Web Scraping Project on Various website

_By Sneha Bhattacharjee

Introduction

The core objective of this project is to demonstrate the versatility and efficiency of Python in web scraping tasks, showcasing its ability to handle diverse data sources and structures. By automating the extraction process, this project aims to provide a comprehensive overview of different sectors, enabling users to stay informed, make data-driven decisions, and gain a deeper understanding of market trends and consumer behavior.

Each section of the project focuses on a specific domain, utilizing tailored scraping techniques to collect and process the data. The collected data is then organized and presented in a user-friendly format, making it accessible and actionable for various applications. Whether it's keeping up with the latest news, finding a new home, exploring job opportunities, checking the weather, discovering movie ratings, or comparing shopping options, this project showcases the potential of web scraping to transform the way we interact with online information.

—

Web Scraping

What is Web Scraping?

Web scraping, also known as web data extraction, is the process of automatically collecting data from the web. It involves fetching web pages and extracting useful information from them, which can be stored and analyzed. This is particularly useful for gathering large amounts of data quickly and efficiently from multiple sources.

Why is Web Scraping Used?

1. **Data Collection:** Web scraping allows the collection of vast amounts of data from websites, which can be used for analysis, research, or reporting.
2. **Market Research:** Businesses can scrape competitors' websites to gather information on pricing, products, and customer reviews.
3. **Content Aggregation:** It can be used to aggregate content from multiple sources, such as news articles, blogs, and forums.
4. **Price Monitoring:** E-commerce sites can monitor prices on competitor websites to adjust their own prices competitively.
5. **Lead Generation:** Companies can collect contact information from websites to generate leads for sales and marketing.
6. **Job Listings:** Aggregating job postings from various job boards into a single platform.

Common Libraries for Web Scraping in Python

Python is one of the most popular languages for web scraping due to its simplicity and the powerful libraries available:

1. **BeautifulSoup:** This library is used for parsing HTML and XML documents and extracting data from them in a readable format. It works well with the `requests` library to fetch web pages.
2. **Requests:** A simple and elegant HTTP library for Python, used to make HTTP requests to retrieve web pages.
3. **Scrapy:** An open-source and collaborative web crawling framework for Python. It provides a comprehensive toolset for scraping and crawling websites.
4. **Selenium:** A browser automation tool that can be used for web scraping dynamic content that requires interaction with JavaScript.

Other Methods for Web Scraping

While Python is a popular choice for web scraping, other programming languages and tools can also be used:

1. **R:**
   - R has packages like `rvest` and `httr` which are used for web scraping.
   - `rvest` simplifies the process of scraping web data, making it similar to `BeautifulSoup`.
2. **JavaScript (Node.js):**
   - Using libraries like `Cheerio` and `Axios`, JavaScript can be used for web scraping.
   - `Puppeteer` is used for headless browser automation, similar to `Selenium`.
3. **Java:**
   - Libraries like `Jsoup` make it easy to scrape web content using Java.
4. **Excel:**
   - Excel can be used for simple web scraping tasks using Power Query.
   - It allows users to import data from web pages into Excel worksheets.

By leveraging these tools and techniques, web scraping can be efficiently performed to gather valuable data from various web sources.

—

1. News Web Scraping Code

In [16]:
```
# News Headlines

import json
import requests
```

```
from bs4 import BeautifulSoup

# Set the URL and request headers
url = "https://www.bbc.com/news"
headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36

# Send an HTTP request and parse the HTML content
response = requests.get(url, headers=headers)
soup = BeautifulSoup(response.content, "html.parser")

# Extract headlines and summaries
articles = soup.find_all("article")
news_data = []

for article in articles:
    if article.find("h3"):
        headline = article.find("h3").text.strip()
    else:
        headline = ""
    summary = article.find("p").text.strip()
    url = article.find("a")["href"]

    news_data.append({
        "headline": headline,
        "summary": summary,
        "url": url
    })

# Save the data to a JSON file
with open("news_data.json", "w") as outfile:
    json.dump(news_data, outfile, indent=2)

#print("Scraping completed! Check the 'news_data.json' file for the results.")

with open('news_data.json', 'r') as f:
    news_data = json.load(f)

# Now you can access the news data from the JSON file
print(news_data)
```

```
[{'headline': '', 'summary': "Claudia Sheinbaum has beaten her rival, Xóchitl Gálvez,
by a landslide to become Mexico's next leader.", 'url': '/news/articles/cp4475gwny1
o'}]
```

—

2. Real Estate Web Scraping Code

In [24]:

```
#Real Estate

# Import libaries
from bs4 import BeautifulSoup
import requests
import pandas as pd
import re

# Set the URL and request headers
url="https://www.nobroker.in/property/rent/hyderabad/multiple?searchParam=W3sibGF0Ijox
page=requests.get(url)
soup=BeautifulSoup(page.content,"html.parser")
```

```
# Iterate through the listings and extract relevant information
listings=soup.find_all("a",class_="overflow-hidden overflow-ellipsis whitespace-nowrap
prices = soup.find_all(["div","span"], class_="font-semi-bold heading-6")
sqfts = soup.find_all(id="unitCode", class_="flex")
typs= soup.find_all("div", class_="font-semibold whitespace-nowrap")


for listing,price,typ,sqft in zip(listings,prices,typs,sqfts):
    print("Basic info:", listing.text)
    print("Price:", price.text)
    print("sqft:", sqft.text)
    print("Rental Yield:", typ.text)
    print()
```

```
Basic info: 2 BHK Flat In Padmavathi for Rent  In Hyderabad
Price: ₹ 20,000No Extra Maintenance
sqft: 450 sqft
Rental Yield: 2 BHK

Basic info: 1 BHK Flat In Uday Chamber for Rent  In Kukatpally
Price: ₹20,000
sqft: 800 sqft
Rental Yield: 1 BHK

Basic info: 2 BHK House for Rent  In Kukatpally
Price: 450 sqft
sqft: 1,500 sqft
Rental Yield: 2 BHK

Basic info: 2 BHK Flat In Star Apartment  for Rent  In Kukatpally
Price: ₹ 12,000 +₹ 1,000Maintenance
sqft: 1,000 sqft
Rental Yield: 2 BHK

Basic info: 1 BHK Flat In Vlcc  for Rent  In Kukatpally
Price: ₹36,000
sqft: 400 sqft
Rental Yield: 1 BHK

Basic info: 1 BHK Flat In Sk Apartments for Rent  In Bhagyanagar Colony, Kukatpally
Price: 800 sqft
sqft: 510 sqft
Rental Yield: 1 BHK

Basic info: 1 BHK Flat for Rent  In Kukatpally
Price: ₹ 23,000No Extra Maintenance
sqft: 400 sqft
Rental Yield: 1 BHK

Basic info: 1 BHK Flat In Anupama Towers for Rent  In Kukatpally
Price: ₹69,000
sqft: 550 sqft
Rental Yield: 1 BHK

Basic info: 2 BHK Apartment In Kasamma Residency for Rent  In Kukatpally
Price: 1,500 sqft
sqft: 670 sqft
Rental Yield: 2 BHK

Basic info: 2 BHK Flat In Standalone Bulinding for Rent  In Kukatpally
Price: ₹ 20,000 +₹ 1,200Maintenance
sqft: 1,000 sqft
Rental Yield: 2 BHK

Basic info: 3 BHK Flat for Rent  In Kukatpally
Price: ₹60,000
sqft: 1,500 sqft
Rental Yield: 3 BHK

Basic info: 1 BHK House for Rent  In Kukatpally,
Price: 1,000 sqft
sqft: 500 sqft
Rental Yield: 1 BHK
```

```
Basic info: 1 RK House for Rent  In Kukatpally
Price: ₹ 20,000 +₹ 1,000Maintenance
sqft: 250 sqft
Rental Yield: 1 RK

Basic info: 3 BHK House for Rent  In Kukatpally
Price: ₹60,000
sqft: 700 sqft
Rental Yield: 3 BHK

Basic info: 1 BHK House for Rent  In Kukatpally
Price: 400 sqft
sqft: 600 sqft
Rental Yield: 1 BHK

Basic info: 2 BHK Flat for Rent  In Kukatpally
Price: ₹ 12,000 +₹ 1,500Maintenance
sqft: 1,200 sqft
Rental Yield: 2 BHK

Basic info: 2 BHK Flat In Harika Nest for Rent  In Kukatpally
Price: ₹24,000
sqft: 950 sqft
Rental Yield: 2 BHK

Basic info: 2 BHK Flat for Rent  In Kukatpally
Price: 510 sqft
sqft: 1,300 sqft
Rental Yield: 2 BHK

Basic info: 2 BHK House for Rent  In Kukatpally
Price: ₹ 16,000 +₹ 1,000Maintenance
sqft: 700 sqft
Rental Yield: 2 BHK

Basic info: 2 BHK House for Rent  In Kukatpally
Price: ₹30,000
sqft: 1,000 sqft
Rental Yield: 2 BHK

Basic info: 4 BHK House for Rent  In Kukatpally
Price: 400 sqft
sqft: 4,500 sqft
Rental Yield: 4 BHK

Basic info: 3 BHK House for Rent  In Kukatpally
Price: ₹ 18,000 +₹ 1,500Maintenance
sqft: 1,300 sqft
Rental Yield: 3 BHK

Basic info: 2 BHK Flat In Ap for Rent  In Kukatpally
Price: ₹36,000
sqft: 1,500 sqft
Rental Yield: 2 BHK

Basic info: 3 BHK Flat In Shankar Sadan Appartment  for Rent  In Kukatpally
Price: 550 sqft
sqft: 1,250 sqft
Rental Yield: 3 BHK
```

```
Basic info: 1 BHK Flat In Standalone Building  for Rent  In Kukatpally
Price: ₹ 24,000 +₹ 1,800Maintenance
sqft: 700 sqft
Rental Yield: 1 BHK

Basic info: 2 BHK Flat In Padmavathi for Rent  In Hyderabad
Price: ₹72,000
sqft: 450 sqft
Rental Yield: 2 BHK
```

