

## Treatment untuk Missing Value

Terdapat beberapa cara untuk mengatasi missing value, antara lain:

1. dibiarkan saja,
2. hapus value itu, atau
3. isi value tersebut dengan value yang lain (biasanya interpolasi, mean, median, etc)

Sebelum melakukan action ke missing value pada data covid diatas, sebaiknya tampilkan beberapa row teratas dari dataset itu

	province_state	country_region	date	latitude	longitude	location_geom	confirmed	deaths	recovered	active	fips	admin2	combined_key
0	NaN	UK	01-02-20	NaN	NaN	NaN	2	0.0	0.0	NaN	NaN	NaN	NaN
1	NaN	UK	18-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN
2	NaN	UK	17-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN
3	NaN	UK	31-01-20	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	UK	19-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN
5	NaN	UK	22-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN
6	NaN	UK	25-02-20	NaN	NaN	NaN	13	0.0	8.0	NaN	NaN	NaN	NaN
7	NaN	UK	16-02-20	NaN	NaN	NaN	9	0.0	8.0	NaN	NaN	NaN	NaN
8	NaN	UK	27-02-20	NaN	NaN	NaN	15	0.0	8.0	NaN	NaN	NaN	NaN
9	NaN	UK	03-02-20	NaN	NaN	NaN	2	0.0	0.0	NaN	NaN	NaN	NaN

dan dilihat kembali jumlah missing value tiap kolomnya

```
Jumlah missing value per kolom:
```

```
province_state      40
country_region       0
date                 0
latitude            126
longitude            126
location_geom        126
confirmed            0
deaths               1
recovered            1
active               51
fips                 51
admin2              158
combined_key         1000
dtype: int64
```

agar dapat ditelaah terlebih dahulu.

Hanya kolom `combine_key` yang keseluruhan barisnya adalah missing value (1000 buah), sementara kolom `country_region`, `date`, dan `confirmed` tidak memiliki missing value. Untuk kolom lainnya terdapat beragam jumlah missing value. Apa yang dapat dilakukan?

Untuk memahami mana kolom yang akan ditreatment dengan tiga perlakuan di atas lihat nature dari data terlebih dahulu. Contohnya pada kolom `death` dan `recovered` jika ada yang missing value maka kemungkinan terbesarnya adalah tidak ada meninggal atau sembuh pada hari tersebut.

Untuk kolom yang seluruhnya missing yaitu `combined_key` dapat dibuang saja satu kolom itu karena tidak ada data yang dapat diketahui dari kolom tersebut.

Sementara, kolom yang lainnya bagaimana? Misal ambil kolom `province_stat`, missing valuenya dapat terjadi bahwa tidak dilaporkan itu berasal dari daerah mana di negara itu. Dapat mengisi misal dengan string `'unknown'` karena tahu kolom tersebut bertipe data string.