Third Project Guidelines

ECO225, Professor Khazra

Please start working on your project ASAP. Web scraping is going to surprise you with errors and failures. Please plan to finish at least one or two days before the deadline to account for unprecedented situations and to get help.

Your third project should include your first and second projects. You will have three main sections for your project (and some subsections that you title yourself). The first main title should be "Project One," and the second main title should be "Project Two", and the third main title should be "Project Three". You can add titles by adding "#" sign before your title in your markdown cell; check this link for more details on titling: https://www.datacamp.com/community/tutorials/markdown-in-jupyter-notebook

Formatting

You will lose **5 points** if your submitted project does not meet any of the following requirements:

1. Submitted projects must be in PDF. Any other format, including a .ipynb file, is not accepted.

2. The projects should have clear "section titles." Besides the subsections you choose to have, you should have three main sections; Project One, Project Two and Project Three.

3. In case you missed this in your first project: choose a title for your paper. What is the question that you would like to answer using this dataset? The question can be the title of your paper.

4. You should write your project in Jupyter Notebook (Python) and submit it in pdf format. If you have problems converting your notebook to pdf, first download it as HTML, then print/save the HTML version in pdf.

5. This is an individual project. However, you are encouraged to check the projects on the Kaggle.com website that use similar data. We have provided some "useful links" on the data list on Quercus. You can use these sources, but the coding and explanations **must be yours**. Do not copy and paste the same chunk of code in your project.

Please note that if you include a graph or a table, you should **explain what you learn from it.** Do not add an output without any explanation. All projects should have an introduction and a conclusion. Suppose you want to send your project to a company or school that you are applying to. The final product should be a clean and comprehensive report.

- Do not include unnecessary chunk of codes or outputs and errors.
- Any graph, summary, or output should have an explanation following. Why do you include it in your report, and what do you understand from it?

Third Project Details

We will build upon the first and second projects towards a full academic paper on a hot topic using real-world data and cool Python techniques, just excellent! Here's what we are going to do for this project.

Part One

1. **(10 points)** Incorporate the comments that you have received in your second report. This is the first part of your third project, which will be graded again. We will check the first section of your projects (your updated previous project) and grade it again. If you have not received any comments, try to improve on your previous report by doing a better job on the introduction or literature review or fine-tuning your visualizations, or adding more meaningful analysis. You will always revisit your past projects until the final project (your masterpiece).

Part Two

The objective of this project is to add more information to your dataset using web-scraping. In practice, finding data with web-scraping can be more complicated than this exercise. Depending on the type of data you want to scrape, you may need to run your scraping program for months to parse data from a website regularly. What we are doing here is a first step and is not going to take long.

1. **(10 points)** Think about the main message/question of your paper. Think about what information you can add to your dataset that is not already in your dataset and requires web-scraping. Know that there are many other available datasets that you can simply download, and you do not need to do web scraping to acquire them. So, focus on the data that you can have only if you generate it with web-scraping.

   If you can find the answer to this question and work on it in the future, you will probably be the first one working on that topic! It can lead to an original paper. So, describe the website that can be an excellent addition to your dataset and clearly explain how it can improve your results. It may be an HTML-based scraping or

API-based scraping (searching for available scraping APIs is a good place to look for new ideas and accessible tools for web-scraping).

Clearly explain (1) why you think this data enhances your paper, (2) the address where you can scrape it from, (3) how you are going to merge the data with your results and how you are going to use this new information in answering your research question.

2. **(10 points)** Do you need to run your program over time (daily, monthly, ...) to generate the data? Can you scrape the data source that you have found for this project? If not, what are the possible challenges that make it hard for you to use it in this project? What do you need to learn to work on that in the future?

3. **(20 points)** If you think you can scrape the website that you have picked in the last part, go ahead and scrape it. I will add extra points for that (based on the originality and quality).

   But, I know that it could be hard at this point to work on that perfect data source. So, instead of working on your proposed website, you can work on the practice website that I have provided for your specific dataset (you can find the practice websites on Quercus module for week 9 in an excel sheet). If you decide to work on the practice website, you still need to do the last parts and develop new ideas.

   You may find it easier to copy and paste the data from the provided websites. But, you should not copy and paste, this is a practice, and I want you to scrape the data as we did in the lab.

   - Scrape the data from the practice website (provided for your project in the excel sheet) or the one that you have found interesting.
   - You should clearly explain each chunk of your code in this part. Maybe divide your scraping code into three (or more) parts like we did in the lab and add explanations for someone new to web-scraping after each part.

4. **(10 points)** Merge the new scraped data with your original data.

5. **(10 points)** Visualize the new data and explain how it is related to your project (use maps, trends, histogram, etc.).

6. **(10 points)** Write a proper summary and conclusion that includes your findings so far.

7. **(20 points)** I will evaluate the quality of your project, explanations, and its originality. You should look at everything from the lens of an economist. All of your plots, figures, and results have to have economic. explanations and interpretations, as I have emphasized before. The appearance of your submission matters as well, fix the typos, have nice headings and take out unnecessary long tables and errors.

Upload your Jupyter Notebook (your code and explanation) in pdf format on Crowdmark for marking.

Enjoy being a creator!