MILESTONE 3

The third milestone contains one set of deliverables and spans one lab period (P5). The work is described as four tasks and included in this document, so you can work at your own pace and see the progression of tasks to complete the project.

Remember that time management is one of the skills you are learning to develop during this project, and as we are approaching the end of the term, you should have gotten better at time management. Therefore, all the tasks for this milestone will be due on the same day: April 12th, the last day of classes with no extensions. We recommend you wrap up the project and submit everything by Friday 8^{th,} April.

File Naming Convention: You are given specific formats for your filenames throughout the project. This format includes your team identifier, e.g., **T021**.

Milestone 3 Objectives
In the first and second milestones, you established a good understanding of the project: the problem domain of dataset analysis, the project goals of building an interactive dataset analyzer program, and the problem solution.

The objectives of the third milestone are:
- Modify the **Txxx_P1_load_data** module to discard duplicated books that appear under the same category.
- Write a README file for the application.
- Write the Project Report.
- Do a final code review for your application

P5 Task 1 Remove duplication.
Use the function design recipe to modify the **Txxx_P1_load_data** module developed in Milestone 1 to remove duplicated values.

In Milestone 1, we did not worry about duplicate entries (e.g., a book can appear twice in the same category). As we are wrapping up the project, we want to ensure that the dictionary returned by the function in the **Txxx_P1_load_data** module does not contain duplicates. Note that if the same book is stored in different categories, it is not considered a duplicate.

Save this new version of the function in a file named **Txxx_P5_load_data.py**, where xxx is your team identifier.

After modifying the function, retest everything, including all the previous functionalities you have implemented in Milestone 1 and Milestone 2. You need to ensure that the modification of the function did not change the required behavior of other functions. Because your function is stored in a file with a new name, you must modify the import statements.
No submission is required at this point.

P5 Task 2: Final Team Code

Your dataset analyzer program is now complete. Well, almost….

As a team, you will release your software. Package your code together.

1. Create a new folder. Required Folder name: **Txxx_data_analyzer**, where xxx is your Brightspace team identifier
2. Copy-paste **Txxx_P5_load_data.py** and the other modules developed during the project into this folder (<u>do not include the old version of load_data</u>). As a team, review the code and ensure that everything works correctly.
3. Remember to include the team members' names at the top of each file. Also, write the date and a version number. We will call this version of the software Version 1.0.
4. Have another look at *Appendix A: Coding "Do's and Don'ts"* from Milestone 1. Does all your new code follow the coding conventions listed in that document? If not, edit the code so that it does. Polish, polish, polish (and while doing this, remember to retest, retest, retest).
5. Make sure you have refracted your code as explained in Milestone 2, and it still compiles and runs.
6. Submit the project folder (i.e., **Txxx_data_analyzer**)—one submission per team. Your submission must include a functional version of your Python program. That means the TA should be able to execute your program and use the interface provided without any additional files other than the ones you submitted. Your submission must include the three modules you previously developed as a team.

P5 Task 3: Code Documentation

At this point, your program should be well-documented, with type annotations, docstrings, and inline comments (where needed). One more piece of documentation is required when delivering software: a README file.

A README file is a text file (called **README.md**) that introduces a project and tells a new user how to run it. Note that your team identifier is **not** part of the file name.

- Read this link for an explanation: https://www.makeareadme.com/
  - (Because no one can read your mind (yet))
- Read this link for tips on how to write a README: https://www.wikihow.com/Write-a-Read-Me
- GitHub is an industry-standard repository for sharing open-source code. Read this link for their advice on writing READMEs: https://guides.github.com/features/wikis/
  - Pay attention only to the README, not the GitHub WIKI
  - GitHub users now format their READMEs in a markup language that allows images and bolding. You will need to introduce these markup items as explained in Lecture 11.

The specific requirements for your team's README are:
- The file must be created using a text editor (for example, Windows Notepad or macOS' equivalent text editor) and be called README.md. It must not be a Word (.docx) file. It must not be a Rich Text Format (.rtf) file. It must not be a PDF file. Hint: run your

README.md file through the text-processing programs presented in class. If the output is garbled, then the file is not a text file.
- Your README must contain the following content, most of which is demonstrated in: https://www.wikihow.com/Write-a-Read-Me
    - o Contact Information (Using the Team Leader's information)
    - o Date
    - o Software's Name and version
    - o Description
    - o Installation
    - o Usage (Not shown in the second link, but shown in the GitHub example. Use shell commands, and do not use images)
    - o Credits (Tell us which team member did what during the project)
    - o License (Your choice. Click on the link in the GitHub example for some options)

The following are posted in Brightspace to help you complete this task
- **README Rubric** – How your report will be marked.

P5 Task 4: Final report

As a team, you must write a **Project Report** that describes your understanding of the problem and how you wrote the program to solve this problem.

Required Filename: **Txxx_report.pdf** where xxx is your team identifier.

The following are posted in Brightspace to help you complete this task
- **Project Report Template** – An empty document skeleton with primary headings and descriptions of what goes in each section. You just need to write each section.
- **Project Report Sample** – A lab re-written in the style of the project report.
- **Project Report Rubric** – How your report will be marked.

The project goal must describe the whole project (until the end of Milestone 3). You must read the provided resources to consider the project *as a whole*.

P5 Task 5: ITP Metrics

You will complete the **Peer Feedback and Team Dynamics assessments** as you did before. You need to complete these assessments individually, but ITP Metrics combines the results into an overall evaluation for each team member.

On your own and without interference from any member of your team:
1. Login into your account on ITP Metrics.
2. Individually, complete the **Peer Feedback** and **Team Dynamics Assessment.**

Optional: There are PDF reports generated by ITP metrics that give you more insight into the feedback that your peers gave to you. These PDF reports are only generated if (1) your group size is three or more and (2) all members of your group complete the task. If these conditions are met, you are free to download these PDF reports and have a look, but it is not required. Please do not email the instructors if a report is not generated – we cannot do anything about it. That is why this is an optional activity.

Brightspace Submission Summary

- P5 – T1: No submission. You need to complete this as it is required for P5-T2
- P5 – T2: **`Txxx_data_analyzer`** (zip folder), which contains all the project code.
- P5 – T3: **`README.md`**
- P5 – T4: **`Txxx_report.pdf`**