

DATA SCIENCE PROJECT



EKSPLORASI & ANALISIS DATA COVID-19 INDONESIA

by Andrean Yonathan

Outline

1

PERKENALAN

2

TENTANG
PROJECT

3

COVID-19 DI
INDONESIA

4

COVID-19 DI
JAWA BARAT

5

PENUTUP

Tentang Saya



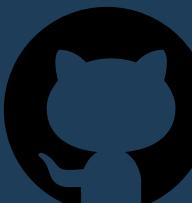
**Andrean
Yonathan**



Tangerang Selatan, Banten



linkedin.com/in/andreanynthn



github.com/andreanynthn

S1 Matematika | 2019 - sekarang

Tentang Project

Covid-19 merupakan virus yang mulai menyebar di akhir tahun 2019. Penyebarannya yang begitu cepat membuat penyakit ini telah ditetapkan sebagai pandemi oleh WHO. Pemerintah di berbagai negara umumnya dengan sigap membuat gugus tugas untuk menangani penyebaran Covid-19 di masyarakat, termasuk pemerintah Indonesia.

Salah satu bentuk aksi yang dilakukan oleh pemerintah adalah dengan mengumpulkan dan menyediakan data pertumbuhan kasus COVID-19 kepada publik. Data pertumbuhan kasus tersebut tidak jarang juga dilengkapi dengan dasbor dan grafik visualisasi pendukung dengan harapan masyarakat dapat memahami informasi dengan lebih mudah.

Dalam project ini, akan dilakukan eksplorasi serta analisis data Covid-19 di Indonesia dan juga Jawa Barat. Data yang digunakan dalam project ini merupakan data *real time* yang diambil langsung dari API yang tersedia di covid19.go.id



COVID-19 DI INDONESIA



DATA SCIENCE PROJECT

Import Library

Sebelum melakukan analisis data, perlu melakukan import library yang dibutuhkan terlebih dahulu. Library yang digunakan dalam project ini dapat dilihat pada gambar di bawah.

Import Library

```
In [1]: import requests
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
```

Import Library

1. Library *requests* digunakan untuk mengirim berbagai request HTTP, dalam hal ini digunakan untuk mengakses data Covid-19 di Indonesia.
2. Library *pandas* digunakan untuk melakukan manipulasi, cleansing, maupun analisis struktur data.
3. Library *numpy* digunakan untuk melakukan proses komputasi numerik terutama dalam bentuk array multidimensional.
4. Library *matplotlib* digunakan untuk membuat grafik.

Mengakses API covid19.go.id

Selanjutnya, untuk mengakses data covid-19 di Indonesia, kita akan melakukannya dengan mengakses API website covid19.go.id dengan menggunakan fungsi `.get()` dari library `requests` seperti gambar di bawah ini

Mengakses API covid19.go.id

```
In [2]: resp = requests.get('https://data.covid19.go.id/public/api/update.json', verify=False)
print(resp)

<Response [200]>
```

Kode Response [200] menandakan bahwa permintaan telah dipenuhi. Informasi lengkap mengenai jenis-jenis status code dapat dilihat di <https://restfulapi.net/http-status-codes/>

Headers API

Headers merupakan salah satu dari tiga informasi utama dalam sebuah response API. Headers umumnya mengandung informasi metadata. Sedangkan dua informasi lainnya yaitu status yang memiliki informasi apakah permintaan telah berhasil atau tidak dan juga body yang berisi konten atas permintaan yang telah kita buat. Untuk mengetahui informasi metada data pada `resp`, kita dapat menggunakan attribut `headers` seperti gambar di bawah

Mengetahui Headers resp

```
In [3]: print(resp.headers)
```

```
{'Server': 'nginx', 'Date': 'Tue, 19 Oct 2021 15:39:08 GMT', 'Content-Type': 'application/json', 'Transfer-Encoding': 'chunked', 'Connection': 'keep-alive', 'Vary': 'Accept-Encoding, Accept-Encoding', 'Last-Modified': 'Tue, 19 Oct 2021 15:38:44 GMT', 'ETag': 'W/"616ee684-38780"', 'X-Content-Type-Options': 'nosniff, nosniff', 'X-XSS-Protection': '1; mode=block, 1; mode=block', 'Strict-Transport-Security': 'max-age=31536000; includeSubDomains; preload', 'Content-Encoding': 'gzip'}
```

Dari hasil di atas, diketahui bahwa konten yang diminta tersedia dalam bentuk json.

Mengekstrak Isi Respon

Untuk mengekstrak konten yang telah diminta, kita dapat menggunakan method `json()` dan menyimpannya dalam `cov_id_raw`. Method `json()` merupakan builtin Json decoder yang digunakan untuk mengekstrak content bertipe json.

Mengekstrak Isi Respon

Jalankan method `json()` pada obyek `resp` dan simpanlah hasilnya sebagai `cov_id_raw`. Dengan menggunakan fungsi `len()` dan method `keys()` cobalah amati ada berapa komponen serta apa saja nama komponen dalam obyek `cov_id_raw` tersebut! Kemudian ekstraklah key ke-2 dan simpan dengan nama `cov_id_update`.

```
In [4]: cov_id_raw = resp.json()
cov_id_raw
```

```
Out[4]: {'data': {'id': 1,
  'jumlah_odp': 6074,
  'jumlah_pdp': 0,
  'total_spesimen': 43863159,
  'total_spesimen_negatif': 34306833},
 'update': {'penambahan': {'jumlah_positif': 903,
  'jumlah_meninggal': 50,
  'jumlah_sembuh': 1530,
  'jumlah_dirawat': -677,
  'tanggal': '2021-10-19',
  'created': '2021-10-19 17:13:38'},
```

Mengekstrak Isi Respon

Selanjutnya dengan menggunakan fungsi `len()` dan `keys()` kita akan mengetahui ada berapa dan apa saja komponen yang terdapat dalam `cov_id_raw` dan kemudian kita akan mengekstrak key ke-2 dan menyimpannya dalam `cov_id_update`.

```
In [5]: print('Length of cov_id_raw : %d.' %len(cov_id_raw))
print('Komponen cov_id_raw : %s.' %cov_id_raw.keys())
Length of cov_id_raw : 2.
Komponen cov_id_raw : dict_keys(['data', 'update']).
```

```
In [6]: cov_id_update = cov_id_raw['update']
```

Analisis Data

Dalam proses analisis data, kita akan fokus pada data yang terdapat di dalam variabel `cov_id_update` untuk menjawab pertanyaan-pertanyaan berikut:

1. Kapan tanggal pembaharuan data penambahan kasus?
2. Berapa jumlah penambahan kasus sembuh?
3. Berapa jumlah penambahan kasus meninggal?
4. Berapa jumlah total kasus positif hingga saat ini?
5. Berapa jumlah total kasus meninggal hingga saat ini?

Analisis Data

- Berdasarkan hasil di atas, dapat diambil kesimpulan sebagai berikut:
- 1.Tanggal pembaharuan data penambahan kasus Covid-19 di Indonesia adalah 20 Oktober 2021.
 - 2.Per tanggal 19 Oktober 2021, terdapat penambahan kasus sembuh sebanyak 1207 orang.
 - 3.Per tanggal 19 Oktober 2021, jumlah kasus meninggal bertambah sebanyak 28 orang.
 - 4.Total kasus positif hingga saat ini (19 Oktober 2021) sebanyak 4237201 orang.
 - 5.Total kasus meninggal hingga saat ini (19 Oktober 2021) sebanyak 143077 orang.

```
In [34]: # Kapan tanggal pembaharuan data penambahan kasus?  
print('Tanggal pembaharuan data penambahan kasus :', cov_id_update['penambahan']['tanggal'])  
  
# Berapa jumlah penambahan kasus sembuh?  
print('Jumlah penambahan kasus sembuh :', cov_id_update['penambahan']['jumlah_sembuh'])  
  
# Berapa jumlah penambahan kasus meninggal?  
print('Jumlah penambahan kasus meninggal :', cov_id_update['penambahan']['jumlah_meninggal'])  
  
# Berapa jumlah total kasus positif hingga saat ini?  
print('Jumlah total kasus positif hingga saat ini :', cov_id_update['total']['jumlah_positif'])  
  
# Berapa jumlah total kasus meninggal hingga saat ini?  
print('Jumlah total kasus meninggal hingga saat ini :', cov_id_update['total']['jumlah_meninggal'])
```

```
Tanggal pembaharuan data penambahan kasus : 2021-10-20  
Jumlah penambahan kasus sembuh : 1207  
Jumlah penambahan kasus meninggal : 28  
Jumlah total kasus positif hingga saat ini : 4237201  
Jumlah total kasus meninggal hingga saat ini : 143077
```

COVID-19 DI JAWA BARAT



DATA SCIENCE PROJECT

Mengakses dan Mengekstrak API

Dengan menggunakan fungsi `.get()` dari library `requests` dan method `json()`, kita akan mengakses dan mengekstrak API data Covid-19 di Jawa Barat.

```
Apa Kabar Jawa Barat?

In [8]: # mengakses dan mengekstrak API
resp_jabar = requests.get('https://data.covid19.go.id/public/api/prov_detail_JAWA_BARAT.json',
                           verify=False)
cov_jabar_raw = resp_jabar.json()
resp_jabar

C:\Users\USER\AppData\Local\Programs\Python\Python39\lib\site-packages\urllib3\connectionpool.py:101
 3: InsecureRequestWarning: Unverified HTTPS request is being made to host 'data.covid19.go.id'. Adding certificate verification is strongly advised. See: https://urllib3.readthedocs.io/en/1.26.x/advanced-usage.html#ssl-warnings
  warnings.warn(
Out[8]: <Response [200]>
```

Status code Response [200] menunjukkan bahwa permintaan kita telah terpenuhi.

Memahami Kasus di Jawa Barat

Selanjutnya, kita akan menggunakan method `keys()` pada `cov_jabar_raw` untuk mengetahui elemen-elemen utama yang tersedia dan kita juga akan menjawab pertanyaan-pertanyaan berikut:

1. Kapan tanggal pembaharuan data COVID-19 di Jawa Barat?
2. Berapa jumlah total kasus COVID-19 di Jawa Barat?
3. Berapa persentase kematian akibat COVID-19 di Jawa Barat?
4. Berapa persentase tingkat kesembuhan dari COVID-19 di Jawa Barat?

Memahami Kasus di Jawa Barat

```
In [27]: # nama-nama elemen utama
print('Nama-nama elemen utama:\n', cov_jabar_raw.keys())

# tanggal pembaharuan data
print('\nTanggal pembaharuan data COVID-19 di Jawa Barat : %s' %cov_jabar_raw['last_date'])

# jumlah total kasus positif di Jawa Barat
print('Jumlah total kasus COVID-19 di Jawa Barat : %d' %cov_jabar_raw['kasus_total'])

# persentase kematian akibat covid-19 di Jawa Barat
print('Persentase kematian akibat COVID-19 di Jawa Barat : %f.2%%' %cov_jabar_raw['meninggal_persen'])

# persentase tingkat kesembuhan covid-19 di Jawa Barat
print('Persentase tingkat kesembuhan dari COVID-19 di Jawa Barat : %f.2%%' %cov_jabar_raw['sembuh_persen'])

Nama-nama elemen utama:
dict_keys(['last_date', 'provinsi', 'kasus_total', 'kasus_tanpa_tgl', 'kasus_dengan_tgl', 'meninggal_persen', 'meninggal_tanpa_tgl', 'meninggal_dengan_tgl', 'sembuh_persen', 'sembuh_tanpa_tgl', 'sembuh_dengan_tgl', 'list_perkembangan', 'data'])

Tanggal pembaharuan data COVID-19 di Jawa Barat : 2021-10-20
Jumlah total kasus COVID-19 di Jawa Barat : 704642
Persentase kematian akibat COVID-19 di Jawa Barat : 2.081766.2%
Persentase tingkat kesembuhan dari COVID-19 di Jawa Barat : 97.667752.2%
```

Berdasarkan hasil pada gambar di atas, didapatkan beberapa informasi berikut:

- 1.Tanggal pembaharuan data Covid-19 di Jawa Barat yaitu pada **20 Oktober 2021**.
- 2.Total kasus Covid-19 di Jawa Barat per 20 Oktober 2021 sebanyak **704642** kasus.
- 3.Persentase kematian akibat Covid-19 di Jawa Barat per 20 Oktober 2021 sebesar **2.081766.2%**.
- 4.Persentase tingkat kesembuhan dari Covid-19 di Jawa Barat per 20 Oktober 2021 sebesar **97.667752.2%**.

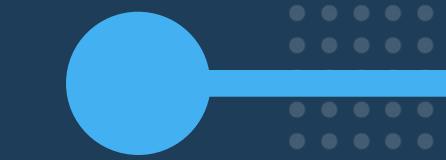
Memperoleh Informasi yang Lebih Lengkap

Data historis perkembangan Covid-19 di Jawa Barat tersimpan dengan nama `list_perkembangan`. Selanjutnya kita akan mengekstrak data tersebut dari `cov_jabar_raw` dan mengubahnya ke bentuk pandas dataframe dan menyimpannya dalam `cov_jabar` dengan menggunakan sintaks berikut

```
In [10]: cov_jabar = pd.DataFrame(cov_jabar_raw['list_perkembangan'])
print('Info cov_jabar:\n', cov_jabar.info())
print('\nLima data teratas cov_jabar:\n', cov_jabar.head())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 596 entries, 0 to 595
Data columns (total 9 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   tanggal          596 non-null    int64  
 1   KASUS             596 non-null    int64  
 2   MENINGGAL         596 non-null    int64  
 3   SEMBUH            596 non-null    int64  
 4   DIRAWAT_OR_ISOLASI 596 non-null    int64  
 5   AKUMULASI_KASUS   596 non-null    int64  
 6   AKUMULASI_SEMBUH  596 non-null    int64  
 7   AKUMULASI_MENINGGAL 596 non-null    int64  
 8   AKUMULASI_DIRAWAT_OR_ISOLASI 596 non-null    int64  
dtypes: int64(9)
memory usage: 42.0 KB
Info cov_jabar:
None
```

Menjinakan Data



Mengingat data yang ada di dalam dataframe tidak semuanya dipakai, maka kita akan "menjinakan" atau merapikan data yang kita miliki agar mempermudah proses analisis data. Ada beberapa tahap yang perlu dilakukan, yaitu

1. Menghapus kolom `DIRAWAT_OR_ISOLASI` dan semua kolom yang berisi nilai kumulatif
2. Merubah format penulisan semua kolom menjadi huruf kecil
3. Mengganti nama kolom `kasus` menjadi `kasus_baru`
4. Memperbaiki tipe data pada kolom `tanggal` menggunakan `pd.to_datetime`

Hasil pengolahan akan disimpan di dalam `cov_jabar_tidy`

Menjinakan Data

Ada beberapa tahapan yang akan Anda lakukan untuk menjinakan data `cov_jabar`, yaitu:

1. Menghapus kolom `DIRAWAT_OR_ISOLASI` dan semua kolom yang berisi nilai kumulatif
2. Merubah format penulisan semua kolom menjadi huruf kecil
3. Mengganti nama kolom `kasus` menjadi `kasus_baru`
4. Memperbaiki tipe data pada kolom `tanggal` menggunakan `pd.to_datetime`

```
In [11]: cov_jabar_tidy = (cov_jabar.drop(columns=[item for item in cov_jabar.columns
                                                if item.startswith('AKUMULASI')
                                                or item.startswith('DIRAWAT')])
                                .rename(columns=str.lower)
                                .rename(columns={'kasus': 'kasus_baru'})
                               )
cov_jabar_tidy['tanggal'] = pd.to_datetime(cov_jabar_tidy['tanggal']*1e6, unit='ns')
print('Lima data teratas:\n', cov_jabar_tidy.head())
```

Lima data teratas:

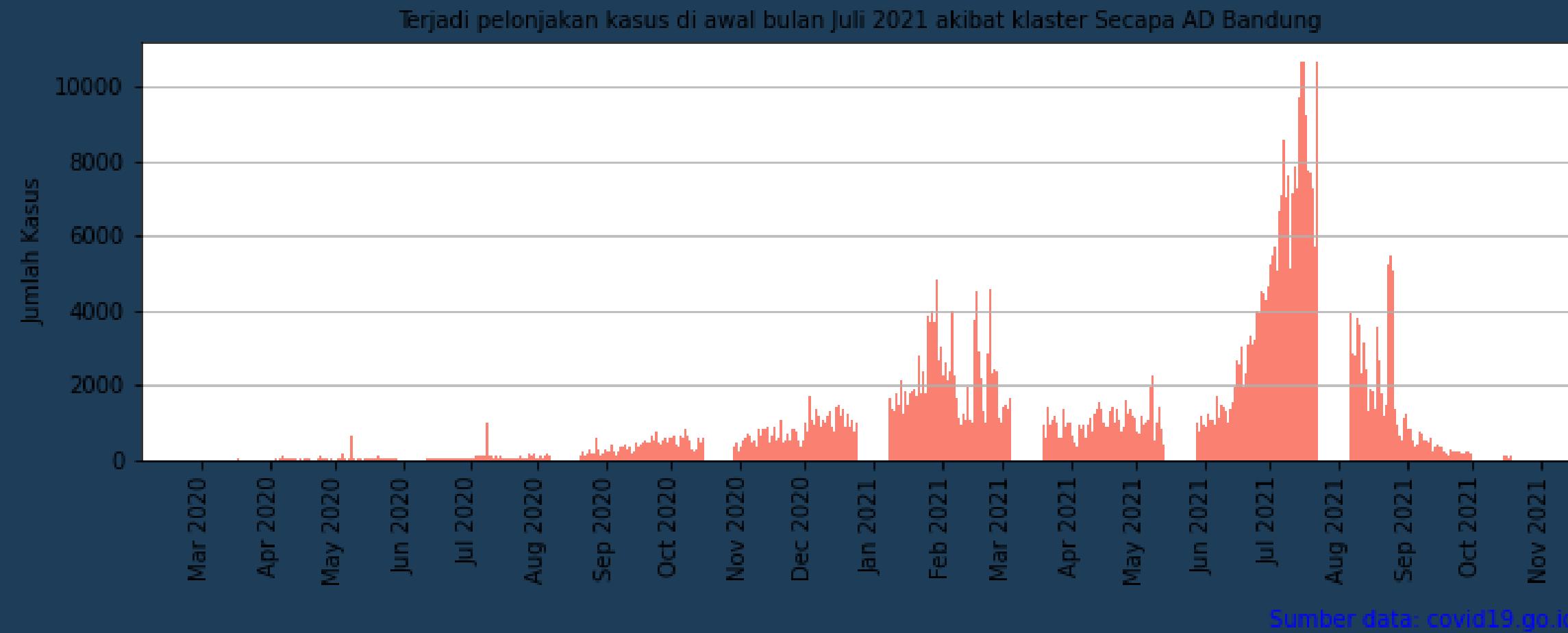
	tanggal	kasus_baru	meninggal	sembuh
0	2020-03-03	1	1	0
1	2020-03-04	1	0	0
2	2020-03-05	1	0	0
3	2020-03-06	1	0	0
4	2020-03-07	0	0	0

Analisis Data

Dalam proses analisis data, kita akan fokus pada data yang terdapat di dalam variabel cov_jabar_tidy.