—

3. Job Portal Web Scraping Code

In [18]:

```
# Job Portal

# Import Libaries
import requests
from bs4 import BeautifulSoup
import re

# Set the URL and request headers
url = "https://internshala.com/jobs/"
headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36

# Send an HTTP request and parse the HTML content
response = requests.get(url, headers=headers)
page=requests.get(url)
soup = BeautifulSoup(page.content, "html.parser")
#print(response)

# Extract job postings
titles = []
companies = []
ctcs =[]
exps = []
locs = []

for company_element in soup.find_all(["div","h3"], class_= "heading_4_5 profile"):
    titles.append(company_element.text.strip())

for company_element in soup.find_all(["div","p"], class_= "heading_6 company_name"):
    companies.append(company_element.text.strip())

for ctc_element in soup.find_all(["span"], class_= "desktop"):
    ctcs.append(ctc_element.text.strip())

for exp_element in soup.find_all(["div"], class_= "item_body desktop-text"):
    exps.append(exp_element.text.strip())

for title, company, ctc, exp in zip(titles, companies, ctcs, exps):
    print(f"Title: {title}")
    print(f"Company: {company}")
    print(f"CTC: {ctc}")
    print(f"Experience: {exp}")
    print()
```

```
Title: Presales Executive - Real Estate
Company: Headstrong HR Consulting Services
CTC: ₹ 2,00,000 - 3,60,000
Experience: 0-1 years

Title: Teacher - English
Company: PlanetSpark
CTC: ₹ 3,00,000
Experience: 1-5 years

Title: Senior Interior Designer
Company: Nettle Creek Interiors
CTC: ₹ 3,00,000 - 4,50,000
Experience: 0-3 years

Title: Senior Social Media Marketing Manager
Company: House Of Amber
CTC: ₹ 5,00,000 - 6,00,000
Experience: 2-4 years

Title: Junior Business Management Associate
Company: Law Gate
CTC: ₹ 2,10,000 - 2,40,000
Experience: 0-2 years

Title: Sales Associate
Company: PlanetSpark
CTC: ₹ 6,00,000 - 7,00,000
Experience: 0-5 years

Title: Audit Associate
Company: Tranistics Data Technologies Private Limited
CTC: ₹ 2,00,000
Experience: 0-1 years

Title: Corporate Sales Manager
Company: Urban NXT
CTC: ₹ 2,50,000 - 7,00,000
Experience: 0-5 years

Title: Junior Content Writer
Company: Growth Accelerators
CTC: ₹ 2,40,000 - 3,00,000
Experience: 0-2 years

Title: English Teacher
Company: PlanetSpark
CTC: ₹ 3,00,000
Experience: 1-5 years

Title: Sales Executive Domestic Tours
Company: Travel Gypsy Pvt. Ltd.
CTC: ₹ 2,00,000 - 2,50,000
Experience: 0-3 years

Title: Video Editor
Company: LearnFinite Edutech Private Limited
CTC: ₹ 2,00,000 - 2,40,000
Experience: 0-2 years
```

```
Title: Academic Trainer ( Computer Science + Maths )
Company: NowIntern
CTC: ₹ 3,00,000 - 5,00,000
Experience: 1-3 years

Title: Digital Marketing Associate
Company: Zama Organics
CTC: ₹ 3,00,000 - 4,00,000
Experience: 1-2 years

Title: Junior Interior Designer
Company: Wattieza Designs
CTC: ₹ 2,00,000 - 3,00,000
Experience: 0-1 years

Title: Senior Telecaller
Company: Adisri Publications Private Limited
CTC: ₹ 2,00,000 - 3,00,000
Experience: 1-4 years

Title: Video Editor
Company: React Labs Private Limited
CTC: ₹ 4,00,000 - 5,00,000
Experience: 1-2 years

Title: Junior Social Media Marketing Manager
Company: Culture Circle
CTC: ₹ 4,50,000 - 6,00,000
Experience: 1 year

Title: Ad Operations Specialist
Company: Unibots
CTC: ₹ 2,50,000 - 3,50,000
Experience: 0-1 years

Title: Junior Graphic Designer
Company: Educase India
CTC: ₹ 2,00,000 - 3,60,000
Experience: 0-2 years

Title: Junior Accountant
Company: Adisri Publications Private Limited
CTC: ₹ 2,00,000 - 2,05,000
Experience: 0-2 years

Title: Human Resources (HR) Associate
Company: LearnFinite Edutech Private Limited
CTC: ₹ 2,00,000 - 2,40,000
Experience: 0-2 years

Title: Junior Media & Public Relations (PR) Specialist
Company: Atelierish Media
CTC: ₹ 2,00,000 - 6,00,000
Experience: 0 years

Title: Video Editor
Company: Big Mount Tv
CTC: ₹ 2,00,000 - 2,50,000
Experience: 1-3 years
```

```
Title: Office Assistant
Company: Veeshan Labs
CTC: ₹ 2,00,000 - 2,50,000
Experience: 1-3 years

Title: Office Assistant
Company: Veeshan Labs
CTC: ₹ 2,00,000 - 2,50,000
Experience: 1-3 years

Title: Market Research Analyst
Company: Sentriscope Data Intelligence Private Limited
CTC: ₹ 3,00,000 - 4,00,000
Experience: 1-3 years

Title: 3D Animator
Company: Multeway
CTC: ₹ 3,00,000 - 4,00,000
Experience: 0 years

Title: Customer Success Associate
Company: Zama Organics
CTC: ₹ 2,50,000 - 3,20,000
Experience: 0-1 years

Title: Influencer Marketing
Company: Friends Media
CTC: ₹ 2,50,000 - 6,00,000
Experience: 1-5 years

Title: Content Writer
Company: Klicksurge
CTC: ₹ 2,00,000 - 4,00,000
Experience: 0-2 years

Title: Junior Social Media Marketing Executive
Company: Grivaa Consultancy Services
CTC: ₹ 2,00,000 - 2,80,000
Experience: 0-3 years

Title: Graphic Designer
Company: Zama Organics
CTC: ₹ 2,00,000 - 2,50,000
Experience: 0-1 years

Title: Digital Marketing Associate
Company: Webmobi
CTC: ₹ 2,40,000 - 3,50,000
Experience: 0-2 years

Title: Business Development Executive
Company: IMI Studios Private Limited
CTC: ₹ 2,16,000 - 3,00,000
Experience: 0-2 years

Title: Operations Manager
Company: Grivaa Consultancy Services
CTC: ₹ 2,00,000 - 2,80,000
Experience: 0-1 years
```

```
Title: Project Executive
Company: Mirats Insights
CTC: ₹ 2,00,000 - 3,00,000
Experience: 0-1 years

Title: Finance Video Creator
Company: Grivaa Consultancy Services
CTC: ₹ 2,00,000 - 3,00,000
Experience: 0-2 years

Title: Human Resources (HR) Associate
Company: Grivaa Consultancy Services
CTC: ₹ 2,00,000 - 2,80,000
Experience: 0-2 years

Title: Data Science Technical Trainer
Company: Allsoft Solutions And Service Private Limited
CTC: ₹ 3,50,000 - 5,00,000
Experience: 1-5 years

Title: Content Marketing
Company: WeMakeScholars
CTC: ₹ 3,00,000 - 3,20,000
Experience: 0-1 years

Title: Digital Marketing Executive
Company: Grivaa Consultancy Services
CTC: ₹ 2,40,000 - 3,00,000
Experience: 0-1 years

Title: Animator
Company: One Media Group
CTC: ₹ 2,16,000 - 3,80,000
Experience: 0-2 years

Title: Telecaller (Tamil)
Company: Astrotalk
CTC: ₹ 2,00,000 - 2,50,000
Experience: 0 years

Title: Corporate Sales Executive
Company: Webmobi
CTC: ₹ 2,00,000 - 3,50,000
Experience: 0-2 years

Title: Business Development Executive (Tamil)
Company: Astrotalk
CTC: ₹ 2,00,000 - 2,50,000
Experience: 0-1 years

Title: Operational Marketing Executive
Company: Swan Environmental Private Limited
CTC: ₹ 2,04,000 - 4,00,000
Experience: 0-5 years

Title: Senior Bussiness Develpoment Service Engineer
Company: Swan Environmental Private Limited
CTC: ₹ 5,00,000 - 7,00,000
Experience: 3-5 years
```

```
Title: Junior Software Tester
Company: Educase India
CTC: ₹ 2,00,000 - 3,00,000
Experience: 0-2 years

Title: Product Developer (Science And Maths Assessments Design) IIT/NIT/IISc/BITS/IIS
ER
Company: Open Door Education
CTC: ₹ 8,00,000 - 12,00,000
Experience: 0-2 years
```

—

4. Weather Web Scraping Code

In [19]:
```
# Weather web Scraping

import requests
from bs4 import BeautifulSoup

# Set the URL and request headers
#url = "https://www.bbc.com/weather/2643743"  #London
url = "https://www.bbc.com/weather/1275004" #Kolkata
headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36

# Send an HTTP request and parse the HTML content
response = requests.get(url, headers=headers)
soup = BeautifulSoup(response.content, "html.parser")
#print(response.content)

# Extract current weather conditions
location = soup.find(["div", "h1"], class_="wr-c-location__name gel-paragon").text
current_temp = soup.find(["span"], class_="wr-value--temperature--c").string.strip()
condition_text = soup.find(["div"], class_= "wr-day__weather-type-description wr-js-da

# Print the collected data
print(f"Location: {location}")
print(f"Current Temperature: {current_temp}")
print(f"Condition: {condition_text}")
```

```
Location: Kolkata - Weather warnings issued
Current Temperature: 27°
Condition: Partly cloudy and a gentle breeze
```

—

5. Movie Rating Web Scraping Code

In [20]:
```
# Movie Ratings

import requests
from bs4 import BeautifulSoup

url = 'https://editorial.rottentomatoes.com/guide/best-movies-of-all-time/'  # IMDb's

# Send a GET request to the IMDb page
response = requests.get(url)
```

```
# Parse the HTML content of the webpage
soup = BeautifulSoup(response.text, 'html.parser')
#print(response)

# Find all movie titles and ratings
titles = []
ratings = []
years = []

for title_element in soup.find_all(["span","a"], class_= "title"):
    titles.append(title_element.text.strip())

for rating_element in soup.find_all(["span"], class_= "score"):
    ratings.append(rating_element.text.strip())

for year_element in soup.find_all(["span"], class_= "year"):
    years.append(year_element.text.strip())

for title, rating, year in zip(titles, ratings, years):
  print(f"Title: {title} {year}" )
  print(f"Rating: {rating}")
  print("--------------------------")
```

```
Title: L.A. Confidential (1997)
Rating: 99%
-------------------------
Title: The Godfather (1972)
Rating: 97%
-------------------------
Title: Casablanca (1942)
Rating: 99%
-------------------------
Title: Seven Samurai (1954)
Rating: 100%
-------------------------
Title: Parasite (2019)
Rating: 99%
-------------------------
Title: Schindler's List (1993)
Rating: 98%
-------------------------
Title: Top Gun: Maverick (2022)
Rating: 96%
-------------------------
Title: Toy Story 2 (1999)
Rating: 100%
-------------------------
Title: Chinatown (1974)
Rating: 98%
-------------------------
Title: On the Waterfront (1954)
Rating: 99%
-------------------------
Title: The Battle of Algiers (1966)
Rating: 99%
-------------------------
Title: Toy Story (1995)
Rating: 100%
-------------------------
Title: Rear Window (1954)
Rating: 98%
-------------------------
Title: Modern Times (1936)
Rating: 98%
-------------------------
Title: How to Train Your Dragon (2010)
Rating: 99%
-------------------------
Title: All About Eve (1950)
Rating: 99%
-------------------------
Title: Spirited Away (2001)
Rating: 96%
-------------------------
Title: Up (2009)
Rating: 98%
-------------------------
Title: The Third Man (1949)
Rating: 99%
-------------------------
Title: Spotlight (2015)
Rating: 97%
-------------------------
```

```
Title: Spider-Man: Into the Spider-Verse (2018)
Rating: 97%
--------------------------
Title: The Philadelphia Story (1940)
Rating: 100%
--------------------------
Title: Finding Nemo (2003)
Rating: 99%
--------------------------
Title: Singin' in the Rain (1952)
Rating: 100%
--------------------------
Title: 12 Angry Men (1957)
Rating: 100%
--------------------------
Title: Toy Story 3 (2010)
Rating: 98%
--------------------------
Title: Sunset Boulevard (1950)
Rating: 98%
--------------------------
Title: Coco (2017)
Rating: 97%
--------------------------
Title: The Godfather, Part II (1974)
Rating: 96%
--------------------------
Title: Three Colors: Red (1994)
Rating: 100%
--------------------------
Title: Selma (2014)
Rating: 99%
--------------------------
Title: Zootopia (2016)
Rating: 98%
--------------------------
Title: Citizen Kane (1941)
Rating: 99%
--------------------------
Title: Annie Hall (1977)
Rating: 97%
--------------------------
Title: Cool Hand Luke (1967)
Rating: 100%
--------------------------
Title: The Holdovers (2023)
Rating: 97%
--------------------------
Title: Inside Out (2015)
Rating: 98%
--------------------------
Title: Dr. Strangelove Or: How I Learned to Stop Worrying and Love the Bomb (1964)
Rating: 98%
--------------------------
Title: Let the Right One In (2008)
Rating: 98%
--------------------------
Title: The Lord of the Rings: The Two Towers (2002)
Rating: 95%
--------------------------
```

```
Title: Knives Out (2019)
Rating: 97%
-------------------------
Title: M (1931)
Rating: 100%
-------------------------
Title: Toy Story 4 (2019)
Rating: 97%
-------------------------
Title: The Wrestler (2008)
Rating: 98%
-------------------------
Title: Goodfellas (1990)
Rating: 95%
-------------------------
Title: The Wizard of Oz (1939)
Rating: 98%
-------------------------
Title: Double Indemnity (1944)
Rating: 97%
-------------------------
Title: Psycho (1960)
Rating: 97%
-------------------------
Title: Paddington 2 (2017)
Rating: 99%
-------------------------
Title: Before Sunrise (1995)
Rating: 100%
-------------------------
Title: The Dark Knight (2008)
Rating: 94%
-------------------------
Title: The Maltese Falcon (1941)
Rating: 99%
-------------------------
Title: It Happened One Night (1934)
Rating: 98%
-------------------------
Title: The Wages of Fear (1953)
Rating: 100%
-------------------------
Title: North by Northwest (1959)
Rating: 97%
-------------------------
Title: Bicycle Thieves (1948)
Rating: 99%
-------------------------
Title: Alien (1979)
Rating: 93%
-------------------------
Title: Argo (2012)
Rating: 96%
-------------------------
Title: Get Out (2017)
Rating: 98%
-------------------------
Title: The Kid (1921)
Rating: 100%
-------------------------
```

```
Title: Mission: Impossible - Dead Reckoning Part One (2023)
Rating: 96%
--------------------------
Title: The Pianist (2002)
Rating: 95%
--------------------------
Title: Kind Hearts and Coronets (1949)
Rating: 100%
--------------------------
Title: The 400 Blows (1959)
Rating: 99%
--------------------------
Title: Grave of the Fireflies (1988)
Rating: 100%
--------------------------
Title: The Big Sick (2017)
Rating: 98%
--------------------------
Title: Minari (2020)
Rating: 98%
--------------------------
Title: Portrait of a Lady on Fire (2019)
Rating: 97%
--------------------------
Title: The Treasure of the Sierra Madre (1948)
Rating: 100%
--------------------------
Title: Apocalypse Now (1979)
Rating: 97%
--------------------------
Title: Mission: Impossible - Fallout (2018)
Rating: 97%
--------------------------
Title: The Last Picture Show (1971)
Rating: 98%
--------------------------
Title: Tampopo (1985)
Rating: 100%
--------------------------
Title: Mad Max: Fury Road (2015)
Rating: 97%
--------------------------
Title: Tokyo Story (1953)
Rating: 100%
--------------------------
Title: A Hard Day's Night (1964)
Rating: 98%
--------------------------
Title: Metropolis (1927)
Rating: 97%
--------------------------
Title: Good Will Hunting (1997)
Rating: 97%
--------------------------
Title: The Gold Rush (1925)
Rating: 100%
--------------------------
Title: Aliens (1986)
Rating: 98%
--------------------------
```

```
Title: Spider-Man: Across the Spider-Verse (2023)
Rating: 95%
--------------------------
Title: The Good, the Bad and the Ugly (1967)
Rating: 97%
--------------------------
Title: Harry Potter and the Deathly Hallows: Part 2 (2011)
Rating: 96%
--------------------------
Title: The Silence of the Lambs (1991)
Rating: 95%
--------------------------
Title: Fanny and Alexander (1982)
Rating: 100%
--------------------------
Title: Laura (1944)
Rating: 100%
--------------------------
Title: The Shop Around the Corner (1940)
Rating: 99%
--------------------------
Title: His Girl Friday (1940)
Rating: 99%
--------------------------
Title: All Quiet on the Western Front (1930)
Rating: 98%
--------------------------
Title: Monsters, Inc. (2001)
Rating: 96%
--------------------------
Title: Nights of Cabiria (1957)
Rating: 100%
--------------------------
Title: Pather Panchali (1955)
Rating: 98%
--------------------------
Title: Meet Me in St. Louis (1944)
Rating: 100%
--------------------------
Title: Witness for the Prosecution (1957)
Rating: 100%
--------------------------
Title: Eighth Grade (2018)
Rating: 99%
--------------------------
Title: Rebecca (1940)
Rating: 98%
--------------------------
Title: Stalker (1979)
Rating: 100%
--------------------------
Title: The Terminator (1984)
Rating: 100%
--------------------------
Title: Memento (2000)
Rating: 94%
--------------------------
Title: The Social Network (2010)
Rating: 96%
--------------------------
```

```
Title: The Hurt Locker (2008)
Rating: 96%
-------------------------
Title: 12 Years a Slave (2013)
Rating: 95%
-------------------------
Title: Catch Me if You Can (2002)
Rating: 96%
-------------------------
Title: Jaws (1975)
Rating: 97%
-------------------------
Title: A Man Escaped (1956)
Rating: 100%
-------------------------
Title: Pan's Labyrinth (2006)
Rating: 95%
-------------------------
Title: The Red Shoes (1948)
Rating: 99%
-------------------------
Title: Anatomy of a Murder (1959)
Rating: 100%
-------------------------
Title: Das Boot (1981)
Rating: 98%
-------------------------
Title: Ikiru (1952)
Rating: 98%
-------------------------
Title: Open City (1945)
Rating: 100%
-------------------------
Title: Lady Bird (2017)
Rating: 99%
-------------------------
Title: Hunt for the Wilderpeople (2016)
Rating: 97%
-------------------------
Title: Hell or High Water (2016)
Rating: 97%
-------------------------
Title: Army of Shadows (1969)
Rating: 97%
-------------------------
Title: 007: Goldfinger (1964)
Rating: 99%
-------------------------
Title: The Lady Eve (1941)
Rating: 99%
-------------------------
Title: Saving Private Ryan (1998)
Rating: 94%
-------------------------
Title: Ratatouille (2007)
Rating: 96%
-------------------------
Title: Star Trek (2009)
Rating: 94%
-------------------------
```

```
Title: The Iron Giant (1999)
Rating: 96%
-------------------------
Title: Monty Python and the Holy Grail (1975)
Rating: 96%
-------------------------
Title: Star Wars: Episode IV - A New Hope (1977)
Rating: 93%
-------------------------
Title: Shadow of a Doubt (1943)
Rating: 100%
-------------------------
Title: WALL-E (2008)
Rating: 95%
-------------------------
Title: Brooklyn (2015)
Rating: 97%
-------------------------
Title: Mr. Smith Goes to Washington (1939)
Rating: 97%
-------------------------
Title: Spider-Man: No Way Home (2021)
Rating: 93%
-------------------------
Title: The Best Years of Our Lives (1946)
Rating: 97%
-------------------------
Title: The Bridge on the River Kwai (1957)
Rating: 96%
-------------------------
Title: Ali: Fear Eats the Soul (1974)
Rating: 100%
-------------------------
Title: Whiplash (2014)
Rating: 94%
-------------------------
Title: The Farewell (2019)
Rating: 97%
-------------------------
Title: Unforgiven (1992)
Rating: 96%
-------------------------
Title: The Adventures of Robin Hood (1938)
Rating: 100%
-------------------------
Title: Pulp Fiction (1994)
Rating: 92%
-------------------------
Title: The King's Speech (2010)
Rating: 94%
-------------------------
Title: Leave No Trace (2018)
Rating: 100%
-------------------------
Title: Star Wars: Episode V - The Empire Strikes Back (1980)
Rating: 95%
-------------------------
Title: The Passion of Joan of Arc (1928)
Rating: 98%
-------------------------
```

```
Title: Quiz Show (1994)
Rating: 97%
-------------------------
Title: Avengers: Endgame (2019)
Rating: 94%
-------------------------
Title: Ran (1985)
Rating: 96%
-------------------------
Title: Safety Last (1923)
Rating: 97%
-------------------------
Title: Iron Man (2008)
Rating: 94%
-------------------------
Title: Moana (2016)
Rating: 95%
-------------------------
Title: Little Women (2019)
Rating: 95%
-------------------------
Title: Puss in Boots: The Last Wish (2022)
Rating: 95%
-------------------------
Title: Casino Royale (2006)
Rating: 94%
-------------------------
Title: The Handmaiden (2016)
Rating: 96%
-------------------------
Title: La Haine (1995)
Rating: 96%
-------------------------
Title: La Strada (1954)
Rating: 97%
-------------------------
Title: Rashomon (1950)
Rating: 98%
-------------------------
Title: Top Hat (1935)
Rating: 100%
-------------------------
Title: The Artist (2011)
Rating: 95%
-------------------------
Title: The Conformist (1970)
Rating: 98%
-------------------------
Title: One Flew Over the Cuckoo's Nest (1975)
Rating: 93%
-------------------------
Title: In the Heat of the Night (1967)
Rating: 96%
-------------------------
Title: Raiders of the Lost Ark (1981)
Rating: 93%
-------------------------
Title: The Peanut Butter Falcon (2019)
Rating: 95%
-------------------------
```

```
Title: Paths of Glory (1957)
Rating: 96%
-------------------------
Title: King Kong (1933)
Rating: 97%
-------------------------
Title: Children of Paradise (1945)
Rating: 98%
-------------------------
Title: A Beautiful Day in the Neighborhood (2019)
Rating: 95%
-------------------------
Title: The LEGO Movie (2014)
Rating: 96%
-------------------------
Title: Before Sunset (2004)
Rating: 94%
-------------------------
Title: Soul (2020)
Rating: 95%
-------------------------
Title: Creed (2015)
Rating: 95%
-------------------------
Title: John Wick: Chapter 4 (2023)
Rating: 94%
-------------------------
Title: The Princess Bride (1987)
Rating: 96%
-------------------------
Title: Sunrise (1927)
Rating: 98%
-------------------------
Title: Before Midnight (2013)
Rating: 98%
-------------------------
Title: Lawrence of Arabia (1962)
Rating: 93%
-------------------------
Title: Strangers on a Train (1951)
Rating: 98%
-------------------------
Title: Sling Blade (1996)
Rating: 97%
-------------------------
Title: Kubo and the Two Strings (2016)
Rating: 97%
-------------------------
Title: Sweet Smell of Success (1957)
Rating: 98%
-------------------------
Title: The Thin Man (1934)
Rating: 98%
-------------------------
Title: Once Upon a Time in the West (1968)
Rating: 96%
-------------------------
Title: Eternal Sunshine of the Spotless Mind (2004)
Rating: 92%
-------------------------
```

```
Title: Sense and Sensibility (1995)
Rating: 97%
-------------------------
Title: BlacKkKlansman (2018)
Rating: 96%
-------------------------
Title: Lost in Translation (2003)
Rating: 95%
-------------------------
Title: Au Hasard Balthazar (1966)
Rating: 100%
-------------------------
Title: Boyhood (2014)
Rating: 97%
-------------------------
Title: The Grapes of Wrath (1940)
Rating: 100%
-------------------------
Title: Sing Street (2016)
Rating: 95%
-------------------------
Title: A Fistful of Dollars (1964)
Rating: 98%
-------------------------
Title: The Truman Show (1998)
Rating: 94%
-------------------------
Title: Life of Brian (1979)
Rating: 96%
-------------------------
Title: 8 1/2 (1963)
Rating: 97%
-------------------------
Title: Marriage Story (2019)
Rating: 95%
-------------------------
Title: Searching for Bobby Fischer (1993)
Rating: 100%
-------------------------
Title: Battleship Potemkin (1925)
Rating: 100%
-------------------------
Title: Sullivan's Travels (1941)
Rating: 100%
-------------------------
