Harshit Mishra (19BCE0799)

CSE-3024 WEB MINING

LAB ASSIGNMENT 4B

Aim: Use BeautifulSoup or Scrapy to crawl any one of the E-commerce websites of your choice and perform the same. The following information needs to be extracted from the page:

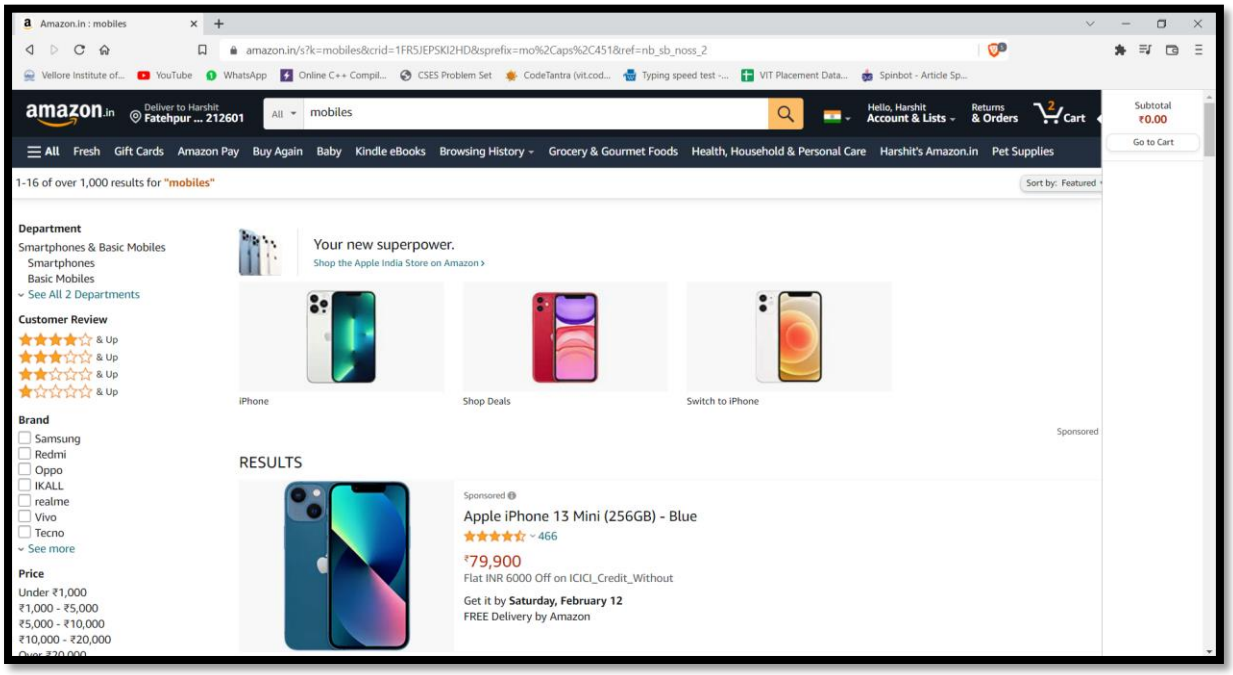
a) Product Name
b) Product Price
c) Product Discount
d) Product Image

Procedure:

- Firstly, we install scrapy package using 'pip install scrapy' in anaconda prompt shell.
- Then we can start shell by using command 'scrapy shell'.
- To initiate the crawler, we run default commands as in scrapy genspider example example.com
- Create a folder named 'mobile' and change the working directory to current folder.
- Create the python (.py) file inside the mobile directory and initiate the scrapy files using command "scrapy genspider example example.com"
- Here I have scrapped Amazon's website featuring mobile phones.
- They Python code is written in the scrapy file.
- Finally, it is exported as csv file using command "scrapy crawl -o mobile data.csv".

