

LAPORAN TUGAS BESAR 2
KU1102 PENGENALAN KOMPUTASI
SEMESTER I 2023/2024

ANALISIS DAN VISUALISASI DATA

“Pendekatan Informatif Menggunakan Python untuk Analisis dan Visualisasi Data COVID-19 Global”



Disusun oleh:

Kelompok 1 K-20

William Anthony	(16523109)
Muhammad Jafar Fadli	(16523137)
Zulfaqqar Nayaka Athadiansyah	(19623116)
Abdullah Farhan	(19623305)

PROGRAM TAHAP PERSIAPAN BERSAMA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
NOVEMBER 2023

PRAKATA

Dengan memanjatkan puja dan puji syukur ke hadirat Tuhan Yang Maha Esa, kami dapat menyelesaikan laporan ini. Laporan ini adalah hasil dari tugas besar 2 KU1102 yang kami susun untuk memenuhi kewajiban akademik dalam mata kuliah KU1102 Pengenalan Komputasi pada semester I 2023/2024.

Pelaksanaan penulisan laporan ini tidak lepas dari bantuan dan arahan dari berbagai pihak. Oleh karena itu, kami mengucapkan terima kasih kepada semua pihak yang telah mendukung penyusunan laporan ini, terutama dosen pengampu mata kuliah KU1102 Pengenalan Komputasi untuk kelas 20: Ibu Dr. Fazat Nur Azizah, S.T., M.Sc. dan Bapak Yuda Sukmana, S.Pd., M.T.

Akhir kata, kami berharap laporan ini dapat menjadi penunjang hasil IP KU1102 kami dan bermanfaat bagi pembaca serta dapat menjadi bahan pelajaran untuk kami selaku mahasiswa Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung.

Dengan penuh kesadaran, kami telah berupaya dengan sepenuh hati untuk mencapai kesempurnaan dalam penyusunan laporan ini. Namun, kami menyadari bahwa masih ada kekurangan dalam penulisan laporan ini. Oleh karena itu, kami sangat menghargai dan berharap untuk menerima saran serta kritik yang konstruktif untuk meningkatkan kualitas penulisan laporan kami di masa mendatang.

DAFTAR ISI

1 Pendahuluan	5
1.1 Latar Belakang	5
1.2 Tujuan	5
1.3 Deskripsi Data dan File	5
2 Analisis Data	5
2.1 Karakteristik Data	6
2.2 Statistik.....	6
3 Visualisasi Data	6
3.1 Perbandingan Kategori	6
3.2 Perubahan terhadap Waktu.....	8
3.3 Hierarki dan Hubungan Keseluruhan-Bagian.....	9
3.4 <i>Plotting Relationships</i>	9
4 <i>Data Cleansing</i>	11
5 Kesimpulan dan Saran	11
6 Lampiran.....	13

Daftar gambar/tabel nanti di sini

1 PENDAHULUAN

1.1 Latar Belakang

Pandemi COVID-19 telah menjadi salah satu tantangan terbesar bagi kesehatan dan kesejahteraan manusia di abad ke-21. Pandemi ini diakibatkan virus SARS-CoV-2 yang muncul pada akhir tahun 2019 di Wuhan, Republik Rakyat Tiongkok. Sejak saat itu, COVID-19 telah menyebar ke lebih dari 200 negara dan wilayah, menginfeksi lebih dari 300 juta orang, menyebabkan lebih dari 5 juta kematian, serta meninggalkan dampak besar terhadap perekonomian dunia. Upaya global untuk mengendalikan dan mengatasi pandemi ini membutuhkan pemahaman yang mendalam dan terkini tentang situasi epidemiologis, dampak sosial-ekonomi, dan respons kebijakan dari berbagai pihak terkait.

Salah satu sumber data yang dapat digunakan untuk mempelajari pandemi COVID-19 adalah Data Science Dojo, sebuah organisasi yang menyediakan data dan kode terbuka tentang berbagai topik global, salah satunya adalah data kasus COVID-19. Data tersebut dapat diakses di [sini](#)¹. Data ini mencakup berbagai indikator, seperti kasus, kematian, pembatasan perjalanan, sumber infeksi, dan masih banyak lagi.

Kami menggunakan Python untuk melakukan analisis pada data kasus COVID-19 yang kami dapatkan. Python adalah salah satu bahasa pemrograman yang paling populer dan serbaguna, yang dapat digunakan untuk berbagai tujuan, termasuk analisis data dan visualisasi. Python memiliki banyak *library* dan modul yang dapat membantu pengguna untuk mengolah, memanipulasi, dan menampilkan data dengan mudah dan efisien. Beberapa *library* dan modul Python yang berguna untuk analisis dan visualisasi data COVID-19 adalah *pandas*, *numpy*, *matplotlib*, dan *seaborn*.

1.2 Tujuan

1.2.1 Melakukan analisis pada data kasus COVID-19 menggunakan Python dengan *library* *pandas*.

1.2.2 Menjelaskan *insight* yang didapat dan keputusan yang bisa dibuat berdasarkan analisis data.

1.2.3 Membuat visualisasi data kasus COVID-19 menggunakan Python dengan *library* *pandas* dan *matplotlib*.

1.3 Deskripsi Data dan File

Data set kasus COVID-19 yang kami gunakan berasal dari Data Science Dojo (lihat catatan kaki 1) yang dikompilasikan dari data set milik Kaggle dan John Hopkins University. Kedua sumber tersebut menggabungkan laporan harian World Health Organization (WHO), European Centre for Disease Prevention and Control (ECDC), US Department of State Travel Advisories, dan banyak sumber lainnya. Data ini kami simpan dalam ekstensi *.csv* dengan nama file "data_corona.csv". Data ini tersusun atas 167 entri dan memuat sepuluh kolom, di antaranya *Country*, *Cases*, *New_cases*, *Deaths*, *New_deaths*, *Transmission*, *Days_last_reported*, *Level*, *Date_last_reported*, dan *Travel_restriction*.

Dalam mengolah data, kami menggunakan *library* *pandas*, *numpy*, *matplotlib*, dan *seaborn*. Data dan *library* yang digunakan di-*import* dengan *command*:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv("data_corona.csv")
```

¹

https://code.datasciencedojo.com/datasciencedojo/datasets/blob/master/Coronavirus/coronavirus_merged_sources_Apr22020.csv

Keterangan kolom data:

- Country : negara
- Cases : banyak kasus (kumulatif)
- New_cases : kasus baru
- Deaths: angka kematian (kumulatif)
- New_deaths : angka kematian baru
- Transmission: Penyebaran (dari domestik atau eksternal)
- Days_last_reported: hari terakhir update data dilaporkan
- Level : tingkat PPKM
- Date_last_reported : tanggal terakhir update data dilaporkan
- Travel_restriction : ada atau tidaknya pembatasan data.

2 ANALISIS DATA

2.1 Karakteristik Data

Blablabla

```
for col in df.columns:
    if df[col].dtype == 'object':
        print(f"{col} adalah atribut kategorikal dengan {len(df[col].unique())}
nilai unik, yaitu: {df[col].unique()}")
    else:
        print(f"{col} adalah atribut kuantitatif dengan tipe data
{df[col].dtype}, memiliki range nilai dari {df[col].min()} sampai
{df[col].max()}")
        print(f"{col} tidak memiliki data yang hilang, yaitu:
{df[col].isna().sum()}")
        print()
```

2.2 Statistik

Blablabla

3 VISUALISASI DATA

3.1 Perbandingan Kategori

Salah satu data yang dapat dibandingkan secara kategorikal adalah data kasus, angka kematian, kasus baru, dan angka kematian baru.

