Author: Shashank Chhoker

Stock Image Scraper

In [ ]:
```
# Import necessary libraries
import numpy as np
import pandas as pd

from bs4 import BeautifulSoup
from selenium import webdriver
import chromedriver_binary
```

In [ ]:
```
# Specify the URL you want to scrape
link = 'https://stock-pictures.netlify.app'
```

In [ ]:
```
# Initialize a Chrome WebDriver
driver = webdriver.Chrome()
driver.get(link)
```

In [ ]:
```
# Parse the webpage content with BeautifulSoup
soup= BeautifulSoup(driver.page_source,"html.parser")
```

In [ ]:
```
# Make the HTML content more readable (prettify)
soup.prettify
```

In [ ]:
```
# Create an empty list to store the scraped data
data=[]
for sp in soup.find_all('div',class_='container'):



    if ('gif' not in sp.find('img').get('src')):

        link=sp.find('img').get('src')

        tags = list(set(sp.find('div',class_ = 'tags').text[7:].split(' ')))

        likes = int(sp.find('div',class_ = 'likes-comments').find_all('span')
        comments =int(sp.find('div',class_ = 'likes-comments').find_all('span

        sp.find(' ')
        data.append([link, ' '.join(tags),likes,comments])
```

In [ ]:
```
# Create an empty list to store the scraped data
data
```

In [ ]:
```
# Loop through all the div elements with the class 'container'
```

In [ ]:
```
"""
for sp in soup.find_all('div',class_ = 'container'):
    tags = list(set(sp.find('div',class_ = 'tags').text[7:].split(' ')))
    print(tags)

"""
```

In [ ]:
```
"""
for sp in soup.find_all('div',class_ = 'container'):
    likes = int(sp.find('div',class_ = 'likes-comments').find_all('span')[0].
    comments =int(sp.find('div',class_ = 'likes-comments').find_all('span')[1
    print(likes,comments)
    #break
"""
```

In [ ]:
```
# Create a Pandas DataFrame from the collected data
df=pd.DataFrame(data,columns = ['Link', 'Tags' ,'Likes','Comments'])
```

In [ ]:
```
# Display the DataFrame
df
```

In [ ]:
```
# Check for any missing data (null values)
df.isnull().sum()
```

In [ ]:
```
# Save the DataFrame to a CSV file
df
```

In [ ]:
```
df.to_csv('data.csv',index=False)
```

In [ ]:
```

```