```
pip install fake_useragent
```

```
    Requirement already satisfied: fake_useragent in /usr/local/lib/python3.7/dist-packages
```

◄ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮ ►

```
import requests
from bs4 import BeautifulSoup
from fake_useragent import UserAgent

ua = UserAgent()
header = {'user-agent':ua.chrome}
google_page = requests.get('https://www.amazon.in/s?k=headphones&crid=1KNIU0KIZ8UF4&sprefix=h
print(google_page.content)

soup = BeautifulSoup(google_page.content,'lxml') # html.parser

#print(soup.prettify())
```

```
[→    b'<!doctype html><html lang="en-in" class="a-no-js" data-19ax5a9jf="dingo"><!-- sp:featu
```

◄ ▮▮ ►

```
print(soup.prettify())
```

```
import pandas as pd
import numpy as np
```

```
def soup_df(soup):
  df = pd.DataFrame(np.zeros(1000))
  df['Product name'] = soup.find('a',class_="a-link-normal s-link-style a-text-normal").text
  df['mrp'] = soup.find('span',class_="a-price a-text-price").span.text
  df['rating'] = (soup.find('div',class_="a-row a-size-small").text).split()[0]
  df['reviews'] = (soup.find('div',class_="a-row a-size-small").text).split()[5]
  df['offer price'] = soup.find('span',class_="a-price-whole").text
  i = 0
  for data in soup.find_all('a',class_="a-link-normal s-link-style a-text-normal"):
    df['Product name'].iloc[i] = data.text
    i=i+1

  i = 0
  for mrp in soup.find_all('span',class_="a-price a-text-price"):
    df['mrp'].iloc[i] = float(mrp.span.text[1:].replace(',',''))
    i=i+1

  i = 0
  for rating in soup.find_all('div',class_="a-row a-size-small"):
    df['rating'].iloc[i] = float((rating.text).split()[0])
    i=i+1
```

```
      i=i+1

  i = 0
  for reviews in soup.find_all('div',class_="a-row a-size-small"):
    df['reviews'].iloc[i] = (reviews.text).split()[5].replace(',','')
    i=i+1

  i = 0
  for offer in soup.find_all('span',class_="a-price-whole"):
    df['offer price'].iloc[i] = float(offer.text.replace(',',''))
    i=i+1

  df = df.drop(0, axis = 1)
  df = df.drop_duplicates(['Product name'])
  return df


df1 = soup_df(soup)
```
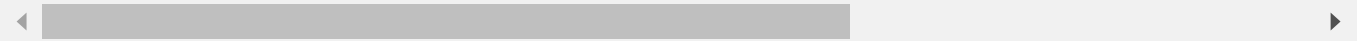
```
    /usr/local/lib/python3.7/dist-packages/pandas/core/indexing.py:670: SettingWithCopyWarni
    A value is trying to be set on a copy of a slice from a DataFrame

    See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user
      iloc._setitem_with_indexer(indexer, value)
```

