

```
pip install requests
```

```
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (2
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local,
```

```
pip install xlswriter
```

```
Collecting xlswriter
  Downloading https://files.pythonhosted.org/packages/3f/c1/2a77723ae03fea7650f8d77a1
  | ████████████████████████████████████████████████████████████████████████████████ | 153kB 5.6MB/s
Installing collected packages: xlswriter
Successfully installed xlswriter-1.4.0
```

```
import requests
import datetime
from tqdm import tqdm
import numpy as np
import pandas as pd
import xlswriter
```

```
pip install newsapi-python
```

```
Collecting newsapi-python
  Downloading https://files.pythonhosted.org/packages/de/9e/9050199ac7cbc755d1c495774
Requirement already satisfied: requests<3.0.0 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local,
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pac
Installing collected packages: newsapi-python
Successfully installed newsapi-python-0.2.6
```

```
from newsapi import NewsApiClient
```

```
newsapi= NewsApiClient(api_key='956160556097434c9cf592566862ea13')
```

```
def get_past_articles(past=100):
    past_articles=dict()
    for past_days in range(1,past):
        from_day=str(datetime.datetime.now()-datetime.timedelta(days=past_days))
        to_day=str(datetime.datetime.now()-datetime.timedelta(days=past_days-1))
        past_articles.update({from_day:to_day})
    return past_articles
```

```
def get_articles(query, past=100):
    past_articles=get_past_articles(past)
    all_articles=[]
    for i,j in tqdm(past_articles.items()):
        for pag in tqdm(range(1,6)):
            pag_articles=newsapi.get_everything(q=query,language='en', from_param='2021-03-27',
            if len(pag_articles)==0:break
            all_articles.extend(pag_articles)
    return all_articles
```

```
articles= get_articles("India")
```

```
100%|██████████| 5/5 [00:00<00:00, 8.87it/s]
91%|██████████| 90/99 [00:53<00:05, 1.74it/s]
 0%|          | 0/5 [00:00<?, ?it/s]
20%|██        | 1/5 [00:00<00:00, 8.73it/s]
40%|████      | 2/5 [00:00<00:00, 8.88it/s]
60%|██████    | 3/5 [00:00<00:00, 8.66it/s]
80%|████████  | 4/5 [00:00<00:00, 8.72it/s]
100%|██████████| 5/5 [00:00<00:00, 8.61it/s]
92%|██████████| 91/99 [00:54<00:04, 1.71it/s]
 0%|          | 0/5 [00:00<?, ?it/s]
40%|████      | 2/5 [00:00<00:00, 10.56it/s]
60%|██████    | 3/5 [00:00<00:00, 9.92it/s]
80%|████████  | 4/5 [00:00<00:00, 9.79it/s]
100%|██████████| 5/5 [00:00<00:00, 9.20it/s]
93%|██████████| 92/99 [00:54<00:04, 1.73it/s]
 0%|          | 0/5 [00:00<?, ?it/s]
20%|██        | 1/5 [00:00<00:00, 8.64it/s]
40%|████      | 2/5 [00:00<00:00, 8.94it/s]
60%|██████    | 3/5 [00:00<00:00, 9.08it/s]
80%|████████  | 4/5 [00:00<00:00, 9.20it/s]
100%|██████████| 5/5 [00:00<00:00, 8.95it/s]
94%|██████████| 93/99 [00:55<00:03, 1.74it/s]
 0%|          | 0/5 [00:00<?, ?it/s]
20%|██        | 1/5 [00:00<00:00, 9.35it/s]
60%|██████    | 3/5 [00:00<00:00, 9.83it/s]
80%|████████  | 4/5 [00:00<00:00, 9.34it/s]
100%|██████████| 5/5 [00:00<00:00, 9.52it/s]
95%|██████████| 94/99 [00:55<00:02, 1.78it/s]
 0%|          | 0/5 [00:00<?, ?it/s]
40%|████      | 2/5 [00:00<00:00, 8.98it/s]
60%|██████    | 3/5 [00:00<00:00, 9.13it/s]
100%|██████████| 5/5 [00:00<00:00, 9.50it/s]
96%|██████████| 95/99 [00:56<00:02, 1.80it/s]
 0%|          | 0/5 [00:00<?, ?it/s]

20%|██        | 1/5 [00:00<00:00, 8.87it/s]
40%|████      | 2/5 [00:00<00:00, 8.87it/s]
80%|████████  | 4/5 [00:00<00:00, 9.06it/s]
100%|██████████| 5/5 [00:00<00:00, 9.08it/s]
97%|██████████| 96/99 [00:56<00:01, 1.80it/s]
 0%|          | 0/5 [00:00<?, ?it/s]
20%|██        | 1/5 [00:00<00:00, 9.89it/s]
40%|████      | 2/5 [00:00<00:00, 9.40it/s]
60%|██████    | 3/5 [00:00<00:00, 9.45it/s]
100%|██████████| 5/5 [00:00<00:00, 9.67it/s]
98%|██████████| 97/99 [00:57<00:01, 1.81it/s]
 0%|          | 0/5 [00:00<?, ?it/s]
20%|██        | 1/5 [00:00<00:00, 9.21it/s]
```

```

40%|███████| 2/5 [00:00<00:00, 8.96it/s]
60%|██████████| 3/5 [00:00<00:00, 8.62it/s]
80%|███████████| 4/5 [00:00<00:00, 8.82it/s]
100%|███████████| 5/5 [00:00<00:00, 8.69it/s]
99%|███████████| 98/99 [00:58<00:00, 1.76it/s]
0%|          | 0/5 [00:00<?, ?it/s]
20%|███| 1/5 [00:00<00:00, 9.48it/s]
40%|█████| 2/5 [00:00<00:00, 8.42it/s]
60%|███████| 3/5 [00:00<00:00, 8.53it/s]
80%|█████████| 4/5 [00:00<00:00, 8.59it/s]
100%|███████████| 5/5 [00:00<00:00, 8.47it/s]
100%|███████████| 99/99 [00:58<00:00, 1.69it/s]

```

```
dataset = pd.DataFrame(articles)
dataset.sample(5)
```

	source	author	title	description	url
4617	{'id': 'techcrunch', 'name': 'TechCrunch'}	Rita Liao	Xiaomi further localizes India supply chain vi...	China's Xiaomi had dominated the Indian smartp...	http://techcrunch.com/2021/02/25/xiaomi-india-...
337	{'id': 'techcrunch', 'name': 'TechCrunch'}	Manish Singh	Google removes millions of negative TikTok rev...	ByteDance's TikTok app, which has gained hundr...	http://techcrunch.com/2020/05/27/google-remove...
	{'id': 'techcrunch'}	Rita	Xiaomi further localizes	China's Xiaomi had	http://techcrunch.com/2021/02/25/xiaomi-

```
dataset.to_excel('news_dataset.xlsx', sheet_name = 'New_sheet')
```

```
from google.colab import files
files.download('news_dataset.xlsx')
```

✓ 0s completed at 13:32 ● ✕