Harshit Mishra (19BCE0799)

Site we are scrapping:

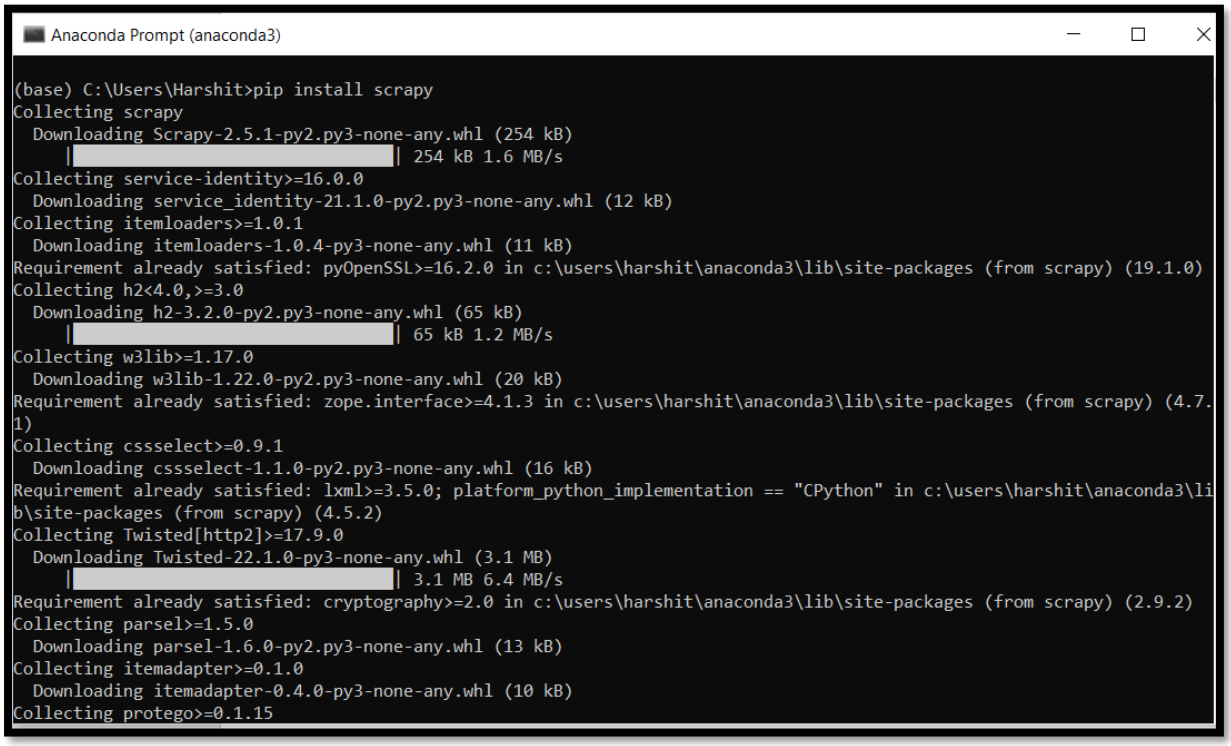


URL of the above website:

https://www.amazon.in/s?k=mobiles&crid=1FR5JEPSKI2HD&sprefix=mo%2Caps%2C451&ref=nb_sb_noss_2

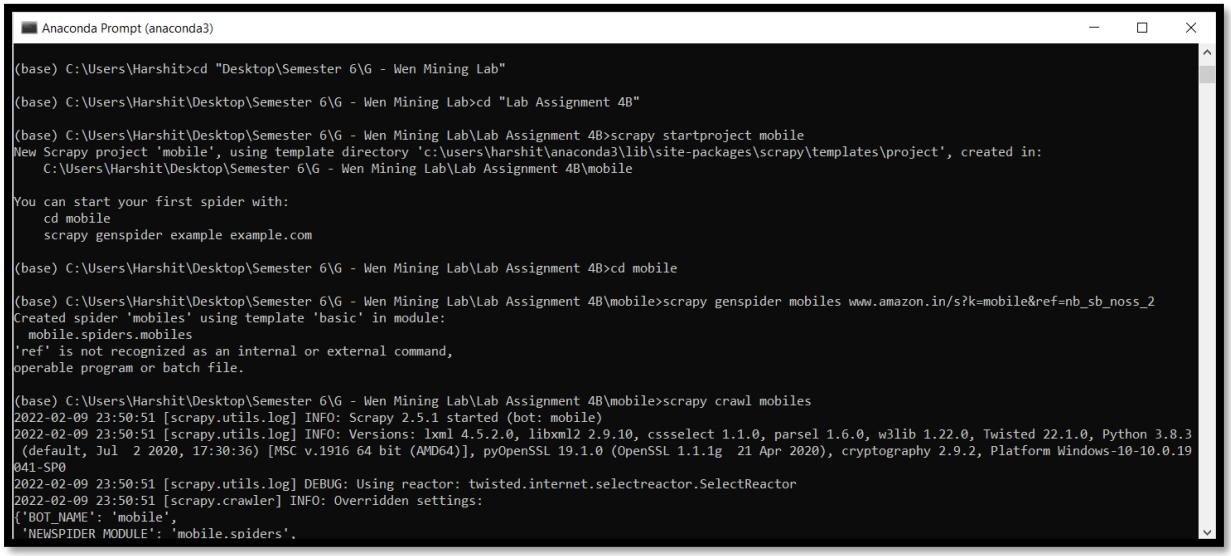
Harshit Mishra (19BCE0799)

