**Rencana Tugas Mandiri Ke-4**

**Mata Kuliah Analisis Data Eksploratif**



**Mata Kuliah Analisis Data Eksploratif**

**Implementasi Analisis Korelasi, Time Series Analysis, Hypothesis**

**Testing dan Regresi Pada Data COVID-19**

**KELAS A**

**DISUSUN OLEH :**

AHMAD ARDHY ANSYAH ( 20083010008 )

**DOSEN PENGAMPU:**

TRESNA MAULANA FAHRUDIN, S.S.T., MT ( 20219930501200 )

PROGRAM STUDI SAINS DATA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”

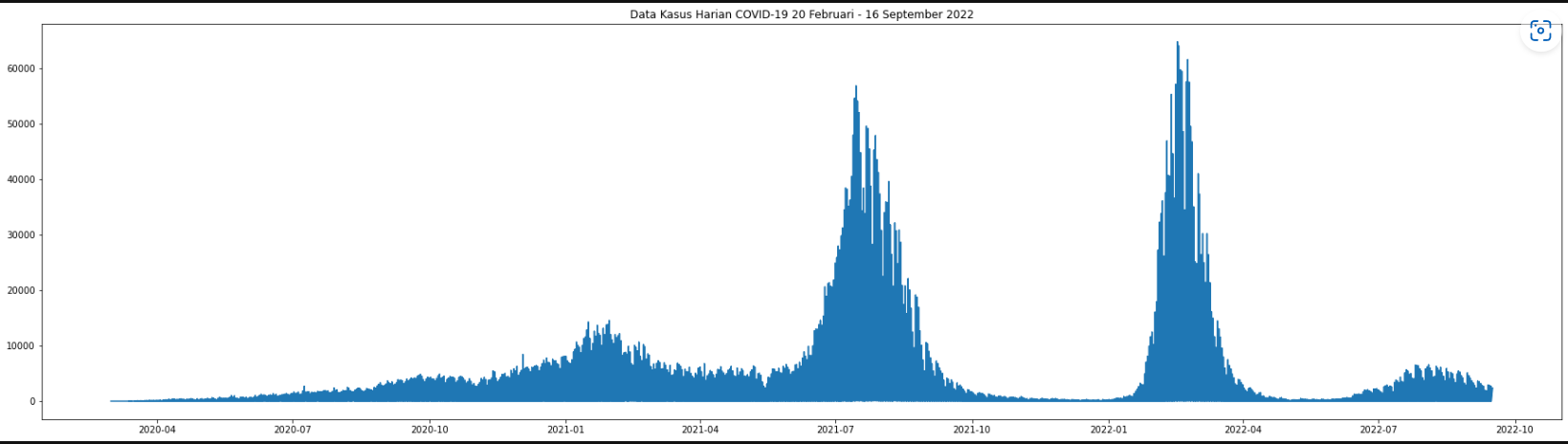
JAWA TIMUR

2022

a. Time Series Analysis

- Visualisasi data selama rentang beberapa tahun

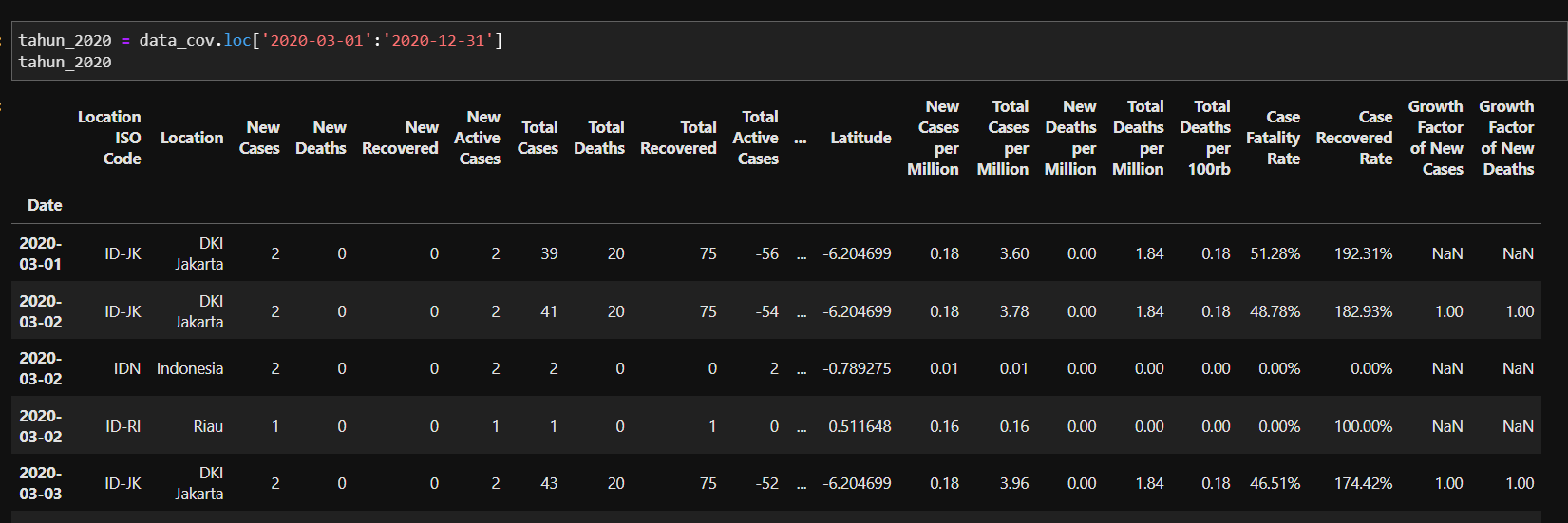
Di sini saya menggunakan data covid columns ‘New Cases’. Pada bagian pertama saya visualisasi data covid ‘New Cases’ dalam rentang beberapa tahun, karena data dimulai dari tahun 2020 bulan Maret hingga tahun 2022 bulan September.



