🕸️ Web Scraping – Lesson 2: Scraping Tables into DataFrames

🎯 What You'll Learn

In this lesson, you'll learn how to scrape structured data from a live website (Wikipedia), convert it into a clean **pandas DataFrame**, and optionally export it to a CSV file.

🧠 Key Concepts

- **Targeting complex tables** in real-world websites using BeautifulSoup
- Navigating through multiple HTML tables and selecting the correct one
- Extracting table headers (`<th>`) and row data (`<td>`) for clean formatting
- Building and populating a pandas DataFrame with scraped content
- Handling edge cases like empty rows or inconsistent structures
- Exporting the final data to a CSV using pandas

📂 Real-World Applications

- Automating data collection from sources like Wikipedia, company listings, or financial data sites
- Creating dynamic datasets for analytics or dashboard tools
- Regularly updating structured CSVs from changing web content

---

✅ Best Practices

- Always inspect HTML structure before scraping
- Use `find_all()` with indexing for pages with multiple tables
- Clean your data (e.g., strip whitespace) before loading it into pandas
- Include `index=False` when exporting to CSV to avoid unwanted row numbers
- Anticipate edge cases like missing rows or extra headers

📌 Recap

✅ Navigated and selected the correct HTML table on a complex webpage
✅ Extracted and cleaned headers and data rows
✅ Loaded data into a pandas DataFrame for analysis
✅ Successfully exported the table to a CSV file