Installing Scrapy in Anaconda:



Here we use the command pip install scrapy

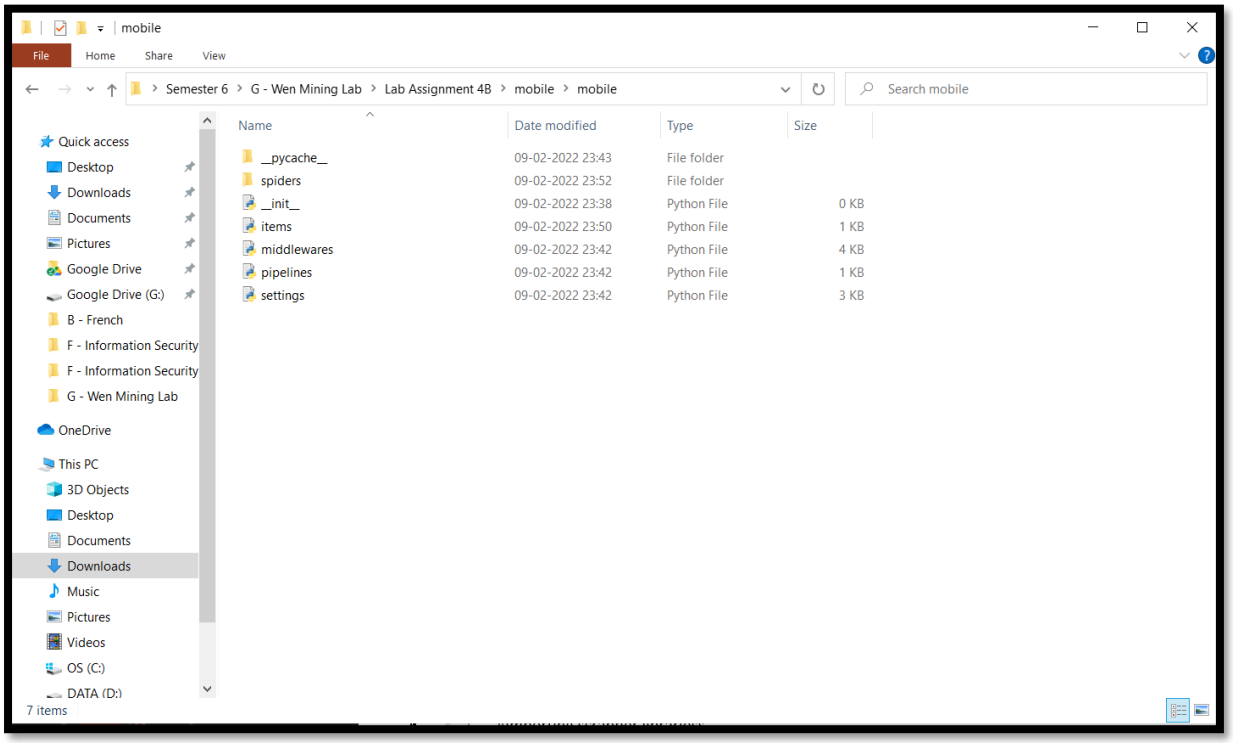
Creating Scrapy Project in mobile directory:

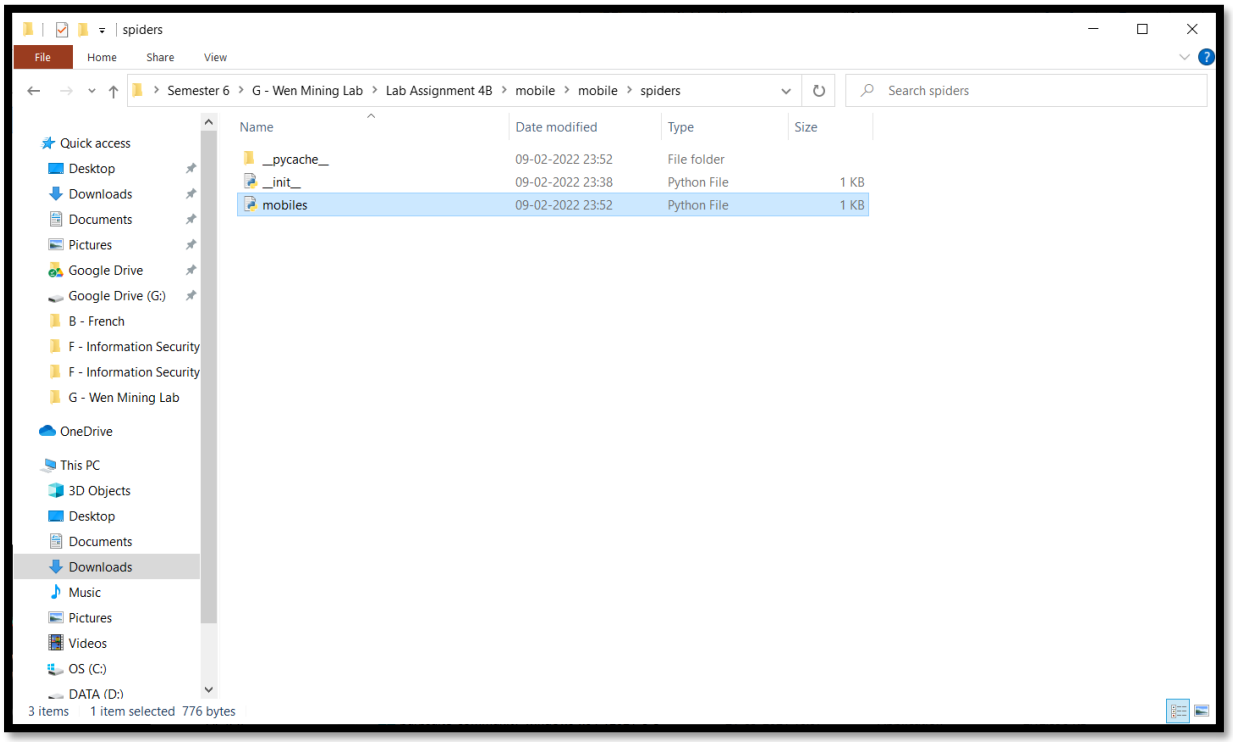


Here we firstly move the directory mobile that we have created.
Then we run the command scrapy genspyder mobiles amazon.com

Harshit Mishra (19BCE0799)

Directory Structure:



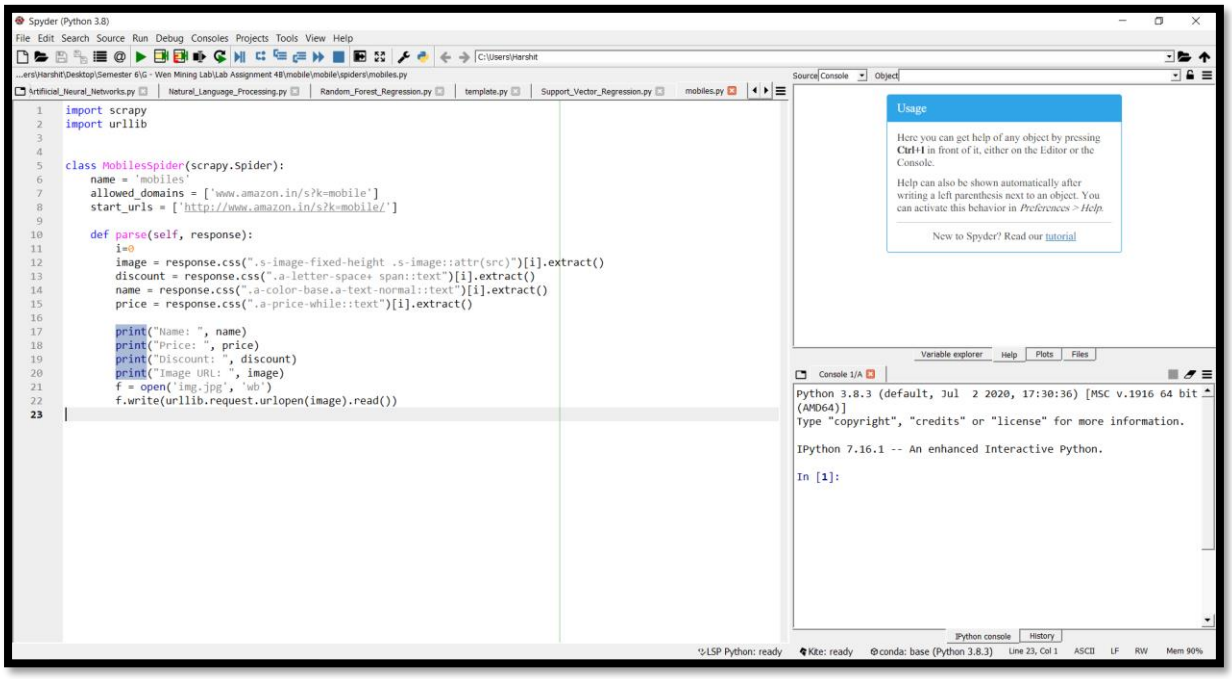


Harshit Mishra (19BCE0799)

Code in mobiles.py:

```
import scrapy
import urllib


class MobilesSpider(scrapy.Spider):
    name = 'mobiles'
    allowed_domains = ['www.amazon.in/s?k=mobile']
    start_urls = ['http://www.amazon.in/s?k=mobile/']

    def parse(self, response):
        i=0
        image = response.css(".s-image-fixed-height .s-image::attr(src)")[i].extract()
        discount = response.css(".a-letter-space+ span::text")[i].extract()
        name = response.css(".a-color-base.a-text-normal::text")[i].extract()
        price = response.css(".a-price-while::text")[i].extract()

        print("Name: ", name)
        print("Price: ", price)
        print("Discount: ", discount)
        print("Image URL: ", image)
        f = open('img.jpg', 'wb')
        f.write(urllib.request.urlopen(image).read())data!! Results in " + Fore.GREEN+"result.csv")
```

Code Snippet:

