

NAME: POOJA MAHENDRA MALUSARE

PROJECT NAME: TAB Web Scraping

Project Guide: Prof. Sameer Warsolkar

1.Aim:

The aim of a web scraping project focused on tablet phones is to automatically extract detailed information about various tablet phone models from multiple websites. This includes specifications, prices, reviews, and availability. The goal is to compile this data into a structured format for comparison, market analysis, and consumer insights, ensuring legal and ethical compliance throughout the process.

2.Objectives:

Data Extraction:

 Automate the retrieval of detailed specifications, prices, reviews, and availability of various tablet phone models from multiple e-commerce and review websites.

> Data Cleaning and Formatting:

• Clean and standardize the extracted data to ensure consistency and accuracy, converting it into a structured format like CSV, JSON, or a database.

Data Integration:

• Integrate the scraped data with existing datasets or databases to enhance the overall dataset, providing a comprehensive view of the tablet phone market.

➤ Analysis and Reporting:

 Analyze the compiled data to identify trends, price comparisons, consumer preferences, and market insights, generating reports to support decision-making.

> Automation and Maintenance:

• Develop automated scripts to regularly update the dataset with new data, ensuring the information remains current and relevant.

➤ Compliance and Ethics:

• Ensure that the web scraping process adheres to legal and ethical standards, including respecting website terms of service and privacy policies. Implement mechanisms to handle any potential issues, such as CAPTCHAs or website blocking.

3. OUTLINE:

From this site, we are going to grab the following information:

- 1. Tab Names
- 2. E-side
- 3. shipping_cost
- 4. Tab_price

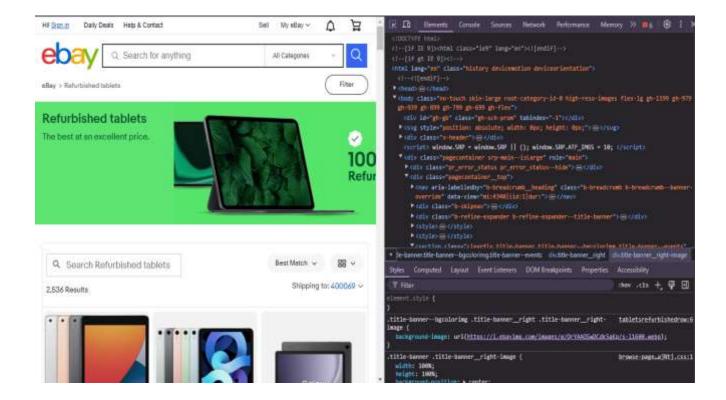
4. STEPS:

Choose the Website and Webpage URL:

➤ The first step is to select the website you want to scrape. We will try to extract data of Tab.

1. Inspect the website:

Now the next step is to understand the website structure. Understand what the attributes of the elements that are of your interest are. Right click on the website to select "Inspect". This will open HTML code. Use the inspector tool to see the name of all the elements to use in the code.



. Installing the important libraries:

Python has several web scrapping libraries. We will use the following libraries:

- > Requests for making HTTP requests to website
- ➤ BeautifulSoup for parsing the HTML code
- > Pandas for storing the scraped data in data frame

3. Write the Python source code:

We'll write the main python code. The code will perform the

following steps:

- ➤ Using requests to send an HTTP GET requests
- ➤ Using BeautifulSoup to parse the HTML code
- > Extracting the required data from the HTML code
- > Store the information in a pandas DataFrame

4. Exporting the extracted data:

We'll export the data as a CSV file. We will use the pandas library. We'll use the pandas library.

5. Benefits:

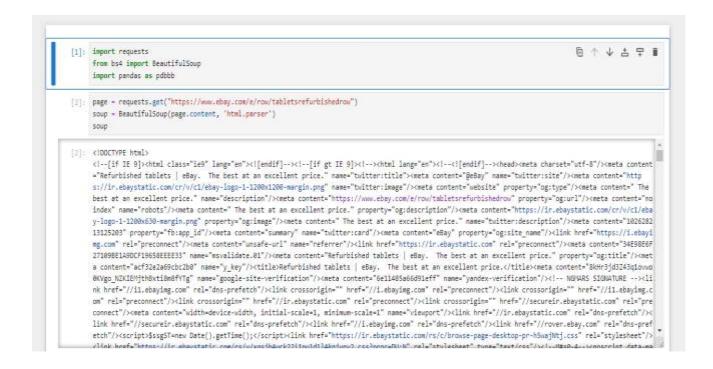
- > Access to valuable data for analysis or research.
- ➤ Automation of data collection, saving time and effort.
- > Stay up to date with change changes on the target websites.

6. Risk:

- Legal issues related to web scraping.
- > Technical challenges due to website changes.