```
num_chunks = 6
df_chunks = np.array_split(df, num_chunks)

for i, df_chunk in enumerate(df_chunks):
    plt.figure(figsize=(10, 10))
```

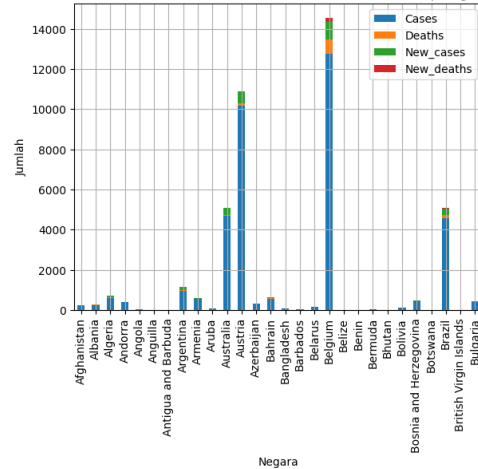
```
df_chunk.groupby('Country')[['Cases', 'Deaths', 'New_cases',
'New_deaths']].sum().plot(kind='bar', stacked=True, xlabel='Negara',
ylabel='Jumlah', title=f'Jumlah kasus, kematian, kasus baru, dan kematian
baru untuk setiap negara (Bagian { i+1}/{ num_chunks})', grid= True, legend=
True, use_index= True)

plt.subplots_adjust(left=0.1, right=0.9, bottom=0.1, top=0.9, wspace=0.2,
hspace=0.2)

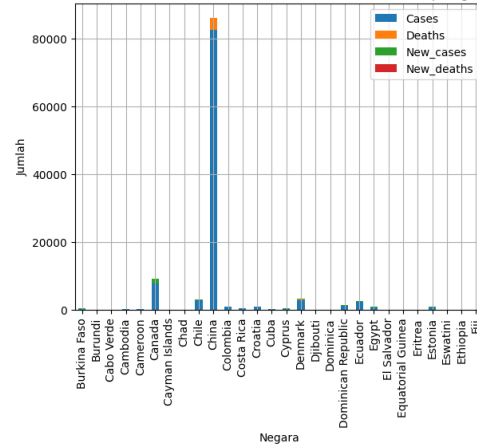
plt.show()
```

Blablabla

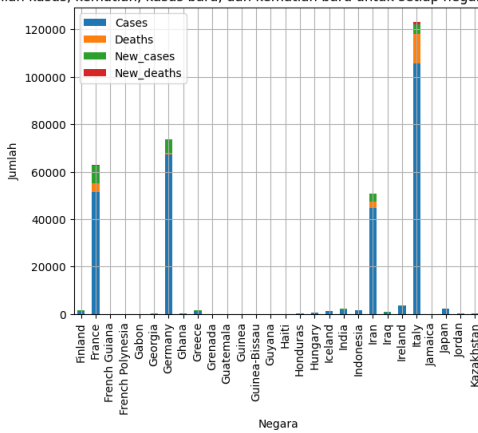
Jumlah kasus, kematian, kasus baru, dan kematian baru untuk setiap negara (Bagian 1/6)



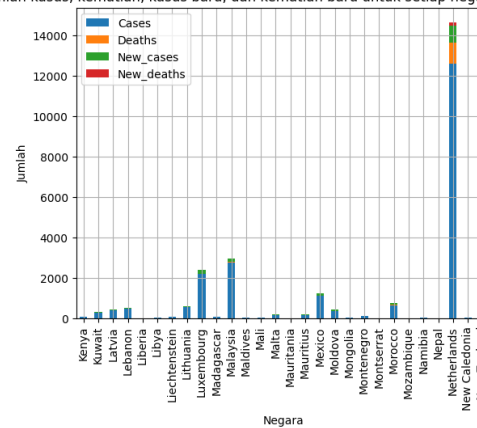
Jumlah kasus, kematian, kasus baru, dan kematian baru untuk setiap negara (Bagian 2/6)



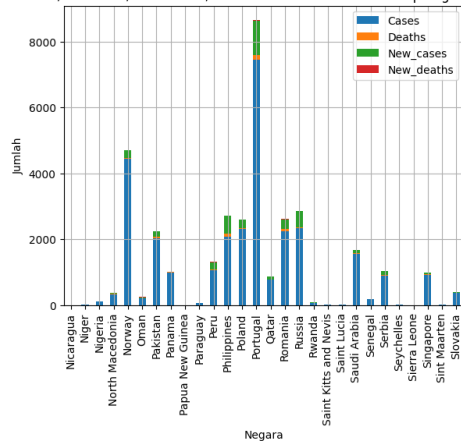
Jumlah kasus, kematian, kasus baru, dan kematian baru untuk setiap negara (Bagian 3/6)



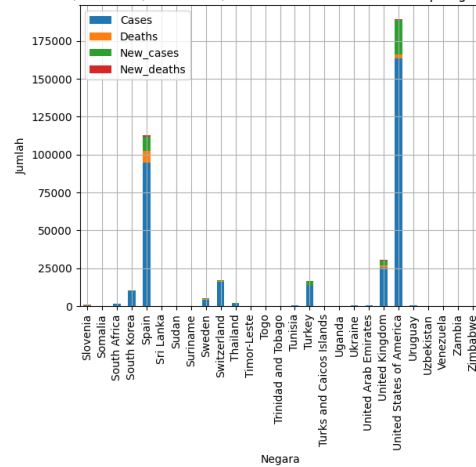
Jumlah kasus, kematian, kasus baru, dan kematian baru untuk setiap negara (Bagian 4/6)



Jumlah kasus, kematian, kasus baru, dan kematian baru untuk setiap negara (Bagian 5/6)



Jumlah kasus, kematian, kasus baru, dan kematian baru untuk setiap negara (Bagian 6/6)



Blablabla

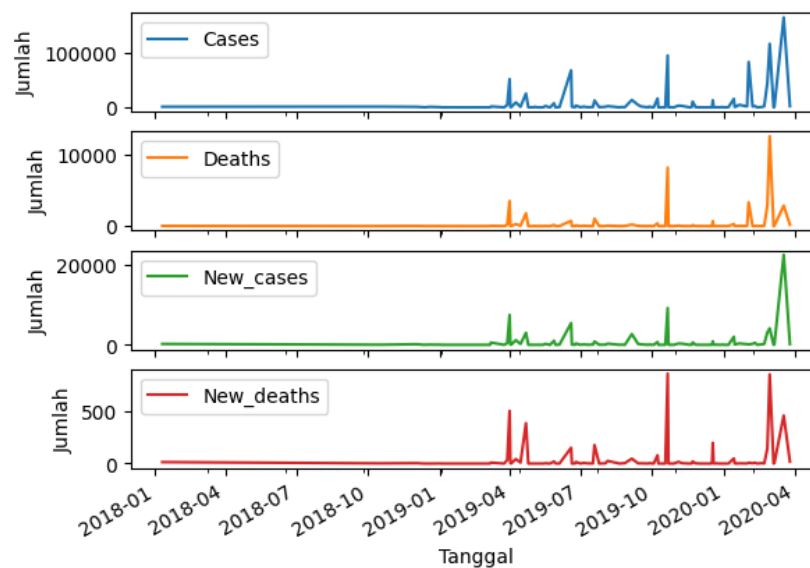
3.2 Perubahan terhadap Waktu

Pada kode ini, kita memvisualisasi kasus dan kematian yang disebabkan oleh COVID-19 di Dunia berdasarkan waktu menggunakan grafik garis.

