Web scraping project of iWatch from Flipkart

Institution Name:- ITvedant Education Pvt.Ltd

Name:- Anjali Ashutosh Tripathi

Email address:- anjaliixyz@gmail.com

Date of Submission:- 30th Nov 2023

AIM OF PROJECT

The aim of a web scraping project focused on the iWatch could be to gather data from flipkart website selling iWatch product. This could include extracting information such as prices, features, customer reviews, availability, and specifications of iWatch model from multiple online retailers or flipkart website. The gathered data could be used for market analysis, price comparison, trend tracking, or to provide insights forpotential buyers or businesses interested in the iWatch market.

DESCRIPTION

This Python-based web scraping project focuses on collecting structured data from Flipkart's website, specifically targeting the iwatch product details. The script employs Requests to fetch the HTML content of the product page and BeautifulSoup to parse and extract relevant information.

The data extracted includes:

- ➢ Product Details: Name and price of the iwatch.
- ➢ Customer Review: Capturing customer names, ratings, reviews, and comments.
- ➢ Product Popularity: Analyzing the number of likes and dislikes for the product.
- ➢ The extracted data is then structured into a Pandas DataFrame and exported to a CSV file for further analysis and insights.

This database contains 4 tables:

- ➢ 1.CName
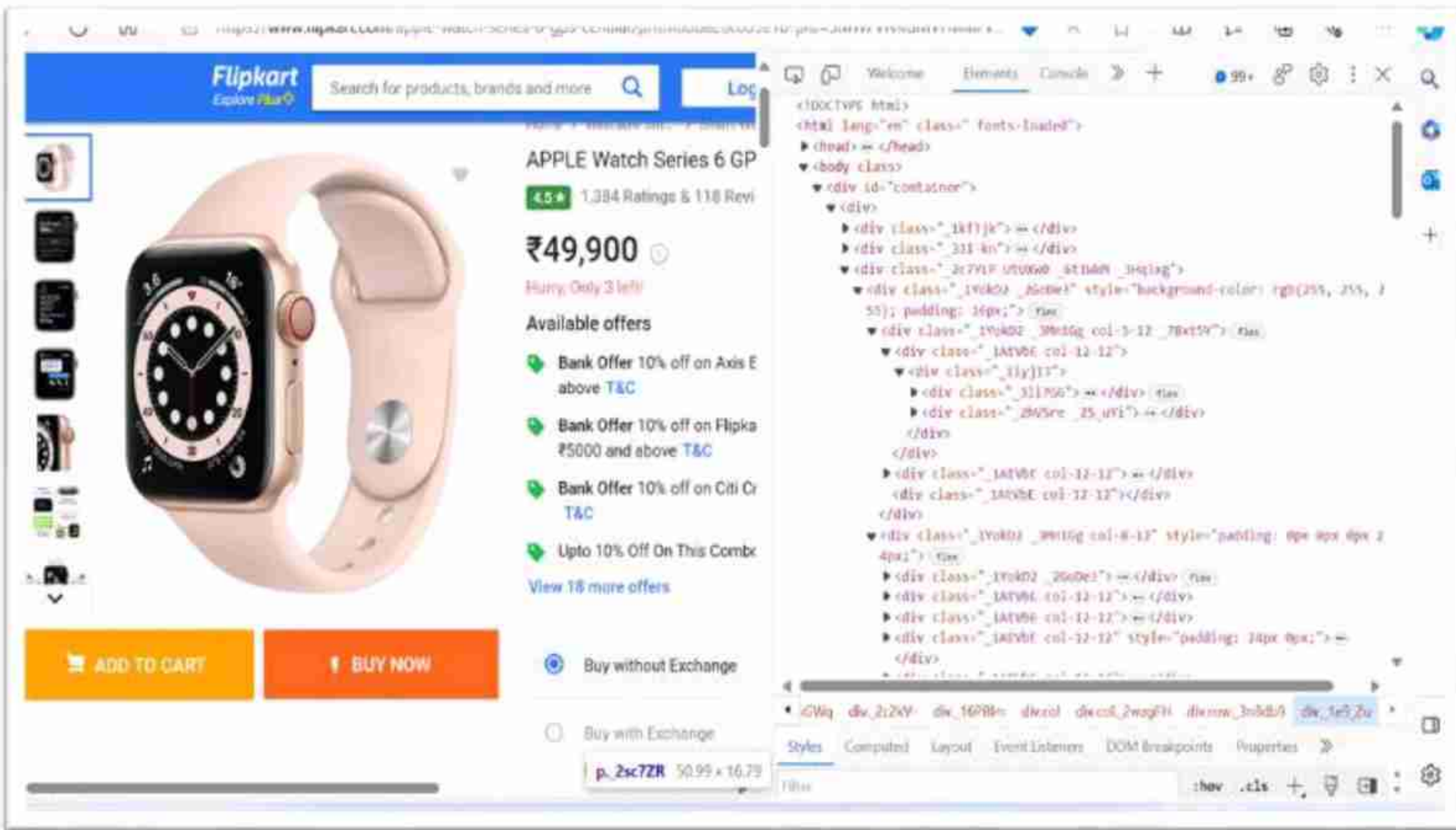- ➢ 2.Reveiw
- ➢ 3.Star
- ➢ 4.Comment

STEPS OF PROJECT:-

1 :- Choose website and web URL

The first step is to select the website you want to scrap. We will try to extract data of smart watch of apple on the website.

Inspect the website

The next step is to understand the website data. Try to understand the structure of the data given on website. Write click on the product and click on "INSPECT" to get the HTML program of the product.We will use the inspector tool to see the name and element of the watch.

Installing the liabraries

Python has several web scraping libraries. We will use the following libraries.

- ➤ **Requests-** for making HTTP requests to website
- ➤ **BeautifulSoup:-** for parsing the HTML code.
- ➤ **Pandas:-**fro storing the scraped data frame

Write the python source code

We'll write the main python code. The code will perform the following steps:-

- ➤ Using requests to send HTTP GET requests
- ➤ Using BeautifulSoup to parse the HTML code
- ➤ Extracting the required data from the HRML code
- ➤ Store the information in pandas dataFrame

Exporting the extracted data

The extracted data will be then structured into pandas DataFrame and exported as CSV file for further analysis and insights.
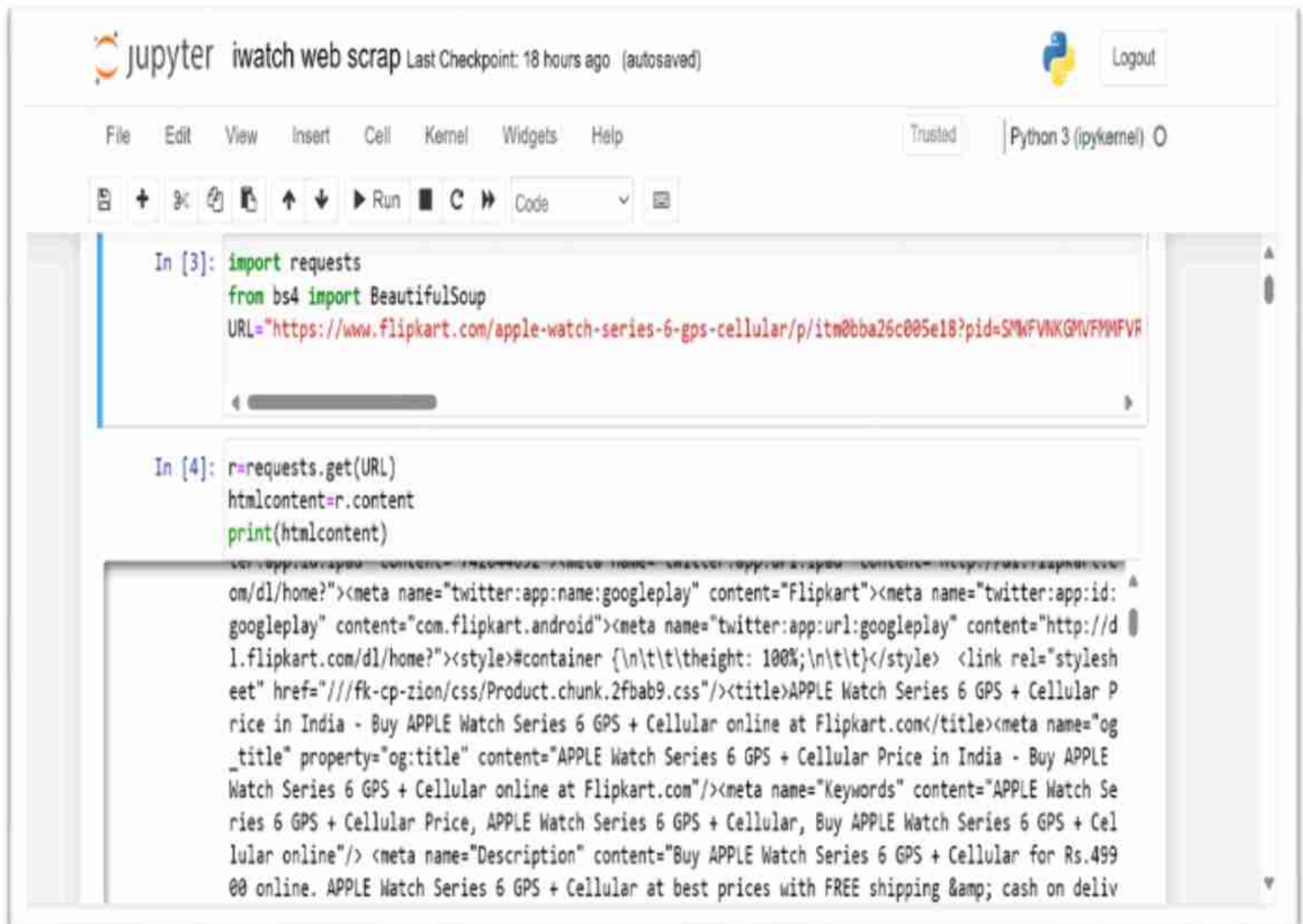
Benefits of data scraping

➢ **Data Collection**: It enables the extraction of structured data from websites, automating the process of gathering information, which can be used for analysis, research, or business intelligence.

➢ **Competitive Analysis**: Helps in monitoring competitors' pricing, product details, marketing strategies, and customer reviews, aiding businesses in making informed decisions and staying competitive.

➢ **Market Research**: Facilitates the collection of market trends, customer sentiments, and demographic information, assisting businesses in understanding consumer behavior and preferences.

➢ **Lead Generation:** Extracts contact information, job postings, or potential sales leads from various websites, aiding sales and marketing efforts.

➢ **Financial Analysis:** Provides access to real-time financial data, stock prices, market indices, and economic indicators crucial for investment decisions and financial modeling

Risk involve in web scraping

➢ **Legal issues related to web scraping**
➢ **Technical challenges due to website changes**

1. Accessing flipkart website by using import requests URL and BeautifulSoup

In [3]: import requests
from bs4 import BeautifulSoup
URL="https://www.flipkart.com/apple-watch-series-6-gps-cellular/p/itm0bba26c005e18?pid=SMWFVNKGMVFMMFVR

In [4]: r=requests.get(URL)
htmlcontent=r.content
print(htmlcontent)

om/dl/home?"><meta name="twitter:app:name:googleplay" content="Flipkart"><meta name="twitter:app:id:
googleplay" content="com.flipkart.android"><meta name="twitter:app:url:googleplay" content="http://d
l.flipkart.com/dl/home?"><style>#container {\n\t\t\theight: 100%;\n\t\t}</style> <link rel="stylesh
eet" href="///fk-cp-zion/css/Product.chunk.2fbab9.css"/><title>APPLE Watch Series 6 GPS + Cellular P
rice in India - Buy APPLE Watch Series 6 GPS + Cellular online at Flipkart.com</title><meta name="og
_title" property="og:title" content="APPLE Watch Series 6 GPS + Cellular Price in India - Buy APPLE
Watch Series 6 GPS + Cellular online at Flipkart.com"/><meta name="Keywords" content="APPLE Watch Se
ries 6 GPS + Cellular Price, APPLE Watch Series 6 GPS + Cellular, Buy APPLE Watch Series 6 GPS + Cel
lular online"/> <meta name="Description" content="Buy APPLE Watch Series 6 GPS + Cellular for Rs.499
00 online. APPLE Watch Series 6 GPS + Cellular at best prices with FREE shipping & cash on deliv

PARSE THE HTML

File   Edit   View   Insert   Cell   Kernel   Widgets   Help   Trusted   Python 3 (ipykernel) O

```python
In [5]: soup=BeautifulSoup(htmlcontent,"html.parser")
        print(soup)
        print(soup.prettify)
```

File   Edit   View   Insert   Cell   Kernel   Widgets   Help   Trusted   Python 3 (ipykernel) O

```html
<!DOCTYPE html>
<html lang="en"><head><link href="https://rukminim2.flixcart.com" rel="preconnect"/><link href="//st
atic-assets-web.flixcart.com/fk-p-linchpin-web/fk-cp-zion/css/app_modules.chunk.905c37.css" rel="sty
lesheet"/><link href="//static-assets-web.flixcart.com/fk-p-linchpin-web/fk-cp-zion/css/app.chunk.61
5ed9.css" rel="stylesheet"/><meta content="text/html; charset=utf-8" http-equiv="Content-type"/><met
a 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" type="application/opensearchdescription+xml"/><meta c
ontent="website" property="og:type"/><meta content="Flipkart.com" name="og_site_name" property="og:s
ite_name"/><link href="/apple-touch-icon-57x57.png" rel="apple-touch-icon" sizes="57x57"/><link href
="/apple-touch-icon-72x72.png" rel="apple-touch-icon" sizes="72x72"/><link href="/apple-touch-icon-1
14x114.png" rel="apple-touch-icon" sizes="114x114"/><link href="/apple-touch-icon-144x144.png" rel
="apple-touch-icon" sizes="144x144"/><link href="/apple-touch-icon-57x57.png" rel="apple-touch-ico
n"/><meta content="app" name="twitter:card"/><meta content="@flipkart" name="twitter:site"/><meta co
ntent="@flipkart" name="twitter:creator"/><meta content="APPLE Watch Series 6 GPS + Cellular Price i
n India - Buy APPLE Watch Series 6 GPS + Cellular online at Flipkart.com" name="twitter:title"/><met
a content="Shop for electronics, apparels &amp; more using our Flipkart app Free shipping &amp; CO
D." name="twitter:description"/><meta content="http://rukninil.flixcart.com/image/300/300/kfeamq80/s
```

Access the price of the product using source code:-

```python
In [6]: price=soup.find_all(class_="_30jeq3 _16Jk6d")
        price
```

```
Out[6]: [<div class="_30jeq3 _16Jk6d">₹49,900</div>]
```

```python
In [7]: price[0].get_text()
```

OUTPUT:-

```
In [?]:  print[?].get_text()
```

Out[7]: '₹49,900'

To access iWatch customer customer name using class data:-

```
In [11]:  name=soup.find_all(class_="_2sc7ZR _2V5EHH")
          name=list(name)
          print(type(name))
          print(name)
          len(name)

          <class 'list'>
          [<p class="_2sc7ZR _2V5EHH">Soubhagya Ranjan </p>, <p class="_2sc7ZR _2V5EHH">Raj Kiran Saraswat</p>,
          <p class="_2sc7ZR _2V5EHH">Rockey Yadav</p>, <p class="_2sc7ZR _2V5EHH">Yogesh Sharma</p>, <p class
          ="_2sc7ZR _2V5EHH">Shashikiran S</p>, <p class="_2sc7ZR _2V5EHH">Flipkart Customer</p>, <p class="_2s
          c7ZR _2V5EHH">Ratnesh Kumar</p>, <p class="_2sc7ZR _2V5EHH">Yaman  Patidar</p>, <p class="_2sc7ZR _2V
          5EHH">Rajendra  Mali</p>, <p class="_2sc7ZR _2V5EHH">Amit  Hingne</p>]

Out[11]:  10
```

```
In [12]:  cust_name=[]
          for i in range(0,len(name)):
              cust_name.append(name[i].get_text())
          cust_name

Out[12]:  ['Soubhagya Ranjan ',
           'Raj Kiran Saraswat',
           'Rockey Yadav',
           'Yogesh Sharma',
           'Shashikiran S',
           'Flipkart Customer',
           'Ratnesh Kumar',
           'Yaman  Patidar',
           'Rajendra  Mali',
           'Amit  Hingne']
```

To access iWatch customer reviews by using this class data

```
In [13]: comment=soup.find_all(class_="t-ZTKy")
         comment
```

```
Out[13]: [<div class="t-ZTKy"><div><div class="">The beautiful watch I ever found ♥ </div><span class="_1H-bm
         y"><span>READ MORE</span></span></div></div>,
          <div class="t-ZTKy"><div><div class="">Red color looks awesome, best amongst all other available col
         ors, it looks trendy and decent at the same time.<br>Awesome delivery and customer support from flipk
         art. I recommend buying from flipkart other e commerce websites are flop in front of flipkart</br></d
         iv><span class="_1H-bmy"><span>READ MORE</span></span></div></div>,
          <div class="t-ZTKy"><div><div class="">Thanks Flipkart for on time delivery and thanks Apple for man
         ufacturing such an awesome watch as usual ! I love navy blue , waiting for this from a long time and
         can't wait to see the same colour for iPhone 12.</div><span class="_1H-bmy"><span>READ MORE</span></s
         pan></div></div>,
          <div class="t-ZTKy"><div><div class="">This is my first smart watch. Earlier I used mi band for acti
         vity tracking.<br/>I'm writing this review after 24 hours of usages. <br/>I received item in good pac
         kaging. Watch came in complete discharge state. So first I charged it full which took almost 1 hour 1
         0 minutes with 18w charger. <br/>Setup process was fairly easy &amp; user friendly. For complete sync
         from iPhone, it took approx 15min but it'll depend on users data. <br/>The watch is amazing with some
         great functional watch faces. Cellular connectivit...</div><span class="_1BWGvX"><span>READ MORE</spa
         n></span></div></div>.
```

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

```
Out[14]: ['The beautiful watch I ever found ♥ READ MORE',
          'Red color looks awesome, best amongst all other available colors, it looks trendy and decent at the
         same time.Awesome delivery and customer support from flipkart. I recommend buying from flipkart other
         e commerce websites are flop in front of flipkartREAD MORE',
          'Thanks Flipkart for on time delivery and thanks Apple for manufacturing such an awesome watch as us
         ual ! I love navy blue , waiting for this from a long time and can't wait to see the same colour for
         iPhone 12.READ MORE',
          'This is my first smart watch. Earlier I used mi band for activity tracking.I'm writing this review
         after 24 hours of usages. I received item in good packaging. Watch came in complete discharge state.
         So first I charged it full which took almost 1 hour 10 minutes with 18w charger. Setup process was fa
         irly easy & user friendly. For complete sync from iPhone, it took approx 15min but it'll depend on us
         ers data. The watch is amazing with some great functional watch faces. Cellular connectivit...READ MO
         RE',
          'The product and the package was Awesome!! Flipkart finally delivered it ON TIME as the promise they
         made. Special thanks to Apple for this great creature and it's simplicity 😊 READ MORE',
```

```
In [18]: Content_Review = []
         for i in range(0,len(Review)):
          Content_Review.append(Review[i].get_text())
         Content_Review

Out[18]: ['Worth every penny',
          'Perfect product!',
          'Wonderful',
          'Highly recommended',
          'Brilliant',
          'Terrific purchase',
          'Perfect product!',
          'Simply awesome',
          'Highly recommended',
          'Pretty good']
```

Access iWatch To review stars using this class

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

Out[16]: ['5', '5', '5', '5', '5', '5', '5', '5', '5', '4']
```

To arrange properly in column all above data by using import Pandas library


```
In [29]: import pandas as pd
         df = pd.DataFrame({"Customer's Name":cust_name,"content_review":comment_review,"Review_Stars":Review_Stars,"Review":Content_Revie
         print(df)

             Customer's Name                               content_review  \
        101  Soubhagya Ranjan      The beautiful watch I ever found ♥READ MORE
        102  Raj Kiran Saraswat  Red color looks awesome, best amongst all othe...
        103       Rockey Yadav   Thanks flipkart for on time delivery and thank...
        104      Yogesh Sharma   This is my first smart watch. Earlier I used m...
        105    Vhashikiran S     The product and the package was Awesome!! flip...
        106  Flipkart Customer                                  SupperREAD MORE
        107     Ratnesh Kumar   I entered in the Apple Eco-system back in 2017,...
        108      Yamen Patidar             Thanks flipkart 💙💙💙💙READ MORE
        109     Rajendra Mali                Awesome iWatchREAD MORE
        110      Amit Mingne    What a beast . ♥ ♥ ♥ loved itREAD MORE

            Review_Stars             Review
        101            5    Worth every penny
        102            5      Perfect product!
        103            5            Wonderful
        104            5    Highly recommended
        105            5            Brilliant
        106            5    Terrific purchase
        107            5      Perfect product!
        108            5        Simply awesome
        109            5    Highly recommended
        110            4          Pretty good
```

To see quick overview of central tendency ,spread ,and distribution of the Data DataFramThen store all the above data in CSV file and again open this file in application.

```
In [30]: df.to_csv("flipkart_iwatch_review.csv")
         p=pd.read_csv("flipkart_iwatch_review.csv")
         p
```

Out[30]:

	Unnamed: 0	Customer's Name	content_review	Review_Stars	Review
0	101	Soubhagya Ranjan	The beautiful watch I ever found ❣️READ MORE	5	Worth every penny
1	102	Raj Kiran Saraswat	Red color looks awesome, best amongst all othe...	5	Perfect product!
2	103	Rockey Yadav	Thanks Flipkart for on time delivery and thank...	5	Wonderful
3	104	Yogesh Sharma	This is my first smart watch. Earlier I used m...	5	Highly recommended
4	105	Shashikiran S	The product and the package was Awesome!! Flip...	5	Brilliant
5	106	Flipkart Customer	SupperREAD MORE	5	Terrific purchase
6	107	Ratnesh Kumar	I entered in the Apple Eco-system back in 2017...	5	Perfect product!
7	108	Yaman Patidar	Thanks Flipkart 😍👍👍👍READ MORE	5	Simply awesome
8	109	Rajendra Mali	Awesome iWatchREAD MORE	5	Highly recommended
9	110	Amit Hingne	What a beast .❤️❤️❤️ loved itREAD MORE	4	Pretty good

Conclusion:-

➢ Analyze Results: Review the scraped data for accuracy, consistency, and relevance to your project goals.
➢ Clean Data: Remove duplicates, handle missing values, and format data appropriately for further analysis.
➢ Visualization and Reporting: Create visualizations or reports to present findings effectively, aiding comprehension and decision-making.
➢ Documentation: Document your process, including the tools used, code snippets, and any challenges faced, for future reference or sharing.
➢ Compliance and Ethics: Ensure adherence to ethical guidelines and legal compliance while scraping data, respecting website terms of service and privacy policies.