Infinite_Image_Scroll_Scrapper_1

February 2, 2025

[21]:
```
#Importing Libraries.
import time
import requests
import pandas as pd
from tqdm import tqdm

from bs4 import BeautifulSoup
from selenium import webdriver

# Selenium
browser = webdriver.Chrome()
browser.get('https://stockmages.netlify.app')

#Now the main point of concern is what should be the stop and for the for loop.␣
 ↪The start should be 0.
#Finding the stop is a manual process. We should use a random number for stops␣
 ↪and see how much scrolling takes place. If it doesn't scroll to the bottom,␣
 ↪we have to increase the value of the stop.
#Infinite Scroll to reach the bottom
for i in tqdm(range(0,5000000,1000)):
    browser.execute_script("window.scrollTo(0,"  + str(i) + ")")
    time.sleep(.1)
#Using this much range will lead ypu to the bottom of the path
```

```
100%|
                    | 5000/5000 [08:56<00:00,  9.32it/s]
```

[29]:
```
# Scraping the Page
soup = BeautifulSoup(browser.page_source,'html.parser')

#Scraping the Image Details
data = []
for sp in tqdm(soup.find_all('div', class_ = 'container')):
    img_link = sp.find('img').get('src')
    tags     = sp.find('span', class_ = 'tag-color').text[7:].strip()
    likes    = int(sp.find('div', class_ = 'likes-comments').
 ↪find_all('span')[0].text.strip()[:-6])
```

```
    comments = int(sp.find('div', class_ = 'likes-comments').
 ↪find_all('span')[1].text.strip()[:-9])
    data.append([img_link, tags, likes, comments])

#Creating a dataframe and saving it
df = pd.DataFrame(data, columns = ['img_link', 'tags', 'likes', 'comments'])
df.to_csv('images.csv', index = False)
```

```
100%|
                    | 481/481 [00:00<00:00, 13143.49it/s]
```

[ ]:

[ ]: