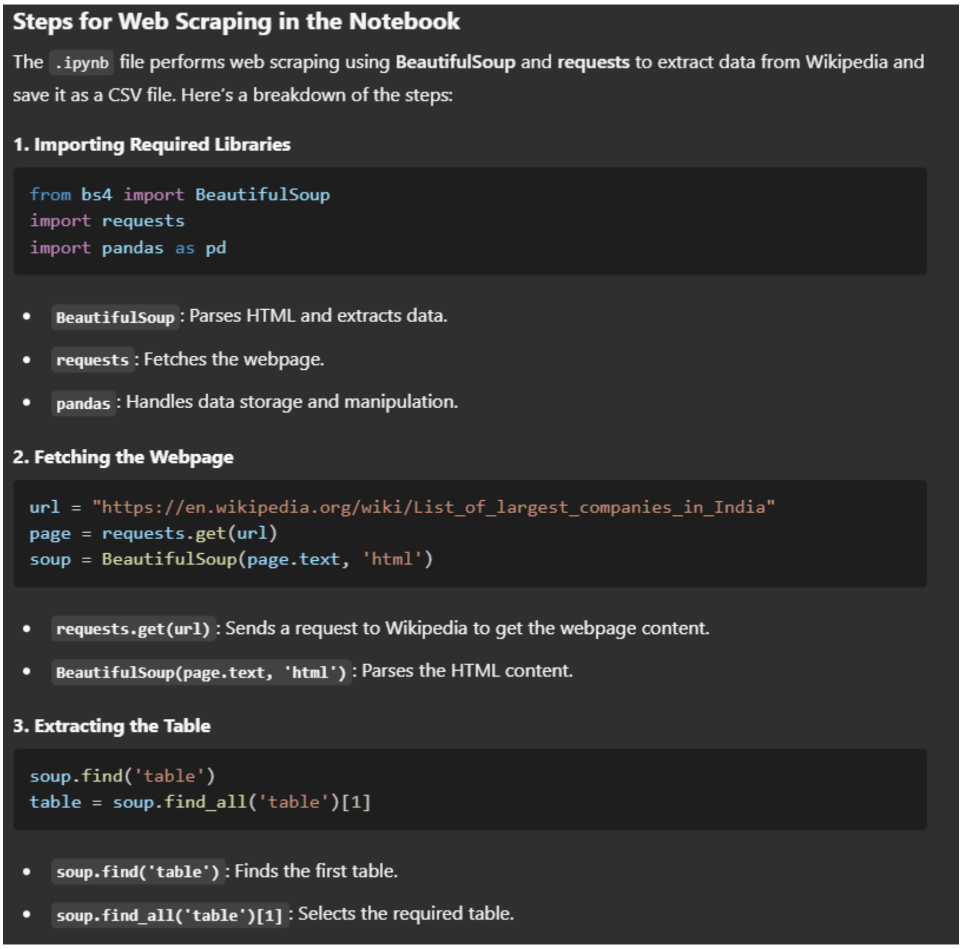
Explanation of the Code: -



Steps for Web Scraping in the Notebook

The `.ipynb` file performs web scraping using **BeautifulSoup** and **requests** to extract data from Wikipedia and save it as a CSV file. Here's a breakdown of the steps:

1. Importing Required Libraries

```
from bs4 import BeautifulSoup
import requests
import pandas as pd
```

- `BeautifulSoup` : Parses HTML and extracts data.
- `requests` : Fetches the webpage.
- `pandas` : Handles data storage and manipulation.

2. Fetching the Webpage

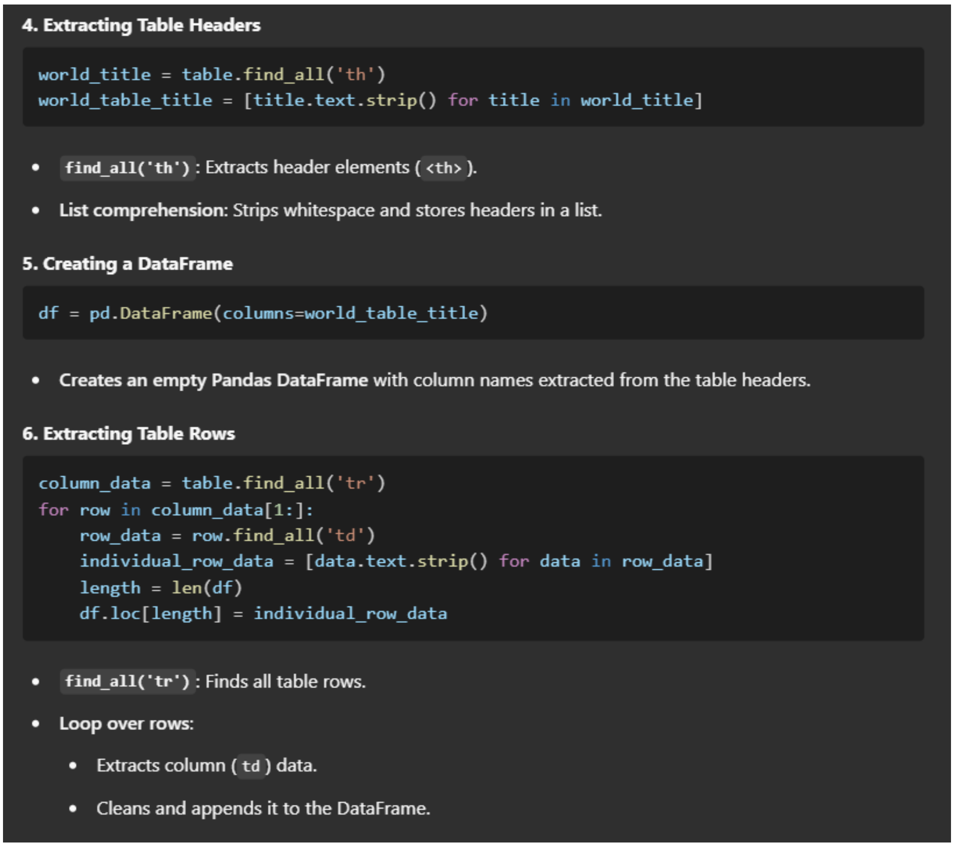
```
url = "https://en.wikipedia.org/wiki/List_of_largest_companies_in_India"
page = requests.get(url)
soup = BeautifulSoup(page.text, 'html')
```

- `requests.get(url)` : Sends a request to Wikipedia to get the webpage content.
- `BeautifulSoup(page.text, 'html')` : Parses the HTML content.

3. Extracting the Table

```
soup.find('table')
table = soup.find_all('table')[1]
```

- `soup.find('table')` : Finds the first table.
- `soup.find_all('table')[1]` : Selects the required table.



4. Extracting Table Headers

```
world_title = table.find_all('th')
world_table_title = [title.text.strip() for title in world_title]
```

- `find_all('th')` : Extracts header elements (`<th>`).
- **List comprehension:** Strips whitespace and stores headers in a list.

5. Creating a DataFrame

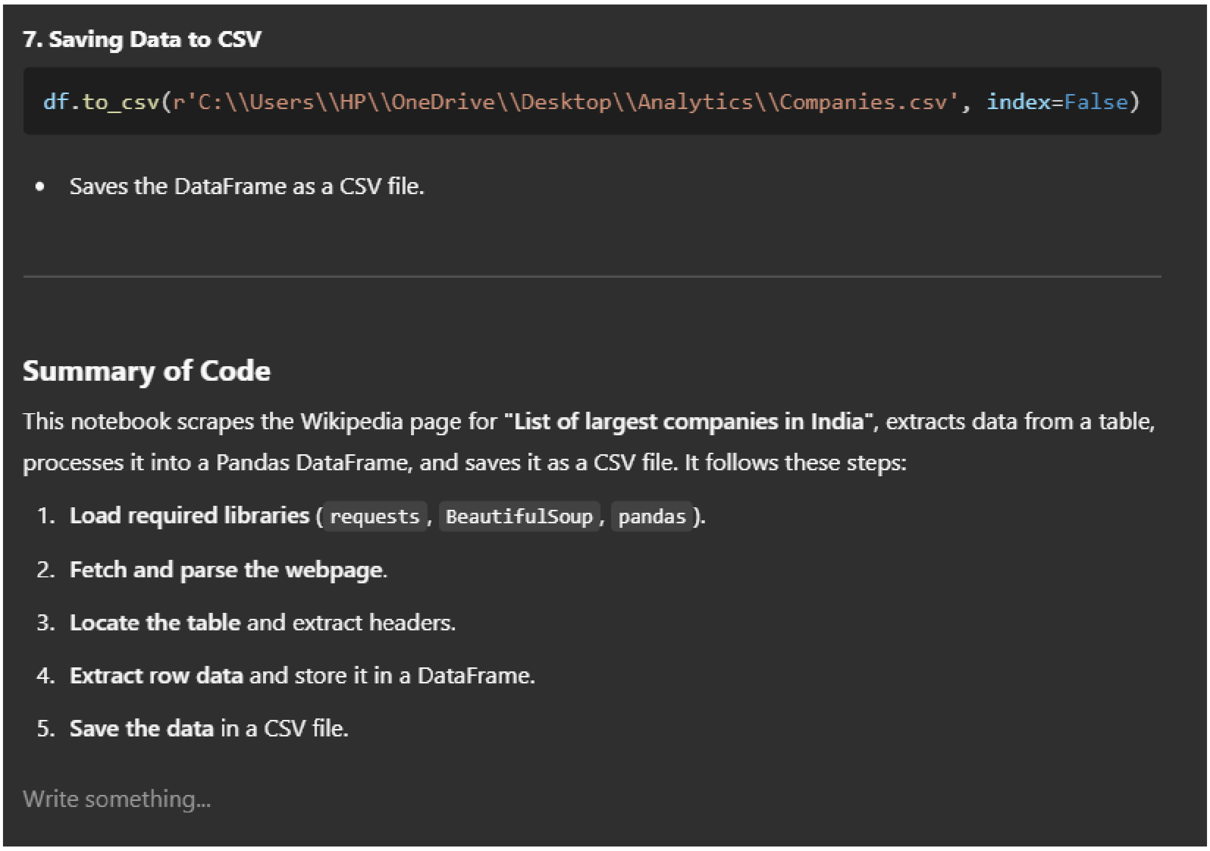
```
df = pd.DataFrame(columns=world_table_title)
```

- **Creates an empty Pandas DataFrame** with column names extracted from the table headers.

6. Extracting Table Rows

```
column_data = table.find_all('tr')
for row in column_data[1:]:
    row_data = row.find_all('td')
    individual_row_data = [data.text.strip() for data in row_data]
    length = len(df)
    df.loc[length] = individual_row_data
```

- `find_all('tr')` : Finds all table rows.
- Loop over rows:
  - Extracts column (`td`) data.
  - Cleans and appends it to the DataFrame.



7. Saving Data to CSV

```
df.to_csv(r'C:\\Users\\HP\\OneDrive\\Desktop\\Analytics\\Companies.csv', index=False)
```

- Saves the DataFrame as a CSV file.

Summary of Code

This notebook scrapes the Wikipedia page for **"List of largest companies in India"**, extracts data from a table, processes it into a Pandas DataFrame, and saves it as a CSV file. It follows these steps:

1. **Load required libraries** (`requests` , `BeautifulSoup` , `pandas`).
2. **Fetch and parse the webpage**.
3. **Locate the table** and extract headers.
4. **Extract row data** and store it in a DataFrame.
5. **Save the data** in a CSV file.

Write something...