Title: The Red Circle (1970)
Rating: 96%
-------------------------
Title: The Lost Weekend (1945)
Rating: 97%
-------------------------
Title: The Nightmare Before Christmas (1993)
Rating: 95%
-------------------------
Title: Oppenheimer (2023)
Rating: 93%
-------------------------
Title: The Discreet Charm of the Bourgeoisie (1972)
Rating: 98%
-------------------------
```

```
Title: The Lord of the Rings: The Fellowship of the Ring (2001)
Rating: 91%
--------------------------
Title: Ford v Ferrari (2019)
Rating: 92%
--------------------------
Title: My Left Foot (1989)
Rating: 98%
--------------------------
Title: Room (2015)
Rating: 93%
--------------------------
Title: The Lord of the Rings: The Return of the King (2003)
Rating: 94%
--------------------------
Title: A Night at the Opera (1935)
Rating: 97%
--------------------------
Title: Halloween (1978)
Rating: 96%
--------------------------
Title: Air (2023)
Rating: 93%
--------------------------
Title: The Sweet Hereafter (1997)
Rating: 98%
--------------------------
Title: Playtime (1967)
Rating: 98%
--------------------------
Title: True Grit (2010)
Rating: 95%
--------------------------
Title: A Quiet Place (2018)
Rating: 96%
--------------------------
Title: Mudbound (2017)
Rating: 97%
--------------------------
Title: Boyz N the Hood (1991)
Rating: 96%
--------------------------
Title: Brazil (1985)
Rating: 98%
--------------------------
Title: Hidden Figures (2016)
Rating: 93%
--------------------------
Title: Grand Illusion (1937)
Rating: 97%
--------------------------
Title: The Conversation (1974)
Rating: 97%
--------------------------
Title: Fargo (1996)
Rating: 94%
--------------------------
Title: Diabolique (1955)
Rating: 95%
--------------------------
```

```
Title: The Apartment (1960)
Rating: 94%
-------------------------
Title: Apollo 13 (1995)
Rating: 96%
-------------------------
Title: Princess Mononoke (1997)
Rating: 93%
-------------------------
Title: Umberto D (1952)
Rating: 98%
-------------------------
Title: Black Panther (2018)
Rating: 96%
-------------------------
Title: Bringing Up Baby (1938)
Rating: 97%
-------------------------
Title: The Sting (1973)
Rating: 93%
-------------------------
Title: Logan (2017)
Rating: 93%
-------------------------
Title: Nightcrawler (2014)
Rating: 95%
-------------------------
Title: The Departed (2006)
Rating: 91%
-------------------------
Title: Juno (2007)
Rating: 94%
-------------------------
Title: Hero (2002)
Rating: 94%
-------------------------
Title: Shaun of the Dead (2004)
Rating: 92%
-------------------------
Title: Stagecoach (1939)
Rating: 100%
-------------------------
Title: Back to the Future (1985)
Rating: 93%
-------------------------
Title: Die Hard (1988)
Rating: 94%
-------------------------
Title: No Country for Old Men (2007)
Rating: 93%
-------------------------
Title: The Lion King (1994)
Rating: 92%
-------------------------
Title: Gravity (2013)
Rating: 96%
-------------------------
Title: The Leopard (1963)
Rating: 98%
-------------------------
```

```
Title: Day for Night (1973)
Rating: 98%
-------------------------
Title: Badlands (1973)
Rating: 97%
-------------------------
Title: Touch of Evil (1958)
Rating: 95%
-------------------------
Title: Yojimbo (1961)
Rating: 96%
-------------------------
Title: A Streetcar Named Desire (1951)
Rating: 97%
-------------------------
Title: Breathless (1959)
Rating: 96%
-------------------------
Title: The Manchurian Candidate (1962)
Rating: 97%
-------------------------
Title: The French Connection (1971)
Rating: 96%
-------------------------
Title: The Bourne Ultimatum (2007)
Rating: 92%
-------------------------
Title: My Fair Lady (1964)
Rating: 95%
-------------------------
Title: It's a Wonderful Life (1946)
Rating: 94%
-------------------------
Title: Some Like It Hot (1959)
Rating: 95%
-------------------------
Title: The Fugitive (1993)
Rating: 96%
-------------------------
Title: Guardians of the Galaxy (2014)
Rating: 92%
-------------------------
Title: Airplane! (1980)
Rating: 97%
-------------------------
Title: Groundhog Day (1993)
Rating: 94%
-------------------------
Title: This Is Spinal Tap (1984)
Rating: 96%
-------------------------
Title: Beauty and the Beast (1991)
Rating: 93%
-------------------------
Title: The Taking of Pelham One Two Three (1974)
Rating: 98%
-------------------------
Title: City Lights (1931)
Rating: 95%
-------------------------
```

```
Title: Kiki's Delivery Service (1989)
Rating: 98%
--------------------------
Title: City of God (2002)
Rating: 91%
--------------------------
Title: Rosemary's Baby (1968)
Rating: 96%
--------------------------
Title: Call Me by Your Name (2017)
Rating: 94%
--------------------------
Title: Aladdin (1992)
Rating: 95%
--------------------------
Title: The Man With a Movie Camera (1929)
Rating: 98%
--------------------------
Title: The Lady Vanishes (1938)
Rating: 98%
--------------------------
Title: The Umbrellas of Cherbourg (1964)
Rating: 97%
--------------------------
Title: Mission: Impossible Rogue Nation (2015)
Rating: 94%
--------------------------
Title: Three Colors: Blue (1993)
Rating: 96%
--------------------------
Title: Milk (2008)
Rating: 93%
--------------------------
Title: Traffic (2000)
Rating: 93%
--------------------------
Title: Invasion of the Body Snatchers (1956)
Rating: 98%
--------------------------
Title: Thor: Ragnarok (2017)
Rating: 93%
--------------------------
Title: The Odd Couple (1968)
Rating: 98%
--------------------------
Title: Bride of Frankenstein (1935)
Rating: 98%
--------------------------
Title: What's Love Got to Do With It (1993)
Rating: 97%
--------------------------
Title: Star Wars: The Force Awakens (2015)
Rating: 93%
--------------------------
Title: Roman Holiday (1953)
Rating: 96%
--------------------------
Title: Amélie (2001)
Rating: 90%
--------------------------
```

```
Title: To Be or Not to Be (1942)
Rating: 96%
-------------------------
Title: All the President's Men (1976)
Rating: 94%
-------------------------
Title: Throne of Blood (1957)
Rating: 96%
-------------------------
Title: Taxi Driver (1976)
Rating: 89%
-------------------------
Title: The Big Sleep (1946)
Rating: 96%
-------------------------
Title: Marvel's the Avengers (2012)
Rating: 91%
-------------------------
Title: Secrets & Lies (1996)
Rating: 96%
-------------------------
Title: Dog Day Afternoon (1975)
Rating: 96%
-------------------------
Title: Being There (1979)
Rating: 95%
-------------------------
Title: Aguirre: The Wrath of God (1972)
Rating: 96%
-------------------------
Title: Arrival (2016)
Rating: 94%
-------------------------
Title: Wings of Desire (1987)
Rating: 95%
-------------------------
Title: Raging Bull (1980)
Rating: 92%
-------------------------
Title: Fruitvale Station (2013)
Rating: 94%
-------------------------
Title: La Dolce Vita (1960)
Rating: 95%
-------------------------
Title: Beauty and the Beast (1946)
Rating: 96%
-------------------------
Title: The Killing (1956)
Rating: 96%
-------------------------
Title: The Rules of the Game (1939)
Rating: 97%
-------------------------
Title: Eyes Without a Face (1960)
Rating: 97%
-------------------------
Title: The Cabinet of Dr. Caligari (1919)
Rating: 96%
-------------------------
```