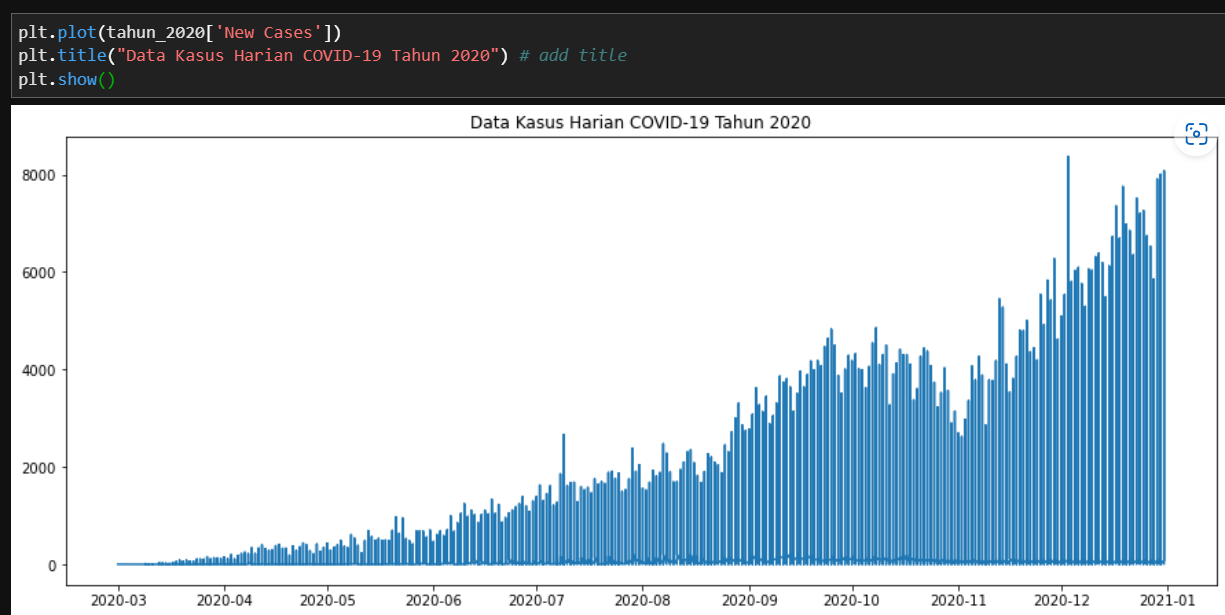
- Pengelompokan (grouping) data time series berdasarkan bulan

Setelah itu saya lakukan visualisasi data tersebut dengan pengelompokan visualisasi 1 tahun yang menunjukkan kasus pertumbuhan setiap bulannya

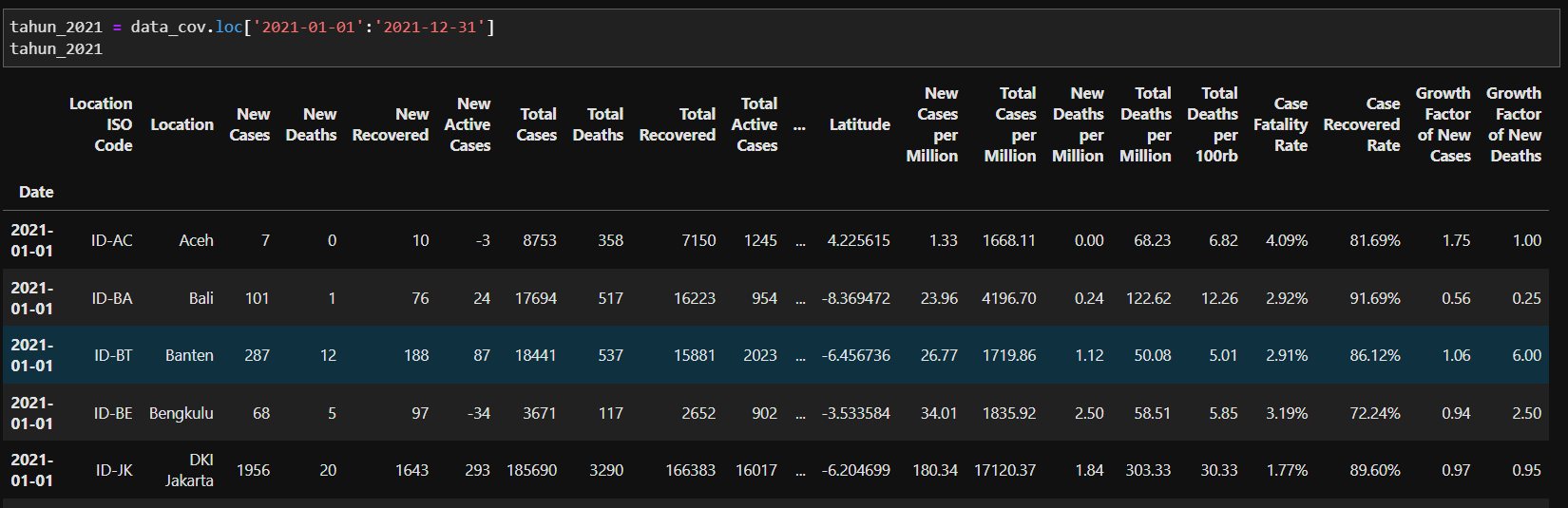
Sebelum kita visualisasi kita buat kelompok data setiap tahunnya



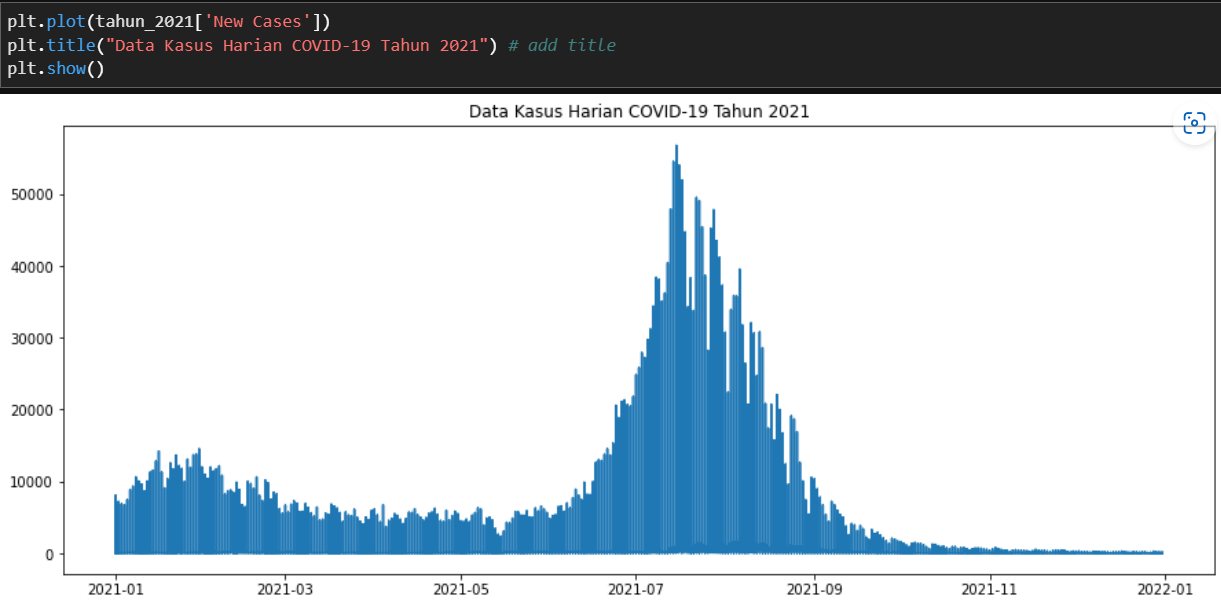
Visualisasi Data Kasus Baru Harian COVID – 19 Pada Tahun 2020 setiap bulannya



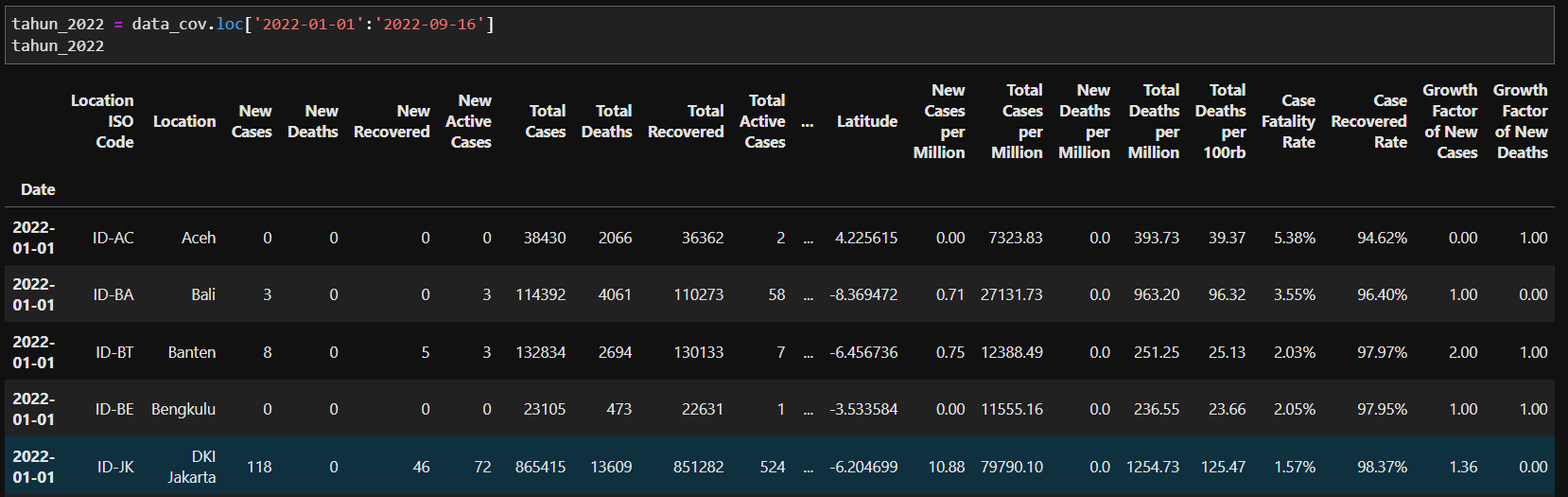
Pengelompokkan Data Pertumbuhan Kasus Baru COVID – 19 pada Tahun 2021



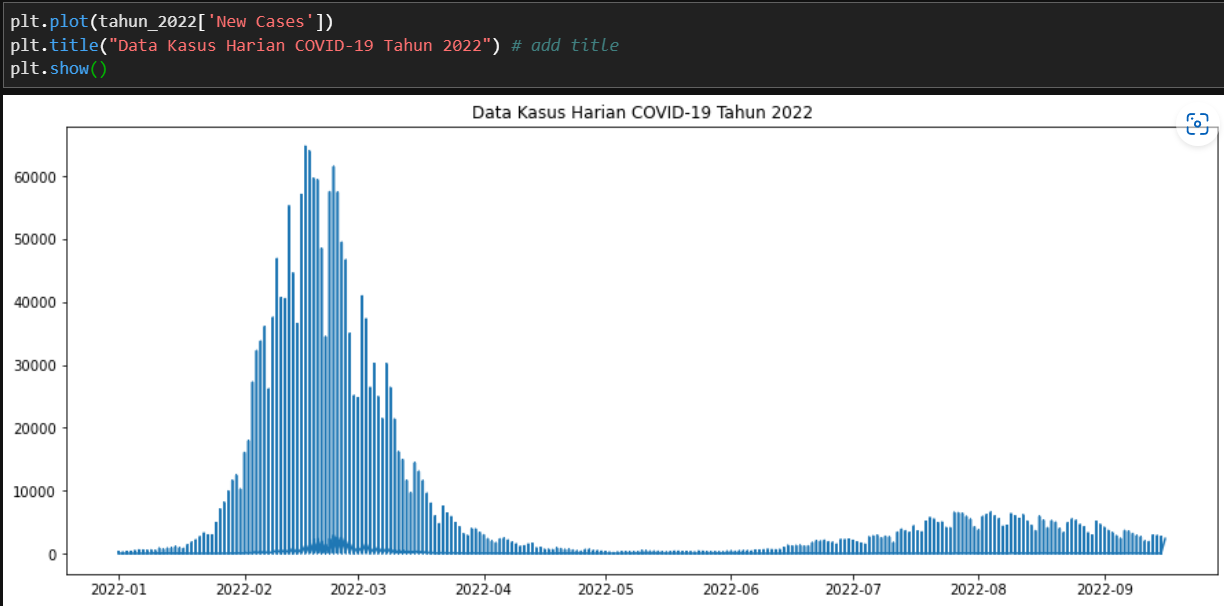
Visualisasi Data Kasus Baru Harian COVID – 19 Pada Tahun 2021 setiap bulannya



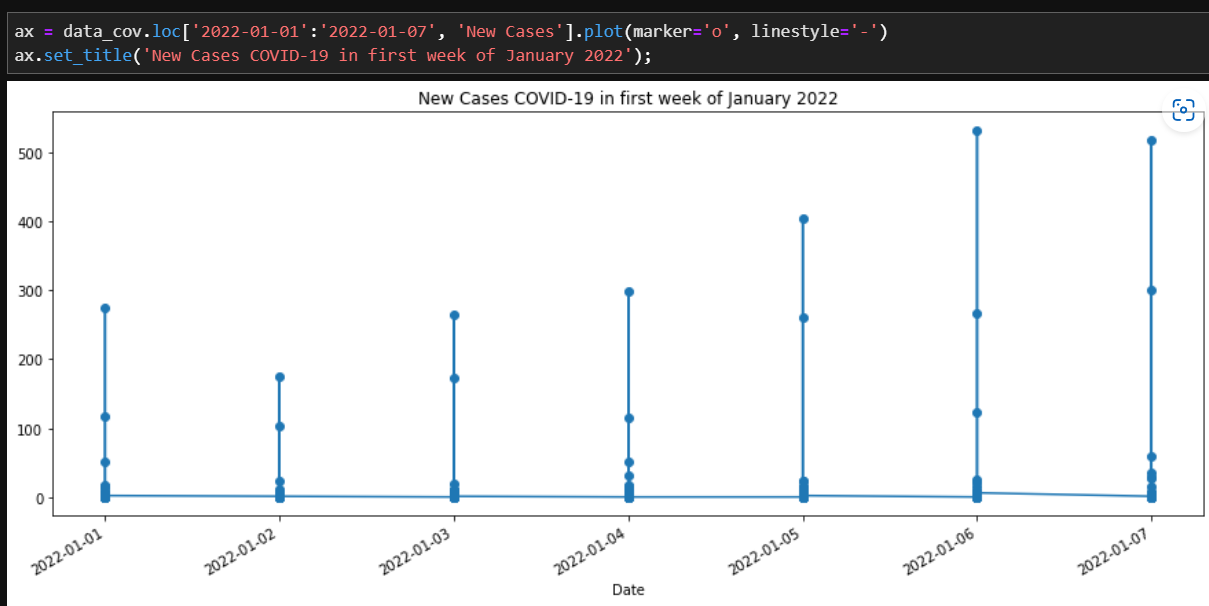
Pengelompokkan Data Pertumbuhan Kasus Baru COVID – 19 pada Tahun 2022



Visualisasi Data Kasus Baru Harian COVID – 19 Pada Tahun 2022 setiap bulannya

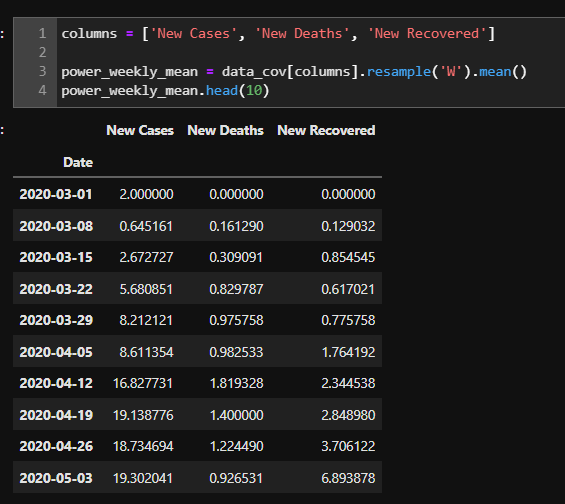


- Pengelompokan (grouping) data time series berdasarkan hari



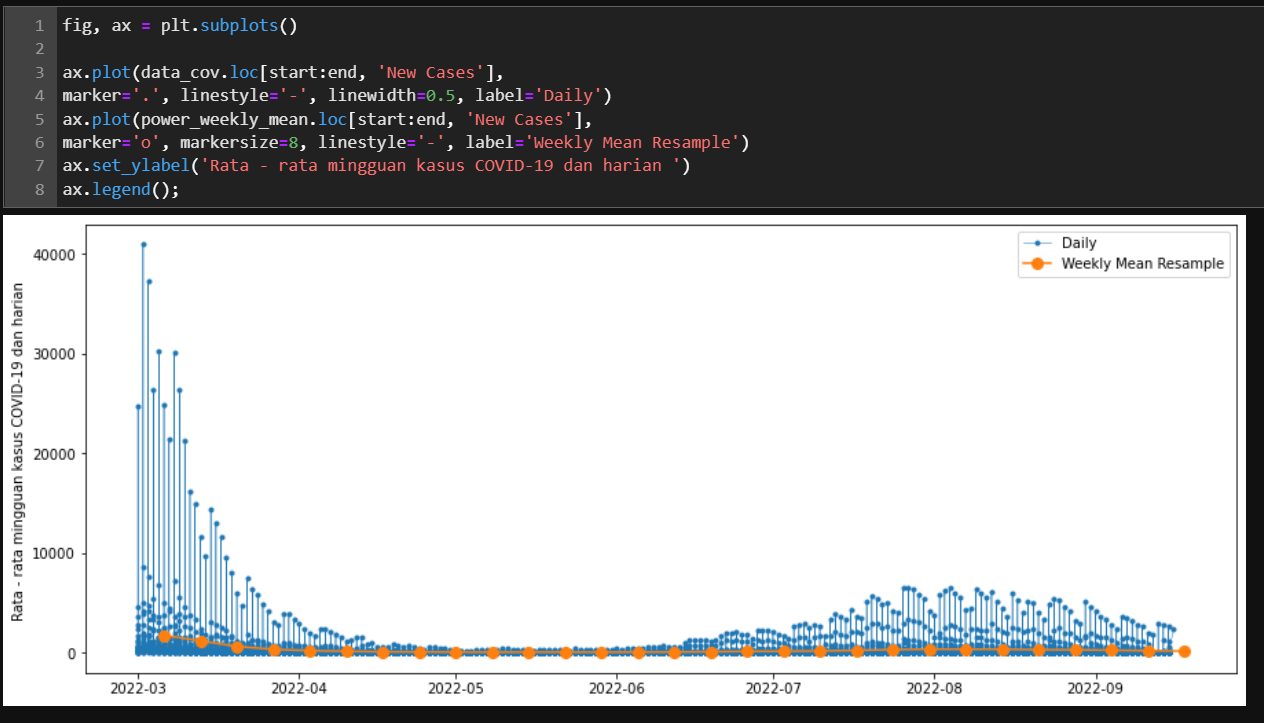
- Visualisasi bersama rata-rata mingguan kasus COVID-19 dan harian