```
plt.figure(figsize=(10, 10))
df.groupby('Date_last_updated')[['Cases', 'Deaths', 'New_cases', 'New_deaths']].
sum().plot(kind='line', xlabel='Tanggal', ylabel='Jumlah', title='Kasus dan
Kematian COVID-19 di Dunia berdasarkan waktu', legend=True, subplots=True)
plt.show()
```

Dapat diamati, tidak ada perkembangan kasus hingga pada 2019-04 dimana mulai muncul kasus.

Kasus dan Kematian COVID-19 di Dunia berdasarkan waktu



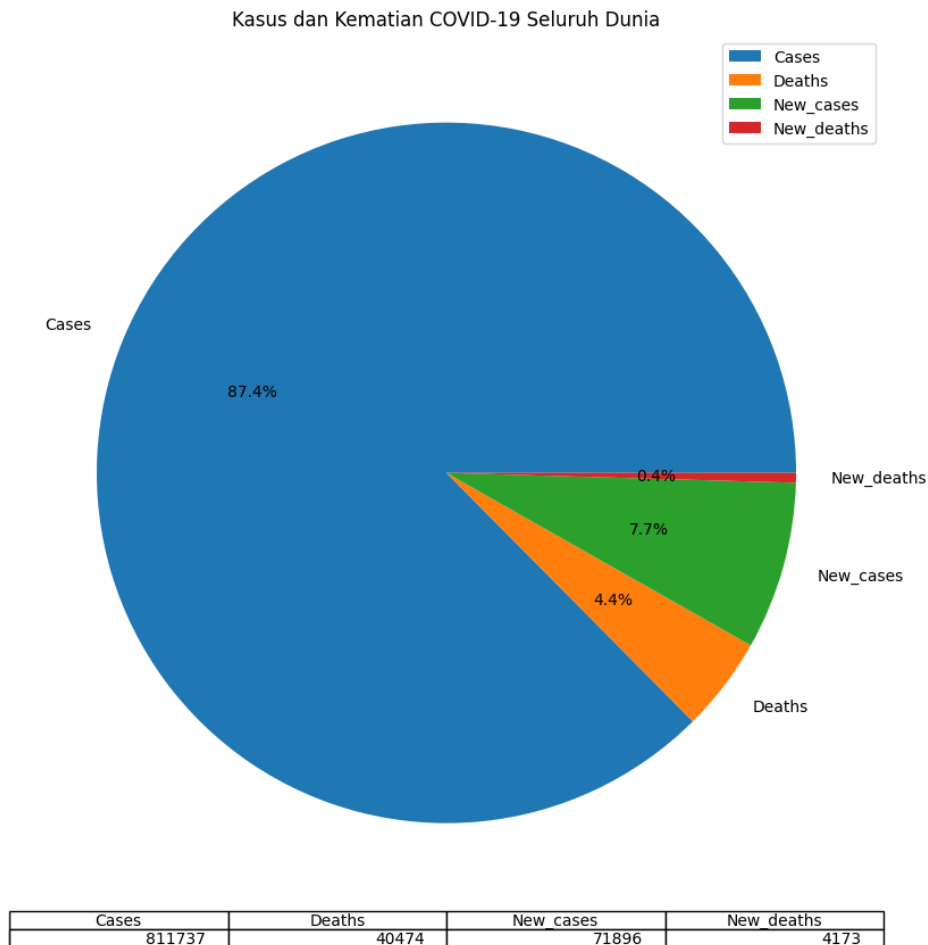
Hal ini disebabkan mulainya didata

3.3 Hierarki dan Hubungan Keseluruhan-Bagian

Pada kode ini, kita memvisualisasi kasus dan kematian yang disebabkan oleh COVID-19 di Dunia berdasarkan negara dengan menggunakan *pie chart*.

