Importing all important libraries

In [301…
```
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
import time
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from selenium.common.exceptions import NoSuchElementException
from selenium.common.exceptions import ElementNotInteractableException
from selenium.common.exceptions import StaleElementReferenceException
from selenium.common.exceptions import TimeoutException
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By
from selenium.webdriver.support.select import Select
```

Open Chrome Webdriver

In [302…
```
# ***********open the webdriver through these commands***********
from selenium.webdriver.chrome.service import Service
s = Service('/Users/ADMIN/.wdm/drivers/chromedriver/win32/95.0.4638.69/chromedriver.exe')
driver = webdriver.Chrome(service = s)
# you can add your web driver location
```

Visit Instagram

In [303…
```
driver.get("https://www.instagram.com/")
driver.maximize_window()
waiter = WebDriverWait(driver,10).until(EC.presence_of_element_located((By.NAME,"username")))
```

In [304…
```
SAMPLE_USERNAME = ''
SAMPLE_PASSWORD = ''
username_locate = driver.find_element_by_name('username')
username_locate.send_keys(SAMPLE_USERNAME)
password_locate = driver.find_element_by_name('password')
password_locate.send_keys(SAMPLE_PASSWORD)
driver.find_element_by_xpath('//*[@id="loginForm"]/div/div[3]').click()
waiter = WebDriverWait(driver,10).until(EC.presence_of_element_located((By.CLASS_NAME,'cmbtv')))
```

```
<ipython-input-304-f804fe441525>:3: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  username_locate = driver.find_element_by_name('username')
<ipython-input-304-f804fe441525>:5: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  password_locate = driver.find_element_by_name('password')
<ipython-input-304-f804fe441525>:7: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  driver.find_element_by_xpath('//*[@id="loginForm"]/div/div[3]').click()
```

In [305…
```
driver.find_element_by_class_name('cmbtv').click()
waiter = WebDriverWait(driver,10).until(EC.presence_of_element_located((By.CLASS_NAME,"aOOlW.HoLwm")))
```

```
<ipython-input-305-99dec4f41042>:1: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  driver.find_element_by_class_name('cmbtv').click()
```

In [306…
```
driver.find_element_by_class_name('aOOlW.HoLwm').click()
waiter = WebDriverWait(driver,10).until(EC.presence_of_element_located((By.CLASS_NAME,'pbgfb')))
```

```
<ipython-input-306-24324b0d2e63>:1: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  driver.find_element_by_class_name('aOOlW.HoLwm').click()
```

Enabling the search-box input

In [307…
```
search_box = driver.find_element_by_class_name("pbgfb").click()
waiter = WebDriverWait(driver,10).until(EC.presence_of_element_located((By.CLASS_NAME,"XTCLo.x3qfX")))
```

```
<ipython-input-307-5496befb0cc4>:1: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  search_box = driver.find_element_by_class_name("pbgfb").click()
```

serach for food

In [308…
```
search_box = driver.find_element_by_class_name("XTCLo.x3qfX")
search_box.send_keys("food")
waiter = WebDriverWait(driver,10).until(EC.presence_of_element_located((By.CLASS_NAME,'_7UhW9.xLCgt.qyrsm.KV-D4
```

```
<ipython-input-308-23f43a02b2bf>:1: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  search_box = driver.find_element_by_class_name("XTCLo.x3qfX")
```

In [309…
```
food_list = []
x = driver.find_elements_by_class_name('_7UhW9.xLCgt.qyrsm.KV-D4.uL8Hv')
for i in x:
    if i.get_attribute('innerHTML')[0] != '#':
        food_list.append(i.get_attribute('innerHTML'))
driver.find_element_by_class_name("aIYm8.coreSpriteSearchClear").click()
```

```
<ipython-input-309-2deea4065499>:2: DeprecationWarning: find_elements_by_* commands are deprecated. Please use
find_elements() instead
  x = driver.find_elements_by_class_name('_7UhW9.xLCgt.qyrsm.KV-D4.uL8Hv')
<ipython-input-309-2deea4065499>:6: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  driver.find_element_by_class_name("aIYm8.coreSpriteSearchClear").click()
```

Find out the followers of the pages

In [310…
```
food_list = food_list[:10]
no_of_follower = []
for i in food_list:
    driver.find_element_by_class_name("pbgfb").click()
    waiter = WebDriverWait(driver,10).until(EC.presence_of_element_located((By.CLASS_NAME,"XTCLo.x3qfX")))
    search_box = driver.find_element_by_class_name("XTCLo.x3qfX")
    search_box.send_keys(i)
    time.sleep(3)
    x = driver.find_elements(By.XPATH,"//a[@class = '-qQT3']/div[1]/div[2]/div[1]/div[1]/div[1]/div")
    for j in x:
        if j.get_attribute('innerHTML') == i:
            j.click()
            break
    time.sleep(3)
    follower = int(driver.find_element(By.XPATH,'//*[@id="react-root"]/section/main/div/header/section/ul/li[2]
    no_of_follower.append(follower)
    driver.back()
    time.sleep(3)
```

```
<ipython-input-310-55dac1e2f8f9>:4: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  driver.find_element_by_class_name("pbgfb").click()
<ipython-input-310-55dac1e2f8f9>:6: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  search_box = driver.find_element_by_class_name("XTCLo.x3qfX")
```

In [311…
```
dict_food_list = {}
for i in range(len(food_list)):
    dict_food_list[food_list[i]] = no_of_follower[i]
dict_food_list
```

Out[311…
```
{'dilsefoodie': 944377,
 'foodie_incarnate': 897267,
 'food_blogger290': 14329,
 'foodies_findings': 279818,
 'foodfusionpk': 1516370,
 'food_lunatic': 124838,
 'food': 151894,
 'yourfoodlab': 1843759,
 'delhifoodwalks': 156558,
 'tasty.foodrecipes': 6880}
```

Top 5 Food pages

In [312…
```
x = list(reversed(sorted(dict_food_list.values())))
top_5 = {}
for i in x[:5]:
    for key in dict_food_list:
        if dict_food_list[key] == i:
            top_5[key] = i
print('Top 5 food pages out of first 10 selected from the "food" list is')
for key in top_5:
    print(key)
```

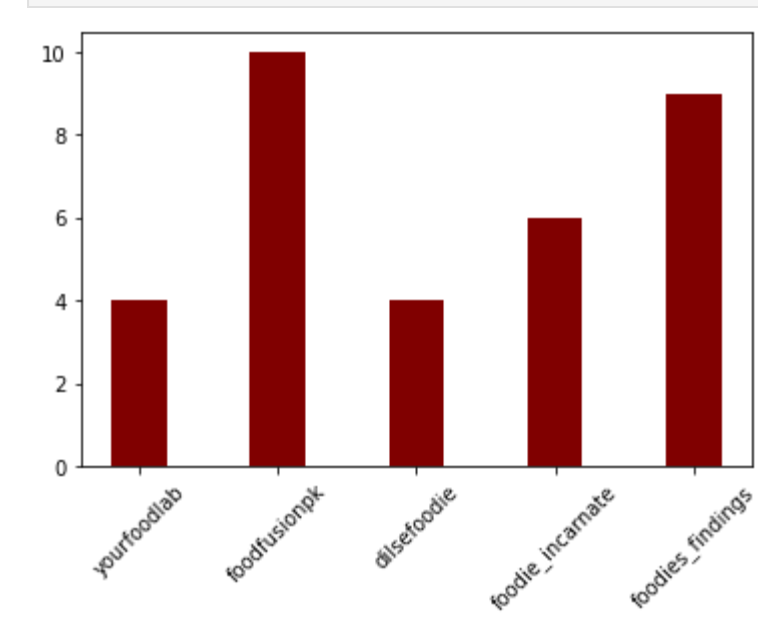
```
Top 5 food pages out of first 10 selected from the "food" list is
yourfoodlab
foodfusionpk
dilsefoodie
foodie_incarnate
foodies_findings
```

In [313…
```
handle = {}
for i in top_5.keys():
    driver.find_element_by_class_name("pbgfb").click()
    waiter = WebDriverWait(driver,10).until(EC.presence_of_element_located((By.CLASS_NAME,"XTCLo.x3qfX")))
    search_box = driver.find_element_by_class_name("XTCLo.x3qfX")
    search_box.send_keys(i)
    time.sleep(3)
    x = driver.find_elements(By.XPATH,"//a[@class = '-qQT3']/div[1]/div[2]/div[1]/div[1]/div[1]/div")
    for j in x:
        if j.get_attribute('innerHTML') == i:
            j.click()
            break
    time.sleep(2)
    for k in range(5):
        driver.execute_script('window.scrollTo(0,document.body.scrollHeight);')
        time.sleep(2)
    driver.execute_script('window.scrollTo(0,0);')
    time.sleep(1)
    all_photos = driver.find_elements(By.XPATH,'//div[contains(@class, "Nnq7C")]/div/a/div[1]')
    photo_count = 0
    for photo in all_photos:
        try:
            photo.click()
        except StaleElementReferenceException:
            continue
        time.sleep(2)
        try:
            x = driver.find_element(By.CLASS_NAME,'_1o9PC.Nzb55').get_attribute('innerHTML')
        except NoSuchElementException:
            continue
        x = list(x.strip().split(' '))
        if x[1] == 'hours':
            photo_count += 1
        elif x[1] == 'hour':
            photo_count += 1
        elif x[1] == 'minutes':
            photo_count +=1
        elif x[1] == 'day':
            photo_count += 1
        elif x[1] == 'days':
            if int(x[0]) <= 3:
                photo_count += 1
            else:
                driver.back()
                break
        else:
            driver.back()
        driver.back()
        time.sleep(2)

    handle[i] = photo_count
```

```
<ipython-input-313-243c90c861e5>:3: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  driver.find_element_by_class_name("pbgfb").click()
<ipython-input-313-243c90c861e5>:5: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  search_box = driver.find_element_by_class_name("XTCLo.x3qfX")
```

Plot the bar graph for pages for posts in last 3 days.

In [314…
```
food_handler = []
posts = []
for key in handle:
    food_handler.append(key)
    posts.append(handle[key])

plt.bar(food_handler,posts,color='maroon',width = 0.4)
plt.xticks(rotation =45)
plt.show()
```



Scrapping Hashtags

In [315…
```
# scraping all the hashtags from these photos
list_of_all_hashtags = []
for i in top_5.keys():
    driver.find_element_by_class_name("pbgfb").click()
    waiter = WebDriverWait(driver,10).until(EC.presence_of_element_located((By.CLASS_NAME,"XTCLo.x3qfX")))
    search_box = driver.find_element_by_class_name("XTCLo.x3qfX")
    search_box.send_keys(i)
    time.sleep(3)
    x = driver.find_elements(By.XPATH,"//a[@class = '-qQT3']/div[1]/div[2]/div[1]/div[1]/div[1]/div")
    for j in x:
        if j.get_attribute('innerHTML') == i:
            j.click()
            break
    time.sleep(2)
    for k in range(5):
        driver.execute_script('window.scrollTo(0,document.body.scrollHeight);')
        time.sleep(2)
    driver.execute_script('window.scrollTo(0,0);')
    time.sleep(2)
    all_photos = driver.find_elements(By.XPATH,'//div[contains(@class, "Nnq7C")]/div/a/div[1]')
    photo_count = 0
    for photo in all_photos:
        if photo_count == 10: # As only latest 10 photos need to be scrapped
            break
        photo.click()
        photo_count += 1
        time.sleep(2)
        all_hashtags = driver.find_elements(By.XPATH,'//a[@class = " xil3i"]')
        for hashtag in all_hashtags:
            list_of_all_hashtags.append(hashtag.get_attribute('innerHTML'))
        driver.back()
        time.sleep(2)
```

```
<ipython-input-315-1cb2d7346559>:4: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  driver.find_element_by_class_name("pbgfb").click()
<ipython-input-315-1cb2d7346559>:6: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  search_box = driver.find_element_by_class_name("XTCLo.x3qfX")
```

In [316…
```
hashtag_freq = {}
for key in list_of_all_hashtags:
    if key in hashtag_freq:
        hashtag_freq[key] += 1
    else:
        hashtag_freq[key] = 1

hashtag = []
frequency = []
for key in hashtag_freq:
    hashtag.append(key)
    frequency.append(hashtag_freq[key])

data = pd.DataFrame(list(zip(hashtag,frequency)),columns =['Hashtags','Frequency'])
data.to_csv('hashtags.csv',index = False)
```