Sebelum melakukan plot pada rata – rata mingguan kasus COVID – 19 perlu dilakukan perhitungan terlebih dahulu untuk menghitung rata – rata dari kasus pertumbuhan setiap minggu dan setiap harinya



dan juga kita tetapkan mulai visualisasi data dimulai dan berakhir pada tanggal berapa

setelah itu baru diplot hasil perhitungan tadi dengan kode di bawah ini

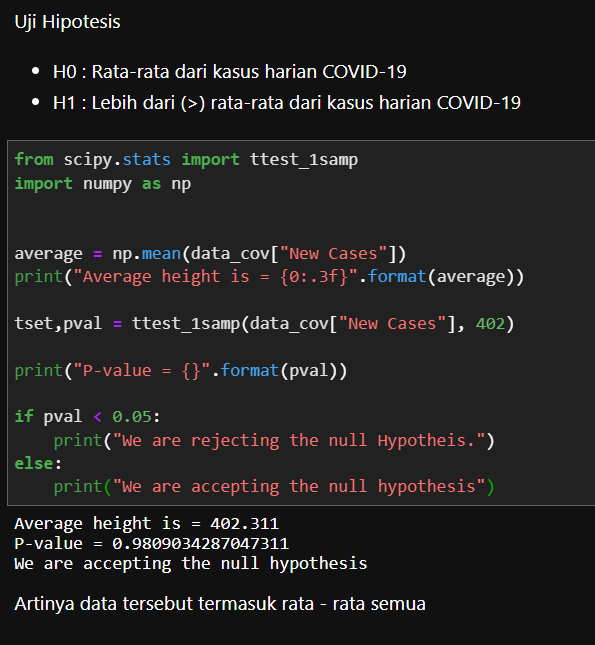


b. Hypothesis Testing

- Gunakan T-test dan H0 = rata-rata dari kasus harian COVID-19, H1 = lebih dari (>) rata-rata dari kasus harian COVID-19

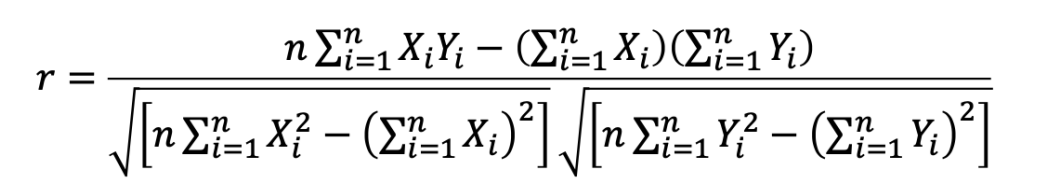
- Simpulkan apakah H0 diterima/ditolak

- Gunakan library Scipy menggunakan modul ttest\_1samp

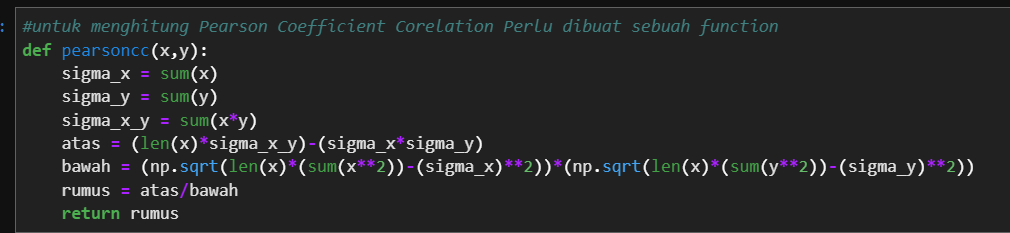


H0 diterima dan berarti semua data tersebut sama dengan rata - rata

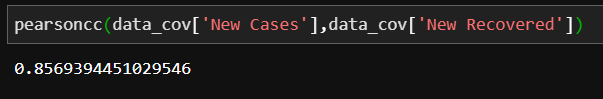
c. Analisis Korelasi Gunakan Pearson untuk mendapatkan nilai koefisian korelasi, implementasikan ke dalam Python secara scratch



