



seaborn



*matplotlib*

# DATA VISUALISATION & ANALYSIS

WITH SEABORN AND MATPLOTLIB

- Josephine XI. MIPA/16 -

# DAFTAR ISI

## MATPLOTLIB

**01** Line Graph (2D)

**02** Histogram (2D)

**03** Pie Chart (2D)

**04** Bar Chart (3D)

**05** Surface (3D)

**06** Scatter Plot (3D)

## SEABORN

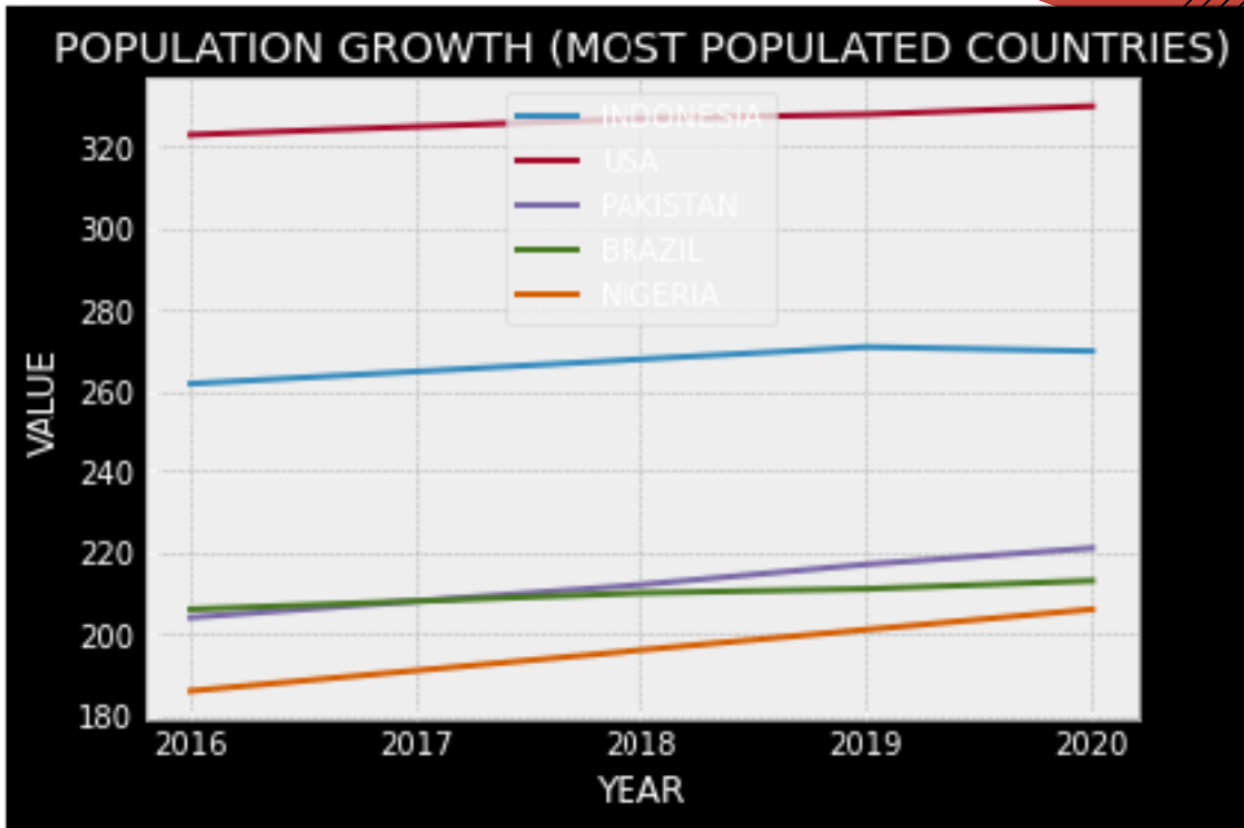
**01** Line Graph

**02** Bar Graph

**03** Scatter Plot

# MATPLOTLIB

## 1. LINE GRAPH (2D)



POPULATION GROWTH ANALYSIS (WITH A LINE GRAPH)

```
In [104]: x = [2016, 2017, 2018, 2019, 2020]
y = [262, 265, 268, 271, 270]

x2 = x
y2 = [323, 325, 327, 328, 330]

x3 = x
y3 = [204, 208, 212, 217, 221]

x4 = x
y4 = [206, 208, 210, 211, 213]

x5 = x
y5 = [186, 191, 196, 201, 206]

plt.title('POPULATION GROWTH (MOST POPULATED COUNTRIES)')
plt.xlabel('YEAR')
plt.ylabel('VALUE')

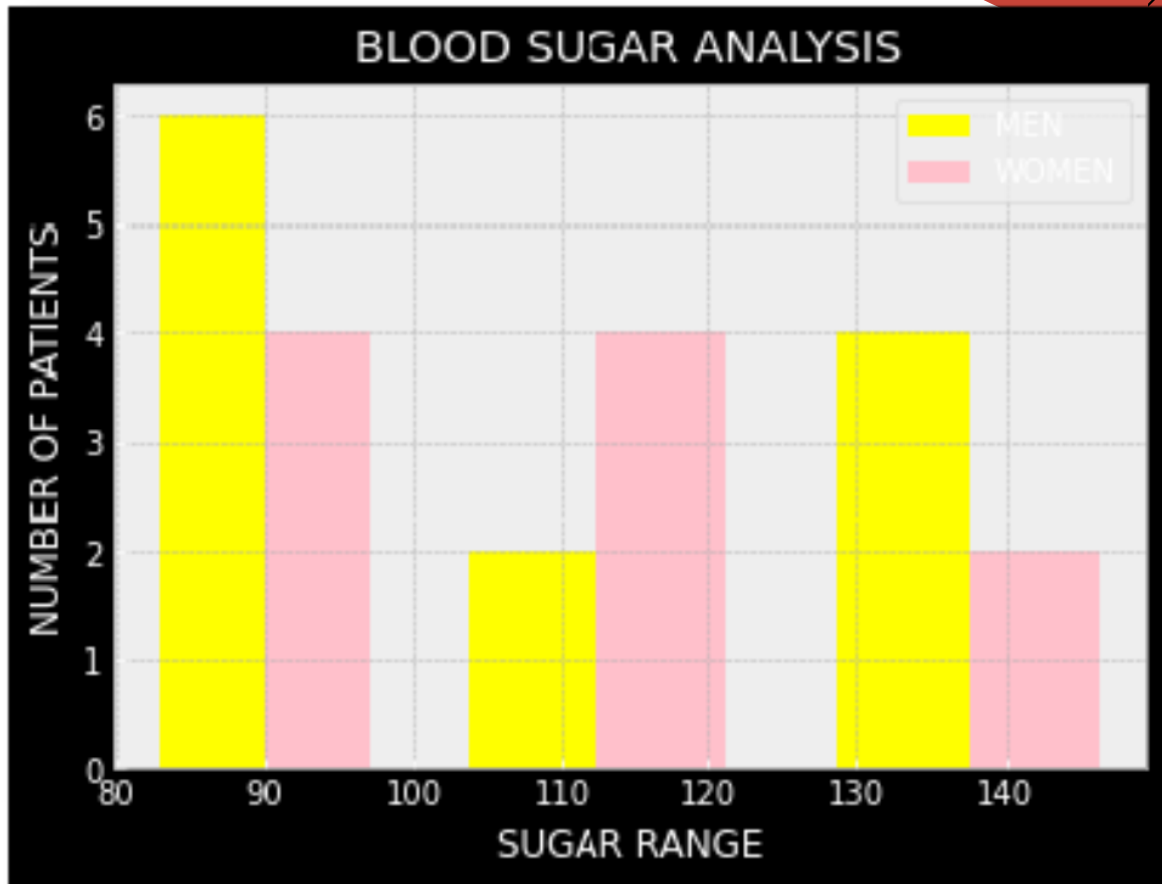
plt.xticks(x)
plt.plot(x, y, label='INDONESIA')
plt.plot(x2, y2, label='USA')
plt.plot(x3, y3, label='PAKISTAN')
plt.plot(x4, y4, label='BRAZIL')
plt.plot(x5, y5, label='NIGERIA')
plt.legend()
plt.show()
```

Grafik pertama yang saya buat menggunakan library Matplotlib adalah Line Graph. Di sini, saya menggunakan data yang saya buat sendiri, yaitu analisa mengenai pertumbuhan penduduk pada negara-negara dengan rata-rata jumlah penduduk terbanyak di dunia, yaitu Indonesia, Amerika Serikat, Pakistan, Brazil, dan Nigeria.

Akan tetapi, garis pada grafik tidak terlalu terlihat dan signifikan dikarenakan pertumbuhan/penurunan penduduk yang terjadi di negara tersebut kecil, serta saya menggunakan jangka waktu 5 tahun terakhir saja (2016, 2017, 2018, 2019, 2020).

# MATPLOTLIB

## 2. HISTOGRAM (2D)



BLOOD SUGAR ANALYSIS (WITH A HISTOGRAM)

```
In [107]: blood_sugar_men = [113, 85, 90, 150, 149, 88, 93, 115, 135, 80, 77, 82, 129]
blood_sugar_women = [67, 98, 89, 120, 133, 150, 84, 69, 89, 79, 120, 112, 100]

plt.xlabel('SUGAR RANGE')
plt.ylabel('NUMBER OF PATIENTS')
plt.title('BLOOD SUGAR ANALYSIS')
plt.hist([blood_sugar_men, blood_sugar_women], bins=[80, 100, 125, 150], rwidth=0.70, color=['yellow', 'pink'],
         label=['MEN', 'WOMEN'])
plt.legend()
plt.show()
```

Grafik kedua adalah Histogram. Saya menggunakan data tekanan gula darah pada beberapa pasien pria dan wanita. Bar berwarna kuning menunjukkan pasien pria, sedangkan bar berwarna pink menunjukkan pasien wanita.

### PASIEN PRIA

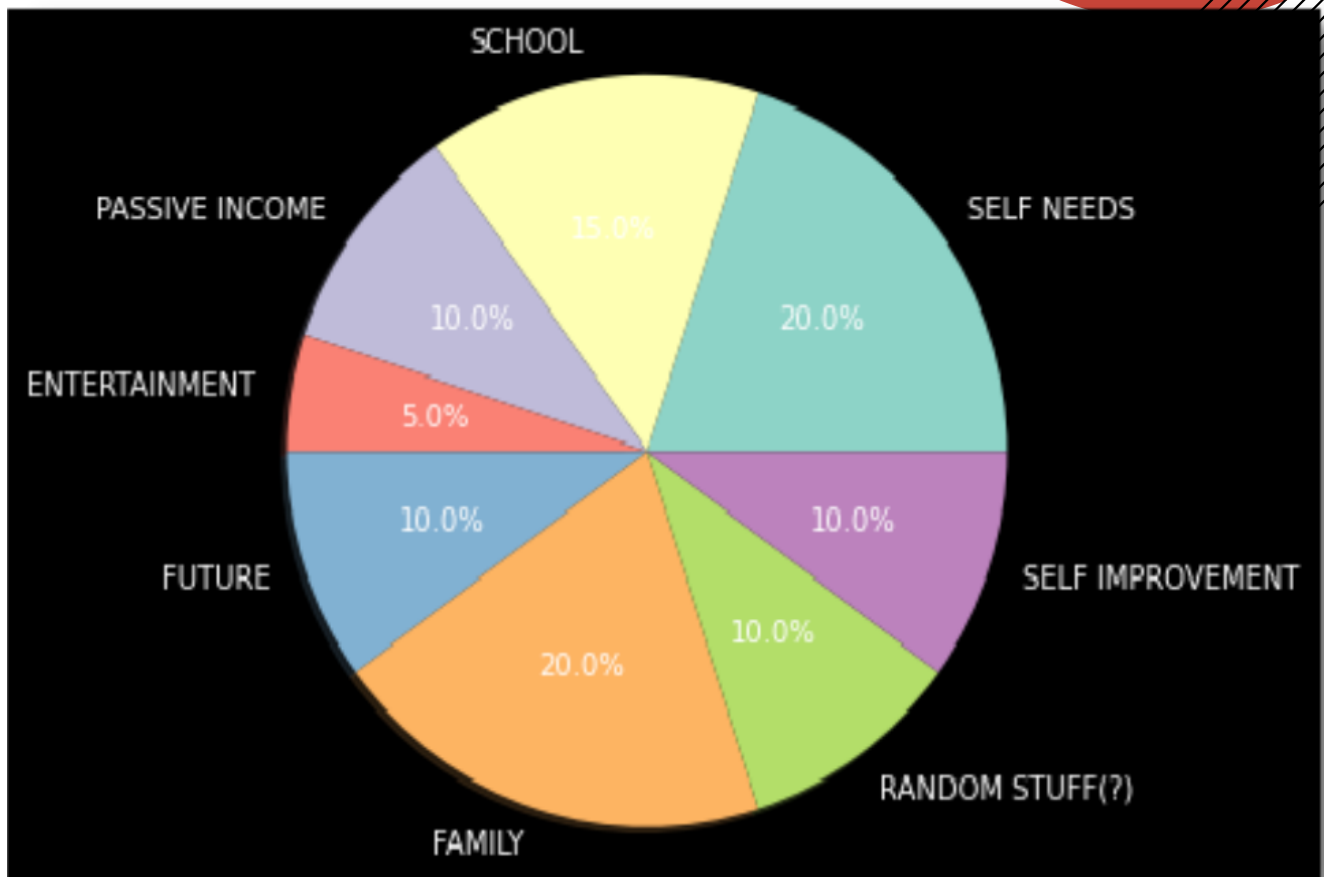
Rentang <80 tekanan gula darah : 1 orang.  
Rentang 80-93 tekanan gula darah : 6 orang.  
Rentang 113-115 tekanan gula darah : 2 orang.  
Rentang 129-150 tekanan gula darah : 4 orang.

### PASIEN WANITA

Rentang <80 tekanan gula darah : 3 orang.  
Rentang 84-98 tekanan gula darah : 4 orang.  
Rentang 100-120 tekanan gula darah : 4 orang.  
Rentang 133-150 tekanan gula darah : 2 orang.

# MATPLOTLIB

## 3. PIE CHART (2D)



WHAT I THINK IN A DAY!<3 (WITH A PIE CHART)

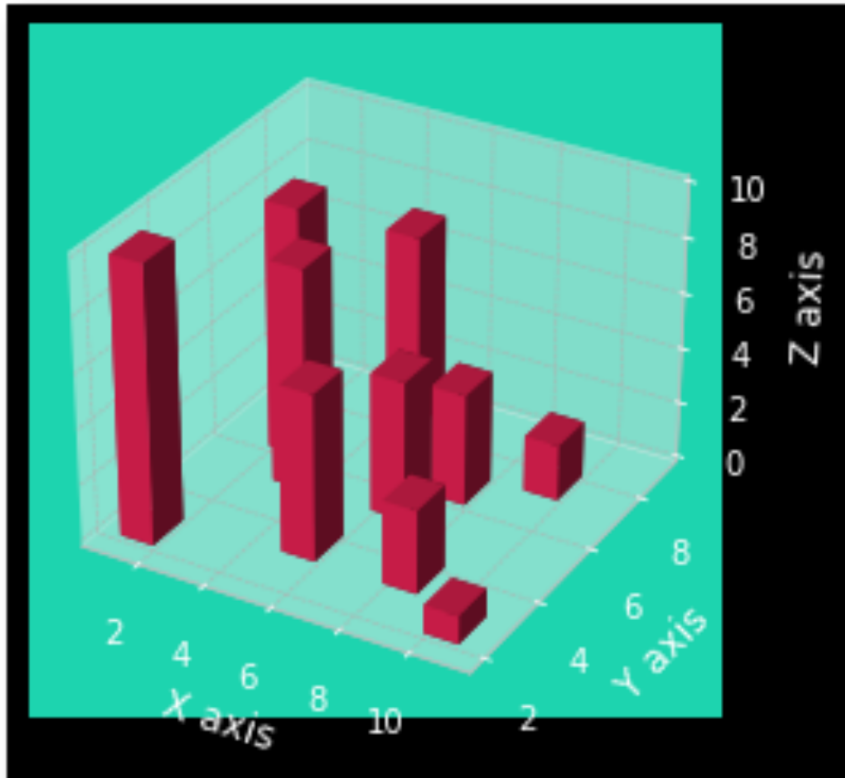
```
In [98]: exp_vals = [20,15,10,5,10,20,10,10]
exp_labels = ["SELF NEEDS","SCHOOL","PASSIVE INCOME","ENTERTAINMENT","FUTURE","FAMILY","RANDOM STUFF(?)","SELF IMPROVEMENT"]
plt.axis("equal")
plt.pie(exp_vals,labels=exp_labels,radius=1.5, autopct='%0.1f%%', shadow=True)
plt.show()
```

Grafik ketiga adalah Histogram. Saya menggunakan data yang saya buat sendiri, yaitu tentang hal-hal yang saya pikirkan setiap harinya. Data tersebut, yaitu :

