**Project Documentation**

Book Data Extraction using Python

# Purpose of the Project

This project aims to extract information such as book titles, prices, and ratings from the website Books to Scrape using web scraping techniques in Python. The data is stored in a structured JSON format for further analysis.

# Challenges Faced

1. Handling Unicode Symbols

* **Issue**: When writing the data to a JSON file, the price appeared as "\u00c2\u00a3" instead of a human-readable £.
* **Solution**: Used json.dump(..., ensure\_ascii=False) to keep Unicode characters like £ readable in the output file.

1. Extracting Ratings
   * **Issue**: Book ratings were embedded as part of a CSS class name (e.g., star-rating Four) instead of being directly available.
   * **Solution**: Retrieved the class list and extracted the second class name as the rating.

# Insights Gained from the Data

* The range of book prices, helping understand affordability.
* The distribution of ratings (e.g., how many books are rated Three vs. Five).