

Web Scraping Using Python & BeautifulSoup

BOAT SMARTWATCH REVIEW

By

Namrata M. Gotmare

Web Scraping Using Python & BeautifulSoup

BOAT SMARTWATCH REVIEW

If you want to scrape a website

- (1) Use the API
 - (2) HTML Web Scraping using some tool like

Step 0:Setting up the environment

Step 1:Get the HTML

Step 2:Parse the HTML

Step 3:HTML tree traversal

STEP 0:-Setting up the environment

```
import requests  
from bs4 import BeautifulSoup  
import pandas as pd  
URL ="https://www.boat-lifestyle.com/products/wave-call-bluetooth-calling-smartwatch"
```

STEP 1:-Get the HTML

STEP 2:- Parse the HTML

```
soup = BeautifulSoup(htmlcontent,"html.parser")
#print(soup)
soup.prettify #Prettify the codes of the HTML page

<bound method Tag.prettify of <!DOCTYPE html>
<html class="no-js" dir="ltr" draggable="false" lang="en">
<head>
<meta charset="utf-8"/>
<meta content="width=device-width, initial-scale=1.0, height=device-height, minimum-scale=1.0, maximum-scale=6.0" name="viewport"/>
<meta content="#ffffff" name="theme-color"/>
<meta content="qad1e3gldwmxjf6luncdygpvcduub" name="facebook-domain-verification"/>
<link href="https://www.boat-lifestyle.com/products/wave-call-bluetooth-calling-smartwatch" rel="canonical"/><link href="//www.boat-lifestyle.com/cdn/shop/files/32x32_256x256.png_32x32_2d0995d9-ec86-4c14-b928-71101777194c_96x96.png?v=1647426716" rel="shortcut icon" type="image/png"/><title>
    boAt Wave Call | Bluetooth Calling Smartwatch with 1.69" (4.29 cm) HD Curved Display
</title><meta content='Shop boat wave call - a bluetooth calling smartwatch having an appropriate 1.69" (4.29 cm) curved display and it has a responsive dialpad for better calling functionality, save 10 contacts on this smartwatch. Order now!' name="description"/><link href="https://cdn.shopify.com" rel="preconnect"/>
<link href="https://www.google-analytics.com" rel="dns-prefetch"/>
<link href="https://fonts.googleapis.com" rel="preconnect"/>
<link crossorigin="" href="https://www.google.com" rel="dns-prefetch"/>
<link crossorigin="" href="https://www.googletagmanager.com" rel="dns-prefetch"/>
.......
```

STEP 3:- HTML tree traversal

```
Commonly Used Types Of Objects:
(1) Tag
(2) Navigablestring
(3) BeautifulSoup
(4) Comment
title = soup.title
print(type(soup)) # BeautifulSoup
print(type(title)) # Tag
print(type(title.string)) # Navigablestring
<class 'bs4.BeautifulSoup'>
<class 'bs4.element.Tag'>
<class 'bs4.element.NavigableString'>
```

TITLE OF THE PRODUCT

```
title=soup.title
title

<title>
    boAt Wave Call | Bluetooth Calling Smartwatch with 1.69" (4.29
cm) HD Curved Display
</title>
```

(1) Check The Product Price:-

```
Price = soup.find_all("div",class_="pdp_price_box")
Price

[<div class="pdp_price_box">
Sale price</span>₹1,199</span>
</div>]
```

(2) Check The Product Price In List Form:-

```
Product_Price = []
for i in range(0,len(Price)):
    Product_Price.append(Price[i].get_text().strip())
Product_Price

['Sale price₹1,199']
```

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

```
names = soup.find_all("span",class_="jdgm-rev__author")
names

[<span class="jdgm-rev__author">Ajay Sen</span>,
<span class="jdgm-rev__author">b venkatesh naik</span>,
<span class="jdgm-rev__author">Ruchi Raj</span>,
<span class="jdgm-rev__author">Faiyaz Ahmad Ahmad</span>,
<span class="jdgm-rev__author">Rahul Modha Rahul Modha</span>,
<span class="jdgm-rev__author">dhruva rupal</span>]
```

(4) Get All Customers Names In List Form:-

```
cust_name = []
for i in range(0,len(names)):
    cust_name.append(names[i].get_text())
cust_name

['Ajay Sen',
 'b venkatesh naik',
 'Ruchi Raj',
 'Faiyaz Ahmad Ahmad',
 'Rahul Modha Rahul Modha',
 'dhruva rupal']
```

(5) Get All Customers Comments On This Product:-

```
Comment = soup.find_all("b",class_="jdgm-rev__title")
Comment

[<b class="jdgm-rev__title">Nice</b>,
 <b class="jdgm-rev__title">5 Star ★★★★☆</b>,
 <b class="jdgm-rev__title">Good Catch</b>,
 <b class="jdgm-rev__title">Best watch</b>,
 <b class="jdgm-rev__title">Fantastic Product</b>,
 <b class="jdgm-rev__title">Fully satisfied with the wave call Smartwatch</b>]
```

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

```
Review_Comment = []
for i in range(0,len(Comment)):
    Review_Comment.append(Comment[i].get_text())
Review_Comment

['Nice',
 '5 Star ★★★★☆',
 'Good Catch',
 'Best watch',
 'Fantastic Product',
 'Fully satisfied with the wave call Smartwatch']
```

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

```
rating_one=soup.find("span",class_="jdgm-rev__rating")["data-score"]
rating_one
```

```
'5'
```

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

```
ratings= [span["aria-label"] for span in soup.find_all("span",class_="jdgm-rev__rating")]
ratings
```

```
['5 star review',
 '5 star review',
 '4 star review',
 '5 star review',
 '5 star review',
 '5 star review']
```

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

```
Review = soup.find_all("div",class_="jdgm-rev__body")
Review

[<div class="jdgm-rev__body"><p>Boat has done brilinet work</p></div>
',
 <div class="jdgm-rev__body"><p>Nice bilt quality, looking good</p></div>,
 <div class="jdgm-rev__body"><p>I'm using this watch and specs are pretty. Battery is lasting good 5 days that's impressive. I'm not a tech geek but this watch is very easy to use for rookies like me 😊</p>
</div>,
 <div class="jdgm-rev__body"><p>Loved this good product great display</p></div>,
 <div class="jdgm-rev__body"><p>I am very happy bcs i gave surprise to my wify and once she opened packed of boat there was incredible smile in her face.</p>
<p>Thank you boat being a such nice product.</p>
<p>Thank you so much</p></div>,
 <div class="jdgm-rev__body"><p>Boat wave call Smartwatch is nice and smart Smartwatch, with nice display screen and it accounts steps , heartbeat, and breathing well . Bluetooth connectivity good... beautiful color strap 😊 😍 overall I am happy with the product... Boat is an affordable and dependable product. I love Boat 💕</p></div>]
```

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

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

['Boat has done brilinet work',
 'Nice bilt quality, looking good',
 "I'm using this watch and specs are pretty. Battery is lasting good 5 days that's impressive. I'm not a tech geek but this watch is very easy to use for rookies like me 😊",
 'Loved this good product great display',
 'I am very happy bcs i gave surprise to my wify and once she opened packed of boat there was incredible smile in her face.\nThank you boat being a such nice product.\nThank you so much',
 'Boat wave call Smartwatch is nice and smart Smartwatch, with nice display screen and it accounts steps , heartbeat, and breathing well . Bluetooth connectivity good... beautiful color strap 😊 😍 overall I am happy with the product... Boat is an affordable and dependable product. I love Boat 💕']
```

(11) Get All The Anchors Tag From This HTML Page:

```
[<a class="visually-hidden skip-to-content" href="#main">skip to content</a>,
<a class="announce" href="/products/airdopes-supreme-long-playback-earbuds">
<div class="announcement-bar_message text--small">
<p><b>Airdopes Supreme</b> is out now! 🎉 Get home the supreme sound<b> at just ₹1299</b></p>
</div>
</a>,
<a aria-controls="desktop-menu-1" aria-expanded="false" class="header_linklist-link link--animated" href="#">
<span class="linklist--title">Categories</span><svg class="icon icon--dropdown-arrow-right icon--direction-aware" focusable="false" height="10" viewBox="0 0 7 10" width="7">
<path clip-rule="evenodd" d="M3.9394 5L0.469727 1.53033L1.53039 0.469666L6.06072 5L1.53039 9.53032L0.469727 8.46967L3.9394 5Z" fill="currentColor" fill-rule="evenodd"></path>
</svg>
</a>,
<a class="header_linklist-link link--animated" href="https://www.boat-lifestyle.com/collections/product-personalization">
<span class="linklist--title">boAt Personalisation</span></a>,
<a class="header_linklist-link link--animated" href="https://www.boat-lifestyle.com/pages/quiz">
<span class="linklist--title">Gift with boAt</span></a>,
<a class="header_linklist-link link--animated" href="/pages/bulk-orders">
<span class="linklist--title">Corporate Orders</span></a>,
```

Applying a Tabular Data Structure

```
import pandas as pd
df = pd.DataFrame({"Customer's Name": cust_name,"Review_Comment":Review_Comment, "rating_one":ratings,"Review":Content_Review })
print(df)
```

	Customer's Name	Review_Comment \
0	Ajay Sen	Nice
1	b venkatesh naik	5 Star ★★★★★
2	Ruchi Raj	Good Catch
3	Faiyaz Ahmad Ahmad	Best watch
4	Rahul Modha Rahul Modha	Fantastic Product
5	dhruba rupal	Fully satisfied with the wave call Smartwatch

	rating_one	Review
0	5 star review	Boat has done brilinet work
1	5 star review	Nice bilt quality, looking good
2	4 star review	I'm using this watch and specs are pretty. Bat...
3	5 star review	Loved this good product great display
4	5 star review	I am very happy bcs i gave surprise to my wif...
5	5 star review	Boat wave call Smartwatch is nice and smart Sm...

Exporting The DataSet Into a CSV File

```
import os
os.getcwd()

'C:\\Users\\Namrata\\Desktop'

df.to_csv("smart_watch_boat.csv")

df1 = pd.read_csv("smart_watch_boat.csv")
df1
```

Unnamed: 0	Customer's Name	Review_Comment	rating_one	Review	
0	Ajay Sen	Nice	5 star review	Boat has done brilinet work	
1	b venkatesh naik	5 Star	★★★★★	5 star review	Nice bilt quality, looking good
2	Ruchi Raj	Good Catch	4 star review	I'm using this watch and specs are pretty. Bat...	
3	Faiyaz Ahmad Ahmad	Best watch	5 star review	Loved this good product great display	
4	Rahul Modha Rahul Modha	Fantastic Product	5 star review	I am very happy bcs i gave surprise to my wif...	
5	dhruba rupal	Fully satisfied with the wave call Smartwatch	5 star review	Boat wave call Smartwatch is nice and smart Sm...	

(12) Change The Column Names Permanently:-

```
col = ["Serial_no.", "Customer_Names", "Comments", "Stars", "Customer_Review"]
df1.columns = col
df1
```

Serial_no.	Customer_Names	Comments	Stars	Customer_Review	
0	Ajay Sen	Nice	5 star review	Boat has done brilinet work	
1	b venkatesh naik	5 Star	★★★★★	5 star review	Nice bilt quality, looking good
2	Ruchi Raj	Good Catch	4 star review	I'm using this watch and specs are pretty. Bat...	
3	Faiyaz Ahmad Ahmad	Best watch	5 star review	Loved this good product great display	
4	Rahul Modha Rahul Modha	Fantastic Product	5 star review	I am very happy bcs i gave surprise to my wif...	
5	dhruba rupal	Fully satisfied with the wave call Smartwatch	5 star review	Boat wave call Smartwatch is nice and smart Sm...	

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

Serial_no.	Customer_Names	Comments	Stars	Customer_Review
0	Ajay Sen	Nice	5 star review	Boat has done brilliant work
1	b venkatesh naik	5 Star ★★★★★	5 star review	Nice build quality, looking good
2	Ruchi Raj	Good Catch	4 star review	I'm using this watch and specs are pretty. Bat...
3	Faiyaz Ahmad Ahmad	Best watch	5 star review	Loved this good product great display
4	Rahul Modha Rahul Modha	Fantastic Product	5 star review	I am very happy bcs i gave surprise to my wif...

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

Serial_no.	Customer_Names	Comments	Stars	Customer_Review
1	b venkatesh naik	5 Star ★★★★★	5 star review	Nice build quality, looking good
2	Ruchi Raj	Good Catch	4 star review	I'm using this watch and specs are pretty. Bat...
3	Faiyaz Ahmad Ahmad	Best watch	5 star review	Loved this good product great display
4	Rahul Modha Rahul Modha	Fantastic Product	5 star review	I am very happy bcs i gave surprise to my wif...
5	dhruba rupal	Fully satisfied with the wave call Smartwatch	5 star review	Boat wave call Smartwatch is nice and smart Sm...

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6 entries, 0 to 5
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Serial_no.       6 non-null      int64  
 1   Customer_Names  6 non-null      object  
 2   Comments         6 non-null      object  
 3   Stars            6 non-null      object  
 4   Customer_Review  6 non-null      object  
dtypes: int64(1), object(4)
memory usage: 372.0+ bytes
```

SORTING THE DATASET FILE

(16) Sorting The Names With Ascending Orders:-

```
df1.Customer_Names.sort_values()

0          Ajay Sen
3      Faiyaz Ahmad Ahmad
4  Rahul Modha Rahul Modha
2          Ruchi Raj
1      b venkatesh naik
5          dhruva rupal
Name: Customer_Names, dtype: object
```

(17) Sorting The Names With Descending Orders:-

```
df1.Customer_Names.sort_values(ascending=False)

5          dhruva rupal
1      b venkatesh naik
2          Ruchi Raj
4  Rahul Modha Rahul Modha
3      Faiyaz Ahmad Ahmad
0          Ajay Sen
Name: Customer_Names, dtype: object
```

FILTERING THE DATA

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

```
df1[df1.Stars=="5 star review"]["Customer_Names"]

0          Ajay Sen
1      b venkatesh naik
3      Faiyaz Ahmad Ahmad
4  Rahul Modha Rahul Modha
5          dhruva rupal
Name: Customer_Names, dtype: object
```

(19) Find The Customers Who Gives This Product greater Than 4 stars:-

```
df1[df1.Stars<="4 star review"]["Customer_Names"]
```

```
2      Ruchi Raj  
Name: Customer_Names, dtype: object
```

WE CAN APPLY STRING METHODS ALSO

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

```
df1.Customer_Names.str.upper()  
0           AJAY SEN  
1           B VENKATESH NAIK  
2           RUCHI RAJ  
3           FAIYAZ AHMAD AHMAD  
4           RAHUL MODHA RAHUL MODHA  
5           DHRUVA RUPAL  
Name: Customer_Names, dtype: object
```

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

```
df1.Customer_Names.str.lower()  
0           ajay sen  
1           b venkatesh naik  
2           ruchi raj  
3           faiyaz ahmad ahmad  
4           rahul modha rahul modha  
5           dhruva rupal  
Name: Customer_Names, dtype: object
```

SLICING OPERATION IN DATASET

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

```
df1.loc[[1,3,5],["Customer_Names","Customer_Review"]]
```

	Customer_Names	Customer_Review
1	b venkatesh naik	Nice bilt quality, looking good
3	Faiyaz Ahmad Ahmad	Loved this good product great display
5	dhruba rupal	Boat wave call Smartwatch is nice and smart Sm...

DATA VISUALIZATION

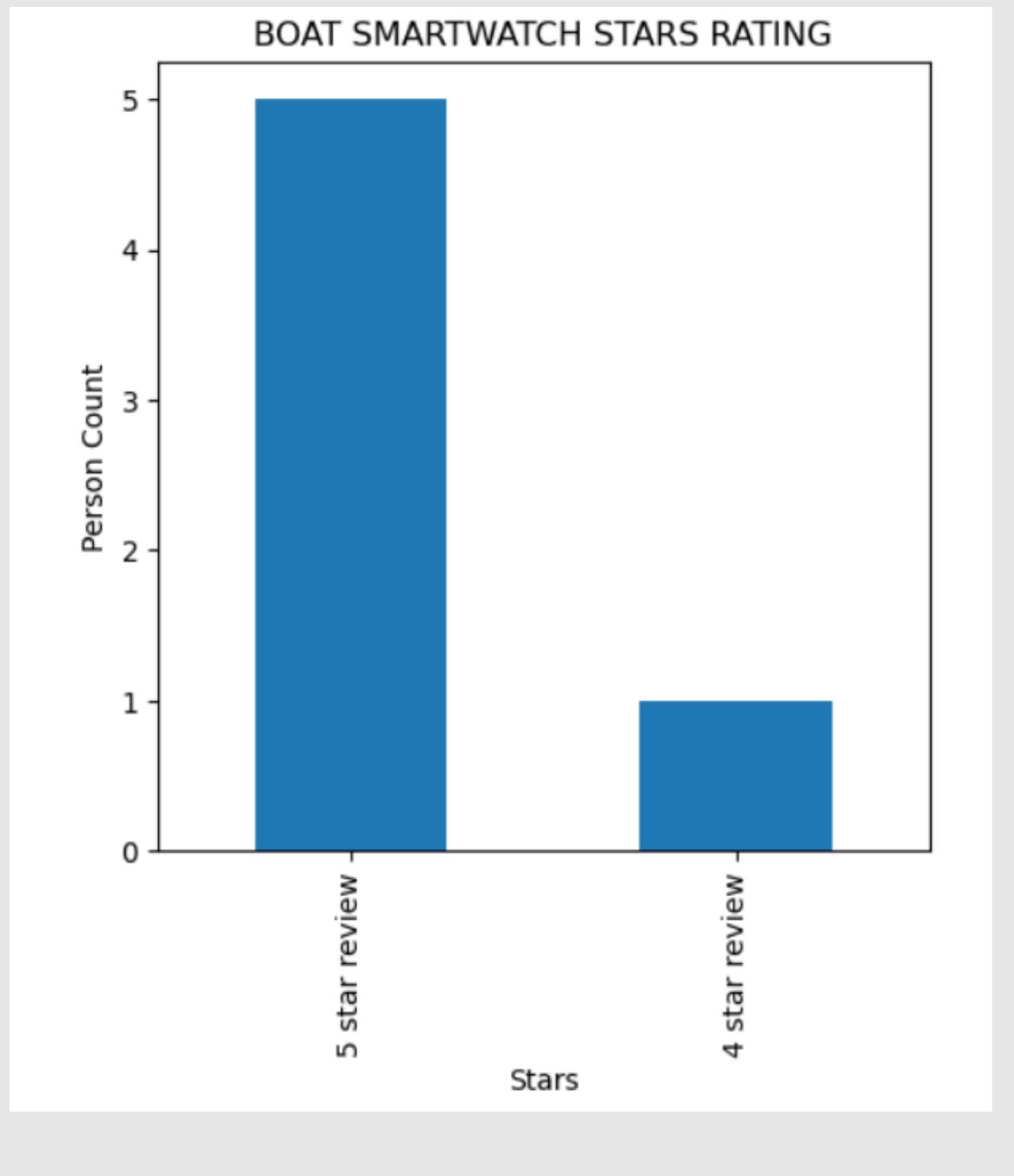
HISTOGRAM

```
import matplotlib.pyplot as plt  
%matplotlib inline  
  
plt.hist(x=df1.Stars)  
plt.show()
```



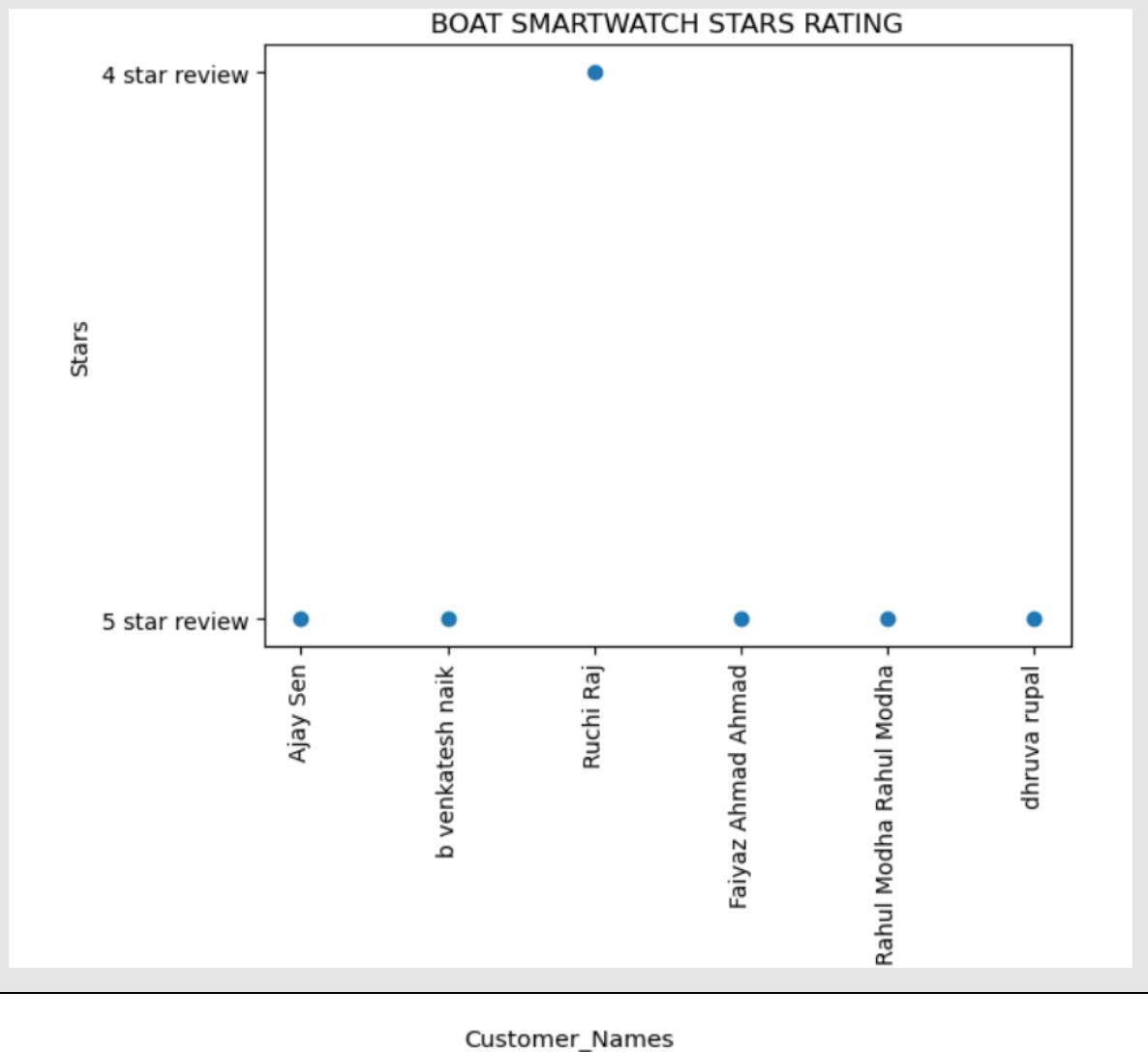
BAR PLOT

```
plt.figure(figsize=(9,9))
df1["Stars"].value_counts().plot(kind="bar")
plt.title("BOAT SMARTWATCH STARS RATING")
plt.ylabel("Person Count")
plt.xlabel("Stars")
plt.show()
```



SCATTER PLOT

```
plt.scatter(x=df1["Customer_Names"],y=df1["Stars"])
plt.title("BOAT SMARTWATCH STARS RATING")
plt.ylabel("Stars")
plt.xlabel("Customer_Names")
plt.xticks(rotation=90)
plt.show()
```



Conclusion:

This project helps to extract information from “Boat” website related to boat smartwatch by making use of various python libraries which include requests, beautifulsoup, pandas numpy. The extracted data is then transferred into dataframe and visualized using different visualization tools.