1. SCHOOL (15%)
2. SELF NEEDS (20%)
3. SELF IMPROVEMENT (10%)
4. RANDOM STUFF (10%)
5. FAMILY (20%)
6. FUTURE (10%)
7. ENTERTAINMENT (5%)
8. PASSIVE INCOME (10%)

# MATPLOTLIB

## 4. BAR CHART (3D)



### 3D BAR CHART

```
In [112]: import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d

fig = plt.figure(facecolor="Black")
ax = plt.axes(projection="3d")

xpos = [1,2,3,4,5,6,7,8,9,10]
ypos = [2,6,5,8,3,5,6,3,7,2]
zpos = np.zeros(10)

xsize = np.ones(10)
ysize = np.ones(10)
zsize = [10,9,8,7,6,5,4,3,2,1]

ax.set_facecolor("#1DD4AF")
ax.set_xlabel("X axis")
ax.set_ylabel("Y axis")
ax.set_zlabel("Z axis")

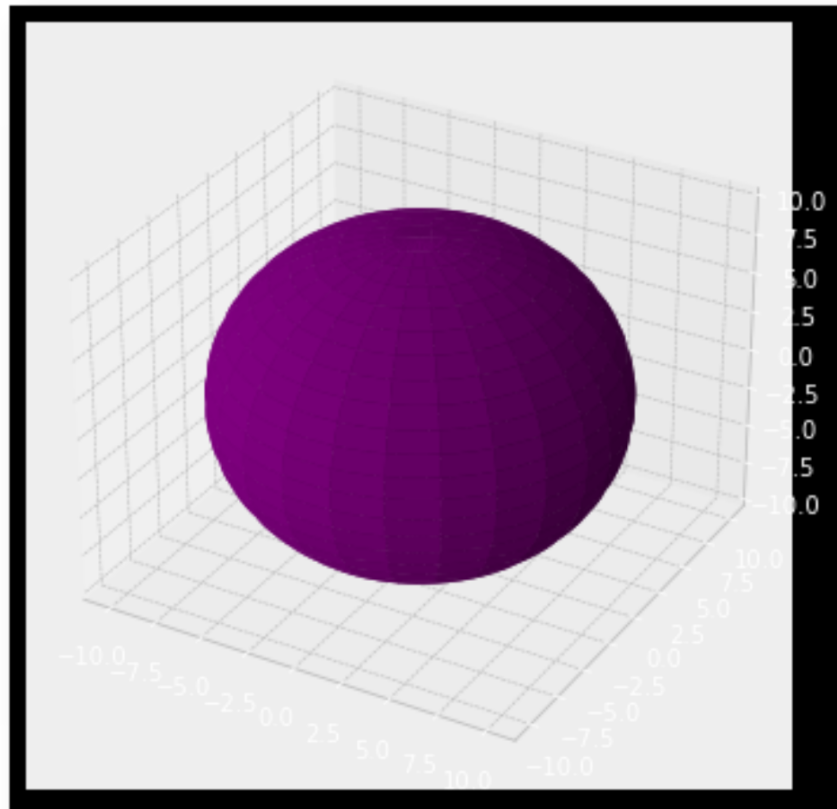
ax.bar3d(xpos, ypos, zpos, xsize, ysize, zsize, color = "#E02050")
plt.show()
```

Grafik selanjutnya adalah grafik 3D dengan bentuk bar chart. Tambahannya adalah kita harus mengimport mplot3d saat ingin membentuk grafik 3D.

Saya menggunakan nilai X = [1,2,3,4,5,6,7,8,9,10], nilai Y = [2,6,5,8,3,5,6,3,7,2], dan nilai Z = [10,9,8,7,6,5,4,3,2,1] kemudian saya mengganti warna bar menjadi pink dan warna backgroundnya menjadi biru muda.

# MATPLOTLIB

## 5. SURFACE (3D)



### 3D SURFACE

```
In [116]: fig = plt.figure(figsize = (6,6))
ax = fig.add_subplot(111, projection='3d')

u = np.linspace(0, 2 * np.pi, 100)
v = np.linspace(0, np.pi, 100)

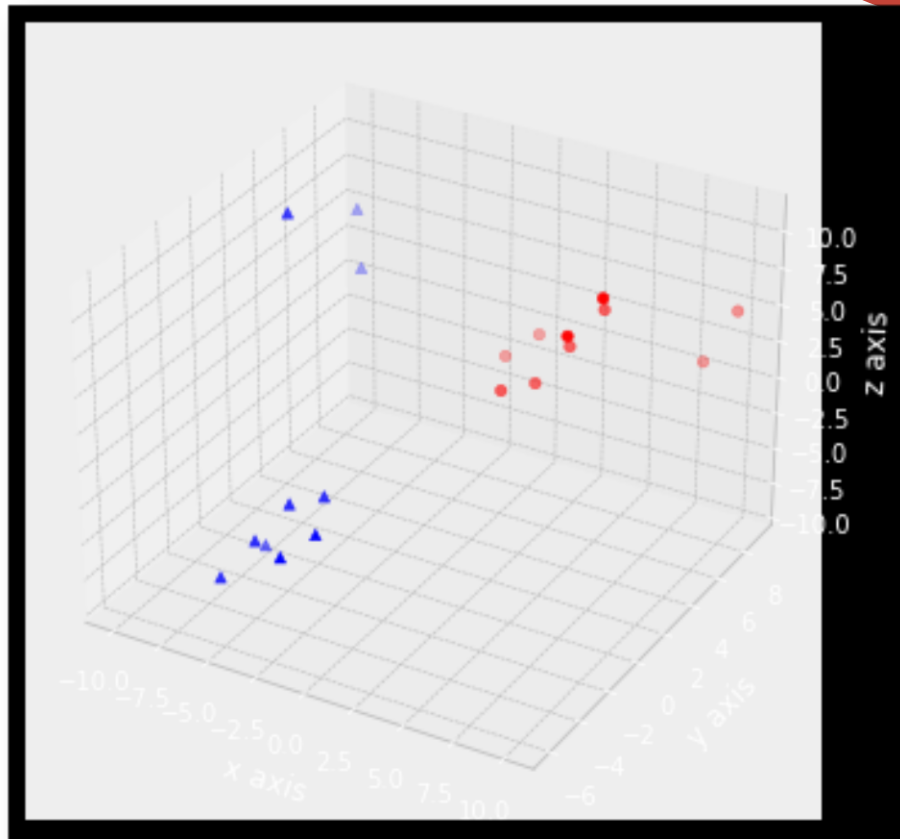
x = 10 * np.outer(np.cos(u), np.sin(v))
y = 10 * np.outer(np.sin(u), np.sin(v))
z = 10 * np.outer(np.ones(np.size(u)), np.cos(v))
ax.plot_surface(x, y, z, rstride=4, cstride=4, color='purple')

plt.show()
```

Grafik selanjutnya adalah grafik surface 3D, dengan bentuk geometri oval, dan saya memberi warna geometrinya menjadi warna ungu.

# MATPLOTLIB

## 6. SCATTER PLOT (3D)



### 3D SCATTER PLOT

```
In [121]: import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

fig = plt.figure(figsize = (6,6))
ax = fig.add_subplot(111, projection='3d')

X = [1,2,3,4,5,6,7,8,9,10]
Y = [5,6,2,3,4,5,1,2,8,9]
Z = [2,3,3,3,5,7,9,11,2,5]

Xs = [-1,-2,-3,-4,-5,-6,-7,-8,-9,-10]
Ys = [-5,-6,-2,-3,-4,-5,-1,2,8,9]
Zs = [-2,-3,-3,-3,-5,-7,-9,11,2,5]

ax.scatter(X, Y, Z, c='r', marker='o')
ax.scatter(Xs, Ys, Zs, c='b', marker='^')

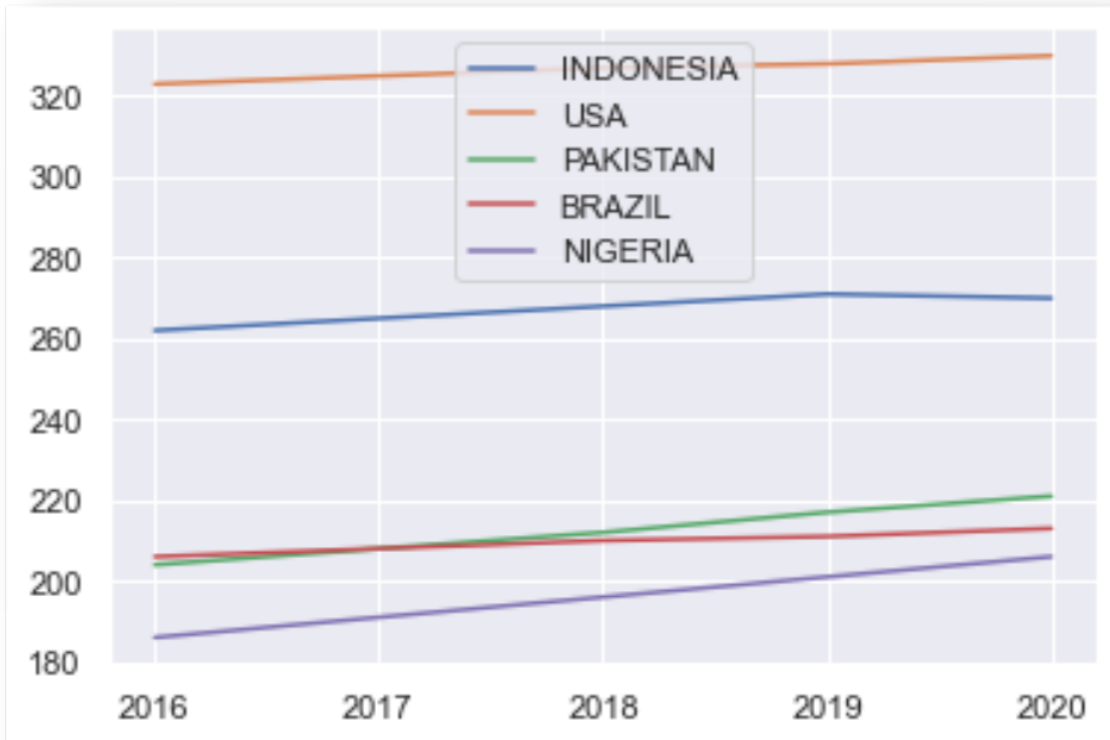
ax.set_xlabel('x axis')
ax.set_ylabel('y axis')
ax.set_zlabel('z axis')
plt.show()
```

Grafik selanjutnya adalah grafik scatter plot 3D. Saya menggunakan 2 jenis variabel, yaitu variabel X,Y,Z dan Xs,Ys,Zs. Pada variabel Xs,Ys, dan Zs, saya mengubah nilai variabel nya menjadi min(-) agar perbedaan di antara keduanya semakin jelas, serta warna kedua variabel juga dibedakan, yakni merah untuk X,Y,Z dan biru untuk Xs,Ys, dan Zs.



# SEABORN

## 1. LINE GRAPH



```
POPULATION GROWTH ANALYSIS (WITH A LINE GRAPH)

In [123]: import matplotlib.pyplot as plt
import seaborn as sns
sns.set_theme(style="darkgrid")

x = [2016, 2017, 2018, 2019, 2020]
y = [262, 265, 268, 271, 270]

x2 = x
y2 = [323, 325, 327, 328, 330]

x3 = x
y3 = [204, 208, 212, 217, 221]

x4 = x
y4 = [206, 208, 210, 211, 213]

x5 = x
y5 = [186, 191, 196, 201, 206]

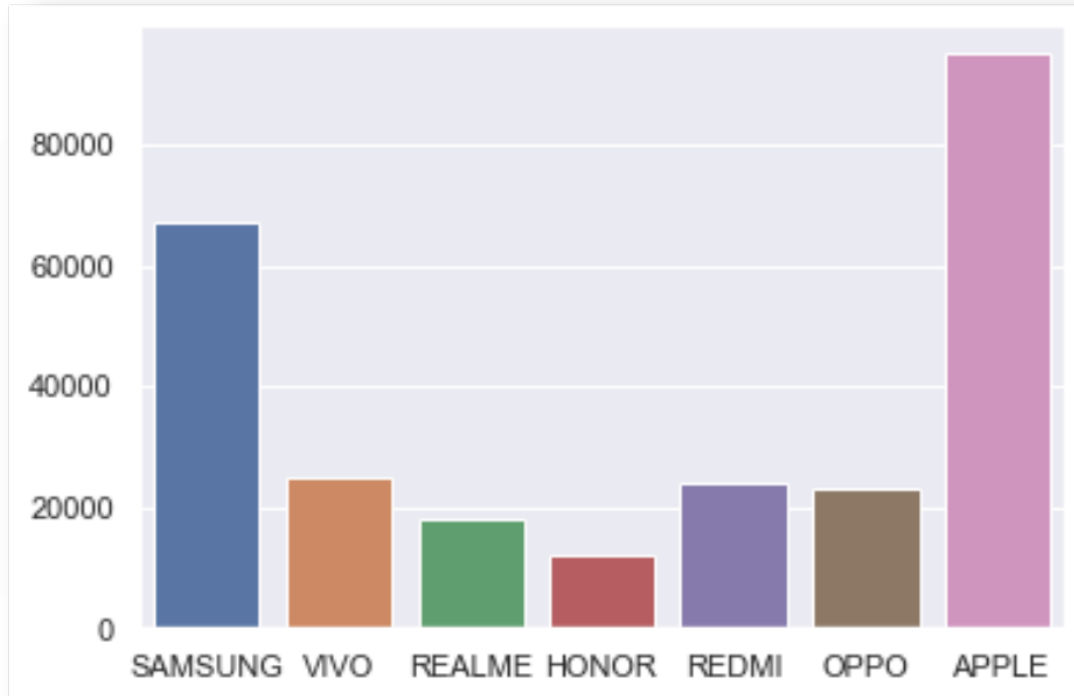
sns.lineplot(x,y, label='INDONESIA')
sns.lineplot(x2,y2, label='USA')
sns.lineplot(x3,y3, label='PAKISTAN')
sns.lineplot(x4,y4, label='BRAZIL')
sns.lineplot(x5,y5, label='NIGERIA')
plt.xticks(x)
plt.legend()
plt.show()

from warnings import filterwarnings
filterwarnings('ignore')
```

Grafik pertama menggunakan library seaborn adalah line graph. Di sini saya menggunakan data sebelumnya, yaitu data pertumbuhan penduduk. Tidak ada perbedaan pada grafik, hanya berbeda warna background.

# SEABORN

## 2. BAR GRAPH



SMARTPHONE PRICES (WITH A BAR GRAPH)

```
In [124]: brand = ['SAMSUNG', 'VIVO', 'REALME', 'HONOR', 'REDMI', 'OPPO', 'APPLE']  
price = [67000, 25000, 18000, 12000, 24000, 23000, 95000]  
  
sns.barplot(x= brand, y= price)
```

Grafik kedua adalah bar graph. Di sini saya menggunakan data harga-harga smartphone (Samsung, Vivo, RealMe, Honor, Redmi, Oppo, Apple), dan rentang harga berkisar antara 18000-95000.

# SEABORN

## 3. SCATTER PLOT



### GDP PER CAPITA & LIFE EXPERIENCE ANALYSIS (WITH A SCATTER PLOT)

```
In [130]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

data_url = 'http://bit.ly/2cLzoxH'
gapminder = pd.read_csv(data_url)

gapminder=gapminder[gapminder.year.isin([2002,1962])]

g = sns.scatterplot(x="gdpPercap", y="lifeExp",
                    hue="continent",
                    data=gapminder);
g.set(xscale="log");
```

Grafik terakhir adalah grafik scatter plot. Di sini, saya menggunakan dataset dari <http://bit.ly/2cLzoxH> dikarenakan sulit untuk membuat sebuah grafik scatter plot yang bagus jika data yang kita miliki hanya sedikit. Data menunjukkan nilai GDP per Capita dan Life Expectation di 5 continent berbeda (Asia, Europe, Africa, Americas, Oceania).

# ANALISIS PERBEDAAN

- **BERDASARKAN HASIL**

Berdasarkan hasil grafiknya, grafik pada library Matplotlib jauh lebih beragam, dikarenakan kita bisa membuat grafik 2D dan 3D di dalamnya. Jika kita ingin membuat grafik yang jauh lebih *Advanced* (professional), maka disarankan untuk memakai library Matplotlib, dan menggunakan library Seaborn jika grafik yang ingin dibuat jauh lebih *simple*.

- **BERDASARKAN PENULISAN CODE**

Berdasarkan penulisan codenya, kita menggunakan code `plt` untuk Matplotlib dan `sns` untuk Seaborn. Seaborn adalah library yang dibentuk di bawah Matplotlib serta jauh lebih *simple* daripada Matplotlib, sehingga code untuk membuat grafik pada Matplotlib jauh lebih rumit dan panjang. Hanya saja, terdapat beberapa kekurangan pada coding untuk Seaborn, sehingga kita harus tetap mengimport Matplotlib.