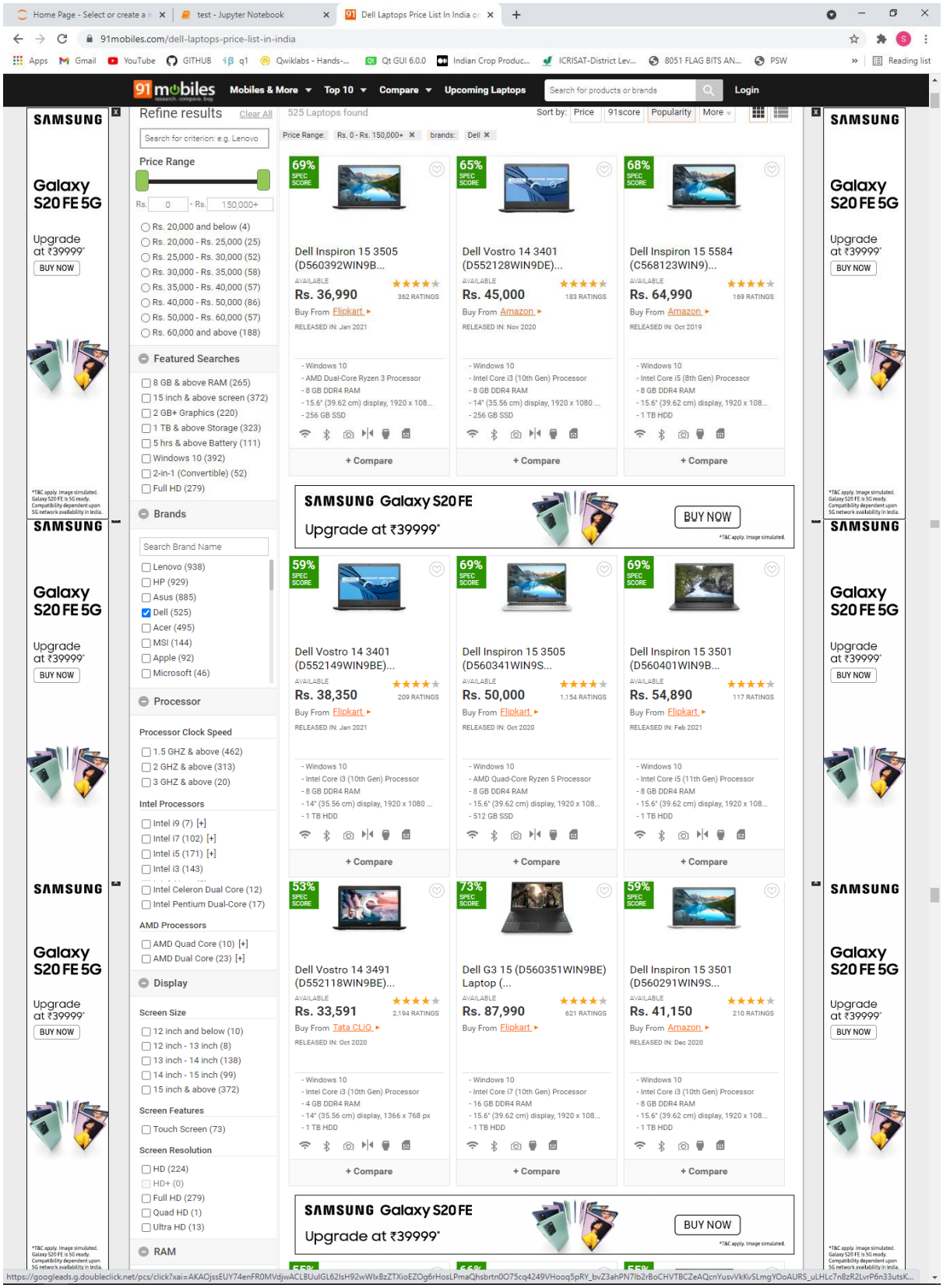
DATA SCIENCE
Web Scraping Using Python

Submitted by:
Samundar Singh
2019272033(SS)

On: ***02-APR -2021***

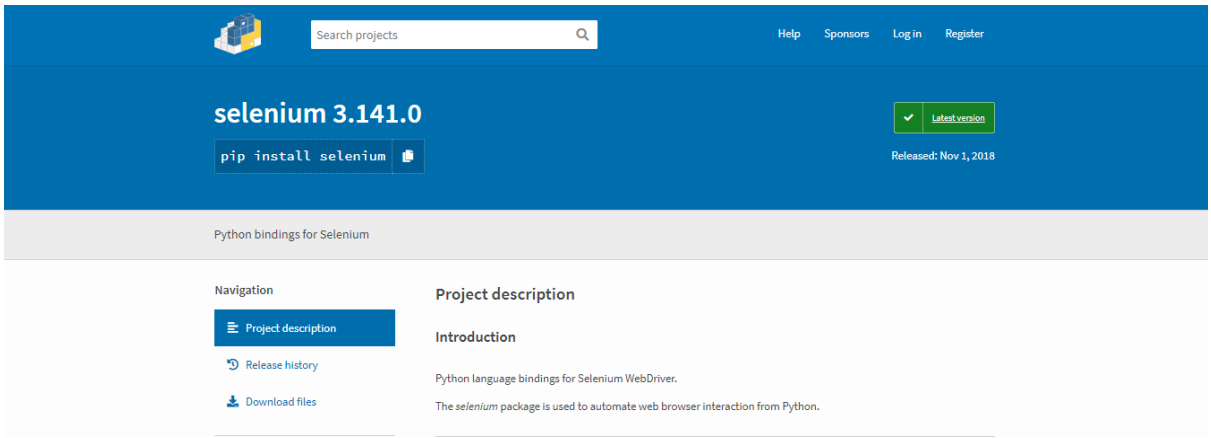
1. Target website Overview:



Link: https://www.91mobiles.com/dell-laptops-price-list-in-india

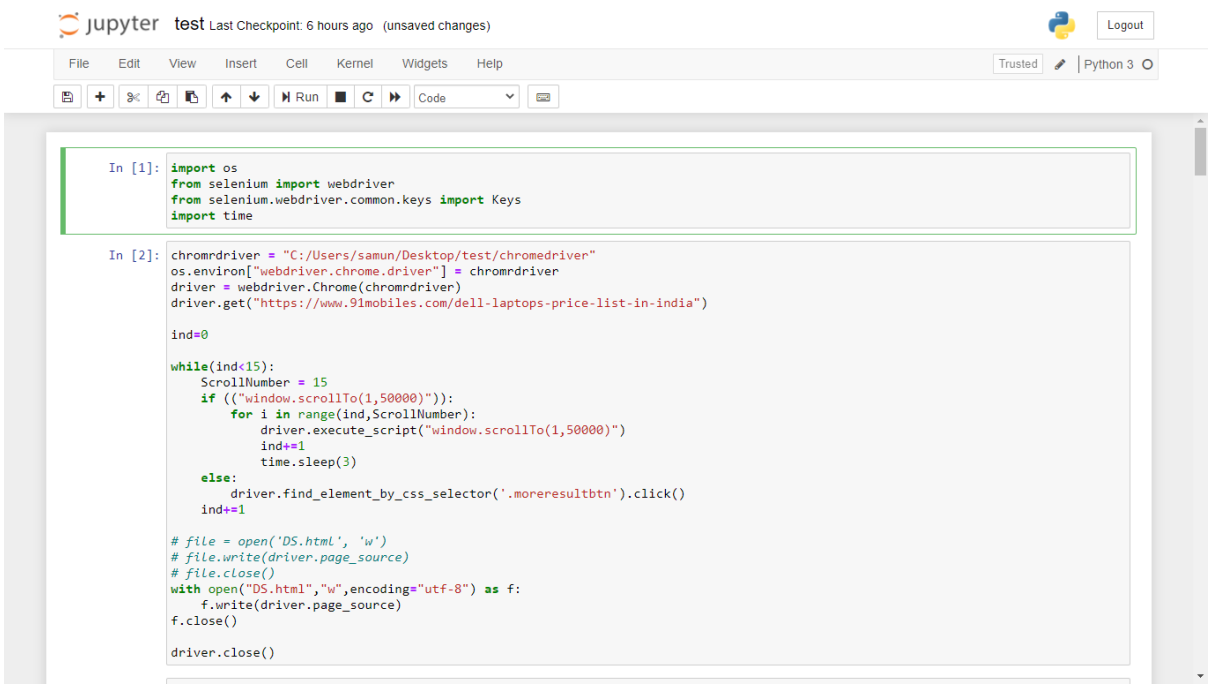
This website include data about various gadgets with the specifications my target is to fetch dell laptop data from here with its price and specification

Step 1: Since this website is based on lazy loading so I am using a automated web browser interaction named selenium 3.141.0 to use auto scroll feature.

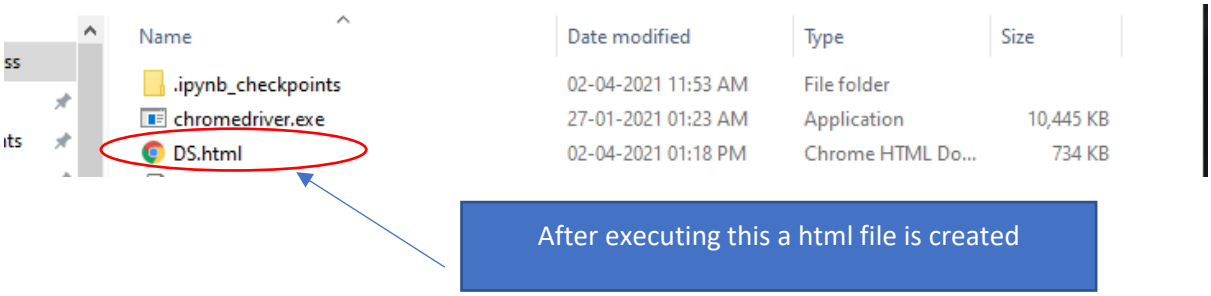


Here I am simply used a scroll_to function whose work is only to scroll the web page and append the data every time to the parent html.

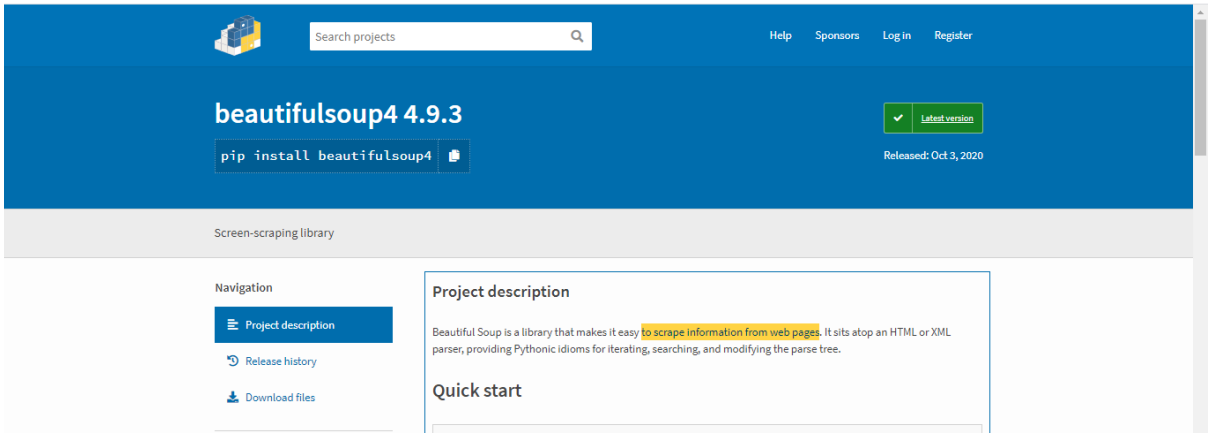
Code:



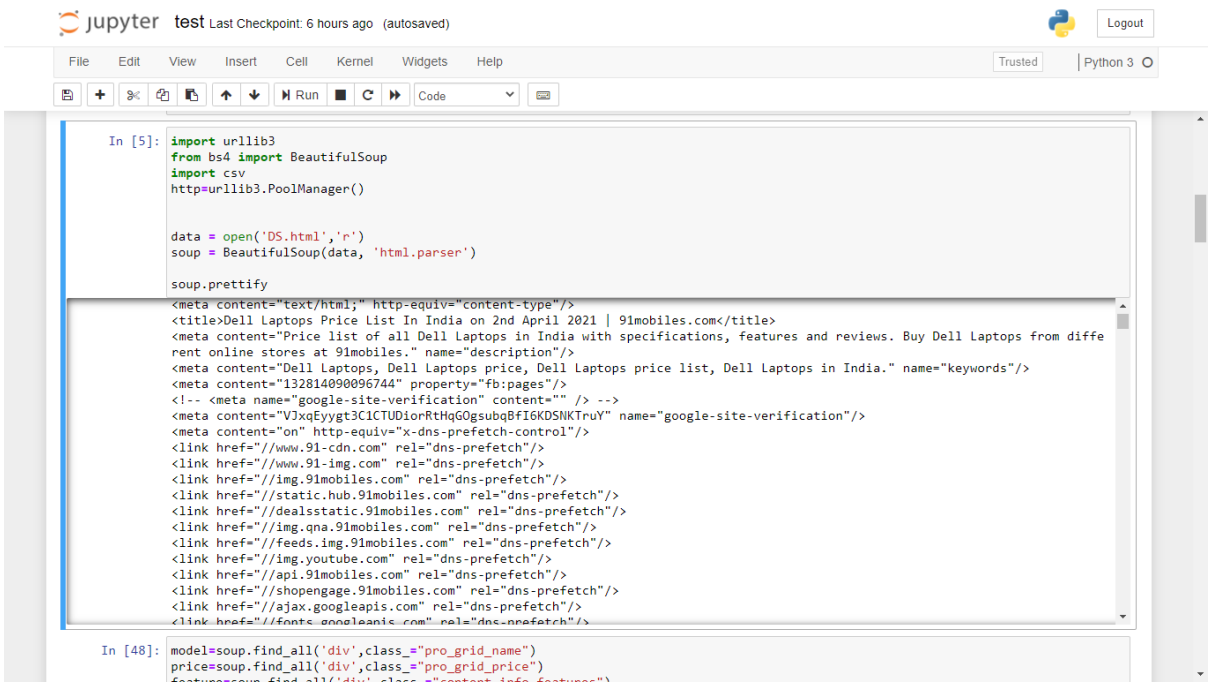
Note : By using sleep function I prevent myself from being blocked by the server.



Step 2: I used a library of python called Beautiful Soup to scrape information from web pages



Code:



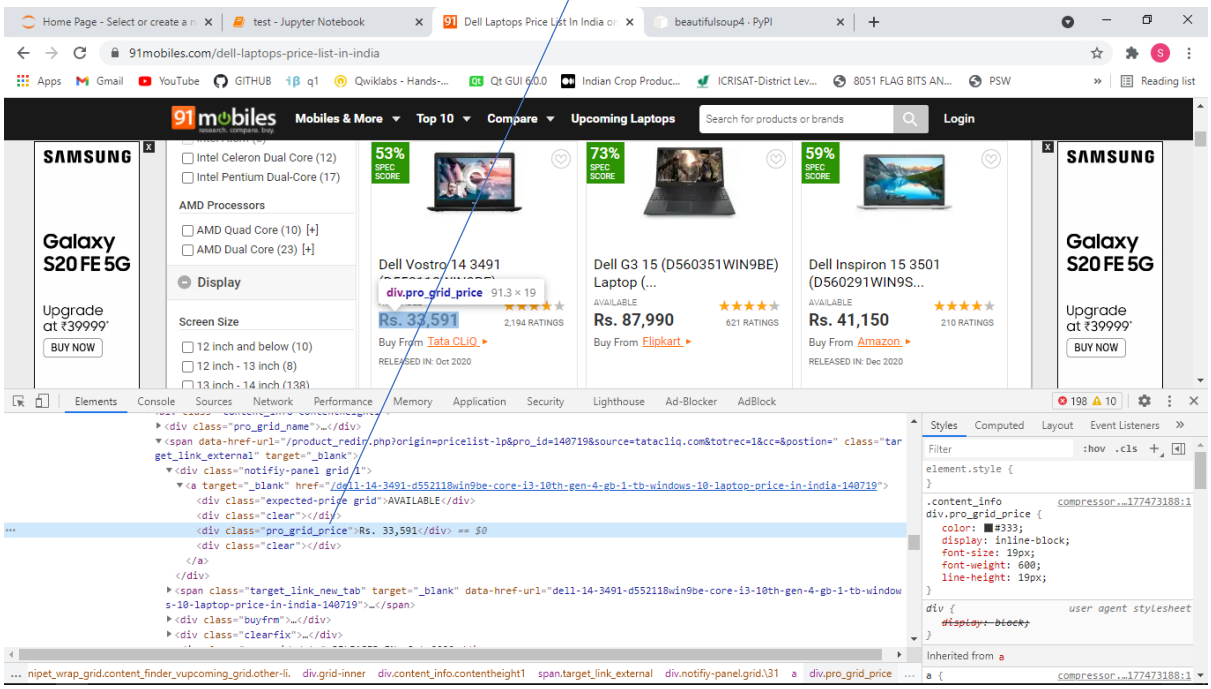
here the DS, html is parsed and in return we get an object named ‘bs4.BeautifulSoup’ .this object include various method from which we can scrape our data .

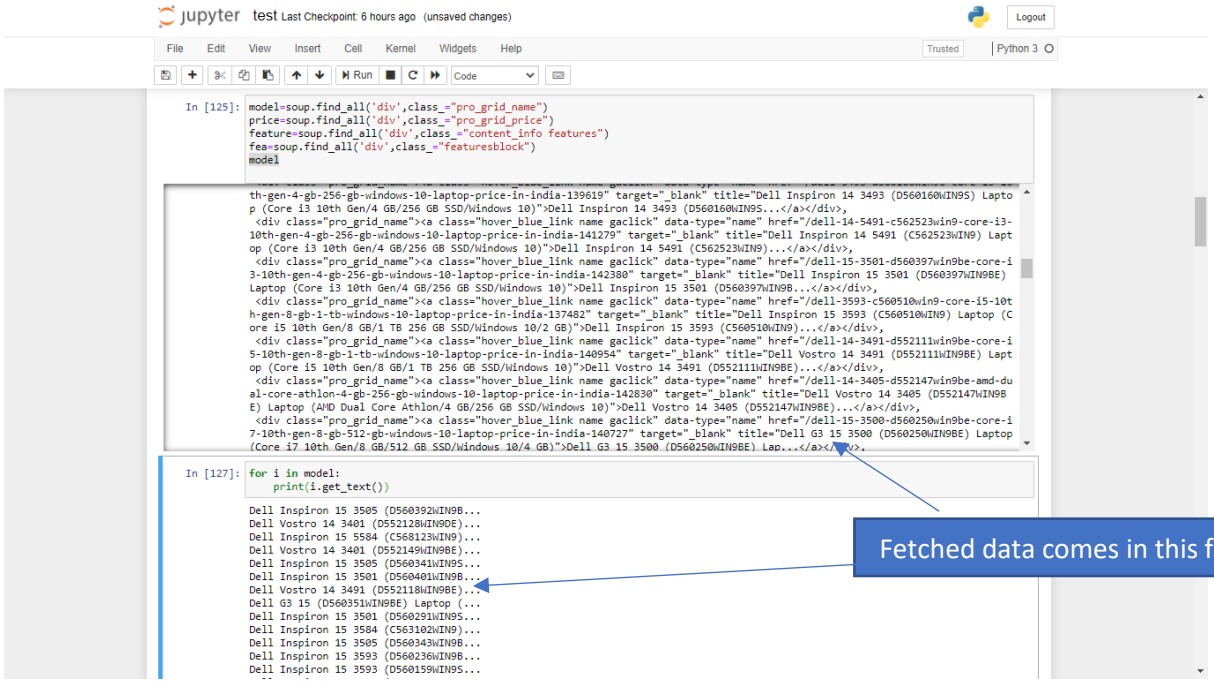
Step 3: Start scrapping the data using Inspecting the web page and finding relevant class.

Code:To fetch all the data from the specified class.

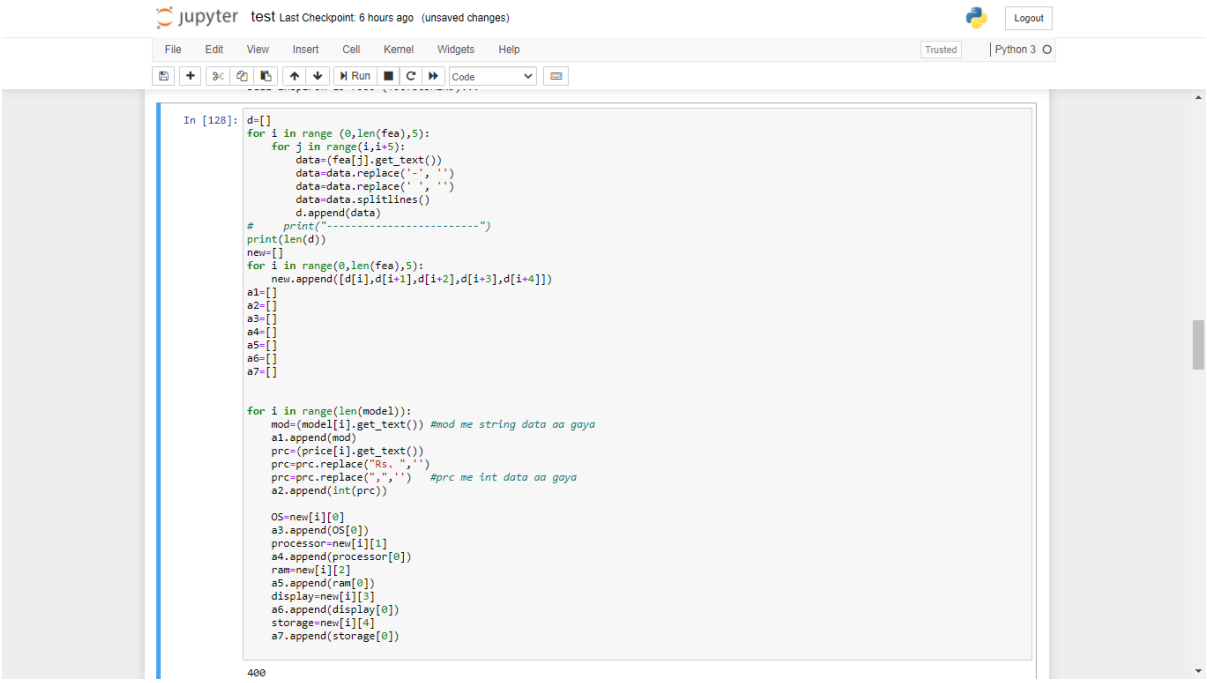
```
In [48]: model=soup.find_all('div',class_="pro_grid_name")
         price=soup.find_all('div',class_="pro_grid_price")
         feature=soup.find_all('div',class_="content_info features")
         fea=soup.find_all('div',class_="featuresblock")
```

Inspecting the website





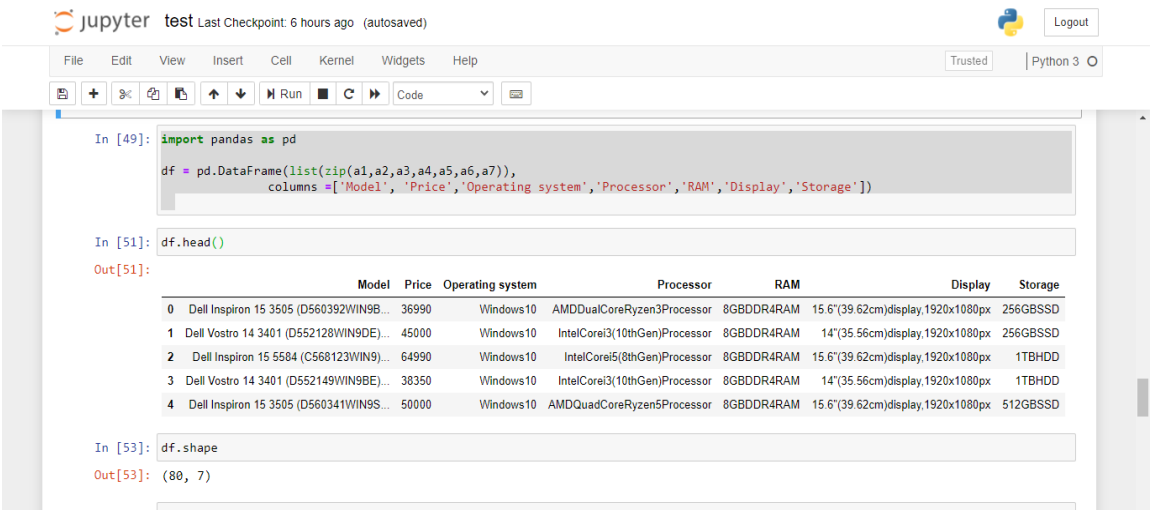
Step 4: Converting unstructured data into structured data



```
In [128]:
d=[]
for i in range (0,len(fea),5):
    for j in range(i,i+5):
        data=(fea[j].get_text())
        data=data.replace('-', '')
        data=data.replace(' ', '')
        data=data.splitlines()
        d.append(data)
#     print("-------------------------")
print(len(d))
new=[]
for i in range(0,len(fea),5):
    new.append([d[i],d[i+1],d[i+2],d[i+3],d[i+4]])
a1=[]
a2=[]
a3=[]
a4=[]
a5=[]
a6=[]
a7=[]


for i in range(len(model)):
    mod=(model[i].get_text()) #mod me string data aa gaya
    a1.append(mod)
    prc=(price[i].get_text())
    prc=prc.replace("Rs. ",'')
    prc=prc.replace(",",'')   #prc me int data aa gaya
    a2.append(int(prc))

    OS=new[i][0]
    a3.append(OS[0])
    processor=new[i][1]
    a4.append(processor[0])
    ram=new[i][2]
    a5.append(ram[0])
    display=new[i][3]
    a6.append(display[0])
    storage=new[i][4]
    a7.append(storage[0])
```

```
400
```

This is a simple code which fetch the data from various object and put it in array ,once every data fetched into array then using pandas we can convert this set of array into a data frame



```
In [49]:
import pandas as pd

df = pd.DataFrame(list(zip(a1,a2,a3,a4,a5,a6,a7)),
                  columns =['Model', 'Price','Operating system','Processor','RAM','Display','Storage'])
```

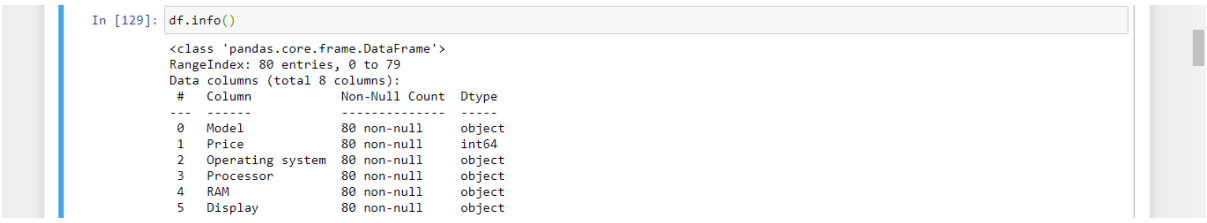
```
In [51]:
df.head()
```

Out[51]:

	Model	Price	Operating system	Processor	RAM	Display	Storage
0	Dell Inspiron 15 3505 (D560392WIN9B...	36990	Windows10	AMDDualCoreRyzen3Processor	8GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	256GBSSD
1	Dell Vostro 14 3401 (D552128WIN9DE)...	45000	Windows10	IntelCorei3(10thGen)Processor	8GBDDR4RAM	14"(35.56cm)display,1920x1080px	256GBSSD
2	Dell Inspiron 15 5584 (C568123WIN9)...	64990	Windows10	IntelCorei5(8thGen)Processor	8GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD
3	Dell Vostro 14 3401 (D552149WIN9BE)...	38350	Windows10	IntelCorei3(10thGen)Processor	8GBDDR4RAM	14"(35.56cm)display,1920x1080px	1TBHDD
4	Dell Inspiron 15 3505 (D560341WIN9S...	50000	Windows10	AMDQuadCoreRyzen5Processor	8GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	512GBSSD

```
In [53]:
df.shape
```

```
Out[53]:
(80, 7)
```



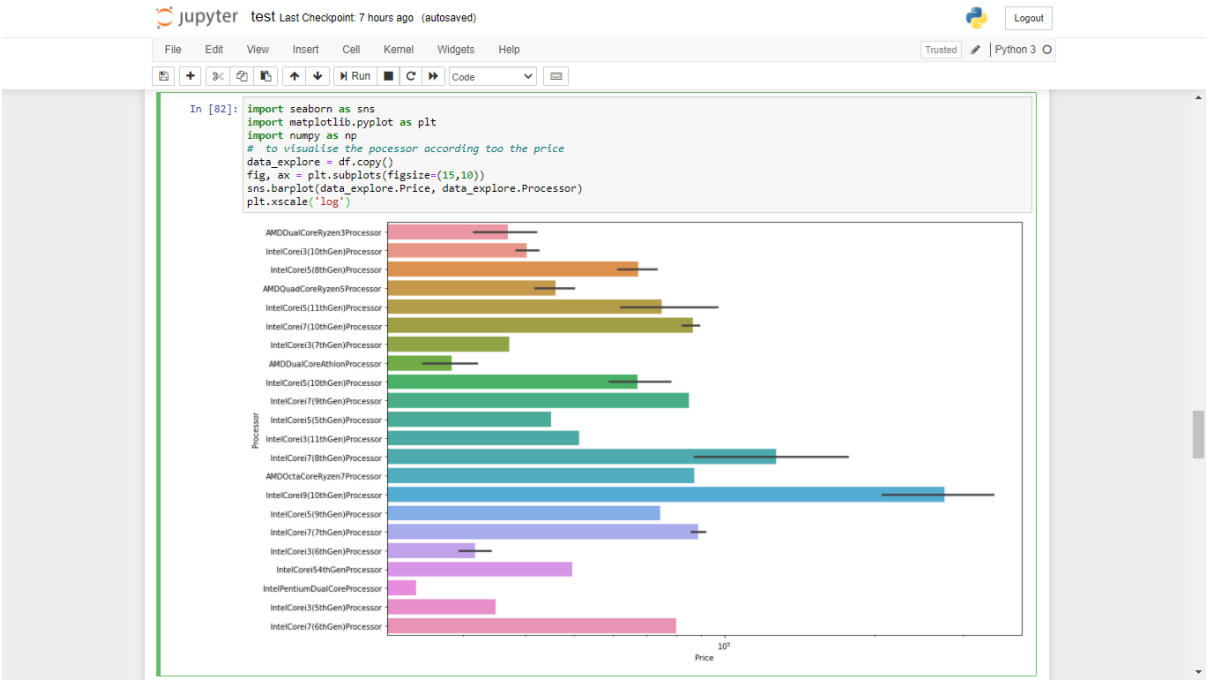
```
In [129]:
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 80 entries, 0 to 79
Data columns (total 8 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Model             80 non-null     object
 1   Price             80 non-null     int64
 2   Operating system  80 non-null     object
 3   Processor         80 non-null     object
 4   RAM               80 non-null     object
 5   Display           80 non-null     object
```

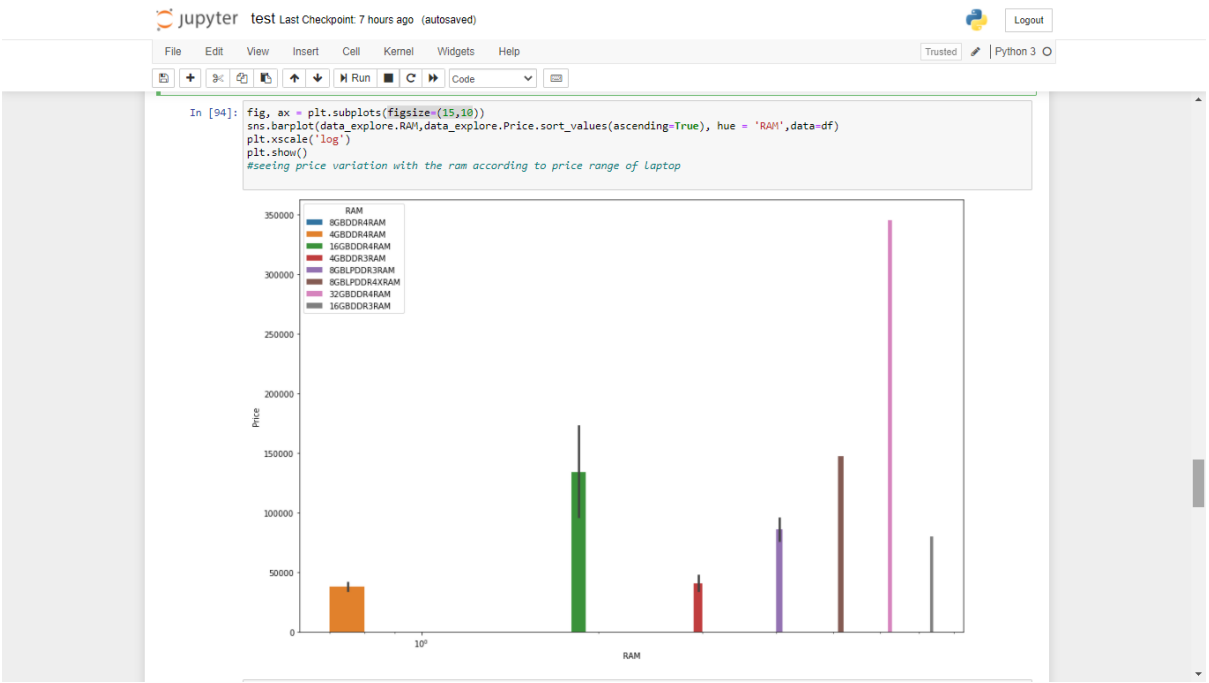
As we can see that the data comes in structured format

Step 5: performing some visualization using the fetched data .

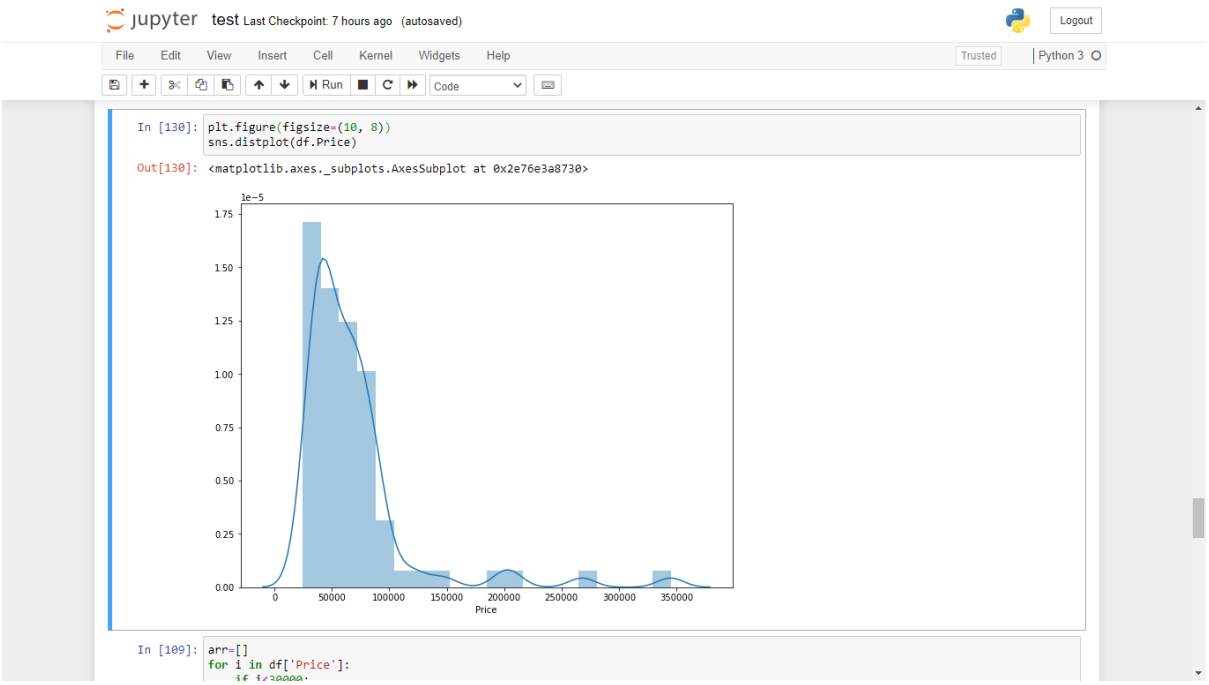
A.) Price visualization of the processor



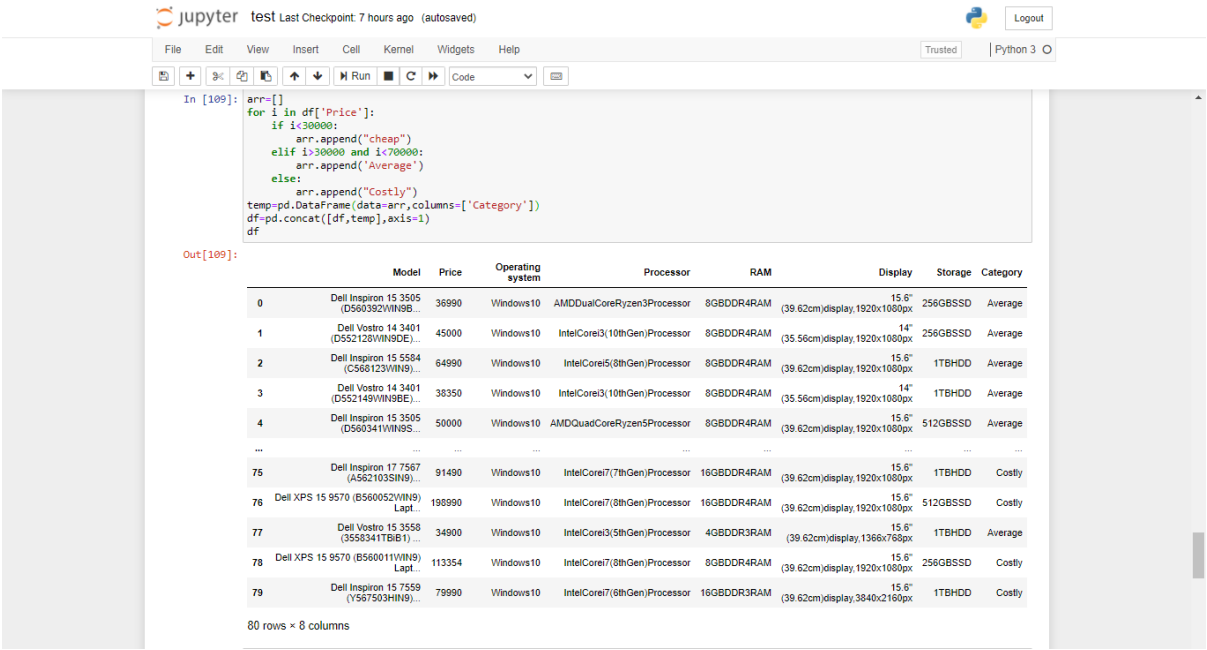
B.) Ram available to various laptop according to the price.



C.) Distplot according to the Price



D.) Adding attribute with categorical feature .



	Model	Price	Operating system	Processor	RAM	Display	Storage	Category
0	Dell Inspiron 15 3505 (D560392WIN9B...	36990	Windows10	AMDDualCoreRyzen3Processor	8GBDDR4RAM	15.6" (39.62cm)display,1920x1080px	256GBSSD	Average
1	Dell Vostro 14 3401 (D552128WIN9DE)...	45000	Windows10	IntelCorei3(10thGen)Processor	8GBDDR4RAM	14" (35.56cm)display,1920x1080px	256GBSSD	Average
2	Dell Inspiron 15 5584 (C568123WIN9)...	64990	Windows10	IntelCorei5(8thGen)Processor	8GBDDR4RAM	15.6" (39.62cm)display,1920x1080px	1TBHDD	Average
3	Dell Vostro 14 3401 (D552149WIN9BE)...	38350	Windows10	IntelCorei3(10thGen)Processor	8GBDDR4RAM	14" (35.56cm)display,1920x1080px	1TBHDD	Average
4	Dell Inspiron 15 3505 (D560341WIN9S...	50000	Windows10	AMDQuadCoreRyzen5Processor	8GBDDR4RAM	15.6" (39.62cm)display,1920x1080px	512GBSSD	Average
...	...	...	...	...	...	...	...	...
75	Dell Inspiron 17 7567 (A562103SIN9)...	91490	Windows10	IntelCorei7(7thGen)Processor	16GBDDR4RAM	15.6" (39.62cm)display,1920x1080px	1TBHDD	Costly
76	Dell XPS 15 9570 (B560052WIN9) Lapt...	198990	Windows10	IntelCorei7(8thGen)Processor	16GBDDR4RAM	15.6" (39.62cm)display,1920x1080px	512GBSSD	Costly
77	Dell Vostro 15 3558 (3558341TBiB1) ...	34900	Windows10	IntelCorei3(5thGen)Processor	4GBDDR3RAM	15.6" (39.62cm)display,1366x768px	1TBHDD	Average
78	Dell XPS 15 9570 (B560011WIN9) Lapt...	113354	Windows10	IntelCorei7(8thGen)Processor	8GBDDR4RAM	15.6" (39.62cm)display,1920x1080px	256GBSSD	Costly
79	Dell Inspiron 15 7559 (Y567503HIN9)...	79990	Windows10	IntelCorei7(6thGen)Processor	16GBDDR3RAM	15.6" (39.62cm)display,3840x2160px	1TBHDD	Costly

E.) See the distribution of ram on newly added feature

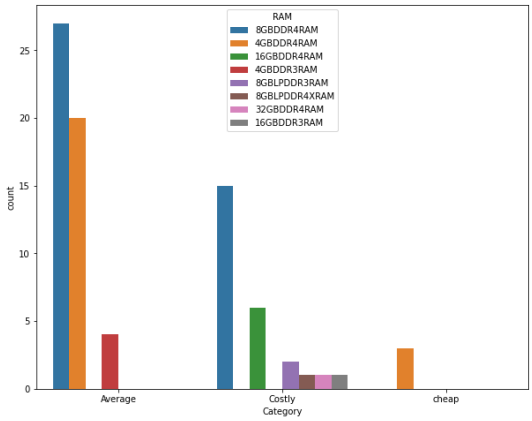
jupyter test Last Checkpoint: 7 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [120]:
```
plt.figure(figsize=(10, 8))
sns.countplot(x='Category',hue='RAM',data=df )
#to see the configuration of ram in category wise data
```

Out[120]: <matplotlib.axes._subplots.AxesSubplot at 0x2e76c2734f0>



In [121]:
```
plt.figure(figsize=(10, 8))
```

F.) See the distribution of Storage on newly added feature :

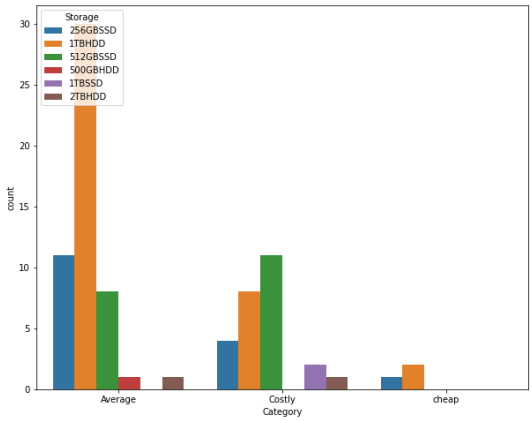
jupyter test Last Checkpoint: 7 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [121]:
```
plt.figure(figsize=(10, 8))
sns.countplot(x='Category',hue='Storage',data=df )
#to see the configuration of Storage in category wise data
```

Out[121]: <matplotlib.axes._subplots.AxesSubplot at 0x2e76c2fd7c0>

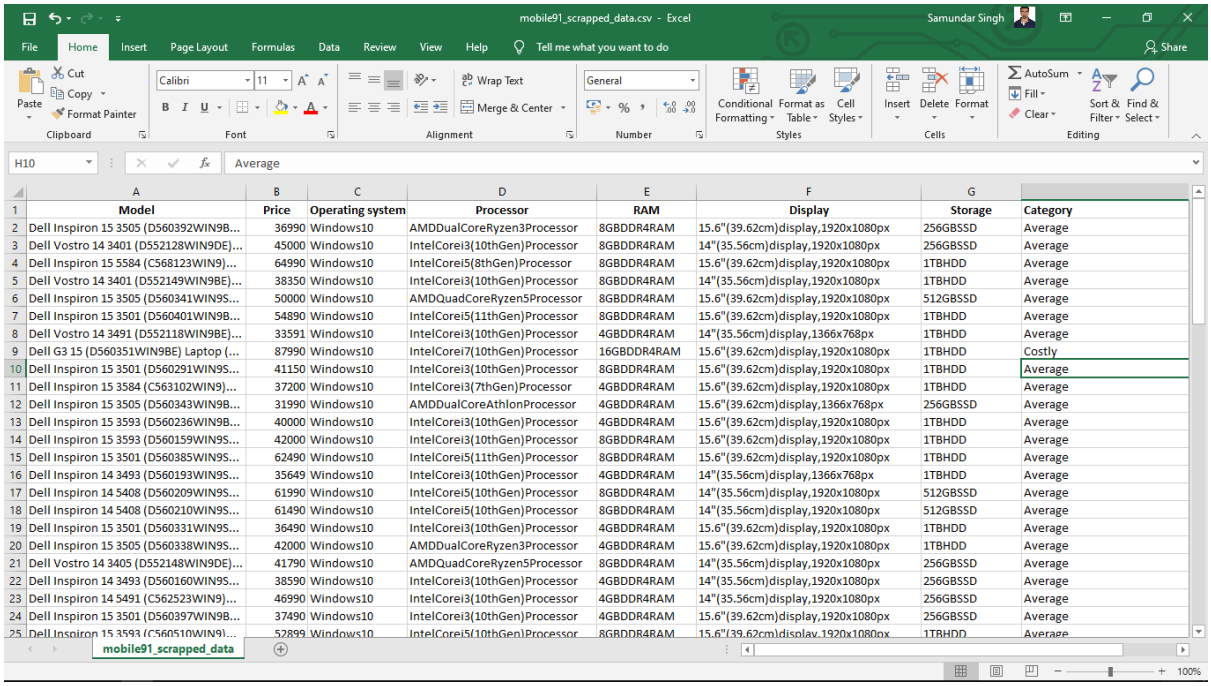


In [122]:
```
#saving the data set into a new csv file

df.to_csv("mobile91_scrapped_data.csv",index=False)
```

Step 5: Save the fetched data into a newely csv file.

Category

```
In [122]: #saving the data set into a new csv file

          df.to_csv("mobile91_scrapped_data.csv",index=False)
```



mobile91_scrapped_data.csv - Excel — Samundar Singh

H10 | Average

	Model	Price	Operating system	Processor	RAM	Display	Storage	Category
2	Dell Inspiron 15 3505 (D560392WIN9B...	36990	Windows10	AMDDualCoreRyzen3Processor	8GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	256GBSSD	Average
3	Dell Vostro 14 3401 (D552128WIN9DE)...	45000	Windows10	IntelCorei3(10thGen)Processor	8GBDDR4RAM	14"(35.56cm)display,1920x1080px	256GBSSD	Average
4	Dell Inspiron 15 5584 (C568123WIN9)...	64990	Windows10	IntelCorei5(8thGen)Processor	8GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD	Average
5	Dell Vostro 14 3401 (D552149WIN9BE)...	38350	Windows10	IntelCorei3(10thGen)Processor	8GBDDR4RAM	14"(35.56cm)display,1920x1080px	1TBHDD	Average
6	Dell Inspiron 15 3505 (D560341WIN9S...	50000	Windows10	AMDQuadCoreRyzen5Processor	8GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	512GBSSD	Average
7	Dell Inspiron 15 3501 (D560401WIN9B...	54890	Windows10	IntelCorei5(11thGen)Processor	8GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD	Average
8	Dell Vostro 14 3491 (D552118WIN9BE)...	33591	Windows10	IntelCorei3(10thGen)Processor	4GBDDR4RAM	14"(35.56cm)display,1366x768px	1TBHDD	Average
9	Dell G3 15 (D560351WIN9BE) Laptop (...	87990	Windows10	IntelCorei7(10thGen)Processor	16GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD	Costly
10	Dell Inspiron 15 3501 (D560291WIN9S...	41150	Windows10	IntelCorei3(10thGen)Processor	8GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD	Average
11	Dell Inspiron 15 3584 (C563102WIN9)...	37200	Windows10	IntelCorei3(7thGen)Processor	4GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD	Average
12	Dell Inspiron 15 3505 (D560343WIN9B...	31990	Windows10	AMDDualCoreAthlonProcessor	4GBDDR4RAM	15.6"(39.62cm)display,1366x768px	256GBSSD	Average
13	Dell Inspiron 15 3593 (D560236WIN9B...	40000	Windows10	IntelCorei3(10thGen)Processor	4GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD	Average
14	Dell Inspiron 15 3593 (D560159WIN9S...	42000	Windows10	IntelCorei3(10thGen)Processor	8GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD	Average
15	Dell Inspiron 15 3501 (D560385WIN9S...	62490	Windows10	IntelCorei5(11thGen)Processor	8GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD	Average
16	Dell Inspiron 14 3493 (D560193WIN9S...	35649	Windows10	IntelCorei3(10thGen)Processor	4GBDDR4RAM	14"(35.56cm)display,1366x768px	1TBHDD	Average
17	Dell Inspiron 14 5408 (D560209WIN9S...	61990	Windows10	IntelCorei5(10thGen)Processor	8GBDDR4RAM	14"(35.56cm)display,1920x1080px	512GBSSD	Average
18	Dell Inspiron 14 5408 (D560210WIN9S...	61490	Windows10	IntelCorei5(10thGen)Processor	8GBDDR4RAM	14"(35.56cm)display,1920x1080px	512GBSSD	Average
19	Dell Inspiron 15 3501 (D560331WIN9S...	36490	Windows10	IntelCorei3(10thGen)Processor	4GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD	Average
20	Dell Inspiron 15 3505 (D560338WIN9S...	42000	Windows10	AMDDualCoreRyzen3Processor	4GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD	Average
21	Dell Vostro 14 3405 (D552148WIN9DE)...	41790	Windows10	AMDQuadCoreRyzen5Processor	8GBDDR4RAM	14"(35.56cm)display,1920x1080px	256GBSSD	Average
22	Dell Inspiron 14 3493 (D560160WIN9S...	38590	Windows10	IntelCorei3(10thGen)Processor	4GBDDR4RAM	14"(35.56cm)display,1920x1080px	256GBSSD	Average
23	Dell Inspiron 14 5491 (C562523WIN9)...	46990	Windows10	IntelCorei3(10thGen)Processor	4GBDDR4RAM	14"(35.56cm)display,1920x1080px	256GBSSD	Average
24	Dell Inspiron 15 3501 (D560397WIN9B...	37490	Windows10	IntelCorei3(10thGen)Processor	4GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	256GBSSD	Average
25	Dell Inspiron 15 3593 (C560510WIN9)...	52899	Windows10	IntelCorei5(10thGen)Processor	8GBDDR4RAM	15.6"(39.62cm)display,1920x1080px	1TBHDD	Average

mobile91_scrapped_data