Midterm Project Visualisasi Data

Sanabila Khoirunnisa 1301204097

Import necessary library

In [660…
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Read the data (Dataset1)

In [661…
```
df_can1 = pd.read_csv('Dataset1.csv')
```

In [662…
```
df_can1
```

Out[662]:

	Entity	Code	Year	Number of executions (Amnesty International)	Deaths - Meningitis - Sex: Both - Age: All Ages (Number)	Deaths - Alzheimer's disease and other dementias - Sex: Both - Age: All Ages (Number)	Deaths - Parkinson's disease - Sex: Both - Age: All Ages (Number)	Deaths - Nutritional deficiencies - Sex: Both - Age: All Ages (Number)	S	(
0	Afghanistan	AFG	2007	15	2933.0	1402.0	450.0	2488.0		
1	Afghanistan	AFG	2008	17	2731.0	1424.0	455.0	2277.0		
2	Afghanistan	AFG	2009	0	2460.0	1449.0	460.0	2040.0		
3	Afghanistan	AFG	2011	2	2327.0	1508.0	473.0	1846.0		
4	Afghanistan	AFG	2012	14	2254.0	1544.0	482.0	1705.0		
...	...	...	...	...	...	...	...	...		
6888	Zimbabwe	ZWE	2015	NaN	1439.0	754.0	215.0	3019.0		
6889	Zimbabwe	ZWE	2016	NaN	1457.0	767.0	219.0	3056.0		
6890	Zimbabwe	ZWE	2017	NaN	1460.0	781.0	223.0	2990.0		
6891	Zimbabwe	ZWE	2018	NaN	1450.0	795.0	227.0	2918.0		
6892	Zimbabwe	ZWE	2019	NaN	1450.0	812.0	232.0	2884.0		

6893 rows × 36 columns

In [663…
```
df_can1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6893 entries, 0 to 6892
Data columns (total 36 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   Entity
6893 non-null   object
 1   Code
6189 non-null   object
 2   Year
6893 non-null   int64
 3   Number of executions (Amnesty International)
267 non-null    object
 4   Deaths - Meningitis - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 5   Deaths - Alzheimer's disease and other dementias - Sex: Both - Age: All Ages
(Number)     6840 non-null   float64
 6   Deaths - Parkinson's disease - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 7   Deaths - Nutritional deficiencies - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 8   Deaths - Malaria - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 9   Deaths - Drowning - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 10  Deaths - Interpersonal violence - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 11  Deaths - Maternal disorders - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 12  Deaths - HIV/AIDS - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 13  Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 14  Deaths - Tuberculosis - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 15  Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 16  Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 17  Deaths - Neonatal disorders - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 18  Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 19  Deaths - Self-harm - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 20  Deaths - Exposure to forces of nature - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 21  Deaths - Diarrheal diseases - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 22  Deaths - Environmental heat and cold exposure - Sex: Both - Age: All Ages (Nu
mber)        6840 non-null   float64
 23  Deaths - Neoplasms - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 24  Deaths - Conflict and terrorism - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 25  Deaths - Diabetes mellitus - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 26  Deaths - Chronic kidney disease - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 27  Deaths - Poisonings - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 28  Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number)
```

```
6840 non-null   float64
 29  Terrorism (deaths)
2532 non-null   float64
 30  Deaths - Road injuries - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 31  Deaths - Chronic respiratory diseases - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 32  Deaths - Cirrhosis and other chronic liver diseases - Sex: Both - Age: All Ag
es (Number)  6840 non-null   float64
 33  Deaths - Digestive diseases - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 34  Deaths - Fire, heat, and hot substances - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
 35  Deaths - Acute hepatitis - Sex: Both - Age: All Ages (Number)
6840 non-null   float64
dtypes: float64(32), int64(1), object(3)
memory usage: 1.9+ MB
```

In [664…

```
df_can1_indo = df_can1[df_can1['Entity']=='Indonesia']
df_can1_indo
```

Out[664]:

	Entity	Code	Year	Number of executions (Amnesty International)	Deaths - Meningitis - Sex: Both - Age: All Ages (Number)	Deaths - Alzheimer's disease and other dementias - Sex: Both - Age: All Ages (Number)	Deaths - Parkinson's disease - Sex: Both - Age: All Ages (Number)	Deaths - Nutritional deficiencies - Sex: Both - Age: All Ages (Number)	D M Se - A (Nu
2684	Indonesia	IDN	2007	1	10001.0	17364.0	5285.0	17906.0	
2685	Indonesia	IDN	2008	10	9608.0	17867.0	5476.0	18472.0	
2686	Indonesia	IDN	2009	0	9183.0	18357.0	5649.0	18940.0	
2687	Indonesia	IDN	2013	5	7915.0	20365.0	6392.0	20669.0	
2688	Indonesia	IDN	2015	14	6946.0	21485.0	6737.0	21526.0	
2689	Indonesia	IDN	2016	4	6469.0	22082.0	6970.0	21104.0	
2690	Indonesia	IDN	1990	NaN	19261.0	9425.0	2481.0	25621.0	
2691	Indonesia	IDN	1991	NaN	18316.0	9809.0	2585.0	24409.0	
2692	Indonesia	IDN	1992	NaN	17432.0	10237.0	2715.0	23425.0	
2693	Indonesia	IDN	1993	NaN	16722.0	10664.0	2854.0	22602.0	
2694	Indonesia	IDN	1994	NaN	16025.0	11101.0	2993.0	21797.0	
2695	Indonesia	IDN	1995	NaN	15296.0	11539.0	3138.0	21032.0	
2696	Indonesia	IDN	1996	NaN	14587.0	11965.0	3289.0	20200.0	
2697	Indonesia	IDN	1997	NaN	14036.0	12438.0	3467.0	20078.0	
2698	Indonesia	IDN	1998	NaN	13744.0	12868.0	3635.0	18823.0	
2699	Indonesia	IDN	1999	NaN	13254.0	13307.0	3801.0	18238.0	
2700	Indonesia	IDN	2000	NaN	12754.0	13792.0	3970.0	17982.0	
2701	Indonesia	IDN	2001	NaN	12514.0	14247.0	4117.0	18035.0	
2702	Indonesia	IDN	2002	NaN	12168.0	14788.0	4291.0	18120.0	
2703	Indonesia	IDN	2003	NaN	11761.0	15328.0	4473.0	18171.0	
2704	Indonesia	IDN	2004	NaN	11292.0	15853.0	4669.0	18205.0	
2705	Indonesia	IDN	2005	NaN	10825.0	16343.0	4877.0	18167.0	
2706	Indonesia	IDN	2006	NaN	10394.0	16797.0	5076.0	17945.0	
2707	Indonesia	IDN	2010	NaN	8637.0	18800.0	5819.0	19607.0	
2708	Indonesia	IDN	2011	NaN	8169.0	19291.0	6032.0	20020.0	
2709	Indonesia	IDN	2012	NaN	8085.0	19831.0	6220.0	20320.0	
2710	Indonesia	IDN	2014	NaN	7184.0	20922.0	6564.0	21152.0	
2711	Indonesia	IDN	2017	NaN	5401.0	22729.0	7168.0	20946.0	
2712	Indonesia	IDN	2018	NaN	5030.0	23591.0	7393.0	20607.0	
2713	Indonesia	IDN	2019	NaN	4715.0	24381.0	7616.0	20348.0	

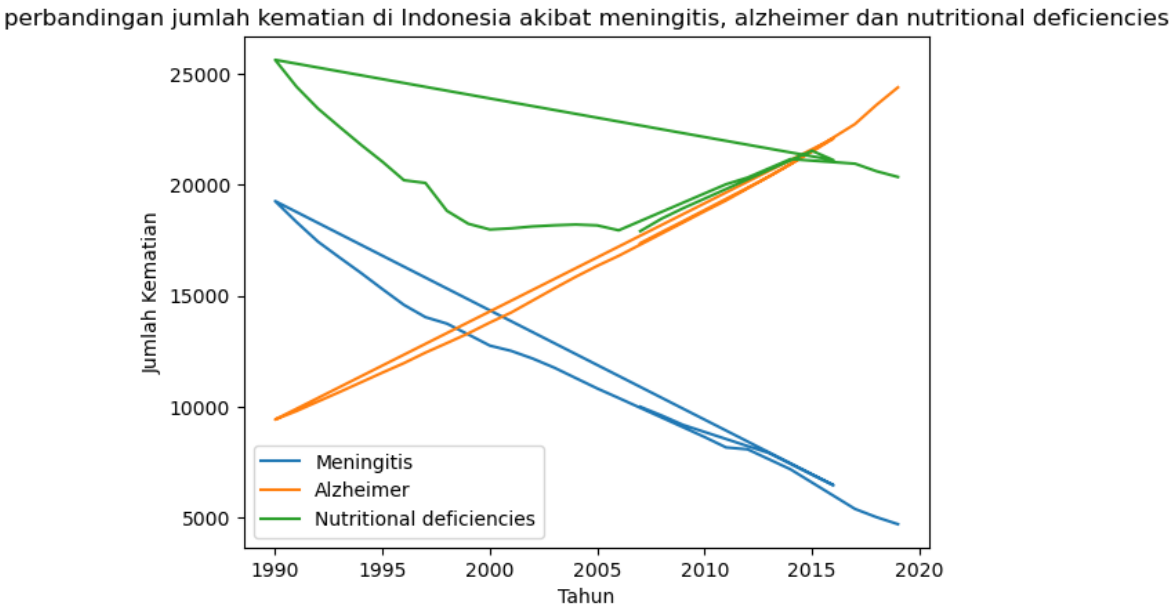
30 rows x 36 columns

specific disease data

In [665...
```
df_can1_meningitis = df_can1_indo[df_can1_indo["Deaths - Meningitis - Sex: Both - A
df_can1_alzheimer = df_can1_indo[df_can1_indo["Deaths - Alzheimer's disease and otl
df_can1_nutrition = df_can1_indo[df_can1_indo["Deaths - Nutritional deficiencies -
```

Nomor 1

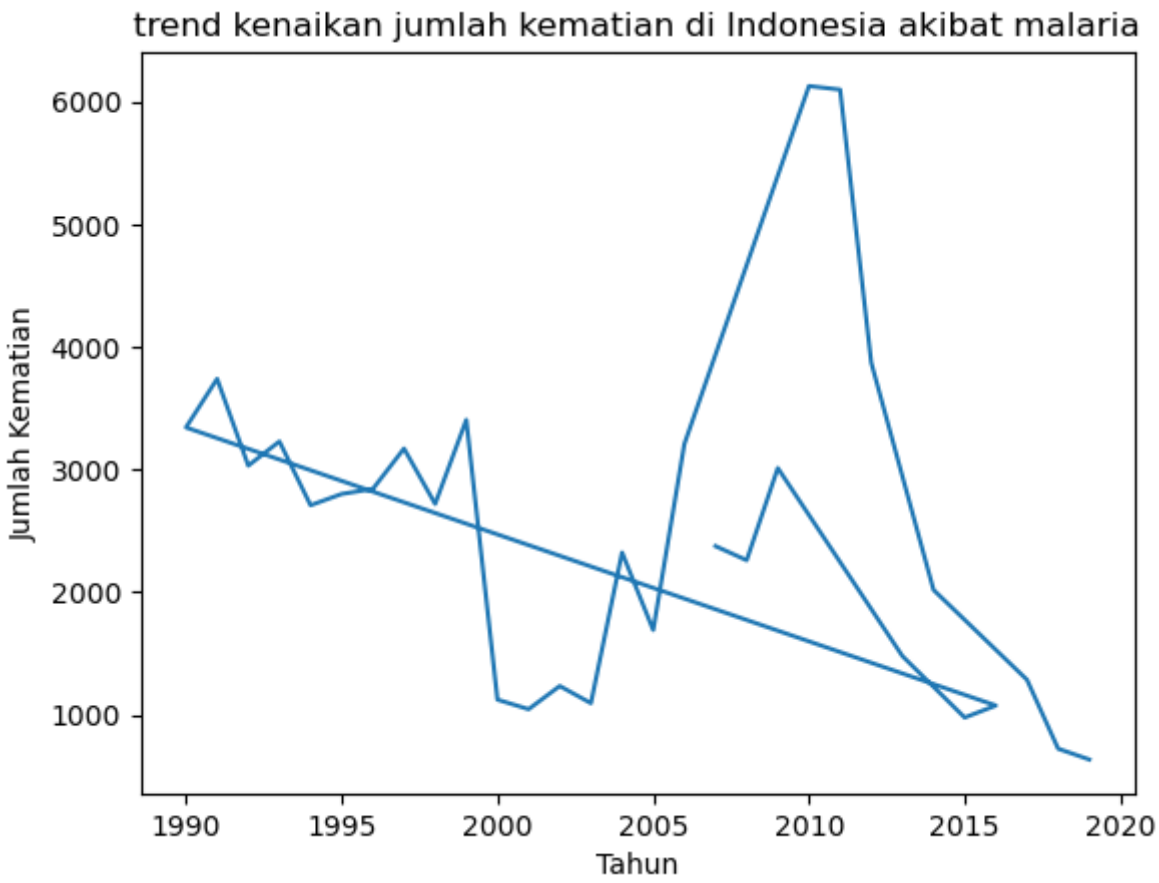
In [666...
```
plt.plot(df_can1_meningitis["Year"], df_can1_meningitis["Deaths - Meningitis - Sex
plt.plot(df_can1_alzheimer["Year"], df_can1_alzheimer["Deaths - Alzheimer's disease
plt.plot(df_can1_nutrition["Year"], df_can1_nutrition["Deaths - Nutritional defici

plt.title('perbandingan jumlah kematian di Indonesia akibat meningitis, alzheimer d
plt.xlabel('Tahun')
plt.ylabel('Jumlah Kematian')
plt.legend()
plt.show()
```



In [667...
```
df_can1_malaria_indo = df_can1[(df_can1['Entity']=='Indonesia') & (df_can1['Deaths
```

Nomor 2

In [668...
```
plt.plot(df_can1_malaria_indo['Year'], df_can1_malaria_indo['Deaths - Malaria - Sex

plt.title('trend kenaikan jumlah kematian di Indonesia akibat malaria')
plt.xlabel('Tahun')
plt.ylabel('Jumlah Kematian')
plt.show()
```



nutritional deficiency data at Indonesia and India

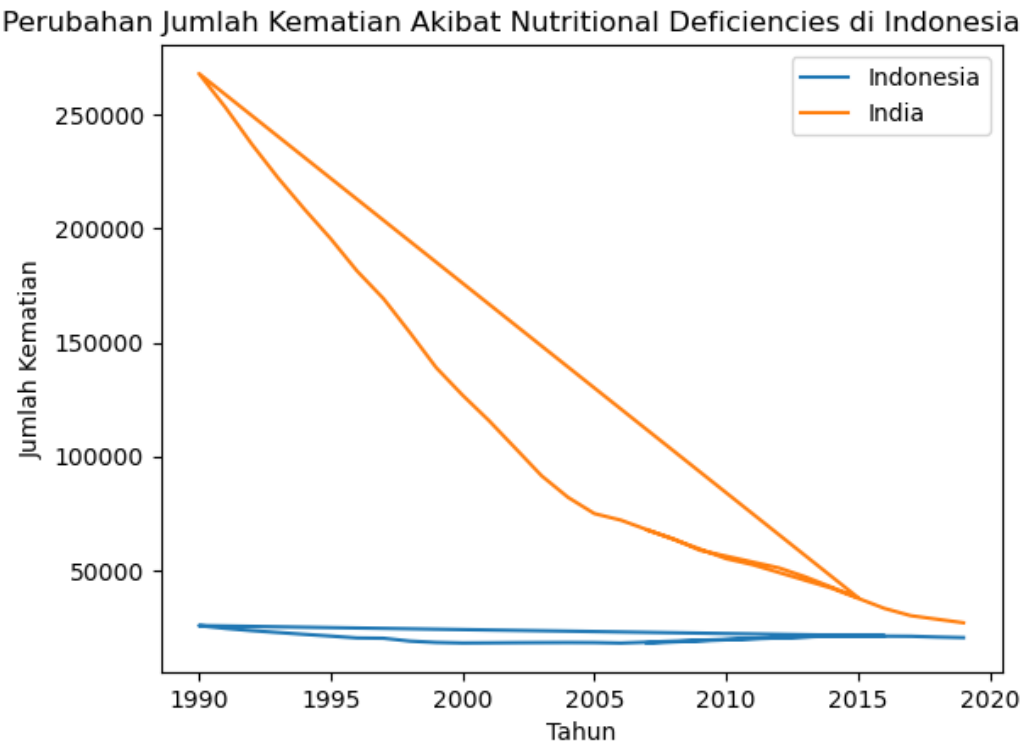
In [669…

```
df_can1_nutrition_indo = df_can1[(df_can1['Entity'] == 'Indonesia') & (df_can1['De

df_can1_nutrition_india = df_can1[(df_can1['Entity'] == 'India') & (df_can1['Death
```

