

Assignment 9

Author Details

1. Author : Varad Girish Mashalkar
2. Roll Number : 33337
3. Batch : M11
4. Date : 11/04/2022
5. Course : DSBDAL

Problem Statement

Create a review scrapper for any ecommerce website to fetch real time comments, reviews, ratings, comment tags, customer name using Python.

Notebook details

1. Number of code cells : 21
2. Python version : 3.7.4
3. Imports
 - A. BeautifulSoup (bs4)
 - B. pandas
 - C. requests

Web page details

1. The web page used for scraping is a FlipKart product page describing headphones from Boat company.
2. The reviews, ratings, customer names, titles have been considered for scraping for the identified problem statement.

Importing required libraries

In [1]:

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

Setting link for scraping

In [2]:

```
page_link = "https://www.flipkart.com/boat-rockerz-255-pro-asap-charge-upto-40-hour
```

Getting source code of target page

In [3]:

```
page = requests.get(page_link)
print("Status code : ", page.status_code)
```

Status code : 200

Note :

-> If the status code is not 200, then the page has not been recieved from the made request properly and further troubleshooting is required.

Initializing BeautifulSoup parser for target web page

In [4]:

```
soup = bs(page.content, 'html.parser')
```

Scraping the web page for different fields

a) Names of customers

The names of the customers have been found to be defined inside a paragraph tag with the class named "_2sc7ZR_2V5EHH"

In [5]:

```
# Finding all paragraph tags with class _2sc7ZR_2V5EHH
customer_names = soup.find_all('p', class_=' _2sc7ZR_2V5EHH')
```

In [6]:

```
customer_names
```

Out[6]:

```
[<p class="_2sc7ZR_2V5EHH">Manash Pratim Saikai</p>,
<p class="_2sc7ZR_2V5EHH">Himanshu Ghotkar</p>,
<p class="_2sc7ZR_2V5EHH">Asad Khan</p>,
<p class="_2sc7ZR_2V5EHH">sujit das</p>,
<p class="_2sc7ZR_2V5EHH">Pankaj Kumar</p>,
<p class="_2sc7ZR_2V5EHH">Bhargav Rajpura</p>,
<p class="_2sc7ZR_2V5EHH">Sagar Mallik</p>,
<p class="_2sc7ZR_2V5EHH">Aftab Alam</p>,
<p class="_2sc7ZR_2V5EHH">Prantik Sutradhar</p>,
<p class="_2sc7ZR_2V5EHH">su hail k</p>]
```

In [7]:

```
# Extracting the customer names from the above result
customer_names_extracted = []
for i in range(0, len(customer_names)):
    customer_names_extracted.append(customer_names[i].get_text())

# printing customer names
customer_names_extracted
```

Out[7]:

```
['Manash Pratim Saikai',
 'Himanshu Ghotkar',
 'Asad Khan',
 'sujit das',
 'Pankaj Kumar',
 'Bhargav Rajpura',
 'Sagar Mallik',
 'Aftab Alam',
 'Prantik Sutradhar',
 'suhail k']
```

b) Ratings provided for the product

The ratings of the product have been found to be defined inside a division tag with the class named "_3LWZIK_1BLPMq"

In [8]:

```
product_ratings = soup.find_all('div', class_="_3LWZlK _1BLPMq")
product_ratings
```

Out[8]:

```
[<div class="_3LWZlK _1BLPMq">5</div>v>,
  <div class="_3LWZlK _1BLPMq">5</div>v>,
  <div class="_3LWZlK _1BLPMq">5</div>v>,
  <div class="_3LWZlK _1BLPMq">5</div>v>,
  <div class="_3LWZlK _1BLPMq">4</div>v>,
  <div class="_3LWZlK _1BLPMq">4</div>v>,
  <div class="_3LWZlK _1BLPMq">5</div>v>,
  <div class="_3LWZlK _1BLPMq">5</div>v>,
  <div class="_3LWZlK _1BLPMq">4</div>v>]
```

```
<div class="_3LWZlK _1BLPMq">4</div>]
```

In [9]:

```
# Extracting the product ratings from the above result
product_ratings_extracted = []
for product_rating in product_ratings:
    product_ratings_extracted.append(product_rating.get_text())

# printing product ratings
product_ratings_extracted
```

Out[9]:

```
['5', '5', '5', '5', '4', '4', '5', '5', '4', '4']
```

c) Product reviews

The reviews of the product have been found to be defined inside a division tag with the class named "t-ZTKy"

In [10]:

```
product_reviews = soup.find_all('div', class_="t-ZTKy")
product_reviews
# class="t-ZTKy"
```

Out[10]:

```
[<div class="t-ZTKy"><div><div class="">1. Material Quality - Not Ba
d...<br/>2. Sound Quality - Awesome...<br/>3. Bass - Decent Quality...
<br/>4. Battery Backup - Very Good...<br/>5. Design & Bulid Qualit
y - Very nice...<br/>6. Fitting - Oky Type...<br/>Overall very nice an
d Velue for money product...<br/>Really i love the product...😄</div>
<span class="_1H-bmy"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">Writing this review after usin
g more than 15 days. 255F pro+ has super sound quality. Bass is decent
and fair enough. Most important highlight of this headset is battery l
ife,it's amazing. Runtime depends on usage but definitely it will give
you more than any other headset.<br/>Should go for it.</div><span clas
s="_1H-bmy"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">One of the best Bluetooth Boa
t. <br/>1- Battery Backup Amazing <br/>2- light wait <br/>3-Awesome So
und quality<br/>4-Bass good<br/>5- Design normal but good not best</di
v><span class="_1H-bmy"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">Nice product and fast delivery
. I am happy nice job Flipkart thanks</div><span class="_1H-bmy"><spa
n>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">Delivered in 20 February.. tod
ay is 28 February and i didn't charge it till now still battery percen
tage is 70%.. i use it daily for online classes, 1 movie and 3 hours F
or music approx 9-10 hours daily so Battery is very Powerfull.it's hou
sing is made by plastic and neck wire of somewhat soft silicone materi
al cable is flat long and strong enough..Bass is above the best if you
use right size of ear buds..even in 90% volume vocals are crystal clea
r treble is well balanced..Call quality is far...</div><span class="_1
BWGvX"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">This product is acctullly bes
t. best coloty and design.<br/>But sound is little low not bed its a p
erfect but little low. bass is very powerful and clear. battery backup
is very best in this price range. definately you purchase this produc
t.</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>
>,
  <div class="t-ZTKy"><div><div class="">Great product., battery backup
up to 14 days use , sound quality is great , <br/>I am happy with the
product 😄😄😄😄</div><span class="_1H-bmy"><span>READ MORE</span></s
pan></div></div>,
  <div class="t-ZTKy"><div><div class="">Battery backup is best and ful
l charge very quickly. Sound quality is best .I love it. You can buy i
t without any hesitation.</div><span class="_1H-bmy"><span>READ MORE</
span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">Just got my hands on this prod
uct. Let me be honest. Build quality pretty good. Everything fine no c
racks or openings. Now the most important point - Bass. I have seen ma
ny people and many YouTubers telling that bass is bit much. For for me
it is pretty well balanced and works out pretty well. I cannot tell mu
ch about battery backup. Let me use this for a couple of days. Then I
will update my post. Sound quality also good and if you are used to li
stening to high volume of sound, you can easil...</div><span class="_1
BWGvX"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">I got this product on 5th oct
2021 , today is 15th october 2021.Iam using this product now with out
```

charging from delivery day...
The battery backup is owsom

Sound quality amd bass are verry nice and morethan worth for your money...
But build quality is not pretty good. But it is also value for money in this range...

The magnetic lock is not perfect... Magnetic power is little lower.
Some features are miss in this headphone ...like in oneplus headphones.. That is automatically off when we...</div>READ MORE</div>]

In [11]:

```
# Extracting product reviews from the above result
product_reviews_extracted = []
for product_review in product_reviews:
    product_reviews_extracted.append(product_review.get_text())

product_reviews_extracted
```

Out[11]:

```
['1. Material Quality - Not Bad...2. Sound Quality - Awesome...3. Bass
- Decent Quality...4. Battery Backup - Very Good...5. Design & Build Q
uality - Very nice...6. Fitting - Okay Type...Overall very nice and Val
ue for money product...Really i love the product...😊READ MORE',
"Writing this review after using more than 15 days. 255F pro+ has sup
er sound quality. Bass is decent and fair enough. Most important highl
ight of this headset is battery life,it's amazing. Runtime depends on
usage but definitely it will give you more than any other headset.Shou
ld go for it.READ MORE",
'One of the best Bluetooth Boat. 1- Battery Backup Amazing 2- light w
ait 3-Awesome Sound quality4-Bass good5- Design normal but good not be
stREAD MORE',
'Nice product and fast delivery . I am happy nice job Flipkart thank
sREAD MORE',
"Delivered in 20 February.. today is 28 February and i didn't charge
it till now still battery percentage is 70%.. i use it daily for onlin
e classes, 1 movie and 3 hours For music approx 9-10 hours daily so Ba
ttery is very Powerfull.it's housing is made by plastic and neck wire
of somewhat soft silicone material cable is flat long and strong enoug
h..Bass is above the best if you use right size of ear buds..even in 9
0% volume vocals are crystal clear treble is well balanced..Call quali
ty is far...READ MORE",
'This product is actually best. best color and design.But sound is
little low not bad it's a perfect but little low. bass is very powerful
and clear. battery backup is very best in this price range. definitely
you purchase this product.READ MORE',
'Great product., battery backup up to 14 days use , sound quality is
great , I am happy with the product 😊😊😊😊READ MORE',
'Battery backup is best and full charge very quickly. Sound quality i
s best .I love it. You can buy it without any hesitation.READ MORE',
'Just got my hands on this product. Let me be honest. Build quality p
retty good. Everything fine no cracks or openings. Now the most import
ant point - Bass. I have seen many people and many YouTubers telling t
hat bass is bit much. For for me it is pretty well balanced and works
out pretty well. I cannot tell much about battery backup. Let me use t
his for a couple of days. Then I will update my post. Sound quality al
so good and if you are used to listening to high volume of sound, you
can easil...READ MORE',
'I got this product on 5th oct 2021 , today is 15th october 2021.Iam
using this product now with out charging from delivery day...The batte
rry backup is owsomSound quality amd bass are verry nice and morethan
worth for your money...But build quality is not pretty good. But it is
also value for money in this range...The magnetic lock is not perfec
t... Magnetic power is little lower.Some features are miss in this hea
dphone ...like in oneplus headphones.. That is automatically off when
we...READ MORE']
```

d) Review titles

The review titles of the product have been found to be defined inside a paragraph tag with the class named "_2-N8zT"

In [12]:

```
title = soup.find_all('p',class_='_2-N8zT')
```

In [13]:

```
title
```

Out[13]:

```
[<p class="_2-N8zT">Just wow!</p>,
 <p class="_2-N8zT">Must buy!</p>,
 <p class="_2-N8zT">Great product</p>,
 <p class="_2-N8zT">Just wow!</p>,
 <p class="_2-N8zT">Good quality product</p>,
 <p class="_2-N8zT">Wonderful</p>,
 <p class="_2-N8zT">Perfect product!</p>,
 <p class="_2-N8zT">Terrific</p>,
 <p class="_2-N8zT">Good choice</p>,
 <p class="_2-N8zT">Worth the money</p>]
```

In [14]:

```
# Extracting review titles from above result
review_title_extracted = []
for i in range(0,len(title)):
    review_title_extracted.append(title[i].get_text())

review_title_extracted
```

Out[14]:

```
['Just wow!',
 'Must buy!',
 'Great product',
 'Just wow!',
 'Good quality product',
 'Wonderful',
 'Perfect product!',
 'Terrific',
 'Good choice',
 'Worth the money']
```

Saving report of the above data in csv file

In [15]:

```
# Initializing empty dataframe
dataframe = pd.DataFrame()
```

In [16]:

```
dataframe
```

Out[16]:

—

In [17]:

```
# Adding customer name to dataframe as a column  
dataframe['Customer Name'] = customer_names_extracted
```

In [18]:

```
dataframe
```

Out[18]:

| | Customer Name |
|---|----------------------|
| 0 | Manash Pratim Saikai |
| 1 | Himanshu Ghotkar |
| 2 | Asad Khan |
| 3 | sujit das |
| 4 | Pankaj Kumar |
| 5 | Bhargav Rajpura |
| 6 | Sagar Mallik |
| 7 | Aftab Alam |
| 8 | Prantik Sutradhar |
| 9 | su hail k |

In [19]:

```
# Adding rest of the columns to the dataframe  
dataframe['Review title'] = review_title_extracted  
dataframe['Ratings'] = product_ratings_extracted  
dataframe['Reviews'] = product_reviews_extracted
```

In [20]:

dataframe

Out[20]:

| | Customer Name | Review title | Ratings | Reviews |
|---|----------------------|----------------------|---------|---|
| 0 | Manash Pratim Saikai | Just wow! | 5 | 1. Material Quality - Not Bad...2. Sound Quali... |
| 1 | Himanshu Ghotkar | Must buy! | 5 | Writing this review after using more than 15 d... |
| 2 | Asad Khan | Great product | 5 | One of the best Bluetooth Boat. 1- Battery Bac... |
| 3 | sujit das | Just wow! | 5 | Nice product and fast delivery . I am happy n... |
| 4 | Pankaj Kumar | Good quality product | 4 | Delivered in 20 February.. today is 28 Februar... |
| 5 | Bhargav Rajpura | Wonderful | 4 | This product is acctully best. best coloty an... |
| 6 | Sagar Mallik | Perfect product! | 5 | Great product., battery backup up to 14 days u... |
| 7 | Aftab Alam | Terrific | 5 | Battery backup is best and full charge very qu... |
| 8 | Prantik Sutradhar | Good choice | 4 | Just got my hands on this product. Let me be h... |
| 9 | su hail k | Worth the money | 4 | I got this product on 5th oct 2021 , today is ... |

In [21]:

```
# Saving the dataframe to csv
dataframe.to_csv('reviews.csv', index=True)
```

Conclusion

1. Implemented web scraper for Flipkart website product

End of Notebook