Myntra_retrieve_buy_links

In [1]:

```
from selenium import webdriver
import urllib.request
import os
import json
import glob
import pandas as pd
from selenium.webdriver.chrome.options import Options
```

In [2]:

```
def retrieve_links(search_string):
    links = []
    options = Options()
    #options.add_argument('--headless')
    #options.add_argument("user-data-dir=selenium")
    driver = webdriver.Chrome('chromedriver',options=options)
    driver.get('https://www.myntra.com/')
    time.sleep(5)
    driver.find_element_by_class_name('desktop-searchBar').send_keys(search_string)
    driver.find_element_by_class_name('desktop-submit').click()
    for i in range(22):
        time.sleep(5)
        for product_base in driver.find_elements_by_class_name('product-base'):
            links.append(product_base.find_element_by_xpath('./a').get_attribute("href")
)
        try:
            driver.find_element_by_class_name('pagination-next').click()
        except:
            pass
    driver.close()
    driver.quit()
    return links
```

In [197]:

```
Fashion = ["Men Casual Shirt",
"Men Casual shoes",
"Men Casual Trouser",
"Men Formal Shirt",
"Men Formal shoes",
"Men formal trouser",
"Men Jeans",
"Men sneaker",
"Men T-Shirts",
"Women boots",
"Women casual shoes",
"Women flats",
"woman heels",
"Women Jeans",
"Women Tops",
"women shorts skirts",
"women-trousers",
"women boots"]

def Dataframe(item):
    DataFrame = pd.read_csv("Myntra.csv")
    DataFrame[item] = item
    DataFrame.to_csv("Myntra.csv",index=False)
    print(DataFrame.shape)


for item in Fashion:
    item = retrieve_links(item)
```

```
    diff = len(item) - 7533
    if diff == 0:
        Dataframe(item)
    elif diff < 0:
        for i in range(abs(diff)):
            item.append("NA")
        Dataframe(item)
    else:
        for i in range(abs(diff)):
            item.pop()
        Dataframe(item)
```

Download Images and image links from Buy_link

In [ ]:

```
import pandas as pd
import os
from selenium import webdriver
import urllib.request
import os
import json
import pandas as pd
from selenium.webdriver.chrome.options import Options
import time
Myntra = pd.read_csv("Myntra.csv",nrows=3000)
fashion = list(Myntra.columns)
```

In [ ]:

```
### for list of buy link
col = fashion[0]
for i in range(0,10):
    url = Myntra[col][i]
    q=0
    image_links = []
    image_names = []
    driver = webdriver.Chrome("chromedriver.exe")
    driver.get(url)
    time.sleep(5)
    base_path = r'C:\Users\SatishMoparthi\Desktop\Python\Myntra'
    for sesid in driver.find_elements_by_class_name('image-grid-col50'):
        for webele in sesid.find_elements_by_class_name('image-grid-imageContainer'):
            for imli in webele.find_elements_by_class_name('image-grid-image'):
                link = imli.get_attribute('style').split("url(\"")[1].split("\")")[0]
                link = link.replace('h_720,q_90,w_540','h_1440,w_1080')
                name = "image_"+str(i)+str(q)+'.jpg'
                image_path = os.path.join(base_path,col,name)

                if len(link)>0:
                    urllib.request.urlretrieve(link,image_path)
                    image_links.append(link)
                    image_names.append(name)

        q+=1
    js = {"url":url,"links":image_links,"name":image_names}
    path = os.path.join(base_path, col,"image_"+str(i)+'.json')
    with open(path, "w") as outfile:
        json.dump(js, outfile)
        outfile.close()

    driver.close()
    driver.quit()
```

In [ ]:

```
## For a single buy link

url = "https://www.myntra.com/tops/forever-new/forever-new-women-teal-solid-empire-top/10
```

```
758292/buy"
q=0
image_links = []
image_names = []
driver = webdriver.Chrome("chromedriver.exe")
driver.get(url)
time.sleep(5)
base_path = r'C:\Users\SatishMoparthi\Desktop'
for sesid in driver.find_elements_by_class_name('image-grid-col50'):
    for webele in sesid.find_elements_by_class_name('image-grid-imageContainer'):
        for imli in webele.find_elements_by_class_name('image-grid-image'):
            link = imli.get_attribute('style').split("url(\"")[1].split("\")")[0]
            link = link.replace('h_720,q_90,w_540','h_1440,w_1080')
            name = "image_1"+str(q)+'.jpg'
            image_path = os.path.join(base_path,name)

            if len(link)>0:
                urllib.request.urlretrieve(link,image_path)
                image_links.append(link)
                image_names.append(name)

    q+=1
js = {"url":url,"links":image_links,"name":image_names}
path = os.path.join(base_path,"image_"+'.json')
with open(path, "w") as outfile:
    json.dump(js, outfile)
    outfile.close()

driver.close()
driver.quit()
```