

amazon listings web scraper

August 10, 2024

```
[6]: from bs4 import BeautifulSoup
import requests
import pandas as pd
import numpy as np
```

```
[4]: # 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.strip()
```

```

        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'}).string.strip()

    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.strip()

    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

```

```

[7]: if __name__ == '__main__':

    page= range(1, 11)

```

```

# 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&page={page}"

# 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)

```

```
[9]: amazon_df
```

[9]:

	title	price \
8	OUBANG 2 Pack Controllers Work with PS4 Contro...	\$37.99
9	PlayStation 4 Slim 1TB Console	
10	PlayStation 4 Slim 1TB Console - Only On PlayS...	
11	Hunting Simulator - PlayStation 4	\$16.61
12	BlueFire Professional 3.5mm PS4 Gaming Headset...	
13	Horizon Zero Dawn Complete Edition Hits - Play...	\$22.45
14	Beastron PS4 Controller Charging Station, PS4 ...	
15	PlayStation 4 Slim 1TB Limited Edition Console...	
16	PlayStation 4 500GB Console [Old Model][Discon...	
17	Sony Playstation PS4 1TB Black Console	\$619.00
18	PlayStation 4 Slim 1TB Gold Console [Discontin...	
19	DualShock 4 Wireless Controller for PlayStatio...	\$58.75
20	DualShock 4 Wireless Controller for PlayStatio...	\$63.00
21	PS4 Stand Cooling Fan Station for Playstation ...	\$25.99
22	Little Witch Nobeta - PlayStation 4	\$49.99
23	PS4 Stand Cooling Fan Station for Playstation ...	\$25.99
24	DualShock 4 Wireless Controller for PlayStatio...	\$61.50
25	Sony PlayStation 4 Dualshock 4 Wireless Contro...	
26	NBA 2K23 - PlayStation 4	\$59.99
27	Upgraded YUYIU Wireless Controller for Ps4 Re...	
28	Hermitshell Hard EVA Travel Case Fits Sony Pla...	\$75.99
29	NHL 23 - PlayStation 4	\$59.99
30	HDMI Cable for Playstation 2 & Playstation 1 C...	
31	Wireless Controller, Console for Ps-4/Pro/Ps-3...	
32	Dying Light 2 Stay Human - PlayStation 4	\$29.99
33	OUBANG 2 Pack Controllers Work with PS4 Contro...	\$37.99
34	NHL 22 (PS4)	\$18.92
35	eXtremeRate Blood Purgatory Patterned Custom F...	
36	Ps4 Controller Skin Silicone Case Grip Glow in...	

	rating	reviews \
8	4.8 out of 5 stars	8 ratings
9	4.7 out of 5 stars	15,164 ratings
10	4.8 out of 5 stars	8,860 ratings
11	4.3 out of 5 stars	937 ratings
12	4.4 out of 5 stars	8,257 ratings
13	4.8 out of 5 stars	10,127 ratings
14	4.4 out of 5 stars	568 ratings
15	4.6 out of 5 stars	711 ratings
16	4.6 out of 5 stars	13,569 ratings
17	4.7 out of 5 stars	853 ratings
18	4.6 out of 5 stars	707 ratings
19	4.7 out of 5 stars	140,711 ratings
20	4.7 out of 5 stars	140,711 ratings
21	4.3 out of 5 stars	6 ratings
22	Previous page of related Sponsored Products	

23	4.3 out of 5 stars	6 ratings
24	4.7 out of 5 stars	140,711 ratings
25	4.5 out of 5 stars	1,545 ratings
26	4.5 out of 5 stars	206 ratings
27	4.4 out of 5 stars	455 ratings
28	4.2 out of 5 stars	111 ratings
29	4.3 out of 5 stars	24 ratings
30	4.5 out of 5 stars	2,302 ratings
31	5.0 out of 5 stars	5 ratings
32	4.6 out of 5 stars	944 ratings
33	4.8 out of 5 stars	8 ratings
34	4.6 out of 5 stars	282 ratings
35	4.2 out of 5 stars	4,357 ratings
36	4.2 out of 5 stars	1,415 ratings

	availability
8	Not Available
9	
10	Only 3 left in stock - order soon
11	Not Available
12	Not Available
13	Not Available
14	Currently unavailable.
15	
16	
17	Not Available
18	Only 1 left in stock - order soon
19	Not Available
20	Not Available
21	Not Available
22	This item will be released on June 30, 2023.
23	Not Available
24	Not Available
25	Not Available
26	In Stock
27	Not Available
28	Not Available
29	In Stock.
30	Not Available
31	Not Available
32	In Stock
33	Not Available
34	Not Available
35	Not Available
36	Not Available

[]: