

**ANALITIK DAN VISUALISASI DATA  
FILTERING DATA MENGGUNAKAN PHYTHON  
(STUDI KASUS DATA COVID 19)**

Disusun untuk memenuhi salah satu tugas mata kuliah Analitik dan Visualisasi  
Data

Dosen pengampu : Yusep Maulana, A.Md.Kom.,S.Kom.,M.Kom.



**Oleh:**

Lindawati : 2107015

**PROGRAM STUDI SISTEM INFORMASI  
INSTITUT TEKNOLOGI GARUT**

**2023**

Dalam filtering data dengan python disini menggunakan aplikasi Anaconda (Jupyter). Data yang digunakan merupakan data Covid 19 di seluruh negara.

Langkah pertama dalam filtering ini yaitu import matplotlib, dan pandas. Matplotlib berguna untuk memvisualisasikan data yang ada sedangkan pandas berguna untuk membaca data, pada kali ini tipe data yang digunakan adalah csv.

```
In [6]: import matplotlib.pyplot as plt
import pandas as pd
```

Berikutnya filelocation digunakan untuk membaca dimana file data yang akan digunakan berada.

```
In [8]: filelocation = 'D:\TUGAS LINDA\MATERI SEMESTER 5\ANALITIK DAN VISUALISASI DATA\covid19.csv'
df = pd.read_csv(filelocation)
```

Untuk menampilkan data yang ada meggunakan syntax sebagai berikut.

```
In [9]: df.head()
```

Berikut merupakan datanya. *(Data lebih lengkap ada di file covid19.csv)*

Out[9]:

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered
0	NaN	Afghanistan	33.0000	65.0000	1/22/20	0	0	0
1	NaN	Albania	41.1533	20.1683	1/22/20	0	0	0
2	NaN	Algeria	28.0339	1.6596	1/22/20	0	0	0
3	NaN	Andorra	42.5063	1.5218	1/22/20	0	0	0
4	NaN	Angola	-11.2027	17.8739	1/22/20	0	0	0

Selanjutnya data yang sudah ada di filter sesuai nama negara(Country/region), sebagai contoh berikut filtering data untuk negara Albania. Disini ada True dan False, nilai bernilai True apabila negara tersebut merupakan Albania.

```
In [11]: df['Country/Region'] == 'Albania'
Out[11]: 0      False
1       True
2      False
3      False
4      False
...
32060    False
32061    False
32062    False
32063    False
32064    False
Name: Country/Region, Length: 32065, dtype: bool
```

Sedangkan untuk menampilkan data hanya negara Albania sebagai berikut. Terdapat 121 rows dan 8 columns. Data tersebut mulai dari awal adanya masyarakat terkena covid sampai tidak ada lagi masyarakat yang terkena covid.

```
In [12]: Albania = df[df['Country/Region'] == 'Albania']
Albania
Out[12]:
```

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered
1	NaN	Albania	41.1533	20.1683	1/22/20	0	0	0
266	NaN	Albania	41.1533	20.1683	1/23/20	0	0	0
531	NaN	Albania	41.1533	20.1683	1/24/20	0	0	0
796	NaN	Albania	41.1533	20.1683	1/25/20	0	0	0
1061	NaN	Albania	41.1533	20.1683	1/26/20	0	0	0
...	...	...	...	...	...	...	...	...
30741	NaN	Albania	41.1533	20.1683	5/17/20	946	31	715
31006	NaN	Albania	41.1533	20.1683	5/18/20	948	31	727
31271	NaN	Albania	41.1533	20.1683	5/19/20	949	31	742
31536	NaN	Albania	41.1533	20.1683	5/20/20	964	31	758
31801	NaN	Albania	41.1533	20.1683	5/21/20	969	31	771

121 rows x 8 columns

Selanjutnya merupakan filtering dimana data terkonfirmasi covid pada negara Albania lebih dari 0.

```
In [13]: kasus = Albania[Albania['Confirmed']>0]
kasus

Out[13]:
```

