

TUGAS LAB PEMBELAJARAN MESIN

Nama : Meina Lisa

NIM : 191402032

TUGAS 4

1. Carilah sebuah dataset

<https://www.kaggle.com/datasets/swashbuckler1/pokemon-gen1gen8>

2. Buatlah dan jelaskan codingan untuk data preprocessing yang berisi:

- **import library**, ada 4 macam library yang digunakan pada file Tugas4 ini, yaitu numpy (numerical python), pandas (python for data analysis), matplotlib untuk visualisasi, dan seaborn.

```
[1] from matplotlib import pyplot as plt
import seaborn as sns
import pandas as pd
import numpy as np
```

- **read dataset**, yaitu dengan perintah berikut dan menyertakan URL dataset.

```
df1 = pd.read_csv('/content/sample_data/pokemon-data.csv')
df1
```

	Pokedex No.	Name	Type	Other Type	HP	Attack	Defense	Special Attack	Special Defense	Speed	Total	Generation	Legendary
0	1	Bulbasaur	Grass	Poison	45	49	49	65	65	45	318	1	0
1	2	Ivysaur	Grass	Poison	60	62	63	80	80	60	405	1	0
2	3	Venusaur	Grass	Poison	80	82	83	100	100	80	525	1	0
3	3	Mega Venusaur	Grass	Poison	80	100	123	122	120	80	625	1	0
4	4	Charmander	Fire	NaN	39	52	43	60	50	65	309	1	0
...

- **eksplorasi data**,

a. Lihat dimensi data, yaitu berapa jumlah baris dan kolom

```
[3] df1.shape

(1045, 13)
```

b. Lihat kolom data, yaitu kolom apa saja yang terdapat pada data

```
[4] df1.columns

Index(['Pokedex No.', 'Name', 'Type', 'Other Type', 'HP', 'Attack', 'Defense',
      'Special Attack', 'Special Defense', 'Speed', 'Total', 'Generation',
      'Legendary'],
      dtype='object')
```

- c. Lihat informasi data, untuk melihat detail informasi dari setiap kolom
- d. Lihat data kosong, mengecek apakah ada data yang kosong di kolom tertentu

```
df1.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1045 entries, 0 to 1044
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Pokedex No.           1045 non-null  int64
1   Name                  1045 non-null  object
2   Type                  1045 non-null  object
3   Other Type            553 non-null   object
4   HP                    1045 non-null  int64
5   Attack                1045 non-null  int64
6   Defense               1045 non-null  int64
7   Special Attack        1045 non-null  int64
8   Special Defense       1045 non-null  int64
9   Speed                 1045 non-null  int64
10  Total                  1045 non-null  int64
11  Generation             1045 non-null  int64
12  Legendary              1045 non-null  int64
dtypes: int64(10), object(3)
memory usage: 106.3+ KB
```

```
[7] df1.isnull().sum()

Pokedex No.      0
Name              0
Type              0
Other Type       492
HP                0
Attack            0
Defense           0
Special Attack    0
Special Defense   0
Speed             0
Total             0
Generation        0
Legendary         0
dtype: int64
```

- e. Lihat ringkasan statistik

```
df1.describe()
```

	Pokedex No.	HP	Attack	Defense	Special Attack	Special Defense	Speed	Total	Generation	Legendary
count	1045.000000	1045.000000	1045.000000	1045.000000	1045.000000	1045.000000	1045.000000	1045.000000	1045.000000	1045.000000
mean	440.769378	70.067943	80.466986	74.661244	73.022010	72.288995	68.807656	439.314833	4.059330	0.121531
std	262.517231	26.671411	32.413665	31.237903	32.724797	28.074148	30.210094	121.970701	2.264232	0.326900
min	1.000000	1.000000	5.000000	5.000000	10.000000	20.000000	5.000000	175.000000	1.000000	0.000000
25%	212.000000	50.000000	55.000000	50.000000	50.000000	50.000000	45.000000	330.000000	2.000000	0.000000
50%	436.000000	68.000000	77.000000	70.000000	65.000000	70.000000	65.000000	458.000000	4.000000	0.000000
75%	670.000000	82.000000	100.000000	90.000000	95.000000	90.000000	90.000000	515.000000	6.000000	0.000000
max	898.000000	255.000000	190.000000	250.000000	194.000000	250.000000	200.000000	1125.000000	8.000000	1.000000

- menangani missing value, yaitu dengan mengganti data kosong menggunakan 0.

```
df1.fillna(0, inplace=True)
df1.isnull().sum()

Pokedex No.      0
Name              0
Type              0
Other Type       0
HP                0
Attack            0
Defense           0
Special Attack    0
Special Defense   0
Speed             0
Total             0
Generation        0
Legendary         0
dtype: int64
```

- **data formatting(optional)**, melakukan formatting data dengan mengatur tipe datanya menjadi int.

```
for i in ['HP', 'Attack', 'Defense', 'Speed', 'Total']:
    df1[i] = df1[i].astype('int')
```

- **menghapus kolom(optional)**.

```
df1.drop('Total', inplace=True, axis=1)
df1.head()
```

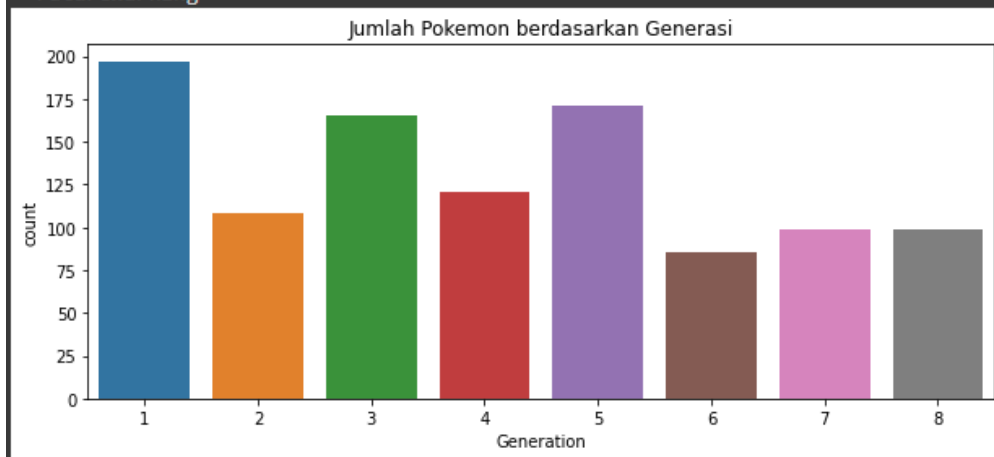
	Pokedex No.	Name	Type	Other Type	HP	Attack	Defense	Special Attack	Special Defense	Speed	Generation	Legendary
0	1	Bulbasaur	Grass	Poison	45	49	49	65	65	45	1	0
1	2	Ivysaur	Grass	Poison	60	62	63	80	80	60	1	0
2	3	Venusaur	Grass	Poison	80	82	83	100	100	80	1	0
3	3	Mega Venusaur	Grass	Poison	80	100	123	122	120	80	1	0
4	4	Charmander	Fire		39	52	43	60	50	65	1	0

3. Buatlah codingan untuk visualisasi data pada:

- **Satu label categorical data**, yaitu mengelompokkan jumlah pada kategori tertentