

```

from bs4 import BeautifulSoup
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
import time
import requests
from bs4 import BeautifulSoup

```

## ▼ Geeks for Geeks

### Problem 2: Geeksforgeeks Data analysis

Scrape the Geeksforgeeks youtube channel videos of the past 6 months' dataset

Submitted by Abhilash V A - 9880507105 - [aabhilash070@gmail.com](mailto:aabhilash070@gmail.com)

## ▼ Scraping the geeks for geeks youtube page and collection the data

+ Code + Text

```

driver_path = "chromedriver_win32_2\\chromedriver.exe"
driver = webdriver.Chrome(executable_path=driver_path)
url = "https://www.youtube.com/@GeeksforGeeksVideos/videos"
driver.get(url)
scrolls = 4

for _ in range(scrolls):
    driver.find_element(By.TAG_NAME, 'body').send_keys(Keys.END)
    time.sleep(2)

video_title_links = []

video_elements = driver.find_elements(By.ID, "video-title-link")

for video_element in video_elements:
    video_title = video_element.text.strip()
    video_url = video_element.get_attribute("href")
    video_text = video_element.get_attribute("aria-label")
    video_data = {'title': video_title, 'url': video_url, 'text': video_text}
    video_title_links.append(video_data)

urls_data=[]
for url in video_title_links:
    try:
        driver.get(url['url'])
        time.sleep(12)
        button=driver.find_element(By.ID,"expand")
        button.click()

```

```

        button1=v_page[2].text
        urls_data.append(button1)
        url["date_data"]=button1
        print(url)
        print("-----")
except Exception as e:
    print(e)
    print("error")
    url["date_data"]="NA"
time.sleep(2)

driver.quit()

<ipython-input-5-1562be072ed6>:2: DeprecationWarning: executable_path has been deprecated, please pass in a Service object
  driver = webdriver.Chrome(executable_path=driver_path)
{'title': '10 TIPS and TRICKS to Crack Internships and Placements | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=2QDMIZGeIC4', 'text': '10 TIPS and TRICKS to Crack Internships and Placements | GeeksforGeeks by GeeksforGeeks 9 days ago 46 seconds'}
-----
{'title': 'Complete School Guide for CBSE | Free Resources For School Students | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=1ECMLxC6jbs', 'text': 'Complete School Guide for CBSE | Free Resources For School Students | GeeksforGeeks by GeeksforGeeks 9 days ago 46 seconds'}
-----
{'title': 'Master System Design | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=vjWLZYeqiN8', 'text': 'Master System Design | GeeksforGeeks by GeeksforGeeks 9 days ago 46 seconds'}
-----
{'title': 'Unraveling the Geek-a-thon: The What, How, and Why | Geek-A-Thon | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=6e0v_NhsFMQ', 'text': 'Unraveling the Geek-a-thon: The What, How, and Why | Geek-A-Thon | GeeksforGeeks by GeeksforGeeks 9 days ago 46 seconds'}
-----
{'title': 'Data Analysis Masterclass | Geek-A-Thon | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=-7Emxpn8DXc', 'text': 'Data Analysis Masterclass | Geek-A-Thon | GeeksforGeeks by GeeksforGeeks 9 days ago 46 seconds'}
-----
{'title': 'GeeksforGeeks Offline Classroom Program | Now Open In Noida & Gurgaon', 'url': 'https://www.youtube.com/watch?v=2bPiFkt3Nmg', 'text': 'GeeksforGeeks Offline Classroom Program | Now Open In Noida & Gurgaon | GeeksforGeeks by GeeksforGeeks 9 days ago 46 seconds'}
-----
{'title': 'Kickstart your DevOps Career | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=8gwwlpBwr_s', 'text': 'Kickstart your DevOps Career | GeeksforGeeks by GeeksforGeeks 9 days ago 46 seconds'}
-----
{'title': 'First Ever KBC at our Offline Classes | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=chYg2EAKW2A', 'text': 'First Ever KBC at our Offline Classes | GeeksforGeeks by GeeksforGeeks 9 days ago 46 seconds'}
-----
{'title': 'Mega Job-a-thon for Working Professionals | 21st July | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=CyOh-WDFUys', 'text': 'Mega Job-a-thon for Working Professionals | 21st July | GeeksforGeeks by GeeksforGeeks 9 days ago 46 seconds'}
-----
{'title': 'Explore GeeksforGeeks Hiring Solutions | For Top Tech Companies | Innovative Hiring Solutions', 'url': 'https://www.youtube.com/watch?v=GDeoPHRqb5w', 'text': 'Explore GeeksforGeeks Hiring Solutions | For Top Tech Companies | Innovative Hiring Solutions | GeeksforGeeks by GeeksforGeeks 9 days ago 46 seconds'}
-----
{'title': 'Geek-O-Lympics 2023 | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=hLqjiIu6LzY', 'text': 'Geek-O-Lympics 2023 | GeeksforGeeks by GeeksforGeeks 1 month ago 46 seconds'}
-----
{'title': 'Web 3.0 Explained | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=zEPRJMK3p90', 'text': 'Web 3.0 Explained | GeeksforGeeks by GeeksforGeeks 1 month ago 17 minutes'}
-----
{'title': 'Geek-O-Lympics 2023 | 1st - 31st July | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=GSqepfZLKU0', 'text': 'Geek-O-Lympics 2023 | 1st - 31st July | GeeksforGeeks by GeeksforGeeks 1 month ago 46 seconds'}
-----
{'title': 'What Is Web Scraping & What It Is Used For | How To Use Web Scraping | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=MK15a0ipg-E', 'text': 'What Is Web Scraping & What It Is Used For | How To Use Web Scraping | GeeksforGeeks by GeeksforGeeks 1 month ago 46 seconds'}
-----
{'title': 'Mega Job-A-Thon | Hire Top Talent for Your Company | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=cWPqVVobUuA', 'text': 'Mega Job-A-Thon | Hire Top Talent for Your Company | GeeksforGeeks by GeeksforGeeks 1 month ago 46 seconds'}
-----
{'title': 'DSA To Development: A Complete Coding Guide | Complete Coding Program By GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=phr81XH2Ee0', 'text': 'DSA To Development: A Complete Coding Guide | Complete Coding Program By GeeksforGeeks by GeeksforGeeks 1 month ago 46 seconds'}
-----
{'title': 'What is Queue (Updated) | Queues Explained | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=ypJwoz_SXTo', 'text': 'What is Queue (Updated) | Queues Explained | GeeksforGeeks by GeeksforGeeks 1 month ago 46 seconds'}
-----
{'title': 'Queue Practice Questions (Updated) | DSA Concepts | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=KG4dbF5xRig', 'text': 'Queue Practice Questions (Updated) | DSA Concepts | GeeksforGeeks by GeeksforGeeks 1 month ago 46 seconds'}
-----
{'title': 'Free Summer Offline Classes on Python Programing | For Students Aged 14-21 | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=OOLXHwZzHfM', 'text': 'Free Summer Offline Classes on Python Programing | For Students Aged 14-21 | GeeksforGeeks by GeeksforGeeks 1 month ago 46 seconds'}
-----
{'title': 'Python Libraries | Data Science | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=nDa75spDmeE', 'text': 'Python Libraries | Data Science | GeeksforGeeks by GeeksforGeeks 1 month ago 46 seconds'}

```

```

-----
{'title': 'ITR Filing Made Simple | Documents require to file ITR | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=SWa-tsbaIpg', 'text': 'ITR Filing Made Simple | Documents'}
-----
{'title': 'Full Stack Development | LIVE Classes | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=cBfC9HLR9Qk', 'text': 'Full Stack Development | LIVE Classes | GeeksforGeeks'}
-----
{'title': 'GeeksforGeeks Classroom Program in NCR', 'url': 'https://www.youtube.com/watch?v=uEhnTpHY26o', 'text': 'GeeksforGeeks Classroom Program in NCR by GeeksforGeeks 2 months ago'}
-----
{'title': 'What is Arrays | Arrays Explained | New Video - GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=3_x_Fb31NLE', 'text': 'What is Arrays | Arrays Explained | New Video'}
-----
{'title': 'Arrays Practice Questions (Updated) | In C++/JAVA/Python | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=J7EhXvniXRm', 'text': 'Arrays Practice Questions (Updated)'}
-----
{'title': 'Gate 2024: Prepare for Success! | GATE CSE 2024 | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=G-3vZKPETKw', 'text': 'Gate 2024: Prepare for Success! | GATE CSE'}
-----
{'title': 'Get Started with CP | Competitive Programming Tips | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=FCEj-mreXvI', 'text': 'Get Started with CP | Competitive Programming'}
-----
{'title': 'React VS Angular VS Vue | GeeksforGeeks', 'url': 'https://www.youtube.com/watch?v=sA2vgEP-f7c', 'text': 'React VS Angular VS Vue | GeeksforGeeks by GeeksforGeeks 2 months ago'}
-----

```

### ▼ saving this data in json format

```

import json

filename = "data_json.json"

with open(filename, "w") as json_file:
    json.dump(video_title_links, json_file)

from datetime import datetime, timedelta

data=video_title_links.copy()

```

### ▼ converting the extracted data into pandas data frame

```

from datetime import datetime
final_list_for_df=[]
for i in data:

    title_of_vedio=i['title']
    url_of_vedio=i['url']
    views_of_vedio=i['date_data'].split()[0]
    views_of_vedio=views_of_vedio.replace(",","")

    views_of_vedio=int(views_of_vedio.strip(","))

    try:
        dmy=[]

        for j in i['date_data'].split()[2:5]:

```

```

        j=j.strip(" ")
        j=j.strip(",")
        dmy.append(j)
        date_of_vedio=int(dmy[1])
        month_of_vedio=dmy[0]
        year_of_vedio=int(dmy[2])

except:
    dmy=[]
    for j in i['date_data'].split()[3:6]:

        j=j.strip(" ")
        j=j.strip(",")
        dmy.append(j)

        date_of_vedio=int(dmy[1])
        month_of_vedio=dmy[0]
        year_of_vedio=int(dmy[2])

text=i['text']
l=text.split("ago")[1].split()[:-2]
l.reverse()

h=0
m=0
s=0
for k in range(len(l)):
    if(l[k].strip(",")==="second" or l[k].strip(",")==="seconds"):
        s=int(l[k+1])
    elif(l[k].strip(',')=='minute' or l[k].strip(',')=='minutes'):
        m=int(l[k+1])
    elif(l[k].strip(",")==='hour' or l[k].strip(",")==='hours'):
        h=int(l[k+1])
date_str = f"{month_of_vedio} {date_of_vedio} {year_of_vedio}"
date_format = "%b %d %Y"
datetime_obj = datetime.strptime(date_str, date_format)
total_seconds = (h * 3600) + (m * 60) + s
final_list=[title_of_vedio,url_of_vedio,views_of_vedio,datetime_obj,total_seconds]
final_list_for_df.append(final_list)


import pandas as pd

df = pd.DataFrame(final_list_for_df, columns=["title", "url", "views", "date_of_rec", "durations"])

```

df

	title	url	views	date_of_rec	durations
0	10 TIPS and TRICKS to Crack Internships and Pl...	https://www.youtube.com/watch?v=2QDMIZGeIC4	604	2023-08-01	593
1	Complete School Guide for CBSE   Free Resource...	https://www.youtube.com/watch?v=1ECMIxC6jbs	456	2023-07-31	56
2	Master System Design   GeeksforGeeks	https://www.youtube.com/watch?v=vjWLZYeqiN8	6624	2023-07-27	46
3	Unraveling the Geek-a-thon: The What, How, and...	https://www.youtube.com/watch?v=6e0v_NhsFMQ	414	2023-07-27	2700
4	Data Analysis Masterclass   Geek-A-Thon   Geek...	https://www.youtube.com/watch?v=-7Emxpn8DXc	568	2023-07-26	5340
...	...	...	...	...	...
	Interview Experience	https://www.youtube.com/watch?			

### Tasks & Questions:

- ▼ 1. Number of videos in the past 6 months from the start date. Must mention the dates in the solution.

```
current_date = datetime.now()
six_months_ago = current_date - timedelta(days=6*30)
filtered_df = df[df['date_of_rec'] >= six_months_ago]
```

```
filtered_df[['title', 'date_of_rec']]
```

	title	date_of_rec
0	10 TIPS and TRICKS to Crack Internships and Pl...	2023-08-01
1	Complete School Guide for CBSE   Free Resource...	2023-07-31
2	Master System Design   GeeksforGeeks	2023-07-27

There are 63 vedios which are posted in last 6 months

58 Meet Shikar Goel. CTO of GeeksforGeeks | Know ... 2023-02-17

## ▼ 2. Create a pandas data frame with columns name as videos title, views, Length of videos, and videos link

60 A.I. Apps That Can Change Your Lives | Geeksfo... 2023-02-15

df

	title	url	views	date_of_rec	durations
0	10 TIPS and TRICKS to Crack Internships and Pl...	https://www.youtube.com/watch?v=2QDMIZGeIC4	604	2023-08-01	593
1	Complete School Guide for CBSE   Free Resource...	https://www.youtube.com/watch?v=1ECMlxC6jbs	456	2023-07-31	56
2	Master System Design   GeeksforGeeks	https://www.youtube.com/watch?v=vjWLZYeqiN8	6624	2023-07-27	46
3	Unraveling the Geek-a-thon: The What, How, and...	https://www.youtube.com/watch?v=6e0v_NhsFMQ	414	2023-07-27	2700
4	Data Analysis Masterclass   Geek-A-Thon   Geek...	https://www.youtube.com/watch?v=-7Emxpn8DXc	568	2023-07-26	5340
...	...	...	...	...	...
	Interview Experience	https://www.youtube.com/watch?			

