**Full Python Project: Advanced Book Scraper**

**Target Website**: https://books.toscrape.com

**Project Goals**

1. Scrape 100 books (5 pages)
2. For each book:
   * Title
   * Price
   * Availability
   * Star rating
   * Category
   * Product description
   * Product page URL
3. Save to:
   * books.csv
   * books.sqlite (SQLite database)

**Tools Required**

pip install requests beautifulsoup4 requests

**Project Structure (Flat File)**

All in one script: bookscrapingapp.py

**📊 Example Output Row**

| **Title** | **Price** | **Rating** | **Category** | **Description** |
| --- | --- | --- | --- | --- |
| It's Only the Himalayas | £45.17 | Two | Travel | An adventure memoir... |

**Pandas queries** based on the **BooksToScrape scraping project**

**Step 1: Load Data First**

**Pandas Queries**

['Title', 'Price', 'Availability', 'Rating', 'Category', 'Description', 'URL']

**1. Convert Price from string to float**

**2. Top 5 most expensive books**

**3. Average price of all books**

**4. Count of books by star rating**

**5. Number of books in each category**

**6. Books with "Python" in the title**

**7. Books currently in stock**

**8. Out-of-stock books**

**9. Count of books missing a description**

**10. Most common book rating**

**11. Number of unique categories**

**12. List all unique categories**

**13. Sort all books alphabetically by title**

**14.n Books with the longest descriptions**

**15. Cheapest book in each category**

**16. Top 3 most expensive books per category**

**17. Books with title length > 50 characters**

**18. Add a column for "is\_expensive" (price > £40)**

**19. Count of expensive books by category**

**20. Save filtered DataFrame of in-stock & expensive books**

**Bonus: Load directly from SQLite (if used)**

Your DataFrame df includes:

['Title', 'Price', 'Availability', 'Rating', 'Category', 'Description', 'URL']

And Price is already converted to float, like:

**21. Find the category with the highest average price**

**22. Category with the lowest average price**

**23. How many books have 5-star ratings**

**24. List all 5-star books priced above £50**

**25. What is the average price of 1-star books?**

**26. Top 10 categories with the most in-stock books**

**27. Which books are completely out of stock?**

**28. Books with descriptions containing the word "mystery"**

**29. Book with the highest price in the entire dataset**

**30.Book with the lowest price**

**31. How many books are priced exactly at £50?**

**32. Rank books by price within each category**

**33. Books with a title starting with 'A'**

**34. Average title length per category**

**35. Most common word in all titles**

**36. Books with descriptions longer than 300 characters**

**37. Distribution of books by rating and category**

**38. Remove duplicate titles**

**39. Find books priced below category average**

**40. Pivot table of average price per category and rating**

**Matplotlib Queries**

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read\_csv("books.csv")

df['Price'] = df['Price'].str.replace('£','').astype(float)

**1. Average Price per Category**

**2. Number of Books by Rating**

**3. Books in Stock vs Out of Stock**

**4. Top 10 Most Expensive Books**

**5. Book Count by Category (Top 10)**

**6. Distribution of Book Prices (Histogram)**

**7. Title Length Distribution**

**8. Boxplot of Prices by Rating**

**9. Number of Books Over Price Thresholds**

**10. Top Categories with Most 5-Star Books**

**12. Top 10 Longest Book Descriptions**

**13. Category Count in Pie Chart (Top 6)**

**14. Books per Rating as Line Chart**

**15. Price Comparison by Rating (Bar)**