

Assignment 2 | Dockerfile and Data Analysis

with Popular Books Dataset

Objective

For this assignment, you will be working with the Popular Books Dataset from Kaggle. The dataset can be accessed from the following link: [[Popular Books Dataset](#)]

Your task is to create a Jupyter Notebook that includes **data cleaning** and **preprocessing steps**, followed by an **analysis** focused on the Harry Potter book series.

Requirments

1. Docker Setup

- Use the Official Jupyter Docker image from the following link: [[Jupyter Docker Image](#)].
- Refer to the documentation provided at this link: [[Jupyter Docker Stacks Documentation](#)] for guidance on selecting and running the Docker image.

2. Dockerfile

- Create a Dockerfile for your project. You can refer to the following reference for assistance: [[Setting Up and Running Jupyter Notebook in a Docker Container](#)].
- Configure the Dockerfile to run Jupyter Notebook when the container launches using the following command:

CMD ["jupyter", "notebook", "--ip='0.0.0.0'", "--port=8888", "--no-browser", "--allow-root"]

3. Notebook Development

- Use the Popular Books Dataset in your Jupyter Notebook.
- Perform data cleaning and preprocessing steps to ensure the dataset is ready for analysis.
- Focus your analysis on the Harry Potter book series.
- Find the most selling books within the Harry Potter series.
- Calculate the average rating of the Harry Potter books.

4. GitHub Repository

- Create a GitHub repository for your project.
- Push the following items to the repository:
 - Dockerfile
 - Jupyter Notebook
 - Dataset (downloaded from Kaggle)
 - PDF file containing the results from the analysis.

Submission Criteria:

- Ensure that your GitHub repository contains all the required files mentioned above.
- Submit the link to your GitHub repository as your assignment submission.

The deadline for the project is **Wednesday 24th April, 11:59 PM.**

Submissions should be made through the provided form [here](#).

If you have any questions or need assistance, please don't hesitate to ask.