◀ ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ ▶

```
df1.head()
```
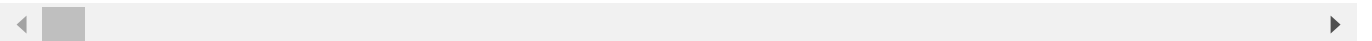
	Product name	mrp	rating	reviews	offer price
0	Hungama HiLife Buzz 101 On Ear Headphones – 9 ...	4999	5	2	2299
1	Infinity (JBL) Glide 510, 72 Hrs Playtime with...	3999	4.2	36225	1699
2	boAt Rockerz 450 Bluetooth Wireless On Ear Hea...	3990	4.1	68409	999
3	boAt Bassheads 900 Wired On Ear Headphones wit...	2490	4.2	71782	579
4	boAt Rockerz 550 Bluetooth Wireless Over Ear H...	4999	4.1	35754	1899

```
def next_page_soup(soup):
  next_page = soup.find('a', class_="s-pagination-item s-pagination-next s-pagination-button
  ua = UserAgent()
  header = {'user-agent':ua.chrome}
  google_page = requests.get(str('https://www.amazon.in'+next_page),headers=header)
  print(google_page.content)

  soup1 = BeautifulSoup(google_page.content,'lxml') # html.parser
  return soup1


soup1 = next_page_soup(soup)
```

```
b'<!doctype html><html lang="en-in" class="a-no-js" data-19ax5a9jf="dingo"><!-- sp:featu
```

```
soup2 = next_page_soup(soup1)
```

```
b'<!doctype html><html lang="en-in" class="a-no-js" data-19ax5a9jf="dingo"><!-- sp:featu
```

```
soup3 = next_page_soup(soup2)
```
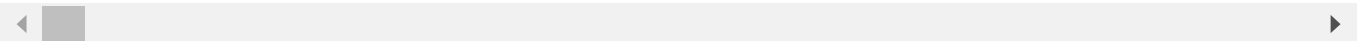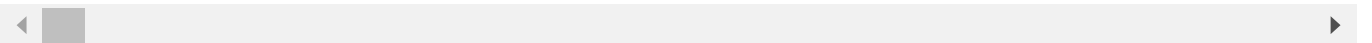
```
b'<!doctype html><html lang="en-in" class="a-no-js" data-19ax5a9jf="dingo"><!-- sp:featu
```

```
soup4 = next_page_soup(soup3)
```

```
b'<!doctype html><html lang="en-in" class="a-no-js" data-19ax5a9jf="dingo"><!-- sp:featu
```

```
soup5 = next_page_soup(soup4)
```

```
b'<!doctype html><html lang="en-in" class="a-no-js" data-19ax5a9jf="dingo"><!-- sp:featu
```

```
df1 = soup_df(soup1)
df2 = soup_df(soup2)
df3 = soup_df(soup3)
df5 = soup_df(soup4)
df6 = soup_df(soup5)
```

```
/usr/local/lib/python3.7/dist-packages/pandas/core/indexing.py:670: SettingWithCopyWarni
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_
  iloc._setitem_with_indexer(indexer, value)
```

```
df4 =pd.concat([df,df1,df2,df3,df5,df6], axis = 0)
df4 = df4.drop_duplicates(['Product name'])
df4
```

