Project for Web Scraping

Web Scraping of Apple Watch From Flipkart

Institute Name:- ITvedant Education Pvt.Ltd
Name:- Raj Ravindra Kadam
Email Address:- siddhay16@gmail.com
Date Of Submission:-30th Nov 2023

AIMOFTHEPROJECT

The aim of this project is to create a Python Web scraping to extract specific details of the Apple Watch from flipkart . Using BeautifulSoup and Requests libraries, the objective is to gather essential information such as product names, prices, customer reviews, ratings, and comments associated with the Apple Watch listings

Description:

The primary goal of this web scraping project is to extract relevant information about the Apple Watch from the specified Flipkart product page for analysis purpose. The key data points include product title, price, comment, and customer reviews.
The extracted data is then structured into a Pandas DataFrame and exported to a CSV file for further analysis and insights.

Project Steps :-

- **URL Retrieval**: You've defined the URL for the Apple Watch product page on Flipkart. HTTP Request: Using the requests library, you've fetched the HTML content from the URL.

- **Parsing HTML:** Utilizing BeautifulSoup, you've parsed the HTML content to navigate and extract specific elements.

- **Data Extraction**: Extracted data includes product prices, customer names, comments, reviews, ratings, filtering.

- **Data Manipulation**: You've stored extracted data into lists and a pandas DataFrame.

- **Data Export**: Exported the DataFrame containing the extracted data to a CSV file.

Project Benefits:-

- **<u>Real-time Insights</u>**: Gather up-to-date Apple Watch prices, reviews, and trends from Flipkart for informed decision-making.

- **<u>Competitive Advantage</u>**: Analyze pricing strategies and customer sentiments, aiding in competitive analysis.

- **<u>Efficiency and Automation</u>**: Streamline data collection processes, enabling efficient market monitoring and operational enhancements.

<u>Challenges Faced:</u>

- **<u>Dynamic Content:</u>** Some parts of the webpage, such as customer reviews, are loaded dynamically using JavaScript. This required additional handling to ensure complete data retrieval.

- **<u>Anti-Scraping Measures:</u>** Flipkart employs anti-scraping measures like rate limiting and IP blocking. The scraping process incorporated delays to mitigate these issues.

<u>Output Format:</u>

The scraped data is stored in a structured format, such as CSV or JSON, for easy integration with downstream processes.

<u>Analysis of project:</u>

Based on our analysis we conclude that this project has most sales in the smartwatch in Flipkart as this product has no negative review and has average of 4.8

Apple Watch Web Scraping from Flipkart Code command:

(1) Access Flipkart web site by using import requests URL and BeautifulSoup library.

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

```python
url ="https://www.flipkart.com/apple-watch-se-gps-2nd-gen-heart-rate-monitor-crash-detection-sleep-tracking/p/itm5484aad9a2dc0?pi
```

```python
r =requests.get(url)
data =r.content
print(data)
```

```
4x114" href="/apple-touch-icon-114x114.png"/><link rel="apple-touch-icon" sizes="144x144" href="/apple-touch-icon-144x144.pn
g"/><link rel="apple-touch-icon" href="/apple-touch-icon-57x57.png"/><meta name="twitter:card" content="app"><meta name="twit
ter:site" content="@flipkart"><meta name="twitter:creator" content="@flipkart"><meta name="twitter:title" content="APPLE Watc
h SE GPS (2nd Gen) Heart Rate Monitor, Crash Detection, Sleep Tracking Price in India - Buy APPLE Watch SE GPS (2nd Gen) Hear
t Rate Monitor, Crash Detection, Sleep Tracking online at Flipkart.com"><meta name="twitter:description" content="Shop for el
ectronics, apparels & more using our Flipkart app Free shipping & COD."><meta property="twitter:image" content="http://rukmin
i1.flixcart.com/image/300/300/xif0q/smartwatch/q/l/y/-original-imaghxg3fsbt9mpa.jpeg"/><meta name="twitter:app:country" conte
nt="in"><meta name="al:ios:app_name" content="Flipkart"><meta name="al:ios:app_store_id" content="742044692"><meta name="twit
ter:app:name:iphone" content="Flipkart"><meta name="twitter:app:id:iphone" content="742044692"><meta name="twitter:app:url:ip
hone" content="http://dl.flipkart.com/dl/home?"><meta name="twitter:app:name:ipad" content="Flipkart"><meta name="twitter:ap
p:id:ipad" content="742044692"><meta name="twitter:app:url:ipad" content="http://dl.flipkart.com/dl/home?"><meta name="twitte
r:app:name:googleplay" content="Flipkart"><meta name="twitter:app:id:googleplay" content="com.flipkart.android"><meta name="t
witter:app:url:googleplay" content="http://dl.flipkart.com/dl/home?"><style>#container {\n\t\t\theight: 100%;\n\t\t}</style>
<link rel="stylesheet" href="///fk-cp-zion/css/Product.chunk.2fbab9.css"/><title>APPLE Watch SE GPS (2nd Gen) Heart Rate Moni
tor, Crash Detection, Sleep Tracking Price in India - Buy APPLE Watch SE GPS (2nd Gen) Heart Rate Monitor, Crash Detection, S
leep Tracking online at Flipkart.com</title><meta name="og_title" property="og:title" content="APPLE Watch SE GPS (2nd Gen) H
eart Rate Monitor, Crash Detection, Sleep Tracking Price in India - Buy APPLE Watch SE GPS (2nd Gen) Heart Rate Monitor, Cras
h Detection, Sleep Tracking online at Flipkart.com"/><meta name="Keywords" content="APPLE Watch SE GPS (2nd Gen) Heart Rate M
onitor, Crash Detection, Sleep Tracking Price, APPLE Watch SE GPS (2nd Gen) Heart Rate Monitor, Crash Detection, Sleep Tracki
```

```python
#prettify the codes of html page
soup =BeautifulSoup(data, "html.parser")
soup.prettify
```

```
<bound method Tag.prettify of <!DOCTYPE html>
<html lang="en"><head><link href="https://rukminim2.flixcart.com" rel="preconnect"/><link href="//static-assets-web.flixcart.
com/fk-p-linchpin-web/fk-cp-zion/css/app_modules.chunk.905c37.css" rel="stylesheet"/><link href="//static-assets-web.flixcar
t.com/fk-p-linchpin-web/fk-cp-zion/css/app.chunk.615ed9.css" rel="stylesheet"/><meta content="text/html; charset=utf-8" http-
equiv="Content-type"/><meta content="IE=Edge" http-equiv="X-UA-Compatible"/><meta content="102988293558" property="fb:page_i
d"/><meta content="658873552,624500995,100000233612389" property="fb:admins"/><meta content="noodp" name="robots"/><link href
="https:///www/promos/new/20150528-140547-favicon-retina.ico" rel="shortcut icon"/><link href="/osdd.xml?v=2" rel="search" ty
pe="application/opensearchdescription+xml"/><meta content="website" property="og:type"/><meta content="Flipkart.com" name="og
_site_name" property="og:site_name"/><link href="/apple-touch-icon-57x57.png" rel="apple-touch-icon" sizes="57x57"/><link hre
f="/apple-touch-icon-72x72.png" rel="apple-touch-icon" sizes="72x72"/><link href="/apple-touch-icon-114x114.png" rel="apple-t
ouch-icon" sizes="114x114"/><link href="/apple-touch-icon-144x144.png" rel="apple-touch-icon" sizes="144x144"/><link href="/a
pple-touch-icon-57x57.png" rel="apple-touch-icon"/><meta content="app" name="twitter:card"/><meta content="@flipkart" name="t
witter:site"/><meta content="@flipkart" name="twitter:creator"/><meta content="APPLE Watch SE GPS (2nd Gen) Heart Rate Monito
r, Crash Detection, Sleep Tracking Price in India - Buy APPLE Watch SE GPS (2nd Gen) Heart Rate Monitor, Crash Detection, Sle
ep Tracking online at Flipkart.com" name="twitter:title"/><meta content="Shop for electronics, apparels &amp; more using our
Flipkart app Free shipping &amp; COD." name="twitter:description"/><meta content="http://rukmini1.flixcart.com/image/300/300/
xif0q/smartwatch/q/l/y/-original-imaghxg3fsbt9mpa.jpeg" property="twitter:image"><meta content="in" name="twitter:app:countr
y"/><meta content="Flipkart" name="al:ios:app_name"/><meta content="742044692" name="al:ios:app_store_id"/><meta content="Fli
pkart" name="twitter:app:name:iphone"/><meta content="742044692" name="twitter:app:id:iphone"/><meta content="http://dl.flipk
```

```python
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'>
```

(2) To access Apple Watch price by using this class.

```
title = soup.title
title
```

```
<title>APPLE Watch SE GPS (2nd Gen) Heart Rate Monitor, Crash Detection, Sleep Tracking Price in India - Buy APPLE Watch SE GPS (2nd Gen) Heart Rate Monitor, Crash Detection, Sleep Tracking online at Flipkart.com</title>
```

```
price =soup.find_all(class_="_30jeq3 _16Jk6d")
price
```

```
[<div class="_30jeq3 _16Jk6d">₹26,999</div>]
```

```
price[0].get_text()
```

```
'₹26,999'
```

(3) To access Apple Watch Customer Name by using the respective class from its HTML content.

```
names =soup.find_all(class_="_2sc7ZR _2V5EHH")
names =list(names)
print (type(names))
print(names)
len(names)
```

```
<class 'list'>
[<p class="_2sc7ZR _2V5EHH">Akshay  Jambhale</p>, <p class="_2sc7ZR _2V5EHH">Madhwesh</p>, <p class="_2sc7ZR _2V5EHH">Flipkart Customer</p>, <p class="_2sc7ZR _2V5EHH">Ravi Pratap  Singh</p>, <p class="_2sc7ZR _2V5EHH">Jyotirmay sardar</p>, <p class="_2s c7ZR _2V5EHH">Avinash Goswami</p>, <p class="_2sc7ZR _2V5EHH">Arizona  Automobiles</p>, <p class="_2sc7ZR _2V5EHH">Vishnu PR</p >, <p class="_2sc7ZR _2V5EHH">Nimmagadda  Hemasri </p>, <p class="_2sc7ZR _2V5EHH">SAI CHARAN VALLURI</p>]

10
```

```
names[0].get_text()
```

```
'Akshay  Jambhale'
```

```
c_names=[]
for i in names:
    c_names.append(i.get_text())
c_names
```

```
['Akshay  Jambhale',
 'Madhwesh',
 'Flipkart Customer',
 'Ravi Pratap  Singh',
 'Jyotirmay sardar',
 'Avinash Goswami',
 'Arizona  Automobiles',
 'Vishnu PR',
```

(4) To access Apple Watch Customer Comment by using the respective class.

```
comment =soup.find_all(class_ ="_2-N8zT")
print(len(comment))
```

```
10
```

```
review_comment =[]
for i in comment:
    review_comment.append(i.get_text())
review_comment
```

```
['Really Nice',
 'Terrific purchase',
 'Just wow!',
 'Best in the market!',
 'Wonderful',
 'Perfect for first time buying an Apple Watch',
 'Great product',
 'Classy product',
 'Fabulous!',
 'Must buy!']
```

(4) To access Apple Watch Customer Review star by using the respective class.

```
stars =soup.find_all('div', class_ ="_3LWZlK _1BLPMq")
print(len(stars))
```

```
10
```

```
s_stars =[]
for i in stars:
    s_stars. append (i.get_text())
s_stars
```

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

(5) To access Apple Watch Customer Comment by using the respective class.

```
review = soup.find_all('div',class_ ="t-ZTKy")
print(len(review))
```

```
10
```

```
cust_review =[]
for i in review:
    cust_review.append(i.get_text())
cust_review
```

```
['Good for An entry Level APPLE WATCH..... Does the job..... Happy With the Purchase 😊READ MORE',
 'Very good productREAD MORE',
 'NiceREAD MORE',
 'The Apple Watch SE (2022) is the most affordable Apple Watch in India right now. Is it a better buy than the Watch Series 8.
I Know Series 8 Might Have more Features but Believe me you are not Going to use any Features Once you Start wearing It . Every
thing is For show Case.It gets most features of the Watch Series 8.The Watch SE is an obvious choice for any iPhone user.READ M
ORE',
 'GoodREAD MORE',
 "It strikes a perfect balance between functionality and affordability. Another impressive feature is the watch's ability to ma
ke and receive calls, send texts, and access Siri, all from your wrist. An ideal choice for anyone looking for a high-quality s
martwatch that won't break the bank.READ MORE",
 'Very nice My Apple Watch Very Good ProductREAD MORE',
 'My first Apple watch in my life. This one is awesome.. The battery stays up to 1 day but for me, not a hard user. Excellent d
isplay, voice call clarity is awesome. Workouts calculation are also great and Fall detection and crash detection are also avai
lable. I plan to go buy dizzo and other watches but when I read the reviews and all, i landed in Apple.READ MORE',
 'Unimaginably good at first I was reluctant to get but after once having I was perfectly fine of having and usage with iPhone
13 ProREAD MORE',
 'Flawless, classy, awesome watchREAD MORE']
```

(8) To see quick overview of the central tendency, spread, and distribution of the data in the DataFrame.

```
df.to_csv("apple_watch.csv")
```

```
import os
os.getcwd
```

```
<function nt.getcwd()>
```

```
df1.describe()
```

	review_stars
count	10.000000
mean	4.900000
std	0.316228
min	4.000000
25%	5.000000
50%	5.000000
75%	5.000000
max	5.000000

This summary can help in identifying outliers, understanding the range of values, and getting a general sense of the data distribution for further analysis

(9) Then store above all data in CSV file and again open this file in application.

```
import os
os.getcwd()
```

```
'C:\\Users\\HP'
```

```
df.to_csv("apple_watch.csv",index= False )
df1= pd.read_csv("apple_watch.csv")
df1
```

	customers_Name	Review_Comment	review_stars	review
0	Akshay Jambhale	Really Nice	4	Good for An entry Level APPLE WATCH..... Does ...
1	Madhwesh	Terrific purchase	5	Very good productREAD MORE
2	Flipkart Customer	Just wow!	5	NiceREAD MORE
3	Ravi Pratap Singh	Best in the market!	5	The Apple Watch SE (2022) is the most affordab...
4	Jyotirmay sardar	Wonderful	5	GoodREAD MORE
5	Avinash Goswami	Perfect for first time buying an Apple Watch	5	It strikes a perfect balance between functiona...
6	Arizona Automobiles	Great product	5	Very nice My Apple Watch Very Good ProductREAD...
7	Vishnu PR	Classy product	5	My first Apple watch in my life. This one is a...
8	Nimmagadda Hemasri	Fabulous!	5	Unimaginably good at first I was reluctant to ...
9	SAI CHARAN VALLURI	Must buy!	5	Flawless, classy, awesome watchREAD MORE

(9) Find max,mean,min etc

```
df.to_csv("apple_watch.csv")
```

```
import os
os.getcwd
```

```
<function nt.getcwd()>
```

```
df1.describe()
```

	review_stars
count	10.000000
mean	4.900000
std	0.316228
min	4.000000
25%	5.000000
50%	5.000000
75%	5.000000
max	5.000000

(10) Find INFO of data

```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 4 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   customers_Name  10 non-null     object
 1   Review_Comment  10 non-null     object
 2   review_stars    10 non-null     int64
 3   review          10 non-null     object
dtypes: int64(1), object(3)
memory usage: 452.0+ bytes
```

(11) sort the name of Customer in Ascending order

```
df1.customers_Name.sort_values()
```

```
0         Akshay  Jambhale
6    Arizona  Automobiles
5         Avinash Goswami
2        Flipkart Customer
4         Jyotirmay sardar
1                 Madhwesh
8    Nimmagadda  Hemasri
3       Ravi Pratap  Singh
9      SAI CHARAN VALLURI
7                Vishnu PR
Name: customers_Name, dtype: object
```

(12) Sort the Names of Customer in Descending order

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

```
7                Vishnu PR
9      SAI CHARAN VALLURI
3       Ravi Pratap  Singh
8    Nimmagadda  Hemasri
1                 Madhwesh
4         Jyotirmay sardar
2        Flipkart Customer
5         Avinash Goswami
6    Arizona  Automobiles
0         Akshay  Jambhale
Name: customers_Name, dtype: object
```

(13) Find who had given full rating

```
df1[df1.review_stars==5]
```

	customers_Name	Review_Comment	review_stars	review
1	Madhwesh	Terrific purchase	5	Very good productREAD MORE
2	Flipkart Customer	Just wow!	5	NiceREAD MORE
3	Ravi Pratap Singh	Best in the market!	5	The Apple Watch SE (2022) is the most affordab...
4	Jyotirmay sardar	Wonderful	5	GoodREAD MORE
5	Avinash Goswami	Perfect for first time buying an Apple Watch	5	It strikes a perfect balance between functiona...
6	Arizona Automobiles	Great product	5	Very nice My Apple Watch Very Good ProductREAD...
7	Vishnu PR	Classy product	5	My first Apple watch in my life. This one is a...
8	Nimmagadda Hemasri	Fabulous!	5	Unimaginably good at first I was reluctant to ...
9	SAI CHARAN VALLURI	Must buy!	5	Flawless, classy, awesome watchREAD MORE

(14) Find who gave less than 5 rating

```
df1[df1.review_stars<5]
```

	customers_Name	Review_Comment	review_stars	review
0	Akshay Jambhale	Really Nice	4	Good for An entry Level APPLE WATCH..... Does ...

<u>CONCLUSION</u>

The web scraping process for the **Apple Watch** Flipkart product page was successful, The extracted data can be used for further analysis and revealed customer sentiments, and product popularity metrics. Regular monitoring and maintenance are recommended to adapt to any changes in the website structure.