Analisis Data - Kasus Baru

Kasus Harian Positif COVID-19 di Jawa Barat

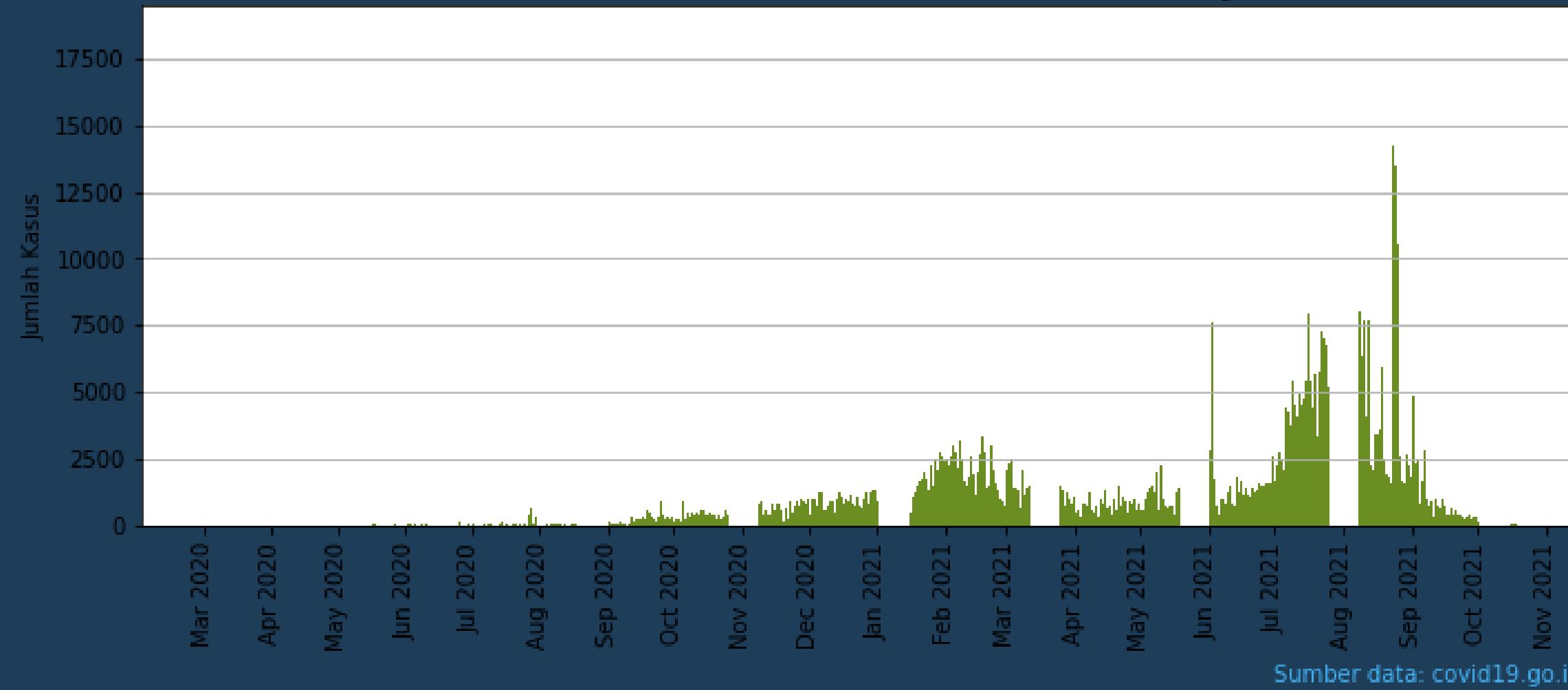


Gambar di atas merupakan grafik kasus baru yang ada di Jawa Barat. Berdasarkan grafik di atas, terjadi pelonjakan kasus di awal bulan Juli 2021 yang diakibatkan oleh klaster Secapa AD Bandung. Informasi selengkapnya tentang klaster Secapa AD Bandung dapat diakses pada <https://news.detik.com/berita/d-5088197/secapa-ad-bandung-jadi-klaster-corona-ini-fakta-dan-perjalanan-kasusnya>

Analisis Data - Kasus Sembuh



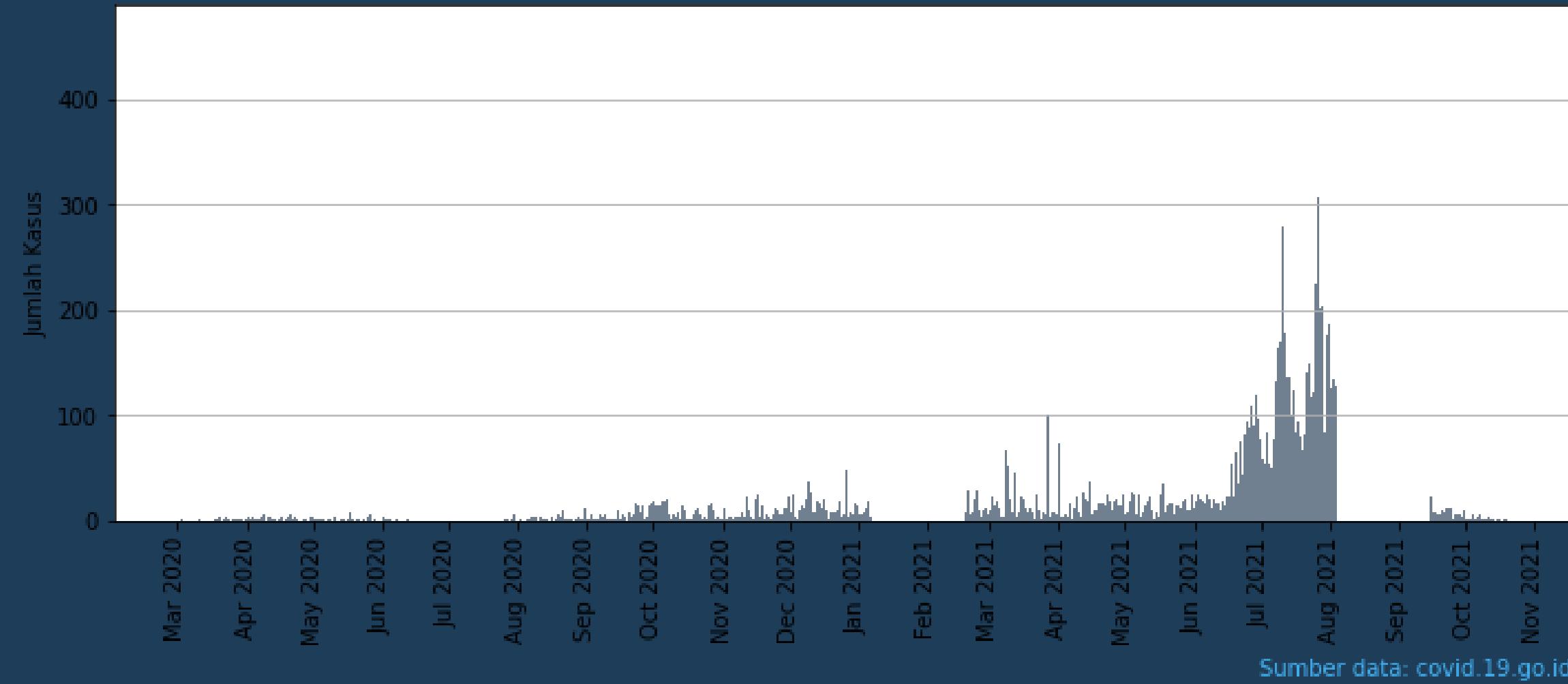
Kasus Harian Sembuh Dari COVID-19 di Jawa Barat



Gambar di atas merupakan grafik kasus sembuh harian yang ada di Jawa Barat. Berdasarkan grafik di atas, pada akhir Juli 2021 sampai akhir Agustus 2021 kasus sembuh harian di Jawa Barat mengalami peningkatan yang cukup tinggi.

Analisis Data - Kasus Meninggal

kasus Harian Meninggal Dari Covid-19 di Jawa Barat



Gambar di atas merupakan grafik kasus meninggal yang ada di Jawa Barat. Berdasarkan grafik di atas, kasus meninggal karena Covid-19 di Jawa Barat mengalami penurunan yang signifikan setelah pada bulan Juli - Agustus mengalami pelonjakan yang cukup tinggi.

Analisis Data - Apakah Pekan Ini Lebih Baik?

Untuk menghitung pertambahan kasus pekanan yang perlu dilakukan yaitu set terlebih dahulu kolom tanggal pada cov_jabar_tidy sebagai index. Kemudian terapkan method pada time series di pandas, resampling per pekan dan gunakan agregasi penjumlahan. Adapun yang akan digunakan untuk mengekstrak informasi pekan dalam satu tahun adalah `weekofyear` dan `year` untuk tahun. Hasil pengolahan akan disimpan di dalam cov_jabar_pekanan.

```
In [26]: cov_jabar_pekanan = (cov_jabar_tidy.set_index('tanggal')['kasus_baru']
                           .resample('W')
                           .sum()
                           .reset_index()
                           .rename(columns={'kasus_baru': 'jumlah'})
                           )
cov_jabar_pekanan['tahun'] = cov_jabar_pekanan['tanggal'].apply(lambda x: x.year)
cov_jabar_pekanan['pekan_ke'] = cov_jabar_pekanan['tanggal'].apply(lambda x: x.weekofyear)
cov_jabar_pekanan = cov_jabar_pekanan[['tahun', 'pekan_ke', 'jumlah']]

print('Info cov_jabar_pekanan:')
print(cov_jabar_pekanan.info())
print('\nLima data teratas cov_jabar_pekanan:\n', cov_jabar_pekanan.head())

Info cov_jabar_pekanan:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 86 entries, 0 to 85
Data columns (total 3 columns):
 #   Column   Non-Null Count  Dtype  
--- 
 0   tahun     86 non-null    int64  
 1   pekan_ke  86 non-null    int64  
 2   jumlah    86 non-null    int64  
dtypes: int64(3)
memory usage: 2.1 KB
None

Lima data teratas cov_jabar_pekanan:
  tahun  pekan_ke  jumlah
0  2020      10       4
1  2020      11       5
2  2020      12      44
3  2020      13      80
4  2020      14      89
```

Analisis Data - Apakah Pekan Ini Lebih Baik dari Pekan Kemarin?

Untuk mengetahui apakah pekan ini lebih baik dari pekan kemarin, maka akan dilakukan kalkulasi sederhana dengan tahapan sebagai berikut:

1. Membuat kolom baru yang berisi jumlah kasus baru dalam satu pekan sebelumnya. Kolom ini diberi nama **jumlah_pekanlalu**.
2. Mengganti nilai **NaN** pada kolom **jumlah_pekanlalu** dengan nilai 0
3. Melakukan komparasi antara kolom **jumlah** dengan kolom **jumlah_pekanlalu**. Hasil komparasi ini disimpan dalam kolom baru dengan nama **lebih_baik**, isinya adalah **True** apabila jumlah kasus baru pekan ini lebih rendah dibandingkan jumlah kasus pekan lalu

Analisis Data - Apakah Pekan Ini Lebih Baik dari Pekan Kemarin?

```
In [49]: cov_jabar_pekanan['jumlah_pekanlalu'] = cov_jabar_pekanan['jumlah'].shift().replace(np.nan, 0).astype(np.int)
cov_jabar_pekanan['lebih_baik'] = cov_jabar_pekanan['jumlah'] < cov_jabar_pekanan['jumlah_pekanlalu']

print('Sepuluh data terbaru:\n', cov_jabar_pekanan.tail(10))
```

	tahun	pekan_ke	jumlah	jumlah_pekanlalu	lebih_baik
76	2021	33	14183	19461	True
77	2021	34	20188	14183	False
78	2021	35	5423	20188	True
79	2021	36	3942	5423	True
80	2021	37	1991	3942	True
81	2021	38	1414	1991	True
82	2021	39	1078	1414	True
83	2021	40	679	1078	True
84	2021	41	654	679	True
85	2021	42	172	654	True

Data pada gambar di atas menunjukkan bahwa 10 pekan terakhir jumlah kasus Covid-19 di Jawa Barat mengalami penurunan tiap pekan kecuali pada pekan 34 yang mengalami peningkatan kasus dari pekan sebelumnya.

Analisis Data - Hingga Saat Ini Ada Berapa Kasus yang Masih Aktif?

Jumlah kasus aktif dapat dihitung dengan cara mengurangi jumlah akumulasi positif dengan jumlah akumulasi sembuh dan jumlah akumulasi meninggal.

note: aktif dalam artian sedang dalam perawatan atau isolasi

```
In [46]: cov_jabar_akumulasi = cov_jabar_tidy[['tanggal']].copy()
cov_jabar_akumulasi['akumulasi_aktif'] = (cov_jabar_tidy['kasus_baru'] -
                                             cov_jabar_tidy['sembuh'] - cov_jabar_tidy['meninggal']).cumsum()
cov_jabar_akumulasi['akumulasi_sembuh'] = cov_jabar_tidy['sembuh'].cumsum()
cov_jabar_akumulasi['akumulasi_meninggal'] = cov_jabar_tidy['meninggal'].cumsum()
print(cov_jabar_akumulasi.tail())
```

	tanggal	akumulasi_aktif	akumulasi_sembuh	akumulasi_meninggal
591	2021-10-15	1709	687919	14664
592	2021-10-16	1738	687997	14665
593	2021-10-17	1726	688079	14665
594	2021-10-18	1706	688165	14666
595	2021-10-19	1770	688204	14668

Dari hasil di atas diperoleh beberapa informasi berikut

1. Per tanggal 19 Oktober 2021, akumulasi kasus aktif di Jawa Barat sebanyak **1770** kasus.
2. Per tanggal 19 Oktober 2021, akumulasi kasus sembuh di Jawa Barat sebanyak **688204** kasus.
3. Per 19 Oktober 2021, akumulasi meninggal di Jawa Barat sebanyak **14668** kasus.

Analisis Data - Pola Kasus Aktif

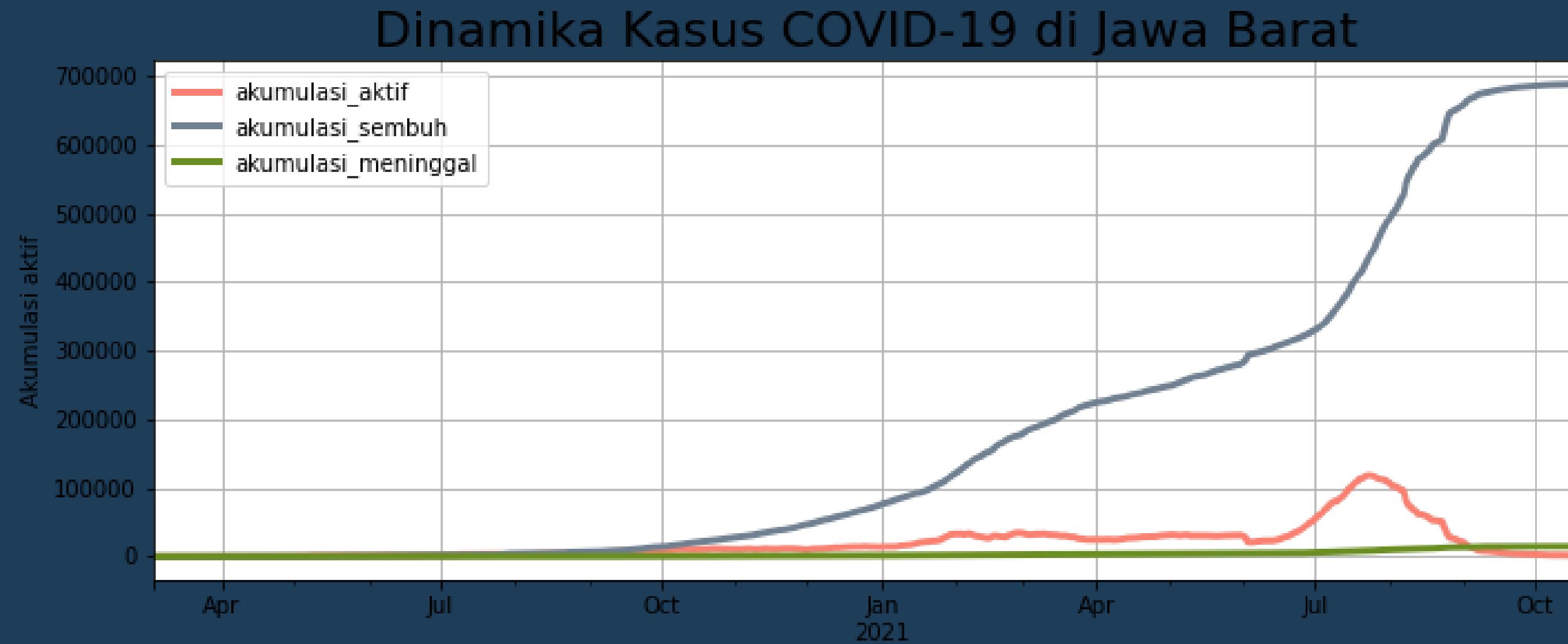


Akumulasi Aktif COVID-19 di Jawa Barat



Grafik di atas menunjukkan perkembangan kasus yang sangat baik di Jawa Barat di mana sejak akhir Juli 2021 hingga saat ini grafik kasus aktif di Jawa Barat terus mengalami penurunan. Pada slide selanjutnya akan ditunjukkan grafik dinamika kasus Covid-19 di Jawa Barat yang kian membaik.

Analisis Data - Dinamika Kasus Covid-19 di Jawa Barat



Sumber data: covid.19go.id

Kesimpulan

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan sebagai berikut:

1. Total kasus Covid-19 di Jawa Barat per 20 Oktober 2021 sebanyak **704642** kasus.
2. Pada tanggal 20 Oktober 2021 terjadi penambahan kasus sembuh sebanyak **1207** orang dan kasus meninggal sebanyak **28** orang
3. Total kasus Covid-19 di Jawa Barat per 20 Oktober 2021 sebanyak **704642** kasus.
4. Dalam 10 pekan terakhir, kasus Covid-19 di Jawa Barat mengalami penurunan.
5. Secara keseluruhan, kasus Covid-19 di Jawa Barat mengalami perkembangan yang baik ditunjukkan oleh grafik akumulasi kasus aktif yang mengalami penurunan dan grafik kasus sembuh yang mengalami peningkatan.