

Hal pertama yang harus kita lakukan adalah mencari data di internet dengan format .txt, .csv, .tsv, namun kali ini saya menggunakan data dengan format .csv yang saya ambil dari halaman “ncov recovered”. Setelah mendapat data yang diinginkan simpan link data tersebut. Setelah itu buka python. Karena kita menggunakan Pandas. Maka ketikkan “import pandas as pd”. Setelah itu klik *run*.

```
In [4]: import pandas as pd
```

Untuk judul yang panjang dan menggunakan spasi, spasi digantikan dengan *underscore* (_). Lalu ketikkan “pd.read_csv”. Untuk yang memakai format selain .csv maka setelah underscore “csv” diubah menjadi format yang dipilih. Lalu, *copy paste* kan link sudah disimpan menggunakan kurung dan tanda petik seperti gambar dibawah ini.

```
In [5]: ncov = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/
```

Setelah itu, cari dataset menggunakan keterangan head, tail, info, dan describe. Lalu satu persatu di klik *run*

```
In [6]: ncov.head()
```

Out[6]:

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	2/27/20	2/28/20	2/29/20
0	Anhui	Mainland China	31.8257	117.2264	0	0	0	0	0	0	...	792	821	
1	Beijing	Mainland China	40.1824	116.4142	0	0	1	2	2	2	...	248	257	
2	Chongqing	Mainland China	30.0572	107.8740	0	0	0	0	0	0	...	401	422	
3	Fujian	Mainland China	26.0789	117.9874	0	0	0	0	0	0	...	228	235	
4	Gansu	Mainland China	36.0611	103.8343	0	0	0	0	0	0	...	81	82	

5 rows x 50 columns

```
In [7]: ncov.tail()
```

Out[7]:

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	...	2/27/20	2/28/20	2/29/20
220	Pierce County, WA	US	47.0676	-122.1295	0	0	0	0	0	0	...	0	0	
221	Plymouth County, MA	US	42.1615	-70.7928	0	0	0	0	0	0	...	0	0	
222	Santa Cruz County, CA	US	36.9741	-122.0308	0	0	0	0	0	0	...	0	0	
223	Tulsa County, OK	US	36.1593	-95.9410	0	0	0	0	0	0	...	0	0	
224	Montgomery County, TX	US	30.3213	-95.4778	0	0	0	0	0	0	...	0	0	

5 rows x 50 columns

```
In [8]: ncov.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 225 entries, 0 to 224
Data columns (total 50 columns):
Province/State      131 non-null object
Country/Region      225 non-null object
Lat                 225 non-null float64
Long                225 non-null float64
1/22/20             225 non-null int64
1/23/20             225 non-null int64
1/24/20             225 non-null int64
1/25/20             225 non-null int64
1/26/20             225 non-null int64
1/27/20             225 non-null int64
1/28/20             225 non-null int64
1/29/20             225 non-null int64
1/30/20             225 non-null int64
```

```
In [9]: ncov.describe()

Out[9]:
```

	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20
count	225.000000	225.000000	225.000000	225.000000	225.000000	225.000000	225.000000	225.000000	225.000000	225.000000
mean	32.296330	-6.115845	0.124444	0.133333	0.160000	0.173333	0.231111	0.271111	0.475556	0.560000
std	19.721523	90.344785	1.866667	1.870829	2.072611	2.142512	2.809409	3.024057	5.364448	5.900757
min	-41.454500	-157.858400	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	27.990400	-86.175200	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	36.974100	4.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	42.695300	84.250000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
max	64.963100	174.886000	28.000000	28.000000	31.000000	32.000000	42.000000	45.000000	80.000000	88.000000

8 rows x 48 columns

Setelah di *Run* akan muncul dataset seperti di atas. Setelah itu ketik “import matplotlib.pyplot as plt”. Library ini digunakan untuk membuat grafik pada data yang sudah ada di atas. Lalu klik *Run*.

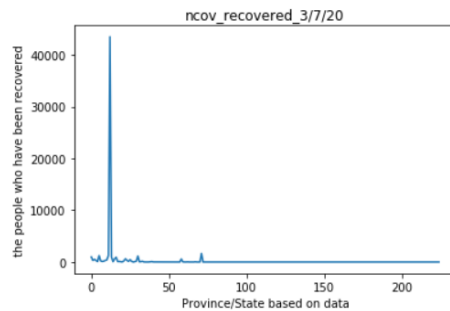
```
In [10]: import matplotlib.pyplot as plt
```

1. Untuk plot, dapat di lihat di dataset info, tentang apa aja yang harus diplot.
2. Label pada sumbu x saya gunakan untuk Province/State based on data
3. Label pada sumbu y saya gunakan untuk the people who have been recovered
4. Untuk judul saya gunakan plot ncov_recovered_3/7/20

Setelah menuliskan program sesuai gambar dibawah dan mengklik run, akan muncul grafik seperti dibawah ini

```
In [12]: plt.plot(ncov['3/7/20'])  
plt.xlabel('Province/State based on data')  
plt.ylabel('the people who have been recovered')  
plt.title('ncov_recovered_3/7/20')
```

```
Out[12]: Text(0.5, 1.0, 'ncov_recovered_3/7/20')
```



Dikarenakan banyak sekali data mengenai perkembangan orang-orang yang telah ter-recovered dari ncov ini, maka untuk plotting grafik, saya hanya menggunakan data paling baru saja, yaitu pada tanggal 7 maret 2020.