.com/robots.txt>)

The task is to visit [https://soundcloud.com/popular/searches](https://soundcloud.com/popular/searchesand%20scrape%20the%20top1) and scrape the top10 most popular searches. The result should be the following pandas DataFrame:

Table

Description automatically generated

Column ***text***: the top10 most popular searches

Column ***link***: the link of the search

Column ***html***: the html code of the search result

Steps: (any other functional solution is also accepted)

1. Get the http code of <https://soundcloud.com/popular/searches> and print the request status, the encoding and the response text.
2. Create a list from the most popular searches, and another list from its search link - which is also available in the response. (help1: Use BeautifulSoup to find all “a” tag) (help2: the search function has a href parameter, and if this is set to True, you can easily get the SoundCloud links out of the result)
3. Create a loop, visit all links and save the html responses to a list. Add it to the DataFrame as a new column.

**Task2** – Scikit-Learn/SciPy example:

Demonstrate one of the applications of SciPy or Scikit-Learn (regression, clustering, classification, ...) on any dataset. Create a plot with matplotlib.

**Task3** – Python package:

Create a python package that includes a function. Import the package and use the function in a Jupyter notebook.

**Submission**: Create a .zip or .rar file from the MSPython folder and send it to [domonkos.febo@gmail.com](mailto:domonkos.febo@gmail.com)

**Deadline**: 2020.05.30, 12:00

**Notes**:

If something is not clear or you got stuck, feel free to reach out with any concrete questions.

Any partial solution will be also evaluated, so please do not skip full sections.

After completing the tasks, the MSPython folder’s structure should look similar to this:

Text

Description automatically generated