Project-2: **200 Points**
Due date: 10/25

Project description: This is a part of a semester-long project. In the future Project-1 and Project-2 will get integrated. In this project you will be scrapping a certain ecommerce website's user reviews for a product. Step to consider before scrapping:

1. At first identify a product for which you want to scrap the user reviews. You also have to scrap the user reviews of the earlier versions of the product. For example, you choose an iphone as your product. The latest is iphone-16, go to an ecommerce website and automatically scrap the user comments for this version of iphone. Now You have to scrap user comments for previous versions of iphone (15, 14, 13). In your project choose a product that has a fairly long history of different versions.
2. Many ecommerce websites may not allow you to scrap the data. You may have to use a browser emulator or use a low key website that may allow you to scrap the website. To begin with, start with amazon and ebay. Ebay should work.
3. After you select your product and the website, copy the URLs of 4-5 different versions on the product and store it in a text file. BUT, nowadays all the user comments are not in a single webpage, you have to click next to load more comments. So, for a single version of the product there may be multiple URLs. So, instead of 4-5 URLs there will be many. If you find some smart way to scrape many user comments for a product, please go for it.
4. Use the above files which contain URLs to the different versions of the product and scrap the content of the URL automatically. There are many python packages out there to scrap a website.
5. Clean the downloaded web pages and extract only the comments from it. Many python packages will do that for you. In fact the package that you are going to use in step-4 above may have some setting to do it.
6. After the extraction of the user comments for a product, version wise, save the comments in separate files. For example, if you have a product and you are considering 4-5 versions of that product then you should have 4-5 different files each containing user comments for each of the versions.

Instructions:

1) Please use the same repo that you used for project-1, just create a separate branch for this project. Name the branch "webScrapping". (IMPORTANT)
2) In this branch push the python files and the requirements.yml.
3) Write an elaborate README.md file. Mentioning what the software does and any other information that you want to mention. Then write instructions on how to use your software. This is important. Think of it as all the steps required to run your software. DO NOT OVERWRITE THE PREVIOUS README FOR PROJECT-1. This is a new branch. Remember that.
4) Also push the output files (4-5 of then) and the input file (containing the URLs) to the repo.

Rubric: 200 points [by interview]

- Able to automatically download and save only the user comments of 4-5 different versions of a product in different files (use github, read instructions above): 150 points
- requirements.yml file is present: 20 points
- README.md is written well: 30 points (this is subjective so try to impress the TAs with an elaborate readme)

If you fail to implement the program I will personally evaluate the project and a substantial amount of points will be taken off, sorry.

How you are going to share the project with the TAs will be announced later. Stay tuned. If you fail to attend the interview in your allotted schedule, substantial points will be deducted, please be there for the online interview, do not forget.

IF YOU HAVE ANY QUESTION ASK ME AFTER THE CLASS OR DURING THE CLASS, I AM HAPPY TO EXPLAIN.