In [1]:

```
import requests
from bs4 import BeautifulSoup
from selenium import webdriver
```

In [2]:

```
#pip install webdriver-manager
#from webdriver_manager.chrome import ChromeDriverManager

#driver = webdriver.Chrome(ChromeDriverManager().install())
```

In [3]:

```
import re

def cleanhtml(raw_html):
    cleanr = re.compile('<.*?>')
    cleantext = re.sub(cleanr, '', raw_html)
    cleantext=  re.sub('^\nRated\xa0\n +', '', cleantext)
    cleantext=  re.sub('\n+', ' ', cleantext)
    cleantext=  re.sub('\.+', '.', cleantext)
    #cleantext=  re.sub('[^A-Za-z0-9 ]+', '', cleantext)


    return cleantext


import string
def get_comment(soup):
    z4=[]
    x=soup.find_all("div",{"class": ["rev-text mbot0","rev-text mbot0 hidden"]})
    for i in range(0,len(x)):
        #print('comment ',i)
        z=cleanhtml(x[i].get_text()).split('.')
        #print(z)
        try:
            rating=x[i].find("div").attrs['aria-label']
        except KeyError:
            break;

        #print(rating)
        z4.append((rating,z))
    return(z4)
```

In [4]:

```
d={1:'hi'}
for i in range(1,3):
    try:
        print(d[i])
    except KeyError:
        break;
print("key not found")
```

```
hi
key not found
```

In [5]:

```
from selenium import webdriver
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.common.by import By
from selenium.webdriver.support import expected_conditions as EC
import pandas as pd
from selenium.common.exceptions import NoSuchElementException
import time
```

```
#list_of_restaurents=['patang','we-idliwale'] #'1bhk-superbar','Effingut-Brewerkz','highland']
#link="https://www.zomato.com/pune/restaurent-baner/reviews"
link="https://www.zomato.com/pune/restaurent-wakad/reviews"

def click_loadmore(l,chrome_options=None):
    if(chrome_options != None):
        driver = webdriver.Chrome(options=chrome_options,executable_path='C:\\Users\\SHAILENDRA\\.wdm\\drivers\\chromedriver\\78.0.3904.70\\win32\\chromedriver.exe')
    else:
        driver = webdriver.Chrome(executable_path='C:\\Users\\SHAILENDRA\\.wdm\\drivers\\chromedriver\\78.0.3904.70\\win32\\chromedriver.exe')

    driver.get(link.replace('restaurent',l))
    time.sleep(100)
    element = driver.find_element_by_xpath("//div[@class='review-sorting text-tabs selectors ui secondary pointing menu mt0']")
    element=element.find_elements_by_xpath("//a[@class='item default-section-title everyone empty']")
        # CLICKING ON ALL REVIEWS WOOHOO
    time.sleep(100)
    element[0].click()
    print(element[0].text)
    time.sleep(15)


    I=0
    while True:
        try:
            loadMoreButton = driver.find_element_by_xpath('//div[@class="load-more bold ttupper tac cursor-pointer fontsize2"]')
            time.sleep(10)
            loadMoreButton.click()
            time.sleep(15)
            I+=1
        except Exception as e:
            print(e)
            break
        #print(I,'\n')
    print("Complete")

    all_df=[]
    content1=driver.page_source
    soup=BeautifulSoup(content1,"html.parser")
    rating=[]
    comment=[]
    res_name=[]
    temp=get_comment(soup)
    #print(temp)
    for j,i in temp:
        #print(i)
        a=i[0].strip()
        if(len(a)!=0):
            #print(a)
            comment.append( ".".join(i))
            rating.append(j)
            res_name.append(l)
        #final=pd.DataFrame([res_name,rating,comment],columns=['res_name','rating','comment']).T
        #all_df.append
    driver.quit()
    time.sleep(15)
    return([res_name,rating,comment])
```

In [ ]:

```
o=1
import requests
from bs4 import BeautifulSoup
from selenium import webdriver
list_of_restaurents=['the-barkhana']#'the-break-room','thikana','highland', # names of the restaurents in baner
PROXY = ['182.74.243.39:3128'] # IP:PORT or HOST:PORT this list needs to be updated from https://free-proxy-list.net/
p={}
for i in range(0,len(PROXY)):
    p[i]=PROXY[i]
p
for i in range(0,len(list_of_restaurents)):
```

```
    chrome_options = webdriver.ChromeOptions()
    chrome_options.add_argument('--proxy-server=%s' % p[i])
    if(i==0):
        z=click_loadmore(list_of_restaurents[i],chrome_options)
    else:
        z=click_loadmore(list_of_restaurents[i],chrome_options)

    df=pd.DataFrame(z).T
    df.columns=['res_name','rating','comment']
    df.to_csv("{}{}.csv".format('C:\\Users\\SHAILENDRA\\',list_of_restaurents[i] ),index=False)
```

```
All Reviews
1292
```

In [ ]:

```
df
```

In [ ]:

```

```

In [ ]:

```

```