

Nama : Ramadhana Noor Salassa Wandya

NIM : 09011282025083

Kelas : SK4A

Pada kesempatan kali ini, saya akan mencoba melakukan web scraping yang mana objek yang akan dilakukan dan diambil sampelnya adalah pada google trends. google trends merupakan website yang menunjukkan search trending pada aplikasi google. disini saya akan mencoba melakukan scraping dengan tujuan untuk melihat daily interest dan history interest yang berspesifik pada region indonesia. disini saya menggunakan library pytrends sebagai library utama, dan library pandas beserta matplotlib sebagai library untuk melakukan visualisasi dan dataframe.

```
In [ ]: from pytrends.request import TrendReq
import pandas as pd
from bs4 import BeautifulSoup
import requests
import matplotlib.pyplot as plt
%matplotlib inline
from pandas.plotting import register_matplotlib_converters
register_matplotlib_converters()
```

```
In [ ]: pytrends = TrendReq(hl = 'en-US')

countries = ['indonesia']

def trending_searches(country):
    global data
    data = pytrends.trending_searches(country)
    print(data.head(5))

for country in countries:
    print(country)
    print('')
    trending_searches(country)
    # print('')
```

indonesia

```
0
0      Sampdoria vs Fiorentina
1      Tomorrow
2      UAS) dideportasi dari Singapura
3      Southampton vs Liverpool
4      Allo Bank
```

```
In [ ]: kumpul = pd.DataFrame(data)
# kumpul
tag = ['Populer']
kumpul.columns = tag
kumpul.to_csv('data.csv', index = False)
```

```
Populer
0      Sampdoria vs Fiorentina
1      Tomorrow
```

```

2   UAS) dideportasi dari Singapura
3       Southampton vs Liverpool
4       Allo Bank
5       Hari Buku Nasional
6       Theo Hernandez
7       Asmaul Husna dan artinya
8       Putri Delina
9       Ashira Zamita
10      Event
11      Apex Legends Mobile
12      Malaysia vs Kamboja
13      Newcastle vs Arsenal
14      Juventus vs Lazio
15      Ucapan Waisak 2022
16      Ppdb.jakarta
17      Kecelakaan Mojokerto
18      Jess No Limit
19      Atlético Madrid vs Sevilla

```

script diatas menampilkan top searching pada hari ini di google region indonesia. data tersebut terurut dari pencarian yang paling banyak ke paling rendah. namun dalam ini terdapat keterbatasan yaitu kita tidak bisa mengetahui exact value dari tiap hasil searching. Tapi data diatas sudah cukup menggambarkan apa saja peristiwa yang terjadi pada hari itu. Data hasil pencarian diatas lalu di export ke dalam bentuk csv.

```

In [ ]: import pandas as pd
        from pytrends.request import TrendReq
        pytrend = TrendReq()

        #get today's trending topics
        trendingtoday = pytrend.today_searches(pn='ID')
        trendingtoday.head(20)
        trendingtoday.to_csv('searchque.csv')

```

script diatas merupakan search query beserta dengan link dan daily search dari google trends.

```

In [ ]: df = pytrend.top_charts(2019, hl='en-US', tz=300, geo='ID')
        df.head()
        kolom = ['kata populer', 'Query']
        df.columns=kolom
        df.to_csv('charts.csv')

```

script diatas akan menampilkan top kata kunci yang dicari pada tahun 2019 beserta querynya. top kata kunci ini diambil dari trending google sehingga bisa menampilkan situasi pada tahun tersebut. hasil dari data ini akan diekspor dalam bentuk csv.

```

In [ ]: pytrends = TrendReq(hl = 'en-us')

        keywords = ['facebook', 'instagram', 'twitter', 'tiktok']
        pytrends.build_payload(keywords, timeframe='today 12-m', geo = 'ID')

        data = pytrends.interest_over_time()
        data

        data.to_csv('sosmed.csv')

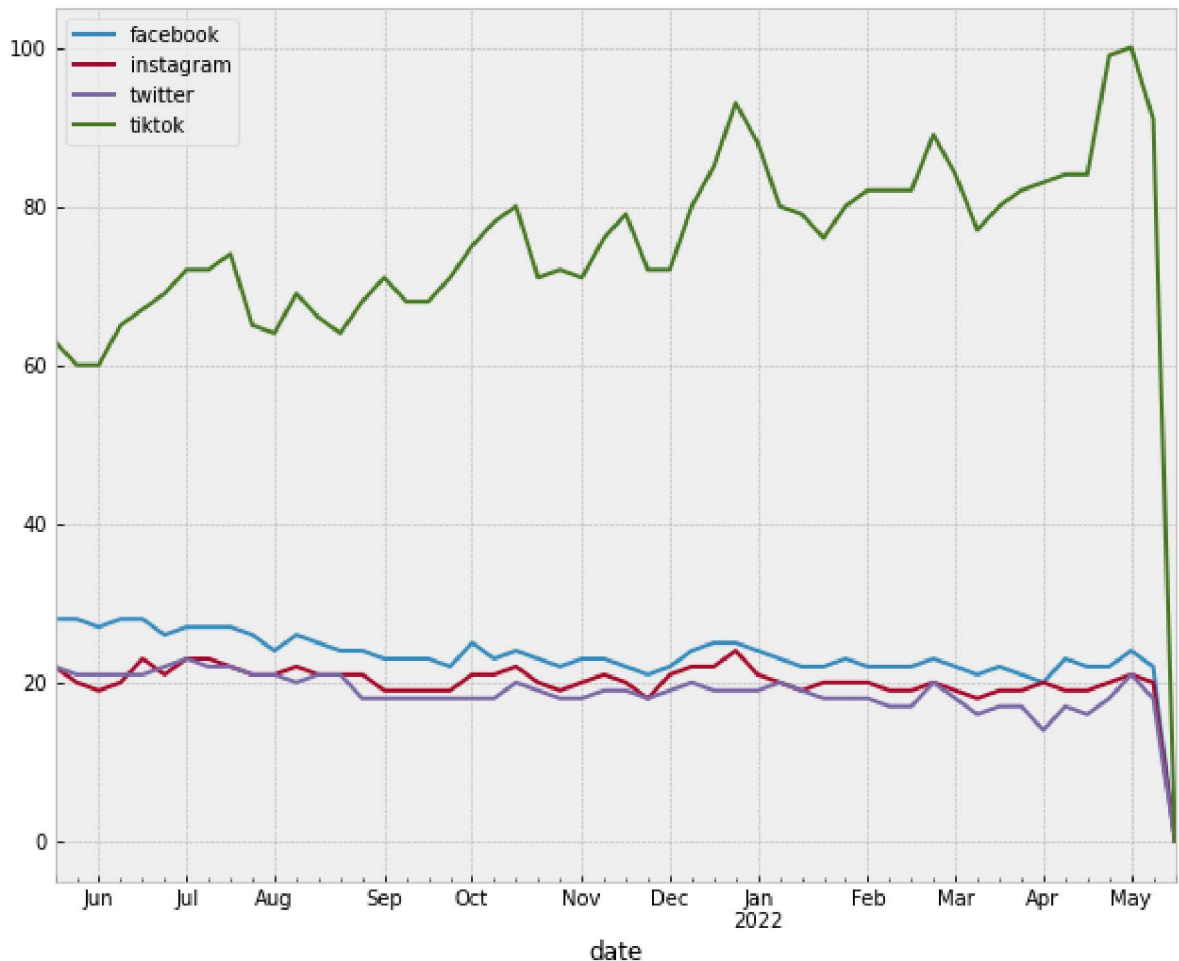
```

```

In [ ]: data.plot(figsize = (10, 8))

```

Out[]: <AxesSubplot:xlabel='date'>



grafik diatas menggambarkan keywords sosial media yang dicari di indonesia pada 1 tahun terakhir. Dengan grafik tersebut, kita dapat membaca dan mengetahui sosial media yang paling populer dan paling banyak digunakan di indonesia. nilai pada sumbu y merupakan nilai interest atau nilai ketertarikan user pada keywords tersebut yang mana data tersebut didapat dengan mengambil data pencarian di google trends.

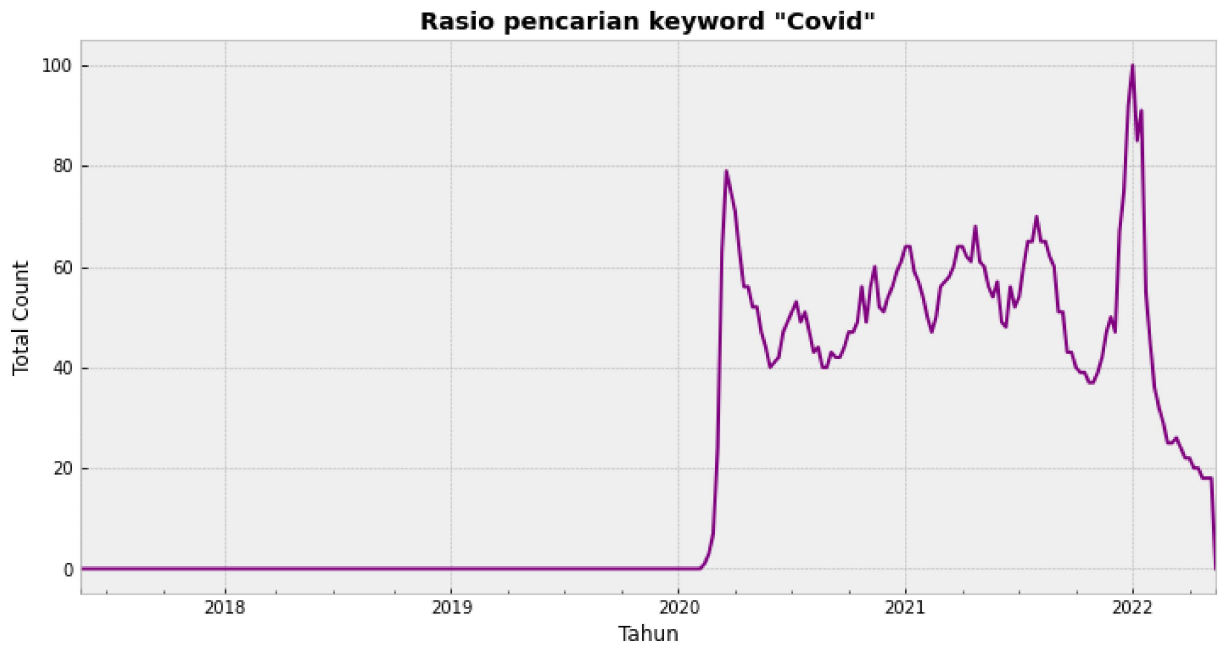
```
In [ ]: plt.style.use('bmh')

# melakukan koneksi ke google
pytrends = TrendReq(hl='id-ID')
# keywords yang akan dicari
pytrends.build_payload(kw_list=['covid'])

# dataframe
time_df = pytrends.interest_over_time()

# Grafik
fig, ax = plt.subplots(figsize=(12, 6))
time_df['covid'].plot(color='purple')
# Judul dan Label
plt.title('Rasio pencarian keyword "Covid"', fontweight='bold')
plt.xlabel('Tahun')
plt.ylabel('Total Count')
```

Out[]: Text(0, 0.5, 'Total Count')



gambar diatas merupakan gambar visualisasi tingkat pencarian dengan kata kunci "covid". bisa dilihat dari grafik tersebut pencarian kata kunci covid melonjak drastis pada awal tahun 2020. karena pada saat itu munculnya kasus pertama covid di indonesia. lalu setelah nya tingkat pencarian dari data tersebut mengalami kenaikan dan penurunan. lalu pada awal tahun 2022, pencarian dengan kata kunci covid kembali naik drastis karena adanya gelombang 2 covid.