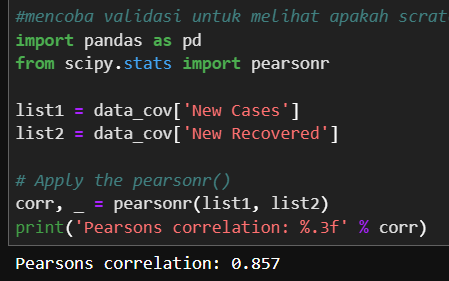
Untuk Menghitung rumus tersebut saya membuat sebuah function scratch dengan inputan rumus di atas. Begini function nya :



Setelah kita buat function nya kita panggil function tersebut dan di-input kan datanya



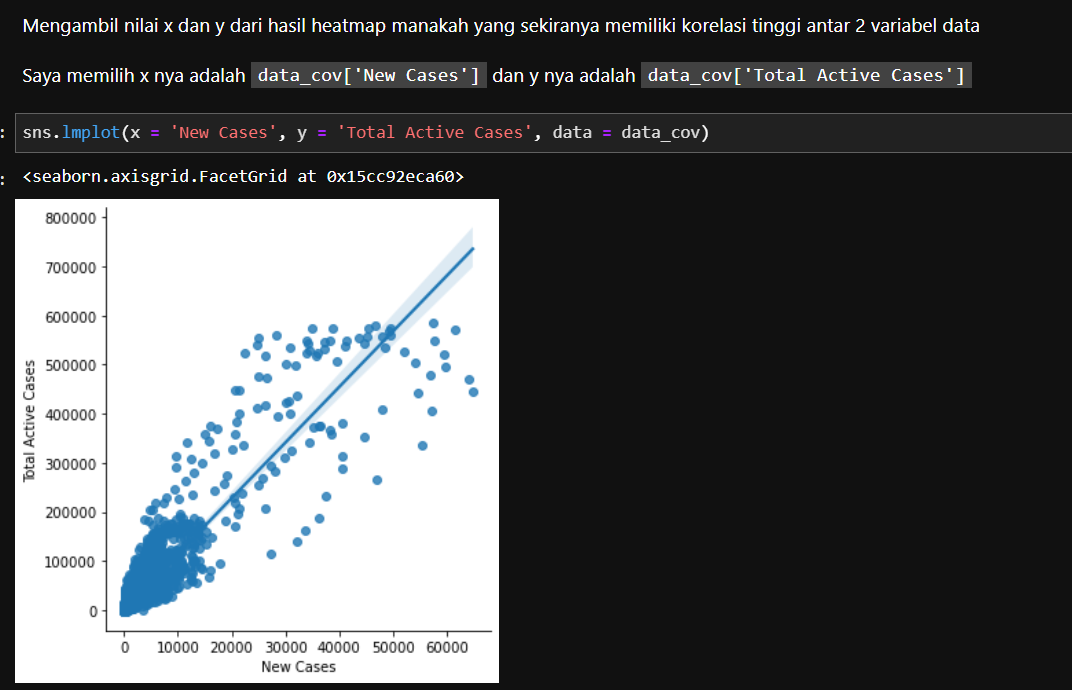
Untuk menguji apakah function yang kita buat sudah benar atau belum, kita coba uji dengan function bawaan dari Python



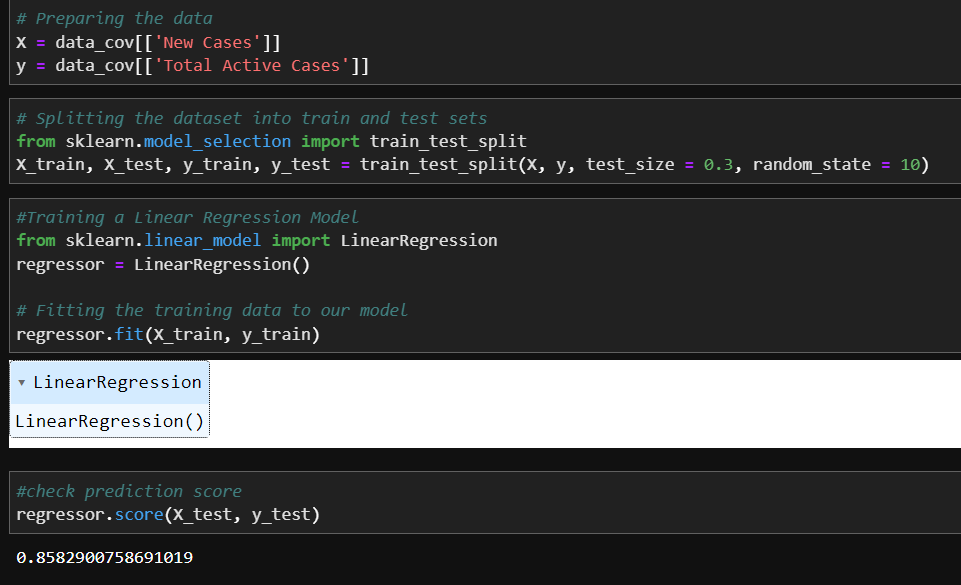
Dan ternyata hasilnya sama. Artinya function yang kita buat benar. Dan hasil koefisien korelasi nya adalah 0.8569… atau dapat dibulatkan ke 0.857

d. Regresi dan evaluasi

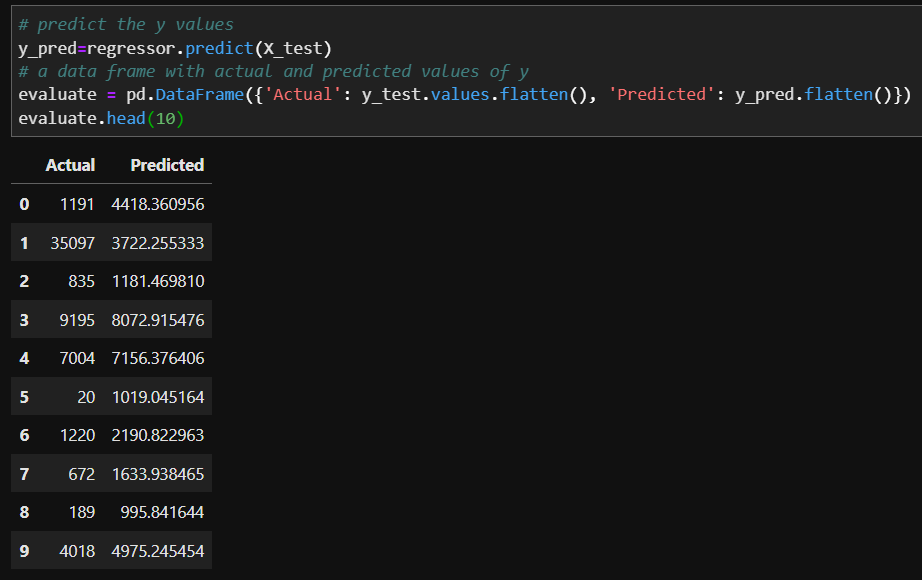
- Cari model regresi yang tepat untuk menyelesaikan data harian COVID-19 di Indonesia



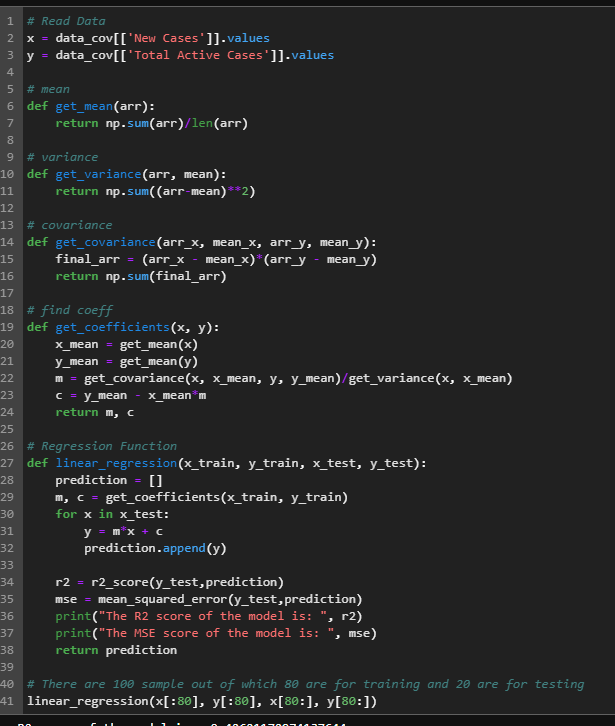
Setelah kita mendapatkan data yang ingin diuji dalam regresi nya, dilakukan split data untuk train dan test data dengan proporsi 70 dan 30. Setelah itu kita lakukan uji dengan Linear Regression. Setelah itu dihitung nilai regresi nya dari data test.



Setelah itu kita simpan nilai y\_prediksi dan y\_test ke dalam tabel untuk kita lihat perbedaan nya



- Setelah kita mendapatkan percobaan menghitung dengan model Linear Regresion bawaan Library python, selanjutnya kita coba membuat dengan fungsi scratch



Namun di sini saya belum bisa mendapatkan hasil prediksi model yang akurat sesuai dengan hasil perhitungan menggunakan library scikit learn



- Evaluasi model regresi menggunakan MAPE menggunakan Python secara scratch

Buat sebuah function untuk menghitung model regresi dengan MAPE nya terlebih dahulu dengan rumus

