


```
In [6]: Soup = BeautifulSoup(data, "html.parser")
        Soup.prettify
```

```
Out[6]: <bound method Tag.prettify of <!DOCTYPE html>
<html lang="en"><head><link href="https://rukminim2.flixcart.com" rel="pre
connect"/><link href="//static-assets-web.flixcart.com/fk-p-linchpin-web/f
k-cp-zion/css/app_modules.chunk.905c37.css" rel="stylesheet"/><link href
="//static-assets-web.flixcart.com/fk-p-linchpin-web/fk-cp-zion/css/app.ch
unk.615ed9.css" rel="stylesheet"/><meta content="text/html; charset=utf-8"
http-equiv="Content-type"/><meta content="IE=Edge" http-equiv="X-UA-Compat
ible"/><meta content="102988293558" property="fb:page_id"/><meta content
="658873552,624500995,100000233612389" property="fb:admins"/><meta content
="noodp" name="robots"/><link href="https://www/promos/new/20150528-14054
7-favicon-retina.ico" rel="shortcut icon"/><link href="/osdd.xml?v=2" rel
="search" type="application/opensearchdescription+xml"/><meta content="web
site" property="og:type"/><meta content="Flipkart.com" name="og_site_name"
property="og:site_name"/><link href="/apple-touch-icon-57x57.png" rel="app
le-touch-icon" sizes="57x57"/><link href="/apple-touch-icon-72x72.png" rel
="apple-touch-icon" sizes="72x72"/><link href="/apple-touch-icon-114x114.p
ng" rel="apple-touch-icon" sizes="114x114"/><link href="/apple-touch-icon-
144x144.png" rel="apple-touch-icon" sizes="144x144"/><link href="/apple-to
uch-icon-57x57.png" rel="apple-touch-icon"/><meta content="app" name="twit
```

STEP 3:- HTML tree traversal

Commonly Used Types Of Objects:

- (1) Tag
- (2) Navigablestring
- (3) BeautifulSoup
- (4) Comment

SAMSUNG Crystal 4K

```
In [7]: title=Soup.title
        title
```

```
Out[7]: <title>SAMSUNG Crystal 4K iSmart Series 138 cm (55 inch) Ultra HD (4K) LED Sm
art Tizen TV 2023 Edition Online at best Prices In India</title>
```

(1) Check The Product Price:-

```
In [8]: Price = Soup.find_all(class_="_30jeq3 _16Jk6d")
        Price[0].get_text()
```

```
Out[8]: '₹45,990'
```

(2) Get All The Customers Names In This HTML Page:-

```
In [9]: names = Soup.find_all(class_="_2sc7ZR _2V5EHH")
names=list(names)
print(type(names))
print(names)

len(names)
```

```
<class 'list'>
[<p class="_2sc7ZR _2V5EHH">Junaid Alam</p>, <p class="_2sc7ZR _2V5EHH">Hars
ha Patle</p>, <p class="_2sc7ZR _2V5EHH">Sanjeet Dash</p>, <p class="_2sc7ZR
_2V5EHH">Rushbh Kambe</p>, <p class="_2sc7ZR _2V5EHH">Sangameshwar</p>, <p cl
ass="_2sc7ZR _2V5EHH">Ravi Kant</p>, <p class="_2sc7ZR _2V5EHH">Rajesh j</p
>, <p class="_2sc7ZR _2V5EHH">MOHAMMED MUKARAM ALI</p>, <p class="_2sc7ZR _2V
5EHH">Amlan Ghosh</p>, <p class="_2sc7ZR _2V5EHH">abhinash Kumar</p>]
```

Out[9]: 10

(3) Get All The Customers Names In List Form:-

```
In [10]: cust_name = []
for i in names:
    cust_name.append(i.get_text())
cust_name
```

Out[10]: ['Junaid Alam',
'Harsha Patle',
'Sanjeet Dash',
'Rushbh Kambe',
'Sangameshwar',
'Ravi Kant',
'Rajesh j',
'MOHAMMED MUKARAM ALI',
'Amlan Ghosh',
'abhinash Kumar']

(4) Get All The Customers Comments On This Product:-

```
In [11]: Comment = Soup.find_all(class_="_2-N8zT")
Comment
```

Out[11]: [<p class="_2-N8zT">Simply awesome</p>,
<p class="_2-N8zT">Terrific purchase</p>,
<p class="_2-N8zT">Super!</p>,
<p class="_2-N8zT">Excellent</p>,
<p class="_2-N8zT">Excellent</p>,
<p class="_2-N8zT">Really Nice</p>,
<p class="_2-N8zT">Best in the market!</p>,
<p class="_2-N8zT">Worth every penny</p>,
<p class="_2-N8zT">Perfect product!</p>,
<p class="_2-N8zT">Fabulous!</p>]

(5) Get All The Customers Comments In List Form:-

```
In [12]: Review_Comment = []  
        for i in Comment:  
            Review_Comment.append(i.get_text())  
        Review_Comment
```

```
Out[12]: ['Simply awesome',  
          'Terrific purchase',  
          'Super!',  
          'Excellent',  
          'Excellent',  
          'Really Nice',  
          'Best in the market!',  
          'Worth every penny',  
          'Perfect product!',  
          'Fabulous!']
```

(6) Check How Much STARS Gives Customers To This Product:-

(STARS RATING OUT OF 5 STARS)

```
In [13]: Stars = Soup.find_all("div",class_="_3LWZlK _1BLPMq")
Stars
```

```
Out[13]: [<div class="_3LWZlK _1BLPMq">5</div>,
  <div class="_3LWZlK _1BLPMq">5</div>,
  <div class="_3LWZlK _1BLPMq">5</div>,
  <div class="_3LWZlK _1BLPMq">5</div>,
  <div class="_3LWZlK _1BLPMq">5</div>,
  <div class="_3LWZlK _1BLPMq">4</div>,
  <div class="_3LWZlK _1BLPMq">5</div>,
  <div class="_3LWZlK _1BLPMq">5</div>,
  <div class="_3LWZlK _1BLPMq">5</div>,
  <div class="_3LWZlK _1BLPMq">5</div>]
```

(7) Check How Much STARS Gives Customers To This Product In List Form:-

(STARS RATING OUT OF 5 STARS)

```
In [14]: Review_Stars = []
for i in range(0,len(Stars)):
    Review_Stars.append(Stars[i].get_text())
Review_Stars
```

```
Out[14]: ['5', '5', '5', '5', '5', '4', '5', '5', '5', '5']
```

(8) Get All The Customers Review To This Product:-

```
In [15]: Review = Soup.find_all("div",class_="t-ZTKy")
Review
```

```
Out[15]: [<div class="t-ZTKy"><div><div class="">Awesome product</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">Good product</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">Awesome.... great product, fast ui, better connectivity options... HDMI Arc has good sound output rather than optical and Bluetooth.... Always have additional two years warranty with Samsung care pack... check with Samsung engineer who comes for installation... Basic wall mounting is free.... For installation call Samsung toll free no.</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">Everything is good but let's see how much its durable.</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">Excellent product from Samsung</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">1. Screen quality is very good even though the SD channels Looking good tested with HD setup-box. I use movie mode with some manual changes.<br><br>2. Sound: according to me vocals are not that much clear and feels like normal speaker.<br><br>3. UI is fast and responsive buy Google Chromecast for Android experience.<br><br>4. Viewing angles are good I worried about this because of VA pannel but it's good in viewing angles.<br><br>5. Service is top notch 😊<br><br>Edit (After uses of 6 months)<br><br>1. I need to replace the Motherboard...<br></div><span class="_1BW GvX"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">Excellent picture quality</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">Very good TV. Very nice sound. awesome to look at</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">Excellent product. Excellent delivery & installation. You just need a good internet connection to enjoy this television. One of the best in the category.</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>,
  <div class="t-ZTKy"><div><div class="">The product is awesome... And the delivery by flipkart was really good..</div><span class="_1H-bmy"><span>READ MORE</span></span></div></div>]
```

(9) Get All The Customers Review To This Product In List Form:-

```
In [16]: Content_Review = []  
for i in range(0,len(Review)):  
    Content_Review.append(Review[i].get_text())  
Content_Review
```

```
Out[16]: ['Awesome productREAD MORE',  
          'Good productREAD MORE',  
          'Awesome.... great product, fast ui, better connectivity options... HDMI Arc  
          has good sound output rather than optical and Bluetooth.... Always have addit  
          ional two years warranty with Samsung care pack... check with Samsung enginee  
          r who comes for installation... Basic wall mounting is free.... For installat  
          ion call Samsung toll free no.READ MORE',  
          "Everything is good but let's see how much its durable.READ MORE",  
          'Excellent product from SamsungREAD MORE',  
          "1. Screen quality is very good even the SD channels Looking good tested wit  
          h HD setup-box. I use movie mode with some manual changes.2. Sound: according  
          to me vocals are not that much clear and feels like normal speaker.3. UI is  
          fast and responsive buy Google Chromecast for Android experience.4. Viewing a  
          ngles are good I worried about this because of VA pannel but it's good in vie  
          wing angles.5. Service is top notch 😊Edit (After uses of 6 months)1. I need  
          to replace the Motherboa...READ MORE",  
          'Excellent picture qualityREAD MORE',  
          'Very good TV. Very nice sound. awesome to look atREAD MORE',  
          'Excellent product. Excellent delivery & installation. You just need a good  
          internet connection to enjoy this television. One of the best in the categor  
          y.READ MORE',  
          'The product is awesome... And the delivery by flipkart was really good..READ  
          MORE']
```

(10) Get All The Likes And Dislikes To Customers Review:-

```
In [17]: LikesDislikes = Soup.find_all(class_="_3c3Px5")
LikesDislikes
```

```
Out[17]: [<span class="_3c3Px5">589</span>,
<span class="_3c3Px5">98</span>,
<span class="_3c3Px5">349</span>,
<span class="_3c3Px5">55</span>,
<span class="_3c3Px5">144</span>,
<span class="_3c3Px5">19</span>,
<span class="_3c3Px5">151</span>,
<span class="_3c3Px5">21</span>,
<span class="_3c3Px5">458</span>,
<span class="_3c3Px5">79</span>,
<span class="_3c3Px5">211</span>,
<span class="_3c3Px5">32</span>,
<span class="_3c3Px5">71</span>,
<span class="_3c3Px5">10</span>,
<span class="_3c3Px5">32</span>,
<span class="_3c3Px5">3</span>,
<span class="_3c3Px5">373</span>,
<span class="_3c3Px5">87</span>,
<span class="_3c3Px5">74</span>,
<span class="_3c3Px5">15</span>,
<span class="_3c3Px5">110</span>,
<span class="_3c3Px5">30</span>,
<span class="_3c3Px5">158</span>,
<span class="_3c3Px5">55</span>,
<span class="_3c3Px5">6</span>,
<span class="_3c3Px5">17</span>,
<span class="_3c3Px5">206</span>,
<span class="_3c3Px5">94</span>,
<span class="_3c3Px5">62</span>,
<span class="_3c3Px5">31</span>,
<span class="_3c3Px5">13</span>,
<span class="_3c3Px5">3</span>,
<span class="_3c3Px5">13</span>,
<span class="_3c3Px5">3</span>,
<span class="_3c3Px5">7</span>,
<span class="_3c3Px5">1</span>,
<span class="_3c3Px5">3</span>,
<span class="_3c3Px5">4</span>,
<span class="_3c3Px5">3</span>,
<span class="_3c3Px5">1</span>]
```

(11) Get All The Likes And Dislikes To Customers Review In List Form:-


```
In [18]: Product_LikesDislikes = []  
         for i in range(10, len(LikesDislikes)):  
             Product_LikesDislikes.append(LikesDislikes[i].get_text())  
         Product_LikesDislikes
```

```
Out[18]: ['211',  
          '32',  
          '71',  
          '10',  
          '32',  
          '3',  
          '373',  
          '87',  
          '74',  
          '15',  
          '110',  
          '30',  
          '158',  
          '55',  
          '6',  
          '17',  
          '206',  
          '94',  
          '62',  
          '31',  
          '13',  
          '3',  
          '13',  
          '3',  
          '7',  
          '1',  
          '3',  
          '4',  
          '3',  
          '1']
```

```
In [19]: like=[]
dislike=[]
for i in range(10,len(Product_LikesDislikes)):
    print(i)
    if i%2==0:
        like.append(Product_LikesDislikes[i])
    elif i%2!=0:
        dislike.append(Product_LikesDislikes[i])

print(like)
print(dislike)
```

```
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
['110', '158', '6', '206', '62', '13', '13', '7', '3', '3']
['30', '55', '17', '94', '31', '3', '3', '1', '4', '1']
```

Applying a Tabular Data Structure

```
In [20]: import pandas as pd
```

```
In [21]: Consolidated_data = pd.DataFrame({"Customers_Name": cust_name, "Review_Comment":
print(Consolidated_data)
```

	Customers_Name	Review_Comment	Review_Stars	\
101	Junaid Alam	Simply awesome	5	
102	Harsha Patle	Terrific purchase	5	
103	Sanjeet Dash	Super!	5	
104	Rushbh Kambe	Excellent	5	
105	Sangameshwar	Excellent	5	
106	Ravi Kant	Really Nice	4	
107	Rajesh j	Best in the market!	5	
108	MOHAMMED MUKARAM ALI	Worth every penny	5	
109	Amlan Ghosh	Perfect product!	5	
110	abhinash Kumar	Fabulous!	5	

		Review	Likes	Dislikes
101		Awesome product	110	30
102		Good product	158	55
103	Awesome.... great product, fast ui, better con...		6	17
104	Everything is good but let's see how much its ...		206	94
105	Excellent product from Samsung	READ MORE	62	31
106	1. Screen quality is very good even the SD cha...		13	3
107	Excellent picture quality	READ MORE	13	3
108	Very good TV. Very nice sound. awesome to look...		7	1
109	Excellent product. Excellent delivery & instal...		3	4
110	The product is awsome... And the delivery by f...		3	1

Exporting The DataSet Into a CSV File

```
In [22]: Consolidated_data.to_csv("SAMSUNG_Crystal_4K_Review.csv")
```

```
In [23]: import os
os.getcwd()
```

```
Out[23]: 'C:\\Users\\dhira'
```

```
In [24]: df1 = pd.read_csv("SAMSUNG_Crystal_4K_Review.csv")
df1.head()
```

```
Out[24]:
```

	Unnamed: 0	Customers_Name	Review_Comment	Review_Stars	Review	Likes	Dislikes
0	101	Junaid Alam	Simply awesome	5	Awesome product READ MORE	110	30
1	102	Harsha Patle	Terrific purchase	5	Good product READ MORE	158	55
2	103	Sanjeet Dash	Super!	5	Awesome.... great product, fast ui, better con...	6	17
3	104	Rushbh Kambe	Excellent	5	Everything is good but let's see how much its ...	206	94
4	105	Sangameshwar	Excellent	5	Excellent product from Samsung READ MORE	62	31

```
In [25]: df1.describe()
```

```
Out[25]:
```

	Unnamed: 0	Review_Stars	Likes	Dislikes
count	10.00000	10.000000	10.000000	10.000000
mean	105.50000	4.900000	58.100000	23.900000
std	3.02765	0.316228	74.557137	30.431891
min	101.00000	4.000000	3.000000	1.000000
25%	103.25000	5.000000	6.250000	3.000000
50%	105.50000	5.000000	13.000000	10.500000
75%	107.75000	5.000000	98.000000	30.750000
max	110.00000	5.000000	206.000000	94.000000

(12) Use .head() To Find First Five Data From DataSet:-

In [26]: df1.head()

Out[26]:

	Unnamed: 0	Customers_Name	Review_Comment	Review_Stars	Review	Likes	Dislikes
0	101	Junaid Alam	Simply awesome	5	Awesome productREAD MORE	110	30
1	102	Harsha Patle	Terrific purchase	5	Good productREAD MORE	158	55
2	103	Sanjeet Dash	Super!	5	Awesome.... great product, fast ui, better con...	6	17
3	104	Rushbh Kambe	Excellent	5	Everything is good but let's see how much its ...	206	94
4	105	Sangameshwar	Excellent	5	Excellent product from SamsungREAD MORE	62	31

(13) Use .tail() To Find Last Five Data From DataSet:-

In [27]: df1.tail()

Out[27]:

	Unnamed: 0	Customers_Name	Review_Comment	Review_Stars	Review	Likes	Dislikes
5	106	Ravi Kant	Really Nice	4	1. Screen quality is very good even the SD cha...	13	3
6	107	Rajesh j	Best in the market!	5	Excellent picture qualityREAD MORE	13	3
7	108	MOHAMMED MUKARAM ALI	Worth every penny	5	Very good TV. Very nice sound. awesome to look...	7	1
8	109	Amlan Ghosh	Perfect product!	5	Excellent product. Excellent delivery & instal...	3	4
9	110	abhinash Kumar	Fabulous!	5	The product is awesome... And the delivery by f...	3	1

(14) Use .info() To Get The Information Of DataSet:-

In [28]: df1.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0             10 non-null    int64
1   Customers_Name         10 non-null    object
2   Review_Comment         10 non-null    object
3   Review_Stars           10 non-null    int64
4   Review                 10 non-null    object
5   Likes                 10 non-null    int64
6   Dislikes               10 non-null    int64
dtypes: int64(4), object(3)
memory usage: 688.0+ bytes
```

SORTING THE DATASET FILE

(15) Sorting The Names With Ascending Orders:-

In [29]: df1.Customers_Name.sort_values()

```
Out[29]: 8          Amlan Ghosh
1          Harsha Patle
0          Junaid Alam
7  MOHAMMED MUKARAM ALI
6          Rajesh j
5          Ravi Kant
3          Rushbh Kambe
4          Sangameshwar
2          Sanjeet Dash
9          abhinash Kumar
Name: Customers_Name, dtype: object
```

(16) Sorting The Names With Descending Orders:-