Nomor 2

In [670…

```
plt.plot(df_can1_nutrition_indo['Year'], df_can1_nutrition_indo['Deaths - Nutritio
plt.plot(df_can1_nutrition_india['Year'], df_can1_nutrition_india['Deaths - Nutrit

plt.title('Perubahan Jumlah Kematian Akibat Nutritional Deficiencies di Indonesia
plt.xlabel('Tahun')
plt.ylabel('Jumlah Kematian')
plt.legend()

# Menampilkan grafik
plt.show()
```

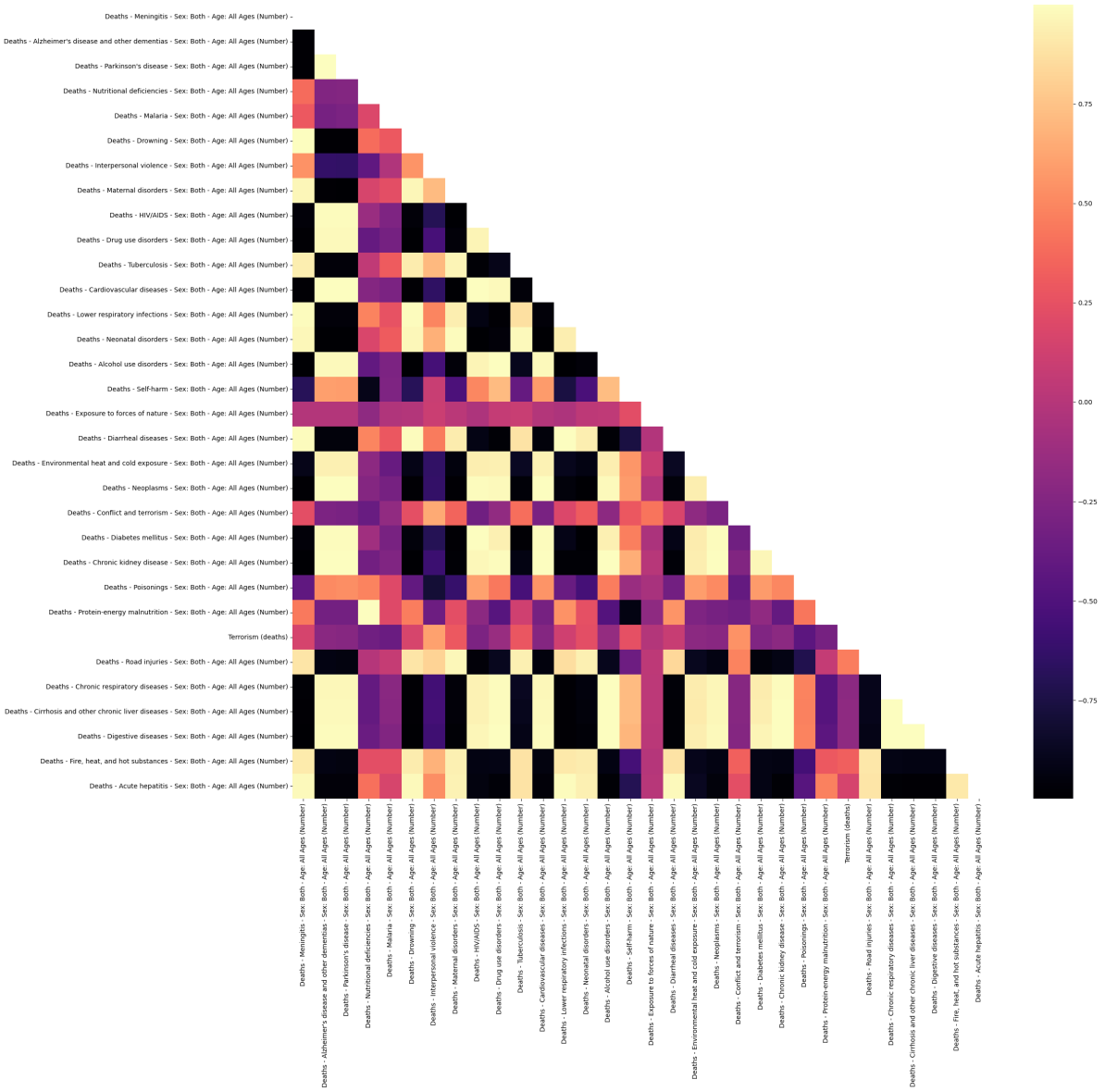


Nomor 3

In [671…

```
df_can1_indo2 = df_can1[df_can1['Entity']=='Indonesia']
df_can1_indo2=df_can1_indo2.drop(['Number of executions (Amnesty International)','

df_can1_indo2.rename(columns = {'Deaths - Meningitis - Sex: Both - Age: All Ages (
df_can1_indo2.rename(columns = {"Deaths - Alzheimer's disease and other dementias
df_can1_indo2.rename(columns = {"Deaths - Parkinson's disease - Sex: Both - Age: A
df_can1_indo2.rename(columns = {'Deaths - Nutritional deficiencies - Sex: Both - Ag
df_can1_indo2.rename(columns = {'Deaths - Malaria - Sex: Both - Age: All Ages (Numb
df_can1_indo2.rename(columns = {'Deaths - Drowning - Sex: Both - Age: All Ages (Nur
df_can1_indo2.rename(columns = {'Deaths - Interpersonal violence - Sex: Both - Age
df_can1_indo2.rename(columns = {'Deaths - Maternal disorders - Sex: Both - Age: Al
df_can1_indo2.rename(columns = {'Deaths - HIV/AIDS - Sex: Both - Age: All Ages (Nur
df_can1_indo2.rename(columns = {'Deaths - Drug use disorders - Sex: Both - Age: Al
df_can1_indo2.rename(columns = {'Deaths - HIV/AIDS - Sex: Both - Age: All Ages (Nur
df_can1_indo2.rename(columns = {'Deaths - Tuberculosis - Sex: Both - Age: All Ages
df_can1_indo2.rename(columns = {'Deaths - Cardiovascular diseases - Sex: Both - Ag
df_can1_indo2.rename(columns = {'Deaths - Lower respiratory infections - Sex: Both
df_can1_indo2.rename(columns = {'Deaths - Neonatal disorders - Sex: Both - Age: Al
df_can1_indo2.rename(columns = {'Deaths - Alcohol use disorders - Sex: Both - Age:
df_can1_indo2.rename(columns = {'Deaths - Self-harm - Sex: Both - Age: All Ages (Nu
df_can1_indo2.rename(columns = {'Deaths - Exposure to forces of nature - Sex: Both
df_can1_indo2.rename(columns = {'Deaths - Diarrheal diseases - Sex: Both - Age: Al
df_can1_indo2.rename(columns = {'Deaths - Environmental heat and cold exposure - Se
df_can1_indo2.rename(columns = {'Deaths - Neoplasms - Sex: Both - Age: All Ages (Nu
df_can1_indo2.rename(columns = {'Deaths - Conflict and terrorism - Sex: Both - Age
df_can1_indo2.rename(columns = {'Deaths - Diabetes mellitus - Sex: Both - Age: All
df_can1_indo2.rename(columns = {'Deaths - Chronic kidney disease - Sex: Both - Age
df_can1_indo2.rename(columns = {'Deaths - Poisonings - Sex: Both - Age: All Ages (M
df_can1_indo2.rename(columns = {'Deaths - Protein-energy malnutrition - Sex: Both
df_can1_indo2.rename(columns = {'Terrorism (deaths)':'Terrorism'}, inplace = True)
df_can1_indo2.rename(columns = {'Deaths - Road injuries - Sex: Both - Age: All Ages
df_can1_indo2.rename(columns = {'Deaths - Chronic respiratory diseases - Sex: Both
df_can1_indo2.rename(columns = {'Deaths - Cirrhosis and other chronic liver disease
df_can1_indo2.rename(columns = {'Deaths - Digestive diseases - Sex: Both - Age: Al
df_can1_indo2.rename(columns = {'Deaths - Fire, heat, and hot substances - Sex: Bot
df_can1_indo2.rename(columns = {'Deaths - Acute hepatitis - Sex: Both - Age: All Ag
```

```
plt.figure(figsize=(24,22))
mask = np.triu(np.ones_like(df_can1_indo2.corr()))

# plotting a triangle correlation heatmap
dataplot = sns.heatmap(df_indo_2.corr(), mask=mask,cmap="magma")

# displaying heatmap
plt.show()
```



A. Tentukan variabel-variabel yang memiliki nilai korelasi paling positif, dan berikan penjelasan terkait nilai korelasi tersebut. --> semakin tinggi nilai dari korelasi maka akan semakin kuat hubungan antara dua variabel tersebut. nilai korelasi paling positif merupakan variabel dengan warna paling cerah, beberapa variabel yang memiliki nilai korelasi yang tinggi adalah seperti parkison dan alzheimer, drowning dan meningitis, maternal disorders dan meningitis. hal ini menunjukan jika salah satu variable meningkat, maka kemungkinan besar akan terjadi peningkatan di variable lainnya.

B. Tentukan variabel-variabel yang memiliki nilai korelasi paling negatif, dan berikan penjelasan terkait nilai korelasi tersebut. --> nilai korelasi paling negatif merupakan variabel dengan warna paling gelap, contohnya ialah acute hepatitis dan HIV/AIDS. ketika salah satu variabel menurun, maka kemungkinan besar variabel lainnya akan meningkat secara linier.

Read the data (Dataset2)

In [672…
```
df_can2 = pd.read_csv('Dataset2.csv')
df_can2
```

Out[672]:

	Area	Year	Unit	Value	Flag	Flag Description
0	Afghanistan	1961	tonnes	319000.00	A	Official figure
1	Afghanistan	1962	tonnes	319000.00	A	Official figure
2	Afghanistan	1963	tonnes	319000.00	A	Official figure
3	Afghanistan	1964	tonnes	380000.00	A	Official figure
4	Afghanistan	1965	tonnes	380000.00	A	Official figure
...	...	...	...	...	...	...
7319	Zimbabwe	2017	tonnes	1326.90	I	Imputed value
7320	Zimbabwe	2018	tonnes	1342.47	I	Imputed value
7321	Zimbabwe	2019	tonnes	1134.00	A	Official figure
7322	Zimbabwe	2020	tonnes	750.00	E	Estimated value
7323	Zimbabwe	2021	tonnes	2908.00	A	Official figure

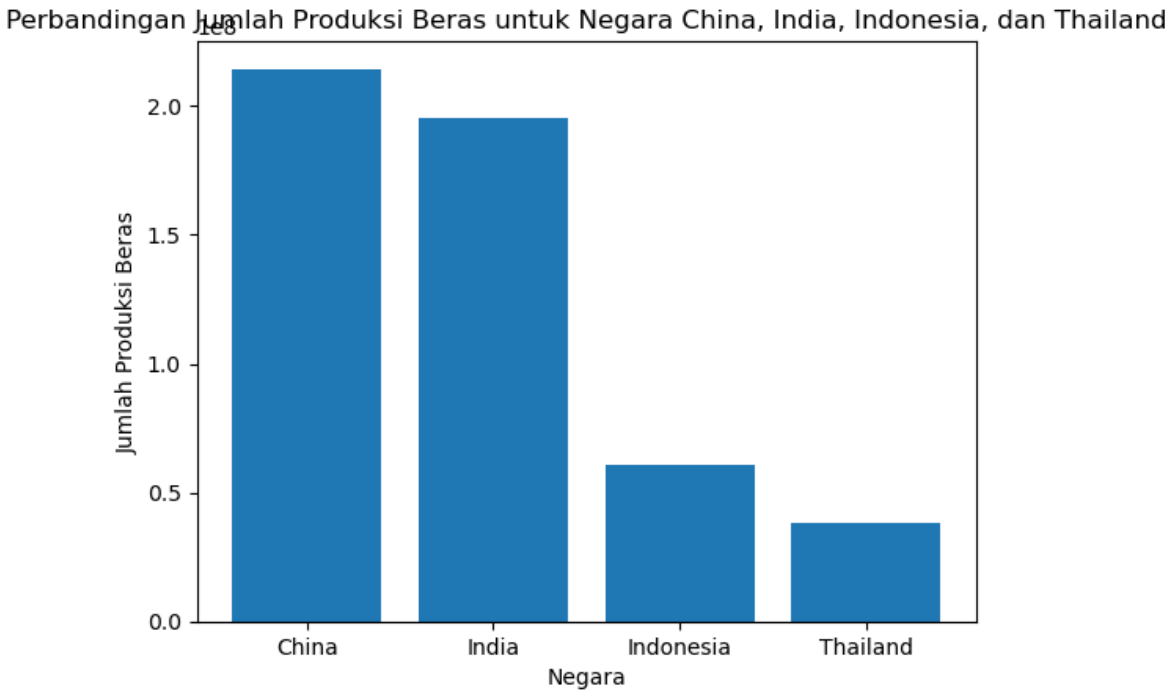
7324 rows × 6 columns

In [673…
```
df_can2_use = df_can2[['Area','Year','Value']]
```

In [674…
```
df_can2_subset = df_can2[df_can2["Area"].isin(["China", "India", "Indonesia", "Tha
```

Nomor 4

In [675…
```
plt.bar(df_can2_subset["Area"], df_can2_subset["Value"])

plt.title("Perbandingan Jumlah Produksi Beras untuk Negara China, India, Indonesia,
plt.xlabel("Negara")
plt.ylabel("Jumlah Produksi Beras")
plt.show()
```