5. Tab web scraping coding command steps

1. Accessing Tab website: .



2. Using BeautifulSoup:

```
page = requests.get("https://www.ebay.com/e/row/tabletsrefurbishedrow")
      soup = BeautifulSoup(page.content, 'html.parser')
      soup
<!--[if IE 9]><html class="ie9" lang="en"><![endif]--><!--[if gt IE 9]><!--><html lang="en"><!--<![endif]--><head>cmeta charset="utf-8"/><meta content
      -"Refurbished tablets | eBay. The best at an excellent price." name-"twitter:title"><mets content-"@eBay" name-"twitter:site"/><mets content-"http
s://ir.ehaystatic.com/cr/v/c1/ebay-logo-1-1200x1200-wargin.png" name-"twitter:image"/><meta content-"website" property-"og:type"/>/meta content-" The
      best at an excellent price," name="description"/>(meta content="https://www.ebay.com/e/row/tabletsrefurbishedron" property="og:url"/>(meta content="no
      index" name="robots"/>cmeta content=" The best at an excellent price." property="og:description"/>cmeta content="https://ir.ebaystatic.com/cr/v/cl/eba
      y-logo-1-1208x638-margin.png" property="og:image"/>cmeta content=" The best at an excellent price." name="twitter:description"/>:meta content="1026282
      13125203" property="fb:app_id"/>:meta content="summary" name="twitter:card"/>:meta content="eBay" property="og:site_name"/>:link href="https://i.ebay1
       mg.com" rel="preconnect"/>(meta content="unsafe-url" name="referrer"/>(link href="https://ir.ebaystatic.com" rel="preconnect"/>(meta content="1459856F
      271098E1A9DCF19658EEEE33" name="msvelidate.01"/>kmeta content="Refurbished tablets | eBay. The best at an excellent price." property="og:title"/>kmet
      a content="acf32e2a69cbc2b8" name="y_key"/><title>Refurbished tablets | eBay. The best at an excellent price.</title>cmeta content="889r3jd3143qlowwo 0KVgo_WZKIEMjthBxti8w8fVTg" name="google-site-verification"/><!-- NGMARS SIGMATURE --><!i
      nk href="//il.ebayimg.com" rel="dns-prefetch"/>{link crossorigin="" href="//i.ebayimg.com" rel="preconnect"/>{link crossorigin="" href="//il.ebayimg.com" rel="preconnect"/>
      om" rel="preconnect"/>klink crossorigin="" href="//ir.ebaystatic.com" rel="preconnect"/klink crossorigin="" href="//secureir.ebaystatic.com" rel="pre
      connect"/>cmeta content="width=device-width, initial-scale=1, minimum-scale=1" name="viewport"/><link href="//ir.ebaystatic.com" rel="dns-prefetch"/)
      link href-"//secureir.ebaystatic.com" rel-"dns-prefetch"/><link href-"//i.ebayimg.com" rel-"dns-prefetch"/><link href-"//rover.ebay.com" rel-"dns-pref
      etch"/><script>$ssgST=new Date().getTime();</script><link href="https://ir.ebaystatic.com/rs/c/browse-page-desktop-pr-h5wajNtj.css" rel="stylesheet"/>
      dink heafs"https://in absortation
                                                     /www.thduck924.toutd1.ldbmio
                                                                                 no recommendative rate fetuliashast tunag tavtines liket ... Head-S---
```

3. Accessing the Tab Names:

```
name_tags=sump.select(".s-time_title")
name_tags:
    new_aspend(i.get_text())
grint(news)

['Apple lPad S (8th Gen) - (2020 Model) - 3168 - Mi-Fi - A2270 - Sood', 'Apple lPad Air 4 (2010) 10.5' SAUB 25668 All Colors (WiFi or Cellular) - Good', 'Samung Galawy Tab Ab 558-XXIAN WIFI 12688 Braphite Wavy Good', 'Apple lPad Pro 11-inch (4th dam) (2022) NG Chip - 12868 - Wi-Fi - A3790 - Good', 'Ni icrosoft Surface Pro 8 13' Touch Tablet, Intel 21 15685, Sapphire', 'CTL Chromebook Touch Tablet Twill 2368 - Wi-Fi - A3790 - Good', 'Ni place 12868 12868 12868 Saug Sold Silver wiFi or Cellular', 'Googla Pixel Tablet 12868 Porcelain WiFi Excellent Condition', 'Apple lPad Fro 3 NI (2021) NI '12868 25008 31306 All Colors (WiFi or Cellular')', 'Apple lPad Fro 3 NI (2021) NI '12868 51/ver Tablet twilfie Cellular') or Cellular', 'Apple lPad Fro 3 NI (2021) NI '12868 51/ver Tablet twilfie Cellular') or Cellular', 'Good', 'Samung Balawy Tab 58 - 50 12.4' 6400 T7580 Unlocked Good', 'Septe 12868 51/ver Tablet twilfie Cellular', 'Good', 'Samung Balawy Tab 58 - 50 12.4' 8400 T7580 Unlocked Ho 5-Pen - Good', 'Samung Balawy Tab 58 - 50 12.4' 8400 T7580 Unlocked Ho 5-Pen - Good', 'Samung Balawy Tab 58 - 50 12.4' 8400 T7580 Unlocked Ho 5-Pen - Good', 'Samung Balawy Tab 58 - 50 12.4' 12008 S1000 Unlocked Work Tablet Wife Tablet Mid 15 Cellular', 'Samung Galawy Tab 58 - 50 12.4' 12008 S1000 Unlocked Work Tablet Wife Tablet Mid 15 Cellular', 'Samung Galawy Tab 58 - 50 12.4' 12008 S1000 Unlocked Work Good', 'Apple 1980 6th 9.7' 2018 Mif or window Mid 15 Cellular', 'Apple 1980 6th 9.7' 2018 Mif or window Mid 15 Cellular', 'Good', 'Apple 1980 6th 9.7' 2018 Mif or window Mid 15 Cellular', 'Apple 1980 6th 9.7' 2018 Mif or window Mid 15 Cellular', 'Apple 1980 6th 9.7' 2018 Mif or Cellula
```

12808, Md-FL 10.2" GRAY SILVES GOLD ", 'Apple 1Pad Pro (2016) 0.7" 1268 12008 25608 All Colors (MIFS or Cellular)", '2018 - Apple IPad 6th Gem 0.7" MR762 LL/A m/1268 & MIFI Only (Silver) - Good", 'Apple 1Pad 7th Gem 10.21m 1010 MIFI Only - Excellent", 'Samming Galaxy Tab A 8.4" 5M-7107U 1268 MIFI + Verizon Clean ESM Good (AWA)", 'Apple 1Pad 2 10.2" 1258 12068 Space Gray Sold Silver MiFI or Cellular - Good", 'Apple 1Pad Pro 2 (2017) 10.5" 6408 12568 51268 All Colors (MIFI or Cellular) Apple 1Pad 8 10.2" 1258 12568 Space Gray Sold Silver MiFI or Cellular - Very Good', 'Samoung Galaxy Tab A7 1760 10.4" 5408 12568 Space Gray Silver Gold MiFI or Cellular - Very Good', 'Samoung Galaxy Tab A7 1760 10.4" 5408 12568 Space Gray Silver Gold MiFI or Cellular - Excellent', 'Samoung Galaxy Tab 27 Plus 1970 12.4" 12568 Blue (MIFI) - Excellent', 'Samoung Galaxy Tab A7 1367U 8.4" Mocha Android Tablet (MIFI + Verizon) - Good', 'Apple 1Pad 8 10.5 MI (2012) 16.9" 6408 12568 All Colors (MIFI or Cellular) - Good', 'Apple 1Pad A15 5 MI (2012) 16.9" 6408 12568 All Colors (MIFI or Cellular) - Cellular - Very Good', 'Apple 1Pad 8 10.5 MIFI - SING 18.5" 12568 Space Gray (MIFI) - Good', 'Apple 1Pad 8 10.5 MIFI - 6408 12568 Gray Silver (MIFI or Cellular) - Excellent', 'Samoung Galaxy Tab 5 Tits Cellular - Silver (MIFI) - Good', 'Apple 1Pad 8 10.21 10.1" 6408 12568 All Colors (MIFI or Cellular) - Excellent', 'Samoung Galaxy Tab 5 Tits Cellular - Silver (MIFI) - Good', 'Apple 1Pad 8 10.21 10.1" 6408 12568 All Colors (MIFI or Cellular) - Silver (MIFI) - Good', 'Apple 1Pad 8 10.21 10.1" 6408 12568 Gray Silver (MIFI) - Cellular - Excellent', 'Apple 1Pad 8 10.21 10.1" 6408 12568 All Colors (MIFI or Cellular) - Silver (MIFI or Cellular) - Silve

[b] Isn(mass)

4. Accessing the Tab_Prices:

5. Accessing the Tab shipping:

```
百个少古里面
[8]: shipping tags = soup.select(".s-item_shipping.s-item_logisticsCost")
      shipping = |
      for tag in shipping tags:
        text = tag.get_text()
         cleaned_text = text.replace('$', '').strip() # Remove the dollar symbol and extra whitespace
         shipping.append(cleaned_text)
     print("shipping:", shipping)
     shipping: ['38.41 shipping', '34.86 shipping', '66.43 shipping', '79.38 shipping', '91.58 shipping', '26.48 shipping', '28.64 shipping', '32.18 shippin
      g', '41.41 shipping', '28.64 shipping', '28.37 shipping', '41.06 shipping', '31.17 shipping', '36.17 shipping', '28.94 shipping', '38.08 shipping', '86.4
      5 shipping", '64.79 shipping', '65.25 shipping', '35.22 shipping', '28.49 shipping', '59.07 shipping', '31.59 shipping', '30.85 shipping', '36.94 shippin
      g', '29.22 shipping', '87.22 shipping', '28.76 shipping', '34.83 shipping', '32.47 shipping', '31.88 shipping', '76.32 shipping', '28.93 shipping', '38.3
      7 shipping', '35.71 shipping', '42.44 shipping', '65.99 shipping', '67.02 shipping', '67.31 shipping', '28.93 shipping', '59.50 shipping', '29.20 shippin
      g", '28.59 shipping', '30.12 shipping', '27.12 shipping', '29.99 shipping', '30.32 shipping', '31.83 shipping', '28.31 shipping', '30.99 shipping', '31.1
      7 shipping', '35.84 shipping', '27.25 shipping', '38.28 shipping', '38.90 shipping', '34.92 shipping', '38.32 shipping', '31.23 shipping', '33.86 shippin
      g', '32.44 shipping']
[9]: len(shipping)
[9]: 60
```

6. Accessing the Tab_E-Side :-

```
cdoy_Refurbished_tags=soup.select(".s-ites_certified-refurbished.s-ites_certified-refurbished-istarge")

eday_Refurbished_append(i.gst_tast())

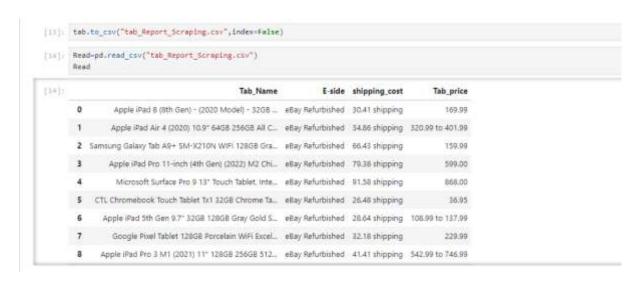
print(eBay_Refurbished)

['eBay_Refurbished', 'eBay_Refurbished', 'eBay_Refurbishe
```

7. Importing the pandas and creating DataFrame

```
[12]: import pandas as pd
       tab = pd.DataFrame({
           "Tab_Name": name,
            "E-side": eBay_Refurbished,
            "shipping_cost": shipping,
            "Tab_price":price
       tab
                                                   Tab Name
                                                                         E-side shipping_cost
                                                                                                      Tab_price
                  Apple iPad 8 (8th Gen) - (2020 Model) - 32GB ... eBay Refurbished 30.41 shipping
                                                                                                         169.99
                 Apple iPad Air 4 (2020) 10.9" 64GB 256GB All C... eBay Refurbished 34.86 shipping 320.99 to 401.99
        2 Samsung Galaxy Tab A9+ SM-X210N WIFI 128GB Gra... eBay Refurbished 66.43 shipping
                                                                                                         159.99
                Apple iPad Pro 11-inch (4th Gen) (2022) M2 Chi... eBay Refurbished 79.38 shipping
                                                                                                         599.00
                  Microsoft Surface Pro 9 13" Touch Tablet, Inte... eBay Refurbished 91.58 shipping
                                                                                                         868.00
            CTL Chromebook Touch Tablet Tx1 32GB Chrome Ta... eBay Refurbished 26.48 shipping
                                                                                                          36.95
               Apple iPad 5th Gen 9.7" 32GB 128GB Gray Gold S... eBay Refurbished 28.64 shipping 108.99 to 137.99
                 Google Pixel Tablet 128GB Porcelain WiFi Excel... eBay Refurbished 32.18 shipping
                                                                                                         229.99
        8
              Apple iPad Pro 3 M1 (2021) 11" 128GB 256GB 512... eBay Refurbished 41.41 shipping 542.99 to 746.99
```

8. Converting and Storing DataFrame in the form of a CSV file and opening the file in application:



6.Conclusion

In conclusion, a web scraping project focused on tablet phones aims to systematically collect detailed and up-to-date information on various models from multiple online sources. By automating the extraction, cleaning, and integration of data, the project ensures a comprehensive and structured dataset that can be used for market analysis, price comparison, and consumer insights. Through adherence to legal and ethical standards, the project not only enhances data collection efficiency but also supports informed decision-making in the tablet phone market.

THANK YOU