```
plt.figure(figsize=(10, 10))
df[['Cases', 'Deaths', 'New_cases',
    'New_deaths']].sum().plot(kind='pie', autopct='%1.1f%%', title='Kasus dan Kematian
    COVID-19 Seluruh Dunia', legend=True, table=True)
plt.show()
```

Dapat diamati, bahwa dengan ini terdapat



adad

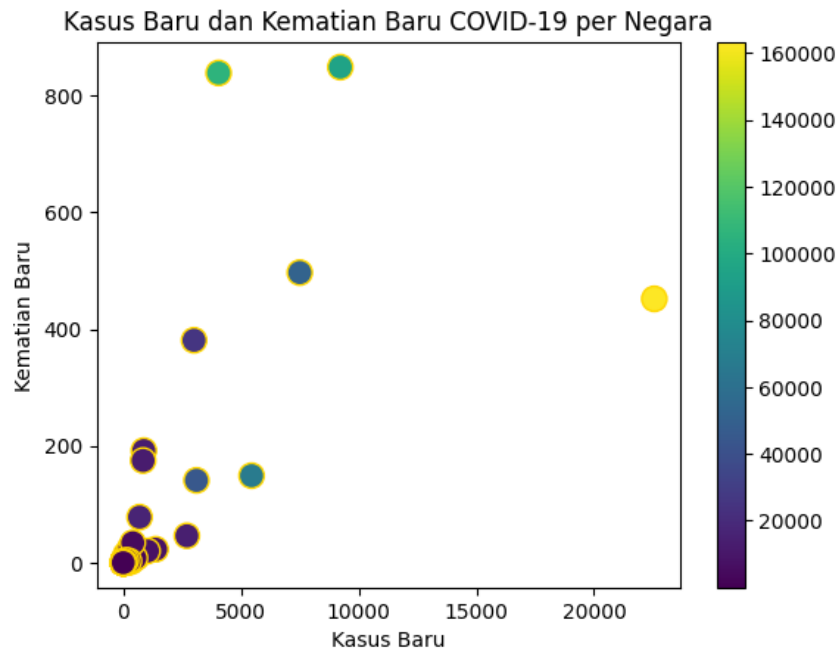
3.4 Plotting Relationships

Pada kode ini, kita memvisualisasi hubungan angka kematian baru terhadap kasus baru yang disebabkan oleh COVID-19 di dunia dengan menggunakan *Plotting*

```
plt.figure(figsize=(20, 20))
df.groupby('Country')[['New_cases', 'New_deaths']].mean().plot(kind='scatter',
    x='New_cases', y='New_deaths', c=df['Cases'], cmap='viridis',
    edgecolor="gold", s=150, xlabel='Kasus Baru', ylabel='Kematian Baru', title='Kasus
    Baru dan Kematian Baru COVID-19 Seluruh Dunia')
```

```
plt.show()
```

Dapat diamati, hubungan keduanya cenderung berbanding lurus. Artinya, semakin banyak kasus baru maka semakin tinggi angka kematian baru



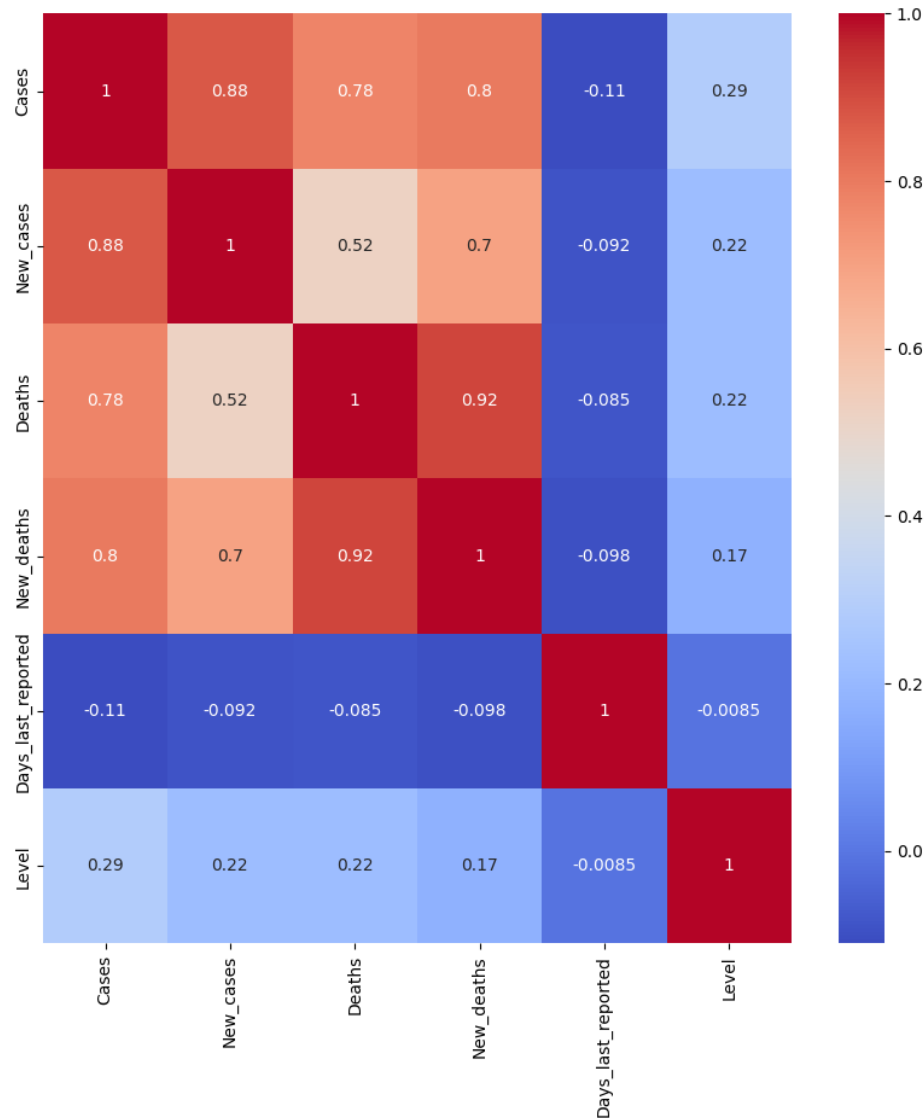
Blablabla

3.5 Korelasi Data

Pada kode ini, kita memvisualisasi nilai korelasi antar data dengan warna merah menunjukkan angka korelasi yang lebih tinggi

```
df.corr(numeric_only = True)
plt.figure(figsize=(10, 10))
sns.heatmap(df.corr(numeric_only = True), annot=True, cmap='coolwarm')
plt.show()
```

Dapat diamati pada diagram korelasi data. Banyak kasus cenderung berkorelasi lurus dengan banyak kasus baru, angka kematian, dan angka kematian baru. Banyak kasus baru cenderung berkorelasi lurus dengan angka kematian baru. Angka kematian berkorelasi dengan angka kematian baru. Sementara itu, level dan tanggal terakhir update data dilaporkan cenderung tidak berkorelasi dengan data lainnya.



Blablabla

4 DATA CLEANSING

- Ubah formatting

5 KESIMPULAN DAN SARAN

Penggunaan `Python`, `Matplotlib`, `Seaborn` dalam analisis dan visualisasi data COVID-19 membantu kita dalam membuat keputusan yang didasarkan pada bukti dan data, bukan asumsi atau spekulasi. Ini sangat penting dalam situasi seperti pandemi, di mana keputusan yang kita buat dapat memiliki dampak yang signifikan pada kesehatan dan kesejahteraan masyarakat. hal ini diterapkan dengan meneliti tren dan menentukan pola dalam data. Visualisasi data memungkinkan kita untuk melihat gambaran besar dan detail sekaligus, serta mengidentifikasi area yang memerlukan perhatian lebih lanjut. Analisis data memberikan kita pemahaman yang lebih mendalam tentang situasi saat ini dan bagaimana hal itu berkembang seiring waktu.

Dengan wawasan ini, kita dapat membuat keputusan yang lebih baik dan lebih tepat waktu untuk merespons pandemi. Misalnya, kita dapat mengidentifikasi daerah yang paling terkena dampak, mengetahui efektivitas langkah-langkah yang telah diambil, dan merencanakan strategi untuk masa depan berdasarkan tren yang diamati.

6 LAMPIRAN

Tabel 1. Pembagian tugas anggota kelompok

<i>Anggota Kelompok</i>	<i>Kontribusi</i>
William Anthony (16523109)	Pembuat visualisasi grafik menggunakan pandas dan seaborn serta menyiapkan Ide Video
Muhammad Jafar Fadli (16523137)	Perencanaan & penulisan laporan ; improvisasi visualisasi data ; menyiapkan Ide Video
Zulfaqqar Nayaka Athadiansyah (19623116)	Perencanaan & penulisan laporan; <i>proof-reading</i> dan improvisasi kode
Abdullah Farhan (19623305)	Pencari Data & Backend engineer (Pembuat Visualisasi Data)