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered
12456	NaN	Albania	41.1533	20.1683	3/9/20	2	0	0
12721	NaN	Albania	41.1533	20.1683	3/10/20	10	0	0
12986	NaN	Albania	41.1533	20.1683	3/11/20	12	1	0
13251	NaN	Albania	41.1533	20.1683	3/12/20	23	1	0
13516	NaN	Albania	41.1533	20.1683	3/13/20	33	1	0
...	...	...	...	...	...	...	...	...
30741	NaN	Albania	41.1533	20.1683	5/17/20	946	31	715
31006	NaN	Albania	41.1533	20.1683	5/18/20	948	31	727
31271	NaN	Albania	41.1533	20.1683	5/19/20	949	31	742
31536	NaN	Albania	41.1533	20.1683	5/20/20	964	31	758
31801	NaN	Albania	41.1533	20.1683	5/21/20	969	31	771

74 rows x 8 columns

Untuk menampilkan grafik masyarakat yang terkonfirmasi covid di negara Albania sebagai berikut. Plt\_show() berguna untuk menampilkan grafik. Dalam menampilkan grafik disini menggunakan matplotlib.

```
In [14]: plt.plot(kasus['Date'], kasus['Confirmed'])
plt.show()
```

Filtering data covid negara Albania hanya pada bulan mei.

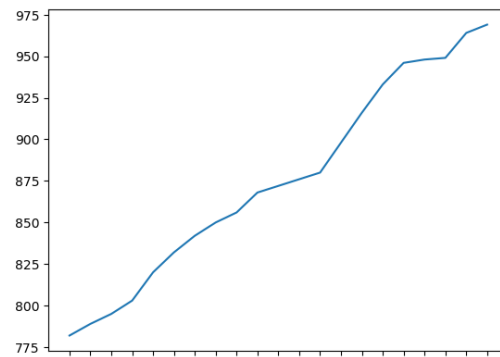
```
In [16]: kasusmei = kasus[kasus.Date >= '5/1/20']
kasusmei

Out[16]:
```

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered
26501	NaN	Albania	41.1533	20.1683	5/1/20	782	31	488
26766	NaN	Albania	41.1533	20.1683	5/2/20	789	31	519
27031	NaN	Albania	41.1533	20.1683	5/3/20	795	31	531
27296	NaN	Albania	41.1533	20.1683	5/4/20	803	31	543
27561	NaN	Albania	41.1533	20.1683	5/5/20	820	31	570
27826	NaN	Albania	41.1533	20.1683	5/6/20	832	31	595
28091	NaN	Albania	41.1533	20.1683	5/7/20	842	31	605
28356	NaN	Albania	41.1533	20.1683	5/8/20	850	31	620
28621	NaN	Albania	41.1533	20.1683	5/9/20	856	31	627
28886	NaN	Albania	41.1533	20.1683	5/10/20	868	31	650
29151	NaN	Albania	41.1533	20.1683	5/11/20	872	31	654
29416	NaN	Albania	41.1533	20.1683	5/12/20	876	31	682
29681	NaN	Albania	41.1533	20.1683	5/13/20	880	31	688
29946	NaN	Albania	41.1533	20.1683	5/14/20	898	31	694

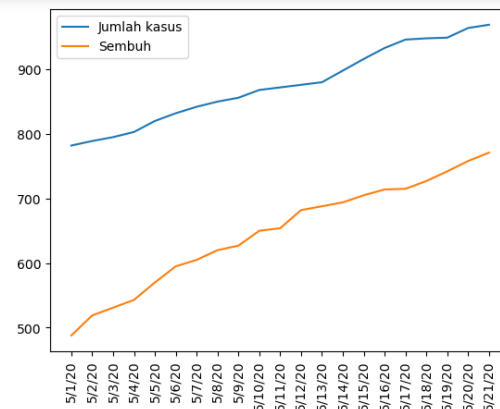
Sedangkan untuk menampilkan grafiknya dengan syntax sebagai berikut.

```
In [17]: plt.plot(kasusmei['Date'], kasusmei['Confirmed'])
plt.show()
```



Jika sebelumnya garis pada grafik tidak ada keterangan maka untuk menambahkan keterangan dapat menggunakan perintah `plt.legend`, sebagai berikut.

```
In [19]: plt.plot(kasusmei.Date, kasusmei.Confirmed)
plt.plot(kasusmei.Date, kasusmei.Recovered)
plt.legend(['Jumlah kasus', 'Sembuh'])
plt.xticks(rotation = 'vertical')
plt.show()
```



Berikut merupakan link untuk melihat syntax dan data yang lebih lengkap :

<https://github.com/lwti18/Analitik-dan-Visualisasi-Data-Lindawati-2107015>