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0    title      150 non-null    object
1    url        150 non-null    object
2    views      150 non-null    int64
3    date_of_rec 150 non-null    datetime64[ns]
4    durations  150 non-null    int64
```

```
dtypes: datetime64[ns](1), int64(2), object(2)
memory usage: 6.0+ KB
```

▼ we have created the data frame with columns title url views , date , durations

▼ 3. Name the most viewed topics in the past 6 months.

```
filtered_df.sort_values(by="views",ascending=False)
```

	title	url	views	date_of_rec	durations
28	GeeksforGeeks Classroom Program   Now in Noida...	<a href="https://www.youtube.com/watch?v=16D2cuRy5JY">https://www.youtube.com/watch?v=16D2cuRy5JY</a>	109383	2023-05-08	38
18	Free Summer Offline Classes on Python Programi...	<a href="https://www.youtube.com/watch?v=OOLXHwZzHfM">https://www.youtube.com/watch?v=OOLXHwZzHfM</a>	106927	2023-06-09	37
21	Full Stack Development   LIVE Classes   Geeksf...	<a href="https://www.youtube.com/watch?v=cBfC9HLR9Qk">https://www.youtube.com/watch?v=cBfC9HLR9Qk</a>	89386	2023-05-30	35
31	Job Fair for Students   Till 25th May Only   G...	<a href="https://www.youtube.com/watch?v=1GEegOb3fHE">https://www.youtube.com/watch?v=1GEegOb3fHE</a>	82851	2023-05-02	39
46	Master DSA Today and be the Masters of Tomorrow	<a href="https://www.youtube.com/watch?v=PsOXP5hg52w">https://www.youtube.com/watch?v=PsOXP5hg52w</a>	73570	2023-03-15	25
...	...	...	...	...	...
	Explore	<a href="https://www.youtube.com/watch?">https://www.youtube.com/watch?</a>			

```
filtered_df.sort_values(by="views",ascending=False).iloc[:10] # top ten
```

	title	url	views	date_of_rec	durations
28	GeeksforGeeks Classroom Program   Now in Noida...	https://www.youtube.com/watch?v=16D2cuRy5JY	109383	2023-05-08	38

```
filtered_df.sort_values(by="views",ascending=False).iloc[:5]['title'] # top 5
```

```
28    GeeksforGeeks Classroom Program | Now in Noida...
18    Free Summer Offline Classes on Python Programi...
21    Full Stack Development | LIVE Classes | Geeksf...
31    Job Fair for Students | Till 25th May Only | G...
46    Master DSA Today and be the Masters of Tomorrow
Name: title, dtype: object
```

```
28    Full Stack Development | LIVE Classes | Geeksf...
url    https://www.youtube.com/watch?v=1GFedOh3ffHF
views  89386
date   2023-05-08
dur    39
```

#### ▼ top 5 viewed topics or vedios in past 6 months

GeeksforGeeks Classroom Program | Now in Noida... - 109383 views

Free Summer Offline Classes on Python Programi... -106927 views

Full Stack Development | LIVE Classes | Geeksf... -89386 views

Job Fair for Students | Till 25th May Only | G... -82851 views

Master DSA Today and be the Masters of Tomorrow -73570 views

#### ▼ 4. Name the topics with the highest video length.

```
df.sort_values(by="durations",ascending=False)
```



	title	url	views	date_of_rec	durations
149	DSA Self-Paced Course Preview   Sandeep Jain  ...	https://www.youtube.com/watch?v=Pt2vYaJgGv8	73875	2022-04-23	7980

```
df.sort_values(by="durations",ascending=False).iloc[:10]
```

	title	url	views	date_of_rec	durations
149	DSA Self-Paced Course Preview   Sandeep Jain  ...	https://www.youtube.com/watch?v=Pt2vYaJgGy8	73875	2022-04-23	7980
4	Data Analysis Masterclass   Geek-A-Thon   Geek...	https://www.youtube.com/watch?v=-7Emxpn8DXc	568	2023-07-26	5340
3	Unraveling the Geek-a-thon: The What, How, and...	https://www.youtube.com/watch?v=6e0v_NhsFMQ	414	2023-07-27	2700
134	Welcome Sandeep Jain at KIET Ghaziabad   Geeks...	https://www.youtube.com/watch?v=UkW6L9uyKJQ	2121	2022-05-25	2640
144	My journey through the tech world from India t...	https://www.youtube.com/watch?v=IXgQGoj0rSk	2205	2022-05-08	2280