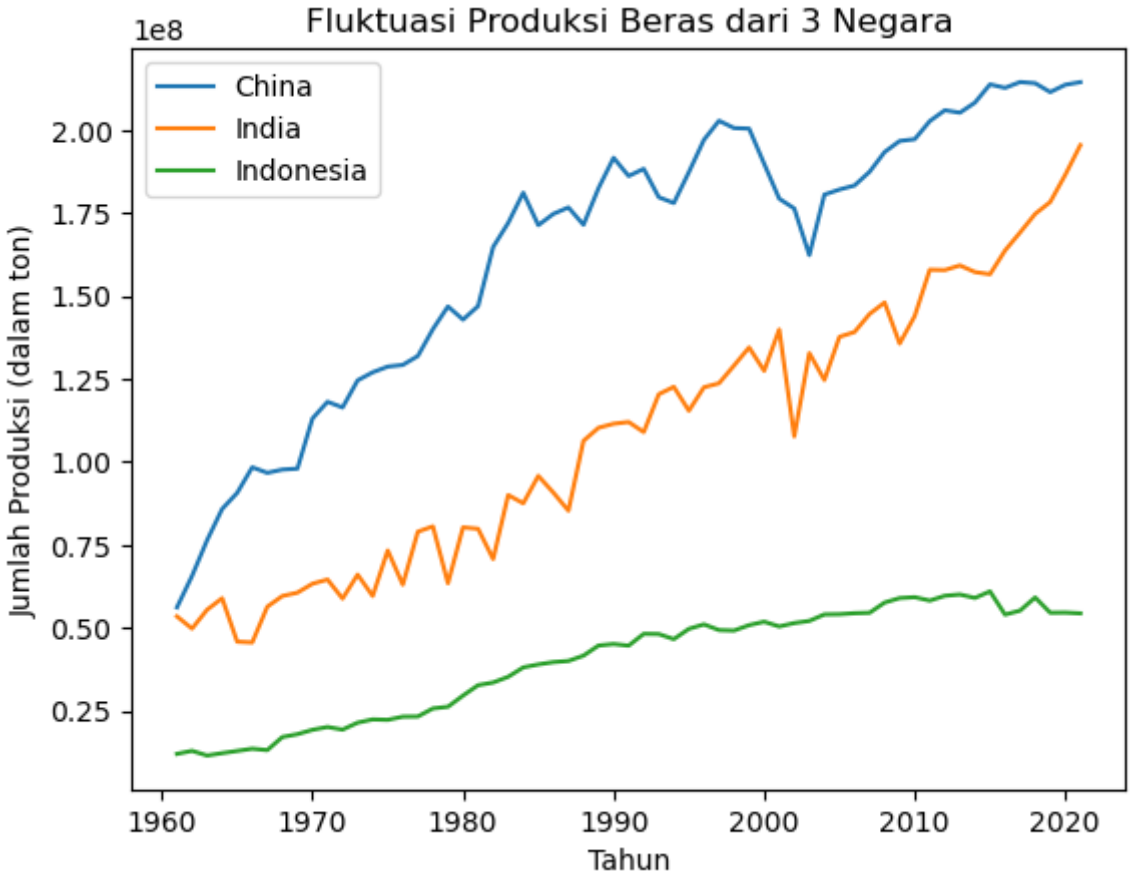
In [676…
```
matrix = df_can2_subset.corr()
print(matrix)
```

```
           Year     Value
Year   1.000000  0.400326
Value  0.400326  1.000000
```

Apakah ada hubungan korelasi jumlah produksi antar negara-negara China, India, Indonesia, dan Thailand? Berikan penjelasan --> tentu terdapat korelasi antara negara - negara tersebut, dapat dilihat pada hasil output diatas bahwa nilai korelasi jumlah produksi beras untuk negara china, india, indonesia, dan thailand sangat tinggi, yaitu mendekati nilai 1. ini menandakan bahwa terdapat hubungan yang kuat antara jumlah produksi beras di 4 negara tersebut, sehingga dapat disimpulkan bahwa hubungan korelasi jumlah produksi beras di antara negara - negara tersebut merupakan korelasi yang positif.

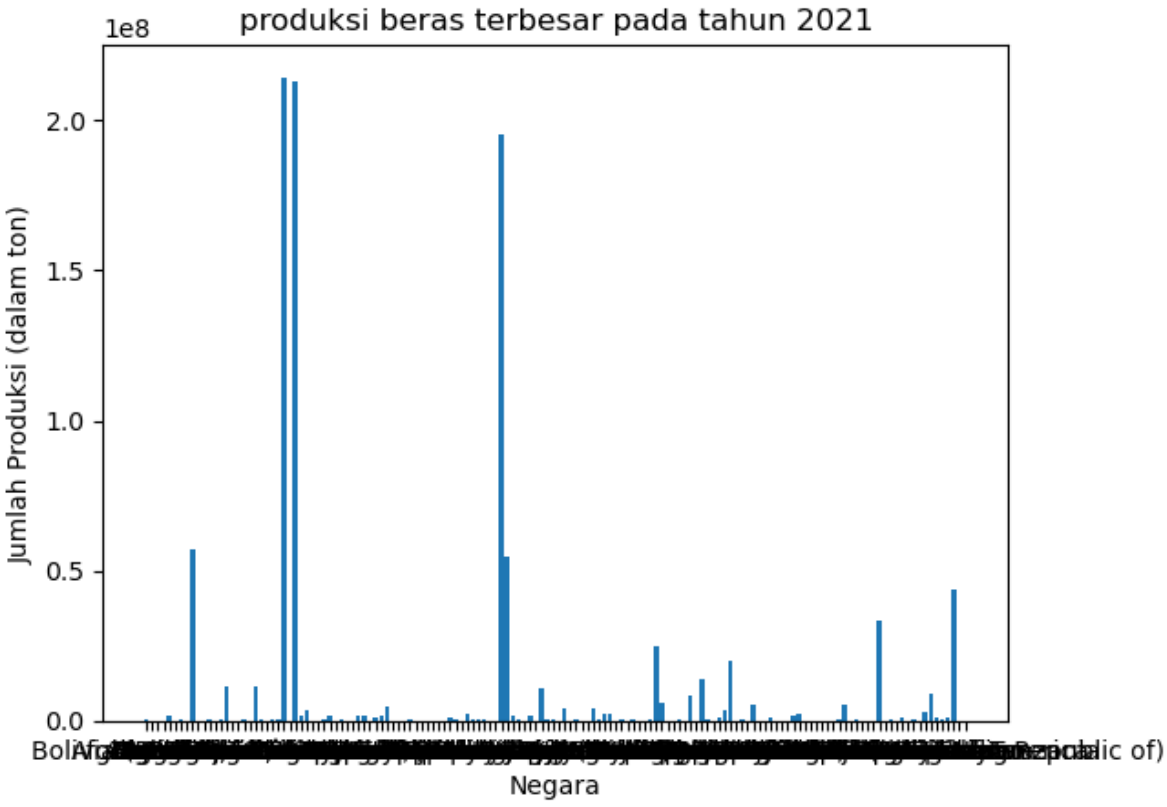
In [677…
```
China = df_can2[df_can2["Area"] == "China"]
India = df_can2[df_can2["Area"] == "India"]
Indonesia = df_can2[df_can2["Area"] == "Indonesia"]
```

