

Persiapan Data

```
#import data dan menyimpan data json ke dalam bentuk dataframe pandas
```

```
import json
```

```
import pandas as pd
```

```
#df germany
```

```
df_germany = pd.read_json('/content/covid - germany.json')
```

```
df_germany = df_germany[['Country', 'Cases', 'Date']]
```

```
print(df_germany)
```

```
#df india
```

```
df_india = pd.read_json('/content/covid - india.json')
```

```
df_india = df_india[['Country', 'Cases', 'Date']]
```

```
print(df_india)
```

```
#df indonesia
```

```
df_indonesia = pd.read_json('/content/covid - indonesia.json')
```

```
df_indonesia = df_indonesia[['Country', 'Cases', 'Date']]
```

```
print(df_indonesia)
```

```
#df usa
```

```
df_usa = pd.read_json('/content/covid-usa.json')
```

```
df_usa = df_usa[['Country', 'Cases', 'Date']]
```

```
print(df_usa)
```

```
#combining dataframe
```

```
all = [df_germany, df_india, df_indonesia, df_usa]
```

```
df_all = pd.concat(all)
```

```
print(df_all)
```

```
#add calculate normalization and normalization column
```

```
df_all['Norm'] = (df_all['Cases'] - df_all['Cases'].min()) / (df_all['Cases'].max() - df_all['Cases'].min())
```

```
df_all['Norm']
```

```
df_all
```

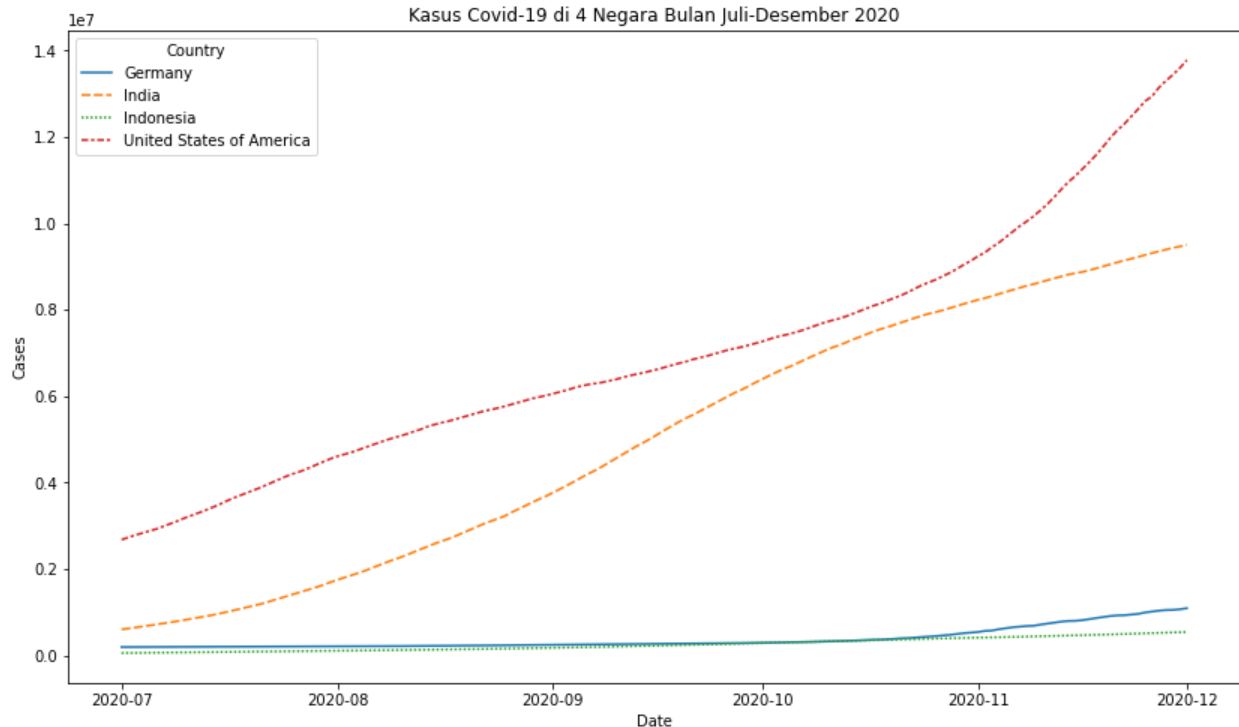
	Country	Cases	Date	Norm
0	Germany	195893	2020-07-01 00:00:00+00:00	0.010069
1	Germany	196370	2020-07-02 00:00:00+00:00	0.010103
2	Germany	196780	2020-07-03 00:00:00+00:00	0.010133
3	Germany	197198	2020-07-04 00:00:00+00:00	0.010164
4	Germany	197523	2020-07-05 00:00:00+00:00	0.010187
...
149	United States of America	13141178	2020-11-27 00:00:00+00:00	0.953727
150	United States of America	13296645	2020-11-28 00:00:00+00:00	0.965060
151	United States of America	13435856	2020-11-29 00:00:00+00:00	0.975208
152	United States of America	13592553	2020-11-30 00:00:00+00:00	0.986631
153	United States of America	13775956	2020-12-01 00:00:00+00:00	1.000000

616 rows x 4 columns

Visualisasi data sebenarnya

```
import seaborn as sns
fig, ax = plt.subplots(figsize = (14,8))

plot_all = sns.lineplot(data= pd.pivot(df_all, index='Date', columns='Country', values='Cases'))
```



Analisis :

Dari grafik dapat dilihat bahwa Amerika Serikat merupakan Negara dengan jumlah kasus terbanyak dari bulan Juli- Desember, dengan jumlah kasus hampir menyentuh 14 juta kasus pada awal Desember 2020, disusul dengan Negara India dengan jumlah kasus mendekati 10 juta kasus pada awal Desember 2020. Jumlah kasus Covid-19 di Jerman dan Indonesia jauh lebih sedikit jika dibandingkan dengan Amerika Serikat dan India, dengan jumlah kasus masih di bawah satu juta kasus hingga awal Desember. Jumlah kasus covid-19 di Indonesia paling sedikit jika dibandingkan dengan 4 negara lainnya.

Visualisasi data normalisasi

```
fig = plt.subplots(figsize = (14,8))  
plot2 = sns.lineplot(data=df_all, x="Date", y="Norm", hue="Country")
```