```
df.sort_values(by="durations",ascending=False).iloc[:5]
```

	title	url	views	date_of_rec	durations
149	DSA Self-Paced Course Preview   Sandeep Jain  ...	https://www.youtube.com/watch?v=Pt2vYaJgGy8	73875	2022-04-23	7980
4	Data Analysis Masterclass   Geek-A-Thon   Geek...	https://www.youtube.com/watch?v=-7Emxpn8DXc	568	2023-07-26	5340
3	Unraveling the Geek-a-thon: The What	https://www.youtube.com/watch?	414	2023-07-27	2700

```
df.sort_values(by="durations",ascending=False).iloc[:5]['title']
```

```
149    DSA Self-Paced Course Preview | Sandeep Jain |...
4      Data Analysis Masterclass | Geek-A-Thon | Geek...
3      Unraveling the Geek-a-thon: The What, How, and...
134    Welcome Sandeep Jain at KIET Ghaziabad | Geeks...
144    My journey through the tech world from India t...
Name: title, dtype: object
```

#### ▼ top 5 highest video length

DSA Self-Paced Course Preview | Sandeep Jain |... 7980 seconds

Data Analysis Masterclass | Geek-A-Thon | Geek... 5340 seconds

Unraveling the Geek-a-thon: The What, How, and... 2700 seconds

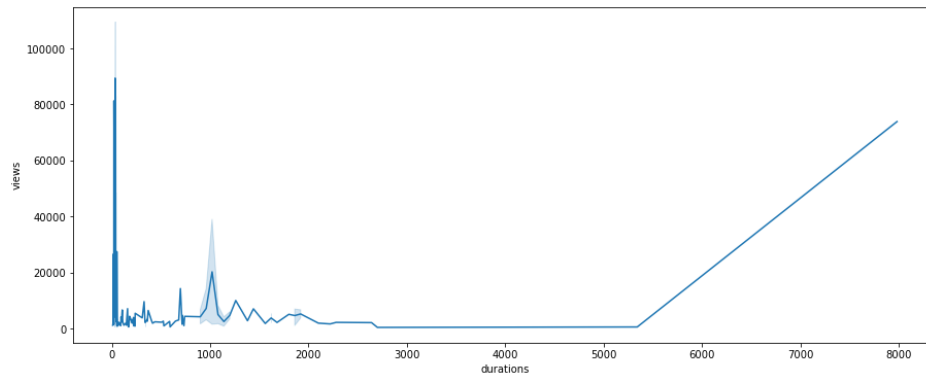
Welcome Sandeep Jain at KIET Ghaziabad | Geeks... 2640 seconds

My journey through the tech world from India t.. 2280 seconds

▼ 5. Make a comparison between the number of views and video length using a Graph.

```
import seaborn as sns
import matplotlib.pyplot as plt
```

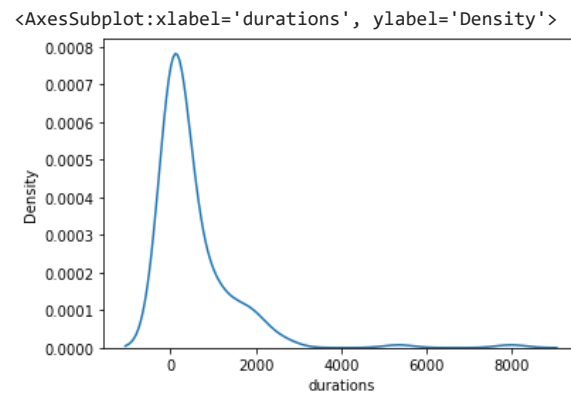
```
plt.figure(figsize=(15, 6))
sns.lineplot(x=df['durations'],y=df['views'])
plt.show()
```



```
plt.figure(figsize=(10, 6))
sns.scatterplot(x=df['durations'],y=df['views'])
plt.show()
```



```
sns.kdeplot(df['durations'])
```



```
df['durations'].mean()
```

```
542.02
```

From above line plot we can say that usually short vedios have much views ie less that 500 seconds

less views for 2000 to 5000 seconds vedios

and the vedios which have high duration above 5000 seconds have good amount of significat views

Note : we have only 2 data points which is above 5000 seconds

▼ Most of the vedios are in range of 0 to 2000 seconds

the mean vedio duration is 542 seconds