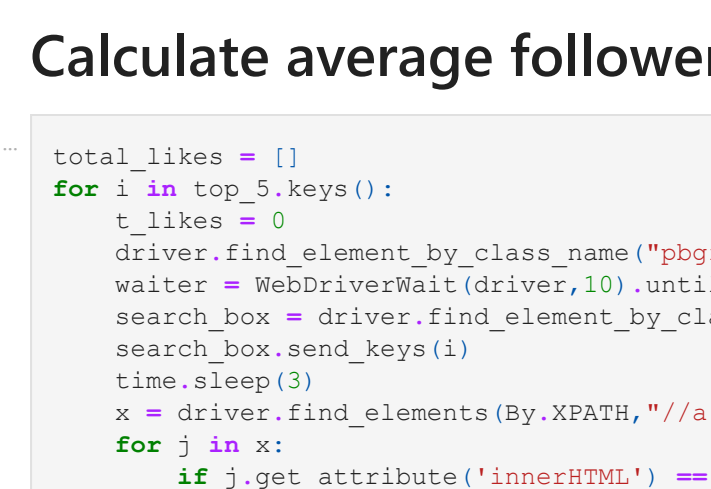
Top 5 Hashtags

In [317…
```
data = data.sort_values(by =['Frequency'],ascending = False)
print("Top 5 Hashtags are:")
for tags in data['Hashtags'][0:5]:
    print(tags)
```

```
Top 5 Hashtags are:
#foodporn
#HappyCookingToYou
#FoodFusion
#recipe
#streetfood
```

Pie graph for these Hashtags

In [318…
```
plt.pie(data['Frequency'][0:5],labels = data['Hashtags'][0:5],autopct = "%.2f%%")
# plt.legend(title = 'Hashtags',loc = 'right')
plt.show()
```



Calculate average followers : likes ratio for the obtained handles

In [320…
```
total_likes = []
for i in top_5.keys():
    t_likes = 0
    driver.find_element_by_class_name("pbgfb").click()
    waiter = WebDriverWait(driver,10).until(EC.presence_of_element_located((By.CLASS_NAME,"XTCLo.x3qfX")))
    search_box = driver.find_element_by_class_name("XTCLo.x3qfX")
    search_box.send_keys(i)
    time.sleep(3)
    x = driver.find_elements(By.XPATH,"//a[@class = '-qQT3']/div[1]/div[2]/div[1]/div[1]/div[1]/div")
    for j in x:
        if j.get_attribute('innerHTML') == i:
            j.click()
            break
    time.sleep(2)
    for k in range(5):
        driver.execute_script('window.scrollTo(0,document.body.scrollHeight);')
        time.sleep(2)
    driver.execute_script('window.scrollTo(0,0);')
    time.sleep(2)
    all_photos = driver.find_elements(By.XPATH,'//div[contains(@class, "Nnq7C")]/div/a/div[1]')
    photo_count = 0
    for photo in all_photos:
        if photo_count == 10: # As only latest 10 photos need to be scrapped
            break
        photo.click()
        photo_count += 1
        time.sleep(2)
        try:
            like = int(driver.find_element(By.XPATH,'//a[@class= "zV_Nj"]/span').get_attribute('innerHTML').rep
        except NoSuchElementException:
            driver.find_element(By.XPATH,'//span[@class = "vcOH2"]').click()
            like = int(driver.find_element(By.XPATH,'//div[@class= "vJRqr"]/span').get_attribute('innerHTML').r

        t_likes += like
        driver.back()
    total_likes.append(t_likes)
```

```
<ipython-input-320-0cee9afcf831>:4: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  driver.find_element_by_class_name("pbgfb").click()
<ipython-input-320-0cee9afcf831>:6: DeprecationWarning: find_element_by_* commands are deprecated. Please use f
ind_element() instead
  search_box = driver.find_element_by_class_name("XTCLo.x3qfX")
```

Total Likes for each handle

In [321…
```
print('Total likes for the top 10 posts of the profiles are:-')
for i in range(len(food_handler)):
    print(food_handler[i],'  ',total_likes[i])
```

```
Total likes for the top 10 posts of the profiles are:-
yourfoodlab    191433
foodfusionpk    35496
dilsefoodie    175837
foodie_incarnate    409417
foodies_findings    99212
```

Average Likes of each Handle

In [322…
```
average_likes = np.array(total_likes)
average_likes = average_likes/10
print('Average likes for the top 10 posts of the profiles are:-')
for i in range(len(food_handler)):
    print(food_handler[i],'  ',average_likes[i])
```

```
Average likes for the top 10 posts of the profiles are:-
yourfoodlab    19143.3
foodfusionpk    3549.6
dilsefoodie    17583.7
foodie_incarnate    40941.7
foodies_findings    9921.2
```

Calculate the like-ratio and plot the bar graph for this

In [323…
```
followers = []
for value in top_5.values():
    followers.append(value)
followers = np.array(followers)
like_ratio = followers/average_likes
plt.bar(food_handler,like_ratio,color = 'green',width = 0.4)
plt.xticks(rotation = 45)
plt.show()
```