```
In [30]: df1.Customers_Name.sort_values(ascending=False)
```

```
Out[30]: 9          abhinash Kumar
          2          Sanjeet Dash
          4          Sangameshwar
          3          Rushbh Kambe
          5          Ravi Kant
          6          Rajesh j
          7  MOHAMMED MUKARAM ALI
          0          Junaid Alam
          1          Harsha Patle
          8          Amlan Ghosh
          Name: Customers_Name, dtype: object
```

FILTERING THE DATA

(17) Find The Customers Who Gives This Product 5 Stars:-

```
In [31]: df1[df1.Review_Stars==5]["Customers_Name"]
```

```
Out[31]: 0          Junaid Alam
          1          Harsha Patle
          2          Sanjeet Dash
          3          Rushbh Kambe
          4          Sangameshwar
          6          Rajesh j
          7  MOHAMMED MUKARAM ALI
          8          Amlan Ghosh
          9          abhinash Kumar
          Name: Customers_Name, dtype: object
```

(18) Find The Customers Who Gives This Product Less Than 5 stars:-

```
In [32]: df1[df1.Review_Stars<5]["Customers_Name"]
```

```
Out[32]: 5          Ravi Kant
          Name: Customers_Name, dtype: object
```

MULTIPLE FILTERING

(19) Check Only The Data Who Gives 5 Stars To This Product:-

```
In [33]: df1[(df1.Review_Stars>=5) & (df1.Customers_Name)]
```

```
Out[33]:
```

	Unnamed: 0	Customers_Name	Review_Comment	Review_Stars	Review	Likes	Dislikes
0	101	Junaid Alam	Simply awesome	5	Awesome product READ MORE	110	30
1	102	Harsha Patle	Terrific purchase	5	Good product READ MORE	158	55
2	103	Sanjeet Dash	Super!	5	Awesome.... great product, fast ui, better con...	6	17
3	104	Rushbh Kambe	Excellent	5	Everything is good but let's see how much its ...	206	94
4	105	Sangameshwar	Excellent	5	Excellent product from Samsung READ MORE	62	31
6	107	Rajesh j	Best in the market!	5	Excellent picture quality READ MORE	13	3
7	108	MOHAMMED MUKARAM ALI	Worth every penny	5	Very good TV. Very nice sound. awesome to look...	7	1
8	109	Amlan Ghosh	Perfect product!	5	Excellent product. Excellent delivery & instal...	3	4
9	110	abhinash Kumar	Fabulous!	5	The product is awsome... And the delivery by f...	3	1

(20) Check Only The Data Who Gives Less Than 5 Stars To This Product:-

```
In [34]: df1[(df1.Review_Stars<5) & (df1.Customers_Name)]
```

```
Out[34]:
```

	Unnamed: 0	Customers_Name	Review_Comment	Review_Stars	Review	Likes	Dislikes
5	106	Ravi Kant	Really Nice	4	1. Screen quality is very good even the SD cha...	13	3

WE CAN APPLY STRING METHODS ALSO

(21) .upper() Method Convert All The LETTERS Into CAPITAL LETTERS:-

```
In [35]: df1.Customers_Name.str.upper()
```

```
Out[35]: 0          JUNAID  ALAM
1          HARSHA  PATLE
2          SANJEET  DASH
3          RUSHBH  KAMBE
4          SANGAMESHWAR
5             RAVI  KANT
6             RAJESH  J
7    MOHAMMED  MUKARAM  ALI
8             AMLAN  GHOSH
9          ABHINASH  KUMAR
Name: Customers_Name, dtype: object
```

(22) .lower() Method Convert All The LETTERS Into Small LETTERS:-

```
In [36]: df1.Customers_Name.str.lower()
```

```
Out[36]: 0          junaid  alam
1          harsha  patle
2          sanjeet  dash
3          rushbh  kambe
4          sangameshwar
5             ravi  kant
6             rajesh  j
7    mohammed  mukaram  ali
8             amlan  ghosh
9          abhinash  kumar
Name: Customers_Name, dtype: object
```

SLICLING OPERATION IN DATASET

(23) Find Names AND Review Of The Location We Have Given:-

```
In [37]: df1.loc[[3,9,6],["Customers_Name","Review_Comment"]]
```

```
Out[37]:
```

	Customers_Name	Review_Comment
3	Rushbh Kambe	Excellent
9	abhinash Kumar	Fabulous!
6	Rajesh j	Best in the market!

