CSE-3024 Web Mining

Lab Assignment 4

Alokam Nikhitha

19BCE2555

Question

Experiment 4 (28.01.2022)

1. Use BeautifulSoup or Scrapy to crawl any one of the E-commerce website of your choice and perform the same. The following information needs to be extracted from the page: (Choose any one product : e.g laptop, smartphone … etc)

a) Product Name

b) Product price

c) Product discount

d) Product image

Problem statement:

To Crawl any of the E-commerce website and extract the data from the page like Name, Price, Discount and image of the Product.

Procedure:

- We will Firstly import our libraries which are necessary in order to scrap the data from the website.
- Later we will declare the variables and also we will initialize the URL and also the beautifulsoup.
- Later we will create the file result.csv and dump the scrapped data into it. Here we will make a made the header with Name, Price, Discount and Image.
- Later we will collect the data of the product and add it to the CSV file.
- On running the python file the results.csv file will be created with scrapped data in it.
- In the result.csv file, the data of the product i.e, Name, Price, Discount and Image will be displayed .The image is returned in the form of a link.

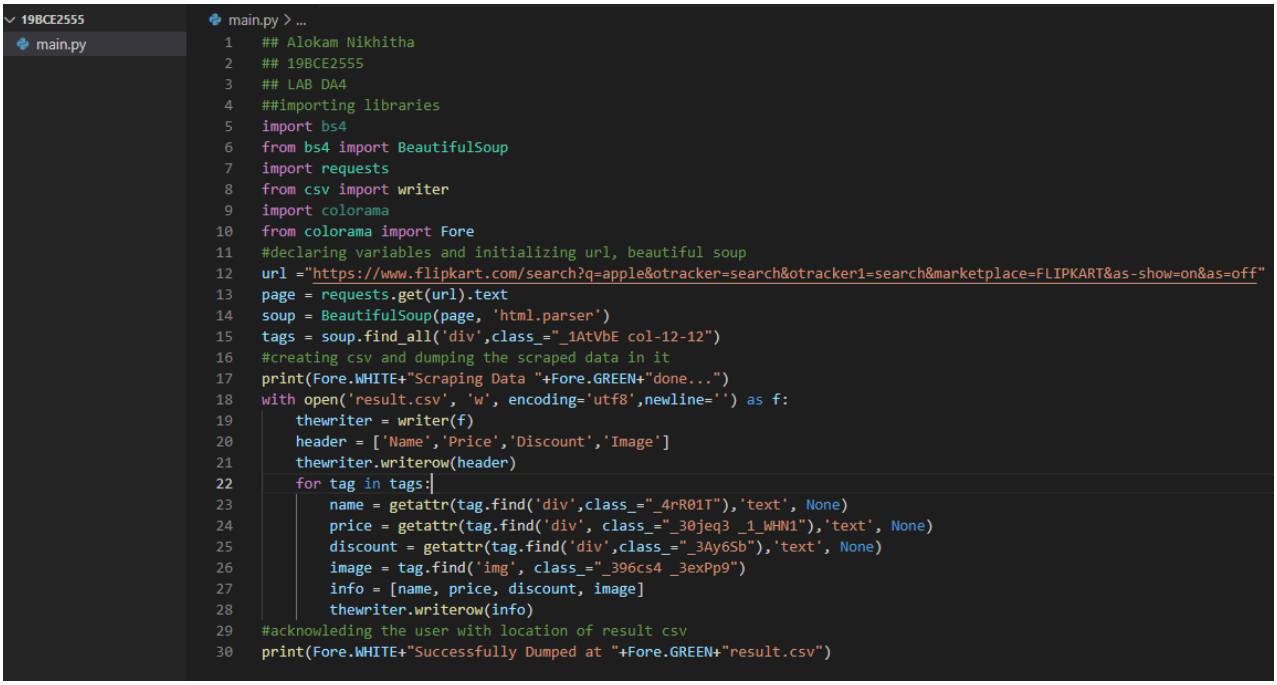
URL of the website from which we are scrapping the data :

"https://www.flipkart.com/search?q=apple&otracker=search&otracker1=search&marketplace=FLIPKART&as-show=on&as=off"

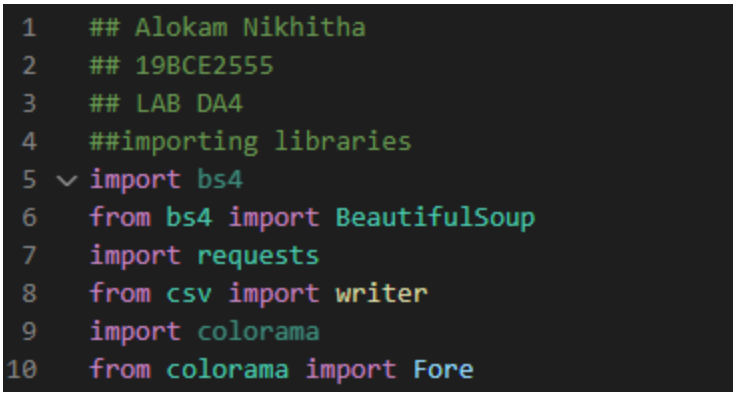
Code:

```
## Alokam Nikhitha
## 19BCE2555
## LAB DA4
##importing libraries
import bs4
from bs4 import BeautifulSoup
import requests
from csv import writer
import colorama
from colorama import Fore
#declaring variables and initializing url, beautiful soup
url ="https://www.flipkart.com/search?q=apple&otracker=search&otracker1=search&marketplace=FLIPKART&as-show=on&as=off"
page = requests.get(url).text
soup = BeautifulSoup(page, 'html.parser')
tags = soup.find_all('div',class_="_1AtVbE col-12-12")
#creating csv and dumping the scraped data in it
print(Fore.WHITE+"Scraping Data "+Fore.GREEN+"done...")
with open('result.csv', 'w', encoding='utf8',newline='') as f:
    thewriter = writer(f)
    header = ['Name','Price','Discount','Image']
    thewriter.writerow(header)
    for tag in tags:
        name = getattr(tag.find('div',class_="_4rR01T"),'text', None)
        price = getattr(tag.find('div', class_="_30jeq3 _1_WHN1"),'text', None)
        discount = getattr(tag.find('div',class_="_3Ay6Sb"),'text', None)
        image = tag.find('img', class_="_396cs4 _3exPp9")
        info = [name, price, discount, image]
        thewriter.writerow(info)
#acknowleding the user with location of result csv
print(Fore.WHITE+"Successfully Dumped at "+Fore.GREEN+"result.csv")
```

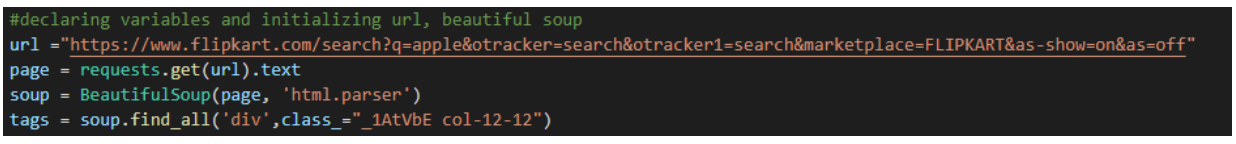
Code Snippets and Outputs:



The python file is stored in the folder named '19BCE2555'.



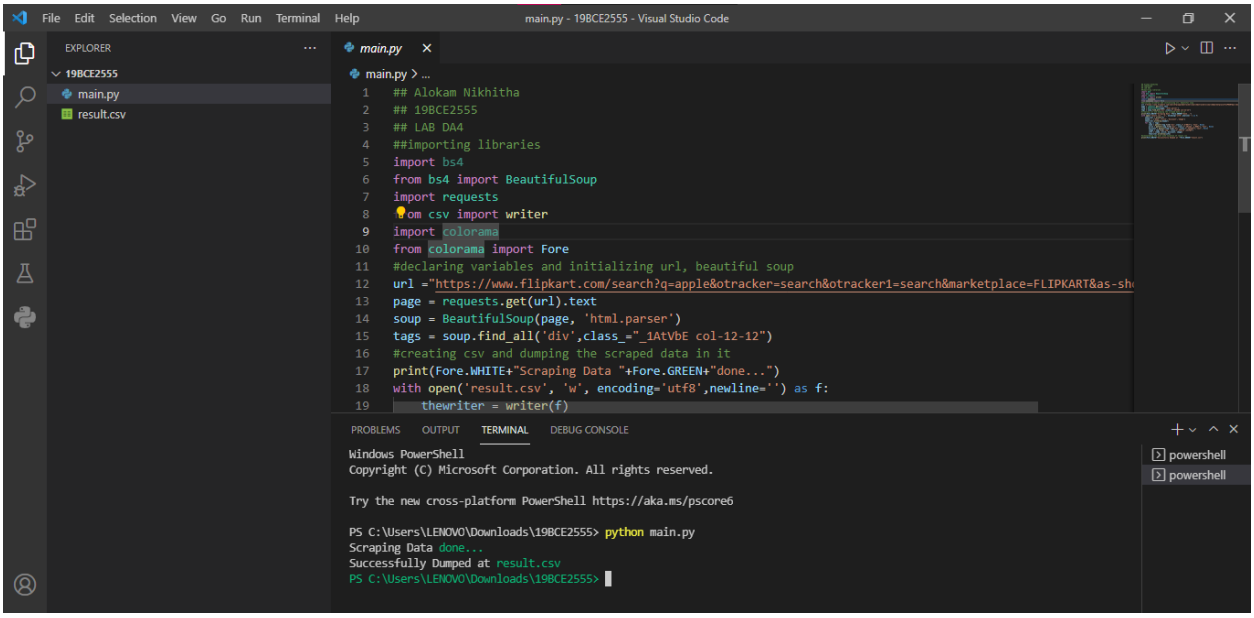
Here we are importing the necessary libraries inorder to Scrap the data.



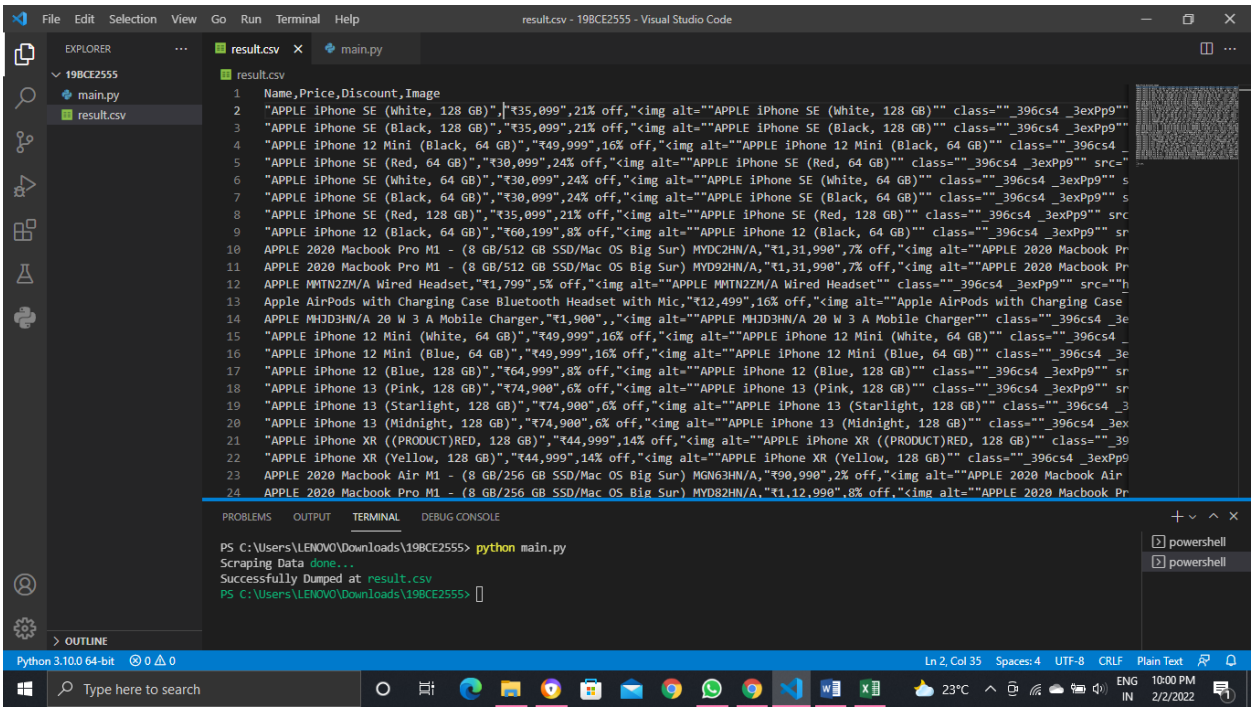
Here we are declaring the variables and also we are Initializing the URL (here url is from 'Flipkart' site). And also initializing the beautifulsoup.

```
#creating csv and dumping the scraped data in it
print(Fore.WHITE+"Scraping Data "+Fore.GREEN+"done...")
with open('result.csv', 'w', encoding='utf8',newline='') as f:
    thewriter = writer(f)
    header = ['Name','Price','Discount','Image']
    thewriter.writerow(header)
    for tag in tags:
        name = getattr(tag.find('div',class_="_4rR01T"),'text', None)
        price = getattr(tag.find('div', class_="_30jeq3 _1_WHN1"),'text', None)
        discount = getattr(tag.find('div',class_="_3Ay6Sb"),'text', None)
        image = tag.find('img', class_="_396cs4 _3exPp9")
        info = [name, price, discount, image]
        thewriter.writerow(info)
```

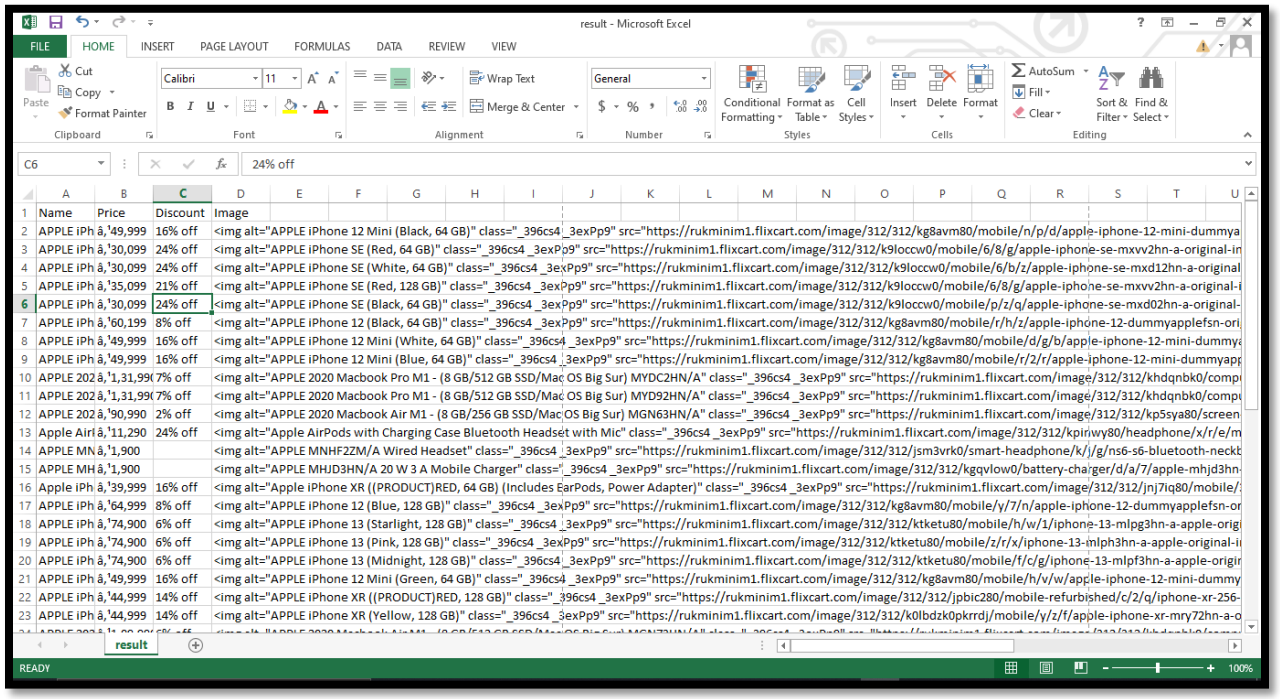
Here we are creating CSV file and dumping the scrapped data in it . Here we are opening the file result.csv and dumping the scrapped data into it. Here we made the header with Name, Price, Discount and Image. Later we are collecting the data of the product and adding it to the CSV file.



Here we are running the code using the command 'python main.py' in the terminal. Here the "result.csv" file got created in the folder after running the code.



In the result.csv we can see the data that is scrapped from the Flipkart website.



Here is the Excel sheet in which we collected the scrapped data. The name of the product, price and Discount are collected. The image is returned in the form of a link.

Results and Output

result.csv | main.py

result.csv

```
Name,Price,Discount,Image
"APPLE iPhone SE (White, 128 GB)","₹35,099",21% off,"<img alt=""APPLE iPhone SE (White, 128 GB)"" class=""_396cs4 _3exPp9""
"APPLE iPhone SE (Black, 128 GB)","₹35,099",21% off,"<img alt=""APPLE iPhone SE (Black, 128 GB)"" class=""_396cs4 _3exPp9""
"APPLE iPhone 12 Mini (Black, 64 GB)","₹49,999",16% off,"<img alt=""APPLE iPhone 12 Mini (Black, 64 GB)"" class=""_396cs4 _
"APPLE iPhone SE (Red, 64 GB)","₹30,099",24% off,"<img alt=""APPLE iPhone SE (Red, 64 GB)"" class=""_396cs4 _3exPp9"" src="
"APPLE iPhone SE (White, 64 GB)","₹30,099",24% off,"<img alt=""APPLE iPhone SE (White, 64 GB)"" class=""_396cs4 _3exPp9"" s
"APPLE iPhone SE (Black, 64 GB)","₹30,099",24% off,"<img alt=""APPLE iPhone SE (Black, 64 GB)"" class=""_396cs4 _3exPp9"" s
"APPLE iPhone SE (Red, 128 GB)","₹35,099",21% off,"<img alt=""APPLE iPhone SE (Red, 128 GB)"" class=""_396cs4 _3exPp9"" src
"APPLE iPhone 12 (Black, 64 GB)","₹60,199",8% off,"<img alt=""APPLE iPhone 12 (Black, 64 GB)"" class=""_396cs4 _3exPp9"" sr
APPLE 2020 Macbook Pro M1 - (8 GB/512 GB SSD/Mac OS Big Sur) MYDC2HN/A,"₹1,31,990",7% off,"<img alt=""APPLE 2020 Macbook Pr
APPLE 2020 Macbook Pro M1 - (8 GB/512 GB SSD/Mac OS Big Sur) MYD92HN/A,"₹1,31,990",7% off,"<img alt=""APPLE 2020 Macbook Pr
APPLE MMTN2ZM/A Wired Headset,"₹1,799",5% off,"<img alt=""APPLE MMTN2ZM/A Wired Headset"" class=""_396cs4 _3exPp9"" src=""h
Apple AirPods with Charging Case Bluetooth Headset with Mic,"₹12,499",16% off,"<img alt=""Apple AirPods with Charging Case
APPLE MHJD3HN/A 20 W 3 A Mobile Charger,"₹1,900",,"<img alt=""APPLE MHJD3HN/A 20 W 3 A Mobile Charger"" class=""_396cs4 _3e
"APPLE iPhone 12 Mini (White, 64 GB)","₹49,999",16% off,"<img alt=""APPLE iPhone 12 Mini (White, 64 GB)"" class=""_396cs4 _
"APPLE iPhone 12 Mini (Blue, 64 GB)","₹49,999",16% off,"<img alt=""APPLE iPhone 12 Mini (Blue, 64 GB)"" class=""_396cs4 _3e
"APPLE iPhone 12 (Blue, 128 GB)","₹64,999",8% off,"<img alt=""APPLE iPhone 12 (Blue, 128 GB)"" class=""_396cs4 _3exPp9"" sr
"APPLE iPhone 13 (Pink, 128 GB)","₹74,900",6% off,"<img alt=""APPLE iPhone 13 (Pink, 128 GB)"" class=""_396cs4 _3exPp9"" sr
"APPLE iPhone 13 (Starlight, 128 GB)","₹74,900",6% off,"<img alt=""APPLE iPhone 13 (Starlight, 128 GB)"" class=""_396cs4 _3
"APPLE iPhone 13 (Midnight, 128 GB)","₹74,900",6% off,"<img alt=""APPLE iPhone 13 (Midnight, 128 GB)"" class=""_396cs4 _3ex
"APPLE iPhone XR ((PRODUCT)RED, 128 GB)","₹44,999",14% off,"<img alt=""APPLE iPhone XR ((PRODUCT)RED, 128 GB)"" class=""_39
"APPLE iPhone XR (Yellow, 128 GB)","₹44,999",14% off,"<img alt=""APPLE iPhone XR (Yellow, 128 GB)"" class=""_396cs4 _3exPp9
APPLE 2020 Macbook Air M1 - (8 GB/256 GB SSD/Mac OS Big Sur) MGN63HN/A,"₹90,990",2% off,"<img alt=""APPLE 2020 Macbook Air
APPLE 2020 Macbook Pro M1 - (8 GB/256 GB SSD/Mac OS Big Sur) MYD82HN/A,"₹1,12,990",8% off,"<img alt=""APPLE 2020 Macbook Pr
```

PROBLEMS OUTPUT TERMINAL DEBUG CONSOLE

```
PS C:\Users\LENOVO\Downloads\19BCE2555> python main.py
Scraping Data done...
Successfully Dumped at result.csv
PS C:\Users\LENOVO\Downloads\19BCE2555>
```

Here is the list of data that is being scrapped from the website and dumped into result.csv file and which is formed in the folder of the code after running the code using beautifulsoup.

	Name	Price	Discount	Image
2	APPLE iPh	â‚¹49,999	16% off	<img alt="APPLE iPhone 12 Mini (Black, 64 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/kg8av
3	APPLE iPh	â‚¹30,099	24% off	<img alt="APPLE iPhone SE (Red, 64 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/k9loccw0/m
4	APPLE iPh	â‚¹30,099	24% off	<img alt="APPLE iPhone SE (White, 64 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/k9loccw0/
5	APPLE iPh	â‚¹35,099	21% off	<img alt="APPLE iPhone SE (Red, 128 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/k9loccw0/n
6	APPLE iPh	â‚¹30,099	24% off	<img alt="APPLE iPhone SE (Black, 64 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/k9loccw0/n
7	APPLE iPh	â‚¹60,199	8% off	<img alt="APPLE iPhone 12 (Black, 64 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/kg8avm80/
8	APPLE iPh	â‚¹49,999	16% off	<img alt="APPLE iPhone 12 Mini (White, 64 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/kg8av
9	APPLE iPh	â‚¹49,999	16% off	<img alt="APPLE iPhone 12 Mini (Blue, 64 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/kg8avn
10	APPLE 202	â‚¹1,31,99(	7% off	<img alt="APPLE 2020 Macbook Pro M1 - (8 GB/512 GB SSD/Mac OS Big Sur) MYDC2HN/A" class="_396cs4 _3exPp9" src="https://rukm
11	APPLE 202	â‚¹1,31,99(	7% off	<img alt="APPLE 2020 Macbook Pro M1 - (8 GB/512 GB SSD/Mac OS Big Sur) MYD92HN/A" class="_396cs4 _3exPp9" src="https://rukmi
12	APPLE 202	â‚¹90,990	2% off	<img alt="APPLE 2020 Macbook Air M1 - (8 GB/256 GB SSD/Mac OS Big Sur) MGN63HN/A" class="_396cs4 _3exPp9" src="https://rukmi
13	Apple Airl	â‚¹11,290	24% off	<img alt="Apple AirPods with Charging Case Bluetooth Headset with Mic" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.c
14	APPLE MN	â‚¹1,900		<img alt="APPLE MNHF2ZM/A Wired Headset" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/jsm3vr
15	APPLE MH	â‚¹1,900		<img alt="APPLE MHJD3HN/A 20 W 3 A Mobile Charger" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/31
16	Apple iPh	â‚¹39,999	16% off	<img alt="Apple iPhone XR ((PRODUCT)RED, 64 GB) (Includes EarPods, Power Adapter)" class="_396cs4 _3exPp9" src="https://rukmi
17	APPLE iPh	â‚¹64,999	8% off	<img alt="APPLE iPhone 12 (Blue, 128 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/kg8avm80/
18	APPLE iPh	â‚¹74,900	6% off	<img alt="APPLE iPhone 13 (Starlight, 128 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/ktketu
19	APPLE iPh	â‚¹74,900	6% off	<img alt="APPLE iPhone 13 (Pink, 128 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/ktketu80/n
20	APPLE iPh	â‚¹74,900	6% off	<img alt="APPLE iPhone 13 (Midnight, 128 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/ktketu
21	APPLE iPh	â‚¹49,999	16% off	<img alt="APPLE iPhone 12 Mini (Green, 64 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/kg8av
22	APPLE iPh	â‚¹44,999	14% off	<img alt="APPLE iPhone XR ((PRODUCT)RED, 128 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/
23	APPLE iPh	â‚¹44,999	14% off	<img alt="APPLE iPhone XR (Yellow, 128 GB)" class="_396cs4 _3exPp9" src="https://rukminim1.flixcart.com/image/312/312/k0lbdzk

This is the view of Excel file in which the data is being stored.