—

6. Shopping Site Web Scraping Code

Pointers: The Below code give an output and make a file in Google Colab, I have downloaded the file and print it as an output to make you understand the output of the code.

In [21]:
```
# Data Scraping from Amazon

from bs4 import BeautifulSoup
import requests
import pandas as pd
import numpy as np
# Function to extract Product Title
def get_title(soup):

    try:
        # Outer Tag Object
        title = soup.find("span", attrs={"id":'productTitle'})

        # Inner NavigatableString Object
        title_value = title.text

        # Title as a string value
        title_string = title_value.strip()

    except AttributeError:
        title_string = ""

    return title_string

# Function to extract Product Price
def get_price(soup):

    try:
        price = soup.find("span", attrs={'id':'priceblock_ourprice'}).string.strip()

    except AttributeError:

        try:
            # If there is some deal price
            price = soup.find("span", attrs={'id':'priceblock_dealprice'}).string.stri

        except:
            price = ""

    return price

# Function to extract Product Rating
def get_rating(soup):

    try:
        rating = soup.find("i", attrs={'class':'a-icon a-icon-star a-star-4-5'}).strir

    except AttributeError:
        try:
            rating = soup.find("span", attrs={'class':'a-icon-alt'}).string.strip()
        except:
```

```
            rating = ""

    return rating

# Function to extract Number of User Reviews
def get_review_count(soup):
    try:
        review_count = soup.find("span", attrs={'id':'acrCustomerReviewText'}).string.

    except AttributeError:
        review_count = ""

    return review_count

# Function to extract Availability Status
def get_availability(soup):
    try:
        available = soup.find("div", attrs={'id':'availability'})
        available = available.find("span").string.strip()

    except AttributeError:
        available = "Not Available"

    return available
if __name__ == '__main__':

    # add your user agent
    HEADERS = ({'User-Agent':'', 'Accept-Language': 'en-US, en;q=0.5'})

    # The webpage URL
    URL = "https://www.amazon.com/s?k=playstation+4&ref=nb_sb_noss_2"

    # HTTP Request
    webpage = requests.get(URL, headers=HEADERS)

    # Soup Object containing all data
    soup = BeautifulSoup(webpage.content, "html.parser")

    # Fetch links as List of Tag Objects
    links = soup.find_all("a", attrs={'class':'a-link-normal s-no-outline'})

    # Store the links
    links_list = []

    # Loop for extracting links from Tag Objects
    for link in links:
            links_list.append(link.get('href'))

    d = {"Title":[], "Price":[], "Rating":[], "Reviews":[],"Availability":[]}

    # Loop for extracting product details from each link
    for link in links_list:
        new_webpage = requests.get("https://www.amazon.com" + link, headers=HEADERS)

        new_soup = BeautifulSoup(new_webpage.content, "html.parser")

        # Function calls to display all necessary product information
        d['Title'].append(get_title(new_soup))
        d['Price'].append(get_price(new_soup))
        d['Rating'].append(get_rating(new_soup))
```

```
        d['Reviews'].append(get_review_count(new_soup))
        d['Availability'].append(get_availability(new_soup))


    amazon_df = pd.DataFrame.from_dict(d)
    amazon_df['Title'].replace('', np.nan, inplace=True)
    amazon_df = amazon_df.dropna(subset=['Title'])
    amazon_df.to_csv("amazon_data.csv", header=True, index=False)
Amazon_data = pd.read_csv(r"C:\Users\Sneha Bhattcharjee\Downloads\amazon_data.csv")
Amazon_data
```

Out[21]:

	Title	Price	Rating	Reviews	Availability
0	PlayStation 4 Slim 1TB Console	NaN	4.5 out of 5 stars	15,704 ratings	In Stock
1	Newest Sony Playstation 4 Slim 1TB SSD Console...	NaN	4.5 out of 5 stars	333 ratings	Only 1 left in stock - order soon.
2	PlayStation®5 Digital Edition (slim)	NaN	4.7 out of 5 stars	5,480 ratings	Not Available
3	Sony PlayStation 4 500GB Premium Bundle (Renewed)	NaN	4.5 out of 5 stars	3 ratings	Only 6 left in stock - order soon.
4	PlayStation 4 Slim 500GB Console [Discontinued...	NaN	4.5 out of 5 stars	1,397 ratings	In Stock
5	PlayStation 4 500GB Console [Old Model] [Discon...	NaN	4.5 out of 5 stars	13,589 ratings	Only 1 left in stock - order soon.

—

Web Scraping: Adapting to Different Websites

If you replace the URL in a web scraping script with another website, it's important to adjust the class selectors and potentially other HTML element selectors accordingly. Each website has its own unique structure and HTML tags, so selectors that work for one site may not work for another. Here's a detailed explanation of why and how to do this:

Importance of Adjusting Class Selectors

1. **Unique Structure:** Every website has a different layout and structure. The HTML tags and classes used to define elements like headings, paragraphs, links, and tables can vary significantly.

2. **Dynamic Content:** Websites often use JavaScript to load content dynamically. The way content is loaded and displayed can affect which elements you need to target in your scraping script.

3. **Consistency:** To ensure you are scraping the correct data, you must identify the specific tags and classes used by the target website for the elements you want to extract.

Steps to Adapt Class Selectors

1. **Inspect the Web Page:** Use browser developer tools (right-click on the web page and select "Inspect" or press `F12`) to examine the HTML structure of the target website.
2. **Identify Relevant Elements:** Look for the HTML tags and classes that contain the data you want to scrape. Pay attention to elements like `<div>`, `<span>`, `<p>`, `<h1>`, etc.
3. **Update Selectors in Your Script:** Modify your web scraping code to use the correct tags and class names identified from the target website.
4. **Test and Iterate:** Run your script and check the output. If the data is not as expected, re-examine the HTML structure and adjust your selectors.

Summary

- **Inspect the Web Page:** Use developer tools to examine the HTML structure.
- **Identify Relevant Elements:** Find the specific tags and classes that contain the data you need.
- **Update Selectors:** Modify your script to use the correct tags and class names.
- **Test and Iterate:** Ensure your script is extracting the correct data.

By following these steps, you can adapt your web scraping scripts to work with different websites, ensuring accurate and efficient data extraction.

While web scraping, it's essential to handle different HTTP response status codes to understand the result of your request. Here is a list of common HTTP response status codes you might encounter, along with a brief explanation of each:

Informational Responses (100–199)

1. **100 Continue**: The server has received the request headers and the client should proceed to send the request body.
2. **101 Switching Protocols**: The requester has asked the server to switch protocols and the server has agreed to do so.

Successful Responses (200–299)

1. **200 OK**: The request was successful, and the server has returned the requested resource.
2. **201 Created**: The request has been fulfilled and resulted in a new resource being created.
3. **204 No Content**: The server successfully processed the request, but there is no content to send in the response.

Redirection Messages (300–399)

1. **301 Moved Permanently**: The requested resource has been assigned a new permanent URI.
2. **302 Found**: The requested resource resides temporarily under a different URI.
3. **304 Not Modified**: The resource has not been modified since the version specified by the request headers.

Client Error Responses (400–499)

1. **400 Bad Request**: The server cannot or will not process the request due to a client error.
2. **401 Unauthorized**: Authentication is required, and the client has not provided valid authentication credentials.
3. **403 Forbidden**: The client does not have access rights to the content.
4. **404 Not Found**: The server cannot find the requested resource.
5. **408 Request Timeout**: The server timed out waiting for the request.
6. **429 Too Many Requests**: The client has sent too many requests in a given amount of time.

Server Error Responses (500–599)

1. **500 Internal Server Error**: An error has occurred on the server side.
2. **501 Not Implemented**: The server does not recognize the request method or lacks the ability to fulfill it.
3. **502 Bad Gateway**: The server, while acting as a gateway or proxy, received an invalid response from the upstream server.
4. **503 Service Unavailable**: The server is currently unavailable (overloaded or down).
5. **504 Gateway Timeout**: The server, while acting as a gateway or proxy, did not receive a timely response from the upstream server.
6. **505 HTTP Version Not Supported**: The server does not support the HTTP protocol version used in the request.

#

—

Conclusion

Through this project, I aim to highlight the importance of data acquisition and its impact on research, business intelligence, and personal decision-making. The project not only underscores the technical aspects of web scraping but also emphasizes ethical considerations and best practices to ensure responsible and compliant data extraction. Ultimately, this project serves as a testament to the power of Python and web scraping in unlocking the full potential of web data.

The above project is done by Sneha Bhattacharjee on 1st June, 2024. Feel free to review and comment.

—