index	Product name	mrp	rating	reviews	offer price
0	Noise Combat with Ultra Low Latency, ENC with Dual mic, Super-Fast Charging and Breathing LED Lights; Bluetooth Headset (Thunderblack)	3999.0	3.3	15	1799.0
1	Boult Audio Bass Buds Q2 Over-Ear Wired Lightweight Stereo Headphones, Deep Bass & in-Built Mic, Headset with Comfortable Ear Cushions, Long Cord (Black)	2490.0	3.9	9822	599.0
2	boAt Rockerz 450 Bluetooth Wireless On Ear Headphone with Mic (Luscious Black)	3990.0	4.1	68408	999.0
3	boAt Bassheads 900 Wired On Ear Headphones with Mic (Carbon Black)	2490.0	4.2	71782	579.0
4	boAt Rockerz 550 Bluetooth Wireless Over Ear Headphone with Mic (Black)	4999.0	4.1	35751	1899.0
5	boAt Rockerz 400 Wireless Bluetooth On Ear Headphones with Mic (Carbon Black)	2990.0	4.1	87781	1399.0
6	JBL Tune 500BT by Harman Powerful Bass Wireless On-Ear Headphones with Mic, 16 Hours Playtime & Multi Connect Connectivity (Black)	3799.0	4.1	40082	3249.0
7	Skullcandy Crusher Wireless Bluetooth Over The Ear Headphone with Mic (Black)	12999.0	4.1	15181	8095.0
8	Sennheiser PXC 550-II Wireless Over The Ear Headphone with Mic (Black)	29990.0	4.1	1188	17990.0
9	Jabra Elite 45h Wireless Bluetooth On Ear Headphones with Mic (Titanium Black)	9999.0	4.2	466	5999.0
10	boAt Rockerz 370 Bluetooth Wireless On Ear Headphone with Mic (Buoyant Black)	2499.0	4.1	18091	899.0
11	boAt Rockerz 510 Wireless Bluetooth On Ear Headphones	3990.0	4.2	58406	1299.0

```
df4['reviews'].unique()
```

```
array(['15', '9822', '68408', '71782', '35751', '87781', '40082', '15181',
       '1188', '466', '18091', '58406', '42960', '79855', '23337',
       '98795', '12892', '7845', '337', '66467', '14442', '4246', '9555',
       '36225', '35754', '14444', '13426', '68409', '44774', '497',
       '4318', '38461', '28337', '53486', '291386', '36510', '6', '5496',
       '3351', '45182', '17', '5875', '1', '10', '4136', '5', '42962',
       '112253', '45,182', '23040', '163673', '12866', '47860', '4',
       '103', '81677', '56599', '10835', '4349', '10246', '12', '36,225',
       '1791', '73466', '709', '449', '7', '2655', '581'], dtype=object)
```

```
df4['reviews'] = df4['reviews'].astype(str)
df4 = df4.reset_index(drop=True)
```

```
i = 0
```

```
for i in range(len(df4['reviews'])):
  df4['reviews'][i] = df4['reviews'][i].replace(',','')
```

```
df4['reviews'] = df4['reviews'].astype(float)
```

```
df4 = df4[df4['reviews']>100]
```

```
df4.shape
```

```
(85, 5)
```

```
df4['discount'] = (1-(df4['offer price']/df4['mrp']))*100
df4.head()
```
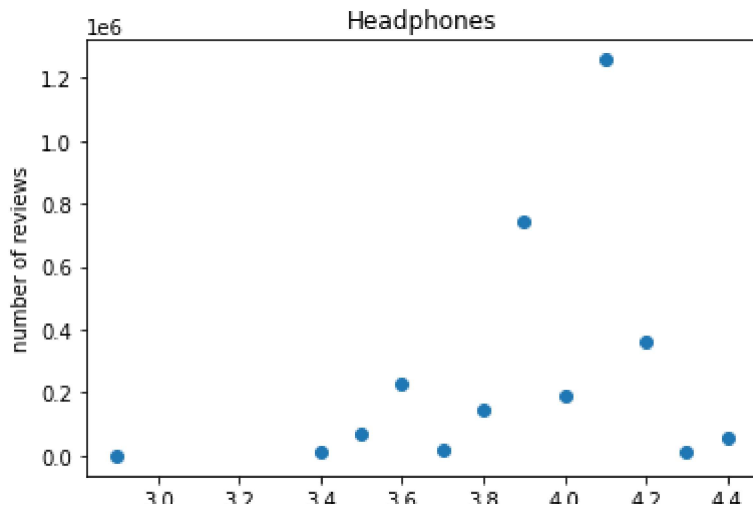
	Product name	mrp	rating	reviews	offer price	discount
1	Boult Audio Bass Buds Q2 Over-Ear Wired Lightw...	2490	3.9	9822.0	599	75.9438
2	boAt Rockerz 450 Bluetooth Wireless On Ear Hea...	3990	4.1	68408.0	999	74.9624
3	boAt Bassheads 900 Wired On Ear Headphones wit...	2490	4.2	71782.0	579	76.747

```
new_df = df4.groupby('rating').sum()
```

```
new_df['reviews']
```

```
rating
2.9        206.0
3.4      12626.0
3.5      66467.0
3.6     226046.0
3.7      20838.0
3.8     144876.0
3.9     746286.0
4.0     190076.0
4.1    1256957.0
4.2     360860.0
4.3      10992.0
4.4      57096.0
Name: reviews, dtype: float64
```

```
import matplotlib.pyplot as plt
plt.scatter(new_df.index,new_df['reviews'])
plt.xlabel('rating')
plt.ylabel('number of reviews')
plt.title('Headphones')
plt.show()
```

Headphones

The below code will recomend the best headphone considering maximum reviews,rating and discount - max(reviews x rating x discount)

```
df4['rating'] = df4['rating'].astype(float)
df4['col'] = df4['reviews']*df4['rating']*df4['discount']

df4[df4['col'] == (df4['reviews']*df4['rating']*df4['discount']).max()]
```

1 entry | Filter

index	Product name	mrp	rating	reviews	offer price	discount	col
38	boAt Bassheads 100 in Ear Wired Earphones with Mic(Black)	999.0	4.1	291386.0	289.0	71.07107107107107	84907371.97197196

It turns out that the same earphone is on amazon's choice


```
df4.drop(['col'], axis = 1, inplace = True)

df4.to csv('Headphone prices')
```

```
df4.shape
df4.head()
```

	Product name	mrp	rating	reviews	offer price	discount
1	Boult Audio Bass Buds Q2 Over-Ear Wired Lightw...	2490	3.9	9822.0	599	75.9438
2	boAt Rockerz 450 Bluetooth Wireless On Ear Hea...	3990	4.1	68408.0	999	74.9624
3	boAt Bassheads 900 Wired On Ear Headphones wit...	2490	4.2	71782.0	579	76.747

✓ 0s   completed at 2:23 PM   ● ✕