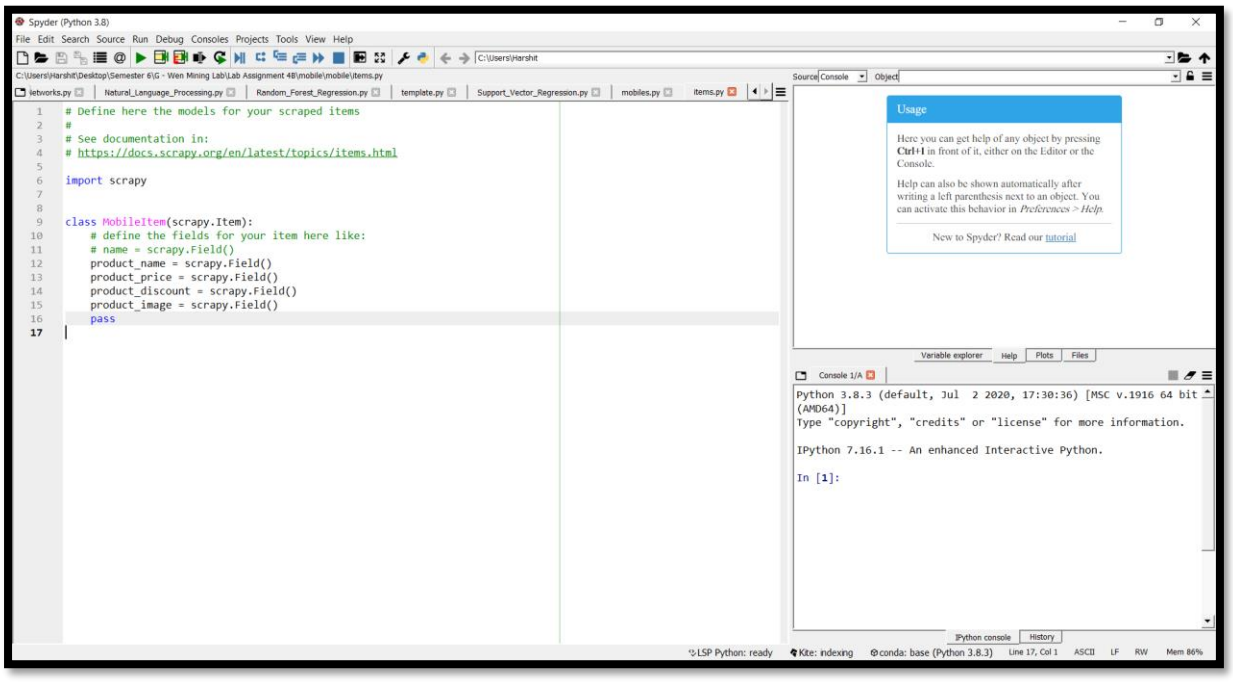


Harshit Mishra (19BCE0799)

Code in Items.py:

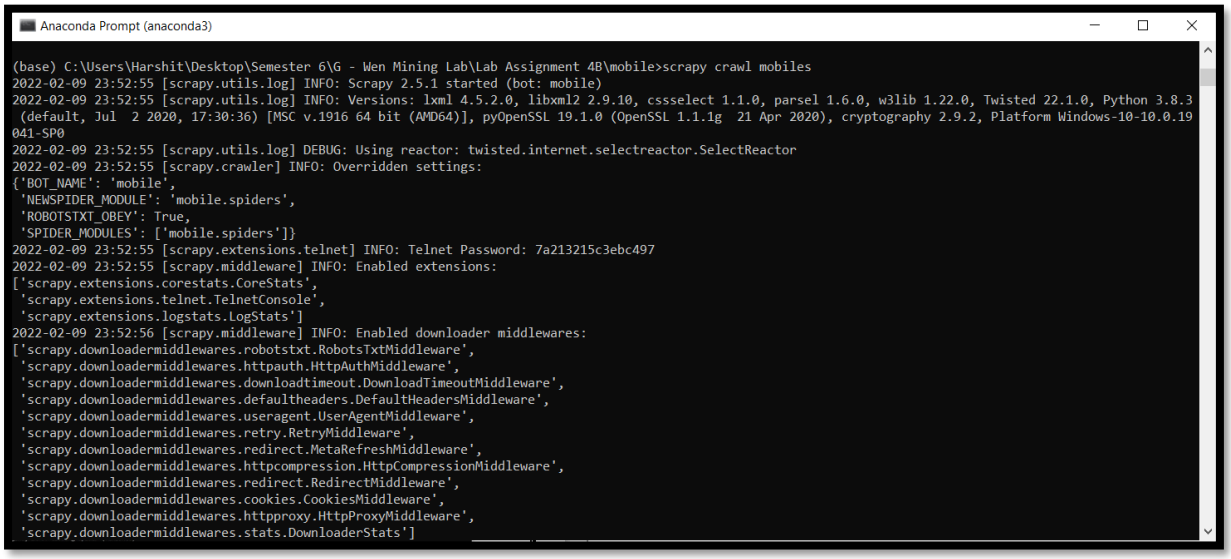
```
# Define here the models for your scraped items
#
# See documentation in:
# https://docs.scrapy.org/en/latest/topics/items.html

import scrapy


class MobileItem(scrapy.Item):
    # define the fields for your item here like:
    # name = scrapy.Field()
    product_name = scrapy.Field()
    product_price = scrapy.Field()
    product_discount = scrapy.Field()
    product_image = scrapy.Field()
```

Code Snippet:



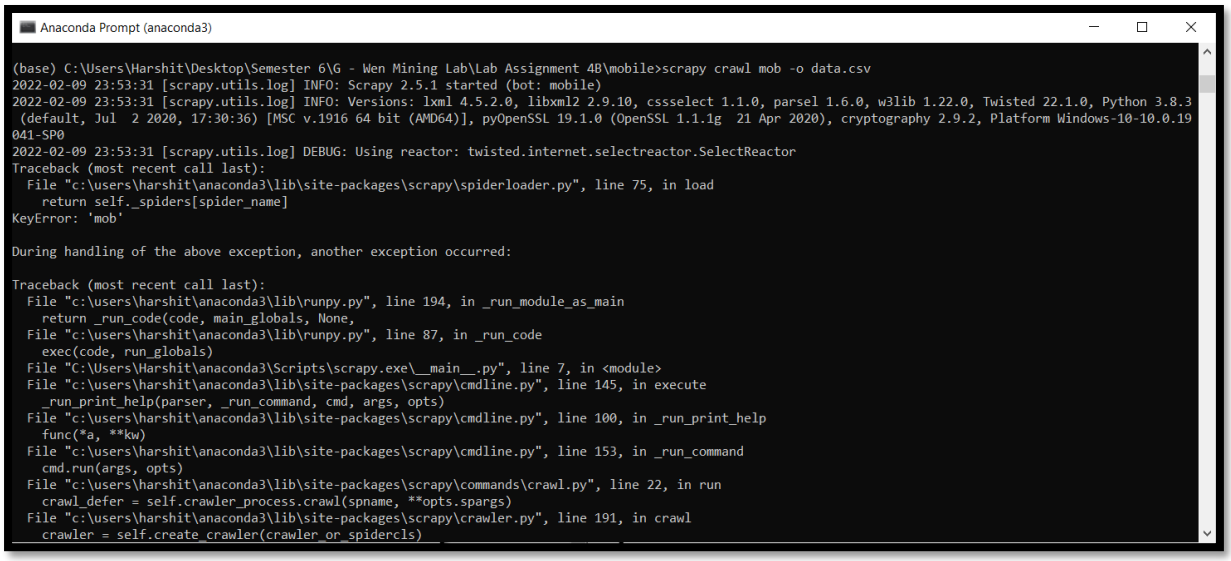
Harshit Mishra (19BCE0799)

Running the code in prompt shell:



Here we use the command ‘scrapy crawl mobiles’ to crawl the web page.

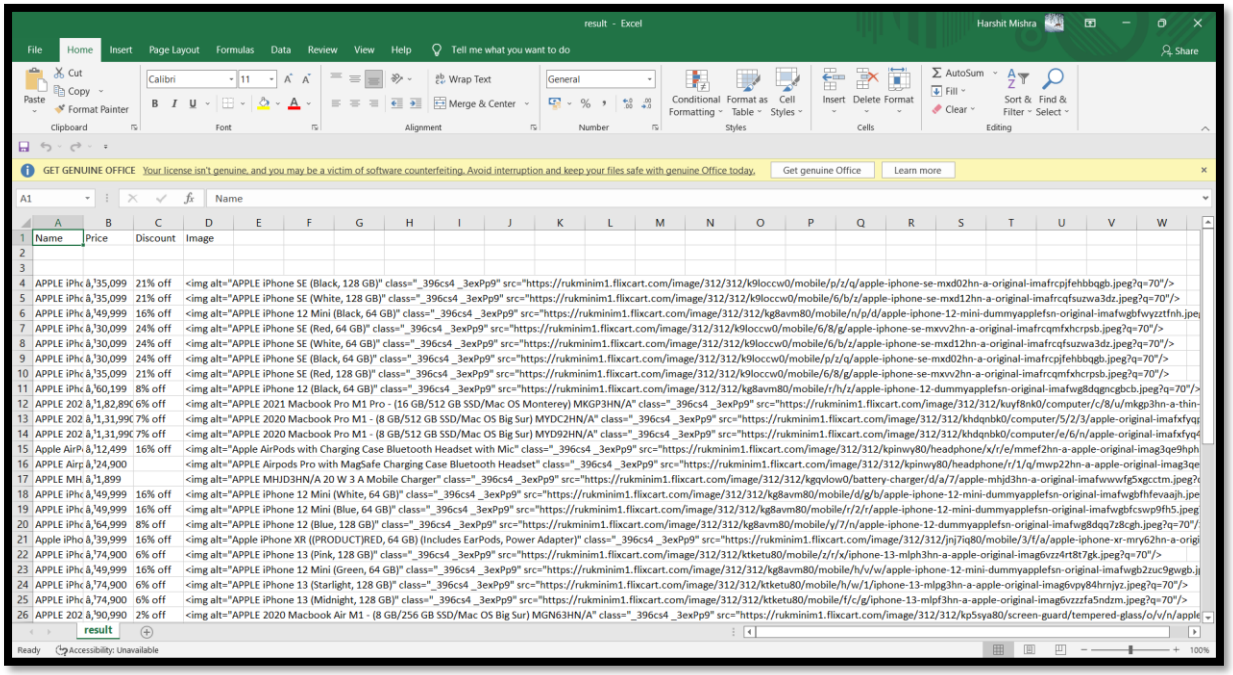
Storing the results of scrapper in a csv file:



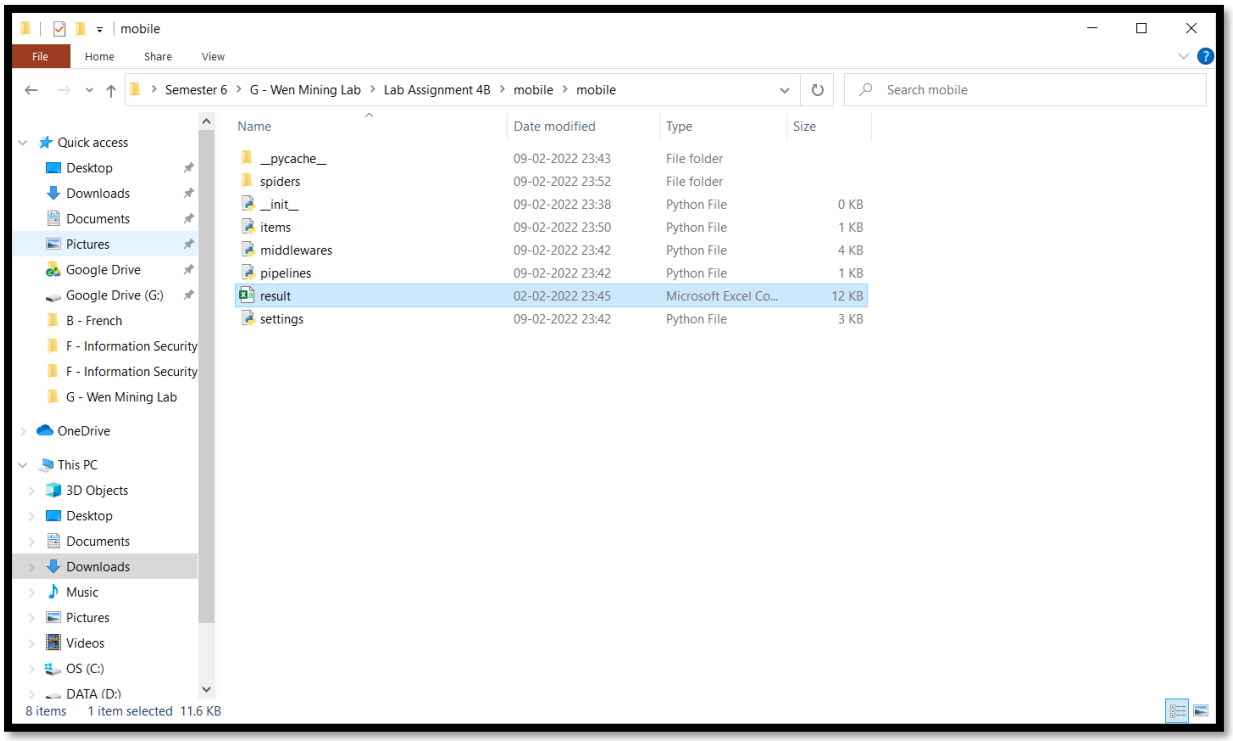
Harshit Mishra (19BCE0799)

Results:

result.csv file:



Directory:



We can see that the result.csv file is dumped in the directory.