

**Name: Vishwajit Shelke**

**roll no: 33172 , Batch : K9**

## **Problem Statement**

Create a review scrapper for any ecommerce website to fetch real time comments, reviews, ratings, comment tags, customer name using Python. Notebook details

## **Imports**

- A. BeautifulSoup (bs4)
- B. pandas
- C. requests

In [3]:

```

# Importing necessary Libraries
import requests
from bs4 import BeautifulSoup
import pandas as pd
import re

# List needed to store all the extracted data
listing=[]

# iterating through 3 pages worth of reviews
for x in range(12,15):

    # request made to receive the Lxml file of that webpage
    r = requests.get(f'https://www.amazon.in/Apple-iPhone-13-128GB-Starlight/product-reviews')

    soup = BeautifulSoup(r.content, 'lxml')

    reviewlist = soup.find('div',class_='a-section a-spacing-none reviews-content a-size-base')

    # extracting all the reviews on that particular page
    review = reviewlist.find_all('div', class_ = "a-section review aok-relative")

    for i in review:

        #####
        # extracting the name of the reviewer
        def reviewer_name():
            try:
                name = i.find('div', class_="a-row a-spacing-mini")
                name_2 = name.find('div',class_='a-profile-content').text
                return name_2
            except:
                return 'none'

        #extracting the rating given by the reviewer
        def reviewer_rating():
            try:
                rating = i.find('span', class_="a-icon-alt").text[:3]
                return rating
            except:
                return 'none'

        # extracting whether the user review has been verified or not
        def reviewer_verified():
            try:
                verified = i.find('span', class_="a-size-mini a-color-state a-text-bold").text
                return verified
            except:
                return 'none'

        # extracting the review title
        def reviewer_comment_title():
            try:
                comment_title = i.find('a', class_="a-size-base a-link-normal review-title a-color-text")
                return comment_title
            except:
                return 'none'

```

```

def reviewer_reviewed():
    try:
        reviewed = i.find('span', class_="a-size-base a-color-secondary review-date")
        return reviewed
    except:
        return 'none'

# additional information like storage size of the phone
def reviewer_extra_info():
    try:
        additional_info = i.find('div', class_="a-row a-spacing-mini review-data review-data-info")
        return additional_info
    except:
        return 'none'

# actual review body
def reviewer_content():
    try:
        content = i.find('span', class_="a-size-base review-text review-text-content")
        return content
    except:
        return 'none'

#extracting linked of images if any of them are present
def reviewer_images():
    try:
        images = i.find('div', class_="a-section a-spacing-top-mini cr-lightbox-image")
        lis=[]
        if(images):
            # for j in images:
            # imi = images.find('img', alt="Customer image")
            # print(imi)
            images_2 = images.find_all('img', class_="cr-lightbox-image-thumbnail")
            for z in images_2:
                lis.append(str(z)[-60:-3])
        if(len(lis)==0):
            return lis
        else:
            return 'none'
    except:
        return 'none'

# storing all of the information in a dictionary format which will be appended in a list
compaining = {
    'name':reviewer_name(),
    'rating':reviewer_rating(),
    'verified': reviewer_verified(),
    'comment_title':reviewer_comment_title(),
    'comment_info':reviewer_content(),
    'extra_info':reviewer_extra_info(),
    'Reviewed or not':reviewer_reviewed(),
    'images_list': reviewer_images()
}

listing.append(compaining)

# print(lis)

df_text = pd.DataFrame(listing)

```

```
df_text.head()
```

Out[3]:

	name	rating	verified	comment_title	comment_info	extra_info	Review or
0	The Tropiquarium	5.0	Verified Purchase	Quality is good but battery life is not so goo...	Don't like Battery life and Charging is very s...	Size name: 128GBColour: BlueVerified Purchase	Review in Indi Novem 2
1	B.CHIDAMBARANATHAN	5.0	Verified Purchase	BEST PHONE	THIS IS THE BEST SMART PHONE I HAVE USED	Size name: 256GBColour: MidnightVerified Purchase	Review in Indi Febr 2
2	Puneet Tiwari	5.0	Verified Purchase	Nice	Nice product and packaging	Size name: 128GBColour: MidnightVerified Purchase	Review in Indi Jan 2
3	A venkateswararao	5.0	Verified Purchase	Fabulous	Fabulous. Especially 4K HDR video recording an...	Size name: 128GBColour: PinkVerified Purchase	Review in Indi Febr 2
4	Praveen kumar	5.0	Verified Purchase	Best ever iPhone	The media could not be loaded.\n ...	Size name: 128GBColour: BlueVerified Purchase	Review in Indi 10 M 2

In [5]:

```
df_text.shape
```

Out[5]:

(30, 8)

In [6]:

```
df_text.to_csv("Extracted_reviews.csv")
```

In [ ]: