

Import libraries

```
from bs4 import BeautifulSoup
import requests
import time
import datetime
```

Connect to Website and pull in data

```
URL = 'https://www.amazon.in/Seek-Buy-Love-Spreadsheet-Accountant/dp/B0CZJ82YT1/ref=sr_1_3?crid=1LUEFJKIBG6VP&dib=eyJ2IjoiMSJ9.Wv-NxNGBS-LNPPftS0JQ6-FYxww2sBXhrGSulbUJv_To6MZTcsWFeG_nj8wKpCk3bd8CbIRPw4LFILXFs6tBAhSPy0e5xWBSxa8K9ramDVfQJU_P1kD8YvU5-xB2kEBSnvFHY3k9_tQrRbXAaC8lz0LMcYi1miYEUyCQ0rANB3yloHWx2wD0gyKNjhq3XKyICH8A6XR3sltRCmsF91DVzSB3NbE898_1pvf6jpaZ8ahBliMSklo-7zf4bQ-0b9aw-ZkpFroNg8-WAmwpp1kDVZgqhRIxF0tfNd38M-aPugc.0prNRSNfcU_iJrkVgiD-90pZm7xgnGuPYIOvdmsi8kk&dib_tag=se&keywords=data+analyst+tshirt&qid=1729252520&sprefix=data+analyst+tshirt%2Caps%2C173&sr=8-3'

headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/128.0.0.0 Safari/537.36 OPR/114.0.0.0", "Accept-Encoding": "gzip, deflate", "Accept": "text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8", "DNT": "1", "Connection": "close", "Upgrade-Insecure-Requests": "1"}

page = requests.get(URL, headers=headers)

soup1 = BeautifulSoup(page.content, "html.parser")

soup2 = BeautifulSoup(soup1.prettify(), "html.parser")
```

Extract title, price and clean it up

```
# Extract title and clean it up
title = soup2.find(id='title_feature_div').get_text(strip=True)

# Extract price and clean it up
price = soup2.find(id='corePriceDisplay_desktop_feature_div').get_text(strip=True)

# Use regex to extract the numeric price
import re
cleaned_price = re.search(r"₹(\d+)", price).group(1)

# Print the cleaned results
print(title)
print(cleaned_price)
```

Seek Buy Love Funny Spreadsheet T-Shirt, Data Analyst Tee,
Accountant Gift, Office Humor Shirt, Casual Graphic Top, Unisex Tee
449

```
# Create a Timestamp for your output to track when data was  
collected
```

```
import datetime  
today = datetime.date.today()  
print(today)
```

2024-10-18

```
# Create CSV and write headers and data into the file
```

```
import csv  
header = ['Title', 'Price', 'Date']  
data = [title, cleaned_price, today]
```

```
with open('WebScrapersDataset.csv', 'w', newline='', encoding='UTF8')  
as f:  
    writer = csv.writer(f)  
    writer.writerow(header)  
    writer.writerow(data)
```

```
import pandas as pd  
df = pd.read_csv('WebScrapersDataset.csv')  
print(df)
```

		Title	Price	
Date				
0	Seek Buy Love Funny Spreadsheet T-Shirt, Data ...	449	2024-10-18	
1	Seek Buy Love Funny Spreadsheet T-Shirt, Data ...	449	2024-10-18	
2	Seek Buy Love Funny Spreadsheet T-Shirt, Data ...	449	2024-10-18	

```
# Now we are appending data to the csv
```

```
with open('WebScrapersDataset.csv', 'a+', newline='',  
encoding='UTF8') as f:  
    writer = csv.writer(f)  
    writer.writerow(data)
```

```
# Combine all of the above code into one function
```

```
def check_price():  
    URL = 'https://www.amazon.in/Seek-Buy-Love-Spreadsheet-  
Accountant/dp/B0CZJ82YT1/ref=sr_1_3?  
crid=1LUEFJKIBG6VP&dib=eyJ2IjoimSJ9.Wv-NxNGBS-LNPPftS0JQ6-  
FYxww2sBXhrGSulbUJv_To6MZTcsWFeG_nj8wKpCk3bd8CbIRPw4LFILXFs6tBAhSPy0  
e5xWBSxa8K9ramDVfQJU_P1kD8YvU5-  
xB2kEBSnvFHY3k9_tQrRbXAaC8lz0LMcYi1miYEUyCQ0rANB3yloHWx2wD0gyKNjhq3X  
KyiCH8A6XR3sltRCmsF91DVzSB3NbE898_1pvf6jpaZ8ahBliMsKlo-7zf4bQ-0b9aw-
```

```
ZkpFroNg8-WAmwpp1kDVZgqhRIxF0tfNd38M-aPugc.0prNRSNfcU_iJrkVgiD-90pZm7xgnGuPYIOvdmsi8kk&dib_tag=se&keywords=data+analyst+tshirt&qid=1729252520&sprefix=data+analyst+tshirt%2Caps%2C173&sr=8-3'
```

```
headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/128.0.0.0 Safari/537.36 OPR/114.0.0.0", "Accept-Encoding": "gzip, deflate", "Accept": "text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8", "DNT": "1", "Connection": "close", "Upgrade-Insecure-Requests": "1"}

page = requests.get(URL, headers=headers)

soup1 = BeautifulSoup(page.content, "html.parser")

soup2 = BeautifulSoup(soup1.prettify(), "html.parser")

# Extract title and clean it up
title = soup2.find(id='title_feature_div').get_text(strip=True)

# Extract price and clean it up
price =
soup2.find(id='corePriceDisplay_desktop_feature_div').get_text(strip=True)

# Use regex to extract the numeric price
import re
cleaned_price = re.search(r"₹(\d+)", price).group(1)

import datetime
today = datetime.date.today()

import csv
header = ['Title', 'Price', 'Date']
data = [title, cleaned_price, today]

with open('WebScraperDataset.csv', 'a+', newline='',
encoding='UTF8') as f:
    writer = csv.writer(f)
    writer.writerow(data)

# Runs check_price after a set time and inputs data into your CSV
counter = 0
while(counter < 20):
    check_price()
    # time.sleep(2)
    counter = counter + 1

import pandas as pd

df = pd.read_csv('WebScraperDataset.csv')

print(df)
```

						Title	Price	
Date								
0	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
1	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
2	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
3	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
4	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
5	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
6	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
7	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
8	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
9	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
10	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
11	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
12	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
13	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
14	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
15	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
16	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
17	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
18	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
19	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
20	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
21	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
22	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
23	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
24	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18
25	Seek	Buy	Love	Funny	Spreadsheet	T-Shirt, Data ...	449	2024-10-18

26	Seek Buy Love Funny Spreadsheet T-Shirt, Data ...	449	2024-10-18
27	Seek Buy Love Funny Spreadsheet T-Shirt, Data ...	449	2024-10-18
28	Seek Buy Love Funny Spreadsheet T-Shirt, Data ...	449	2024-10-18