

# PROJECT FOR WEB SCRAPING

## **Web Scrapping of Apple Macbook From Flipkart**

**Institute Name:- ITvedant Education Pvt.Ltd**

**Name:- Shreyash Narendra Sakpal**

**Email Address:- shreyashsakpal19@gmail.com**

**Date Of Submission:-03 Dec 2023**

## AIM OF THE PROJECT

The aim of this project is to create a Python Web scraping to extract specific details of the Apple Macbook 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 Macbook listings



## Description:-

Web scraping of Flipkart products involves programmatically extracting information about various products listed on the Flipkart e-commerce platform. This process allows users to gather data such as product names, prices, ratings and reviews available on Flipkart's product pages.

To perform web scraping on Flipkart, you'd typically utilize tools, programming languages, or libraries like Python's BeautifulSoup. These tools can simulate human browsing behavior, sending requests to Flipkart's servers, retrieving the HTML content of product pages, and then parsing the HTML to extract the desired product information.

## Project Steps :-

- **URL Retrieval:** You've defined the URL for the Apple Macbook 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:-

Market Analysis and Pricing Strategy: Extracting product details, pricing information, discounts, and stock availability helps in analyzing market trends, understanding competitor pricing strategies, and adjusting your own pricing strategy accordingly.

Competitor Monitoring: Regularly scraping Flipkart's product listings allows for tracking competitors' product ranges, availability, and pricing changes. This information helps in identifying gaps in the market or potential niches to explore.

Product Research and Development: Scraping Flipkart can provide insights into customer reviews, ratings, and feedback on products. This data aids in understanding customer preferences, identifying popular features, and guiding product development decisions.

Sales and Marketing Insights: Extracting data on product descriptions, specifications, and customer reviews can guide marketing campaigns, content creation, and SEO strategies by understanding what features or aspects customers value the most.

Price Comparison and Adjustment: Scraping Flipkart allows for comparing prices of similar products across different vendors. This enables businesses to adjust their pricing strategies to stay competitive in the market.

Consumer Behavior Analysis: Analyzing data from Flipkart's product reviews and ratings can provide valuable insights into consumer behavior, preferences, and sentiments toward certain products or brands.

Personalized Customer Experience: Insights gained from scraping Flipkart can be used to tailor product offerings, customer service, and shopping experiences to meet specific customer needs and preferences.

## Output Format:

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

## Analysis of project:

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

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

```
In [11]: 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 Macbook price by using this class.

```
In [12]: title = soup.title
         title

Out[12]: <title>Apple 2022 Macbook Air M2 8 Gb 256 Gb Ssd Mac Os Monterey Mly33hn A Reviews: Latest Review of Apple 2022 Macbook Air M2
8 Gb 256 Gb Ssd Mac Os Monterey Mly33hn A | Price in India | Flipkart.com</title>

In [13]: price = soup.find_all(class_="_30jeq3")
         price

Out[13]: [<div class="_30jeq3">₹1,04,990</div>]

In [14]: price[0].get_text()

Out[14]: '₹1,04,990'
```

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

```
In [15]: names = soup.find_all(class_="_2sc7ZR _2V5EHH")
         names = list(names)
         print(type(names))
         print(names)
         len(names)

Out[15]: 10

In [16]: names[0].get_text()

Out[16]: 'Piyush Khare'

In [17]: c_names = []
         for i in names:
             c_names.append(i.get_text())
         c_names

Out[17]: ['Piyush Khare',
          'Vinu Viswan',
          'Mohit Anand',
          'Yash Patil',
          'Riya Singh',
          'Hariprasad Mandal',
          'Pratyush Sahu',
          'Shatakshi Pathak',
          'kartik charande',
          'Rahul Anand']
```

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

```
In [18]: comment = soup.find_all(class_="_2-N8zT")
         print(len(comment))

10

In [19]: review_comment = []
         for i in comment:
             review_comment.append(i.get_text())
         review_comment

Out[19]: ['Perfect product!',
          'Super!',
          'Awesome',
          'Mind-blowing purchase',
          'Fabulous!',
          'Just wow!',
          'Brilliant',
          'Good',
          'Worth every penny',
          'Highly recommended']
```



**(5)To access Apple Macbook Customer Review star by using the respective class.**

```
In [36]: stars=soup.find_all('div', class_="_3LWZlK _1BLPMq")
print(stars)
```

<div class=

```
In [21]: s_stars = []
         for i in stars:
             s_stars.append(i.get_text())
         s_stars
```

```
Out[21]: ['5', '5', '5', '5', '5', '5', '5', '3', '5', '5']
```

**(6)To access Apple Macbook Customer Comment by using the respective class.**

```
In [37]: review = soup.find_all('div',class_ ="t-ZTKy")
print(review)
```

<div class="t-ZTKy"><div>div class="">I loved it, it's my first MacBook that I earned from my hardwork 🍀❤️</div><span class="">"\_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div>div class="">Super product</div><span class="">"\_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div>div class="">When i ordered and came to know about seller review,i was scared that in what condition i will get macbook. but i was wrong.<br/><br/>This is my first macbook and i am fully satisfied with the product with original packaging of apple and got genuine product.<br/><br/>It's very thin and light and looks very premium.<br/><br/>As A software developer, i installed intellij idea ide and guess what.<br/>It's magic,it is opening just in a click and within 5 seconds.<br/>Great performance, no heating issue, sound quality awesome...</div><span class="">"\_1BWGVX"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div>div class="">just love how apple is in their own league!</div><span class="">"\_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div>div class="">Such a great deal.. very happy with the performance and battery life..Original product..🍀❤️<br/>Thank you Flipkart..!!!</div><span class="">"\_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div>div class="">Its really very good and compact device.</div><span class="">"\_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div>div class="">Its very good product.</div><span class="">"\_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div>div class="">Note-<br/>It's not for everyone.<br/>You can't attach dual monitor setup easily<br/>You need to buy 3rd party adpoter at y our risk. I talked to Apple their port support only mirroring, not extended screen feature.<br/><br/>Related to Ms Excel, normal excel will work<br/>But will not be able to run high macro based files.<br/><br/>For port and accessories, there is heating issues, hence be ready for another 7000 for display adpoter and 2000-5000 for usb hub<br/><br/>Rest UI and using experience is v ery smooth, unbelievable when i ...</div><span class="">"\_1BWGVX"><span>READ MORE</span></span></div></div>, <div class="t-ZTKy"><div>div class="">Worth every penny u spend</div><span class="">"\_1H-bmy"><span>READ MORE</span></span></div></div>, <div class="">"t-ZTKy"><div>div class="">This is the first MacBook I ever have. And I must say no one can match Apple product.</div><span class="">"\_1H-bmy"><span>READ MORE</span></span></div></div>



```
In [23]: cust_review = []
         for i in review:
             cust_review.append(i.get_text())
         cust_review

Out[23]: ["Loved it, it's my first MacBook that I earned from my hardwork 🥳❤️ READ MORE",
          'Super productREAD MORE',
          "When i ordered and came to know about seller review,i was scared that in what condition i will get macbook. but i was wrong.T his is my first macbook and i am fully satisfied with the product with original packaging of apple and got genuine product.It's very thin and light and looks very premium.As A software developer, i installed intelliJ idea ide and guess what.It's magic,it is opening just in a click and within 5 seconds.Great performance, no heating issue, sound quality awesome an...READ MORE",
          'just love how apple is in their own league!!READ MORE',
          'Such a great deal.. very happy with the performance and battery life..Original product..🥳❤️Thank you Flipkart!!!!READ MORE',
          'Its really very good and compact device.READ MORE',
          'Its very good product.READ MORE',
          "Note--It's not for everyone.You can't attach dual monitor setup easilyYou need to buy 3rd party adopter at your risk. I talked to Apple their port support only mirroring, not extended screen feature.Related to Ms Excel, normal excel will workBut will not be able to run high macro based files.For port and accessories, there is heating issues, hence be ready for another 7000 for display adopter and 2000-5000 for usb hubRest UI and using experience is very smooth, unbelievable when i c...READ MORE",
          'Worth every penny u spendREAD MORE',
          'This is the first MacBook I ever have. And I must say no one can match Apple product.READ MORE']
```

## (7) To arrange properly in column all above data by using -Pandas library

```
In [24]: import pandas as pd

In [25]: df= pd.DataFrame({"customers_Name":c_names,"Review_Comment":review_comment,"review_stars":s_stars,"review":cust_review})
         df

Out[25]:
```

	customers_Name	Review_Comment	review_stars	review
0	Piyush Khare	Perfect product!	5	Loved it, it's my first MacBook that I earned ...
1	Vinu Viswan	Super!	5	Super productREAD MORE
2	Mohit Anand	Awesome	5	When i ordered and came to know about seller r...
3	Yash Patil	Mind-blowing purchase	5	just love how apple is in their own league!!RE...
4	Riya Singh	Fabulous!	5	Such a great deal.. very happy with the perfor...
5	Hariprasad Mandal	Just wow!	5	Its really very good and compact device.READ MORE
6	Pratyush Sahu	Brilliant	5	Its very good product.READ MORE
7	Shatakshi Pathak	Good	3	Note--It's not for everyone.You can't attach d...
8	kartik charande	Worth every penny	5	Worth every penny u spendREAD MORE
9	Rahul Anand	Highly recommended	5	This is the first MacBook I ever have. And I m...

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

```
In [28]: df.to_csv("apple_macbook.csv")

In [29]: import os
         os.getcwd()

Out[29]: <function nt.getcwd()>

In [30]: dfi.describe()

Out[30]:
```

	review_stars
count	10.000000
mean	4.800000
std	0.632456
min	3.000000
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.**

```
In [26]: import os  
os.getcwd()
```

```
Out[26]: 'C:\\Users\\HP'
```

```
In [27]: df.to_csv("apple_macbook.csv",index= False )  
df1= pd.read_csv("apple_macbook.csv")  
df1
```

```
Out[27]:
```

	customers_Name	Review_Comment	review_stars	review
0	Piyush Khare	Perfect product!	5	Loved it, it's my first MacBook that I earned ...
1	Vinu Viswan	Super!	5	Super productREAD MORE
2	Mohit Anand	Awesome	5	When i ordered and came to know about seller r...
3	Yash Patil	Mind-blowing purchase	5	just love how apple is in their own league!!RE...
4	Riya Singh	Fabulous!	5	Such a great deal.. very happy with the perfor...
5	Hariprasad Mandal	Just wow!	5	Its really very good and compact device.READ MORE
6	Pratyush Sahu	Brilliant	5	Its very good product.READ MORE
7	Shatakshi Pathak	Good	3	Note--It's not for everyone.You can't attach d...
8	kartik charande	Worth every penny	5	Worth every penny u spendREAD MORE
9	Rahul Anand	Highly recommended	5	This is the first MacBook I ever have. And I m...

**(10) Find max,mean,min etc**

```
In [30]: df1.describe()
```

```
Out[30]:
```

	review_stars
count	10.000000
mean	4.800000
std	0.632456
min	3.000000
25%	5.000000
50%	5.000000
75%	5.000000
max	5.000000

**(11) Find INFO of data**

```
In [31]: 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
```

### (13) sort the name of Customer in Ascending order

```
In [32]: df1.customers_Name.sort_values()
```

```
Out[32]: 5    Hariprasad Mandal
         2      Mohit Anand
         0      Piyush Khare
         6    Pratyush Sahu
         9      Rahul Anand
         4      Riya Singh
         7    Shatakshi Pathak
         1      Vinu Viswan
         3      Yash Patil
         8    kartik charande
         Name: customers_Name, dtype: object
```

### (14) Sort the Names of Customer in Descending order

```
In [33]: df1.customers_Name.sort_values(ascending=False)
```

```
Out[33]: 8    kartik charande
         3      Yash Patil
         1      Vinu Viswan
         7    Shatakshi Pathak
         4      Riya Singh
         9      Rahul Anand
         6    Pratyush Sahu
         0      Piyush Khare
         2      Mohit Anand
         5    Hariprasad Mandal
         Name: customers_Name, dtype: object
```

### (15) Find who had given full rating

```
In [34]: df1[df1.review_stars==5]
```

```
Out[34]:
```

	customers_Name	Review_Comment	review_stars	review
0	Piyush Khare	Perfect product!	5	Loved it, it's my first MacBook that I earned ...
1	Vinu Viswan	Super!	5	Super productREAD MORE
2	Mohit Anand	Awesome	5	When i ordered and came to know about seller r...
3	Yash Patil	Mind-blowing purchase	5	just love how apple is in their own league!!RE...
4	Riya Singh	Fabulous!	5	Such a great deal.. very happy with the perfor...
5	Hariprasad Mandal	Just wow!	5	Its really very good and compact device.READ MORE
6	Pratyush Sahu	Brilliant	5	Its very good product.READ MORE
8	kartik charande	Worth every penny	5	Worth every penny u spendREAD MORE
9	Rahul Anand	Highly recommended	5	This is the first MacBook I ever have. And I m...

## (16) Find who gave less than 5 rating

```
In [35]: df1[df1.review_stars<5]
```

```
Out[35]:
```

	customers_Name	Review_Comment	review_stars	review
7	Shatakshi Pathak	Good	3	Note--It's not for everyone.You can't attach d...

## CONCLUSION

The web scraping process for the **Apple Macbook** 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.