Projet roadmap

The work required is divided into four stages, each of which is described as follows:

1) Pre-project: define an issue related directly or indirectly to tourism issues. Describe your use case and prove the feasibility of your idea by specifying the source sites for your data. To prove feasibility, I recommend that you test scraping data from the sites identified and checking whether the data obtained is interesting.

2) Scraping the data: in this step, you describe the structure of the data you put in place following the scrapping and its origins. You should also describe the quality of your data in terms of representativeness, the completeness of the values, the cleaning and the homogenization of the data (which structures, databases, etc.). The aim of this stage is to present the difficulties encountered during the scrapping, the solutions found and the presentation of the final structure of your data.

3) Describe the machine learning problem (the machine learning task(s) you wish to implement as part of your project). Visualizations, or the reduction of data in order to understand it, are considered to be tasks.

4) The final prototype you are proposing in terms of a web application, or an exploratory notebook, an information search interface, etc

Some requirements:

-Your project must include textual data and a related problem. A bonus is awarded if you manage to connect this project to the second project in the NLP module,

-the project must deal with a subject related to tourism

-the project must take into account an ecological and responsible dimension (social, personal, etc.)

-the data collected must be obtained using at least one of the following techniques: a) scraping via API , b)scraping via selenium/beautifulsoup. The data is therefore obtained by a)+b) minimum.

Deliverables:

13/12: the final version of the pre-project with the first scrapping notebooks

26/12: the link to the data collected in csv format and the scrapping notebook (github)

15/01: the link to the Machien learning notebook

20/01: a user manual for illustrating use cases for your prototype.

Defense: date to be defined (in progress) during exam week.

-At the end of the project, you are invited to submit on DVO a) your final Python code, b) your production, c) your oral presentation (the slides)

-Prepare an oral presentation describing your project (aim, hypothesis, method, problems. solutions, results (as a demo), conclusions) in 5 minutes.