In [678…
```
plt.plot(China["Year"], China["Value"], label="China")
plt.plot(India["Year"], India["Value"], label="India")
plt.plot(Indonesia["Year"], Indonesia["Value"], label="Indonesia")

plt.title("Fluktuasi Produksi Beras dari 3 Negara")
plt.xlabel("Tahun")
plt.ylabel("Jumlah Produksi (dalam ton)")
plt.legend()
plt.show()
```



In [679…

```
df_can2_2021 = df_can2[df_can2["Year"] == 2021]

plt.bar(df_can2_2021["Area"], df_can2_2021["Value"])
plt.title("produksi beras terbesar pada tahun 2021")
plt.xlabel("Negara")
plt.ylabel("Jumlah Produksi (dalam ton)")
plt.show()
```



In [680…
```
asia_Tenggara = ['Indonesia', 'Malaysia', 'Philippines', 'Singapore', 'Thailand',
df_can2_asia_Tenggara = df_can2[df_can2['Area'].isin(asia_Tenggara)]
df_can2_asia_Tenggara
```

Out[680]:

	Area	Year	Unit	Value	Flag	Flag Description
1036	Cambodia	1961	tonnes	2383000.0	A	Official figure
1037	Cambodia	1962	tonnes	2039000.0	A	Official figure
1038	Cambodia	1963	tonnes	2622000.0	A	Official figure
1039	Cambodia	1964	tonnes	2760000.0	A	Official figure
1040	Cambodia	1965	tonnes	2500000.0	A	Official figure
...	...	...	...	...	...	...
6442	Thailand	2017	tonnes	32898903.0	A	Official figure
6443	Thailand	2018	tonnes	32348114.0	A	Official figure
6444	Thailand	2019	tonnes	28617948.0	A	Official figure
6445	Thailand	2020	tonnes	30231025.0	A	Official figure
6446	Thailand	2021	tonnes	33582000.0	A	Official figure

366 rows × 6 columns

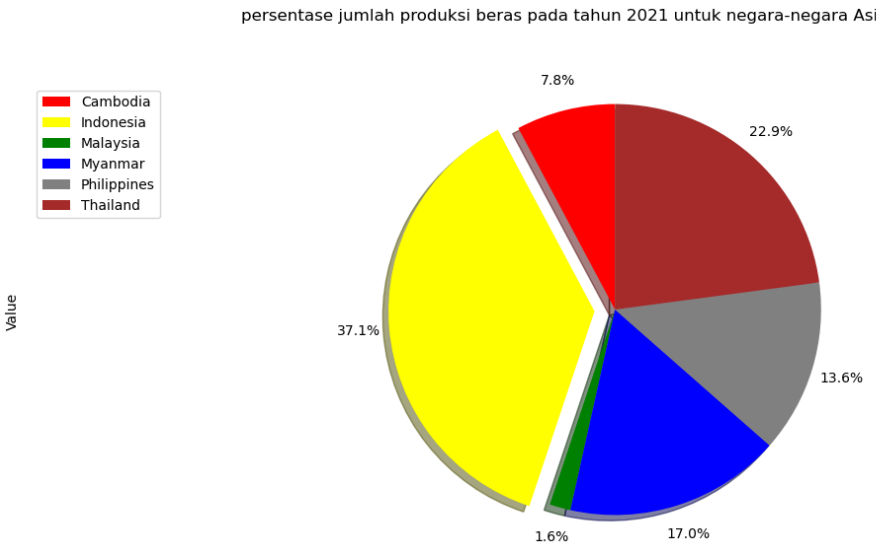
In [681…
```
df_can2_asia_Tenggara = df_can2_asia_Tenggara[df_can2_asia_Tenggara['Year']==2021]
df_can2_asia_Tenggara = df_can2_asia_Tenggara[['Area','Value']]
df_can2_asia_Tenggara
```

Out[681]:

	Area	Value
1096	Cambodia	11410000.0
3395	Indonesia	54415294.0
4142	Malaysia	2418148.0
4598	Myanmar	24910000.0
5273	Philippines	19960170.2
6446	Thailand	33582000.0

In [682…
```
colors_list = ['red', 'yellow', 'green', 'blue', 'grey', 'brown']
explode_list = [0, 0.1, 0, 0, 0, 0]
df_can2_asia_Tenggara['Value'].plot(kind='pie',
                            figsize=(15, 6),
                            autopct='%1.1f%%',
                            startangle=90,
                            shadow=True,
                            labels=None,
                            pctdistance=1.15,
                            colors=colors_list,
                            explode=explode_list
                            )
plt.title('persentase jumlah produksi beras pada tahun 2021 untuk negara-negara As
plt.axis('equal')

# add legend
plt.legend(labels=df_can2_asia_Tenggara['Area'], loc='upper left')
```

```
plt.show()
```


persentase jumlah produksi beras pada tahun 2021 untuk negara-negara Asia Tenggara
Cambodia
Indonesia
Malaysia
Myanmar
Philippines
Thailand
Value
7.8%
22.9%
37.1%
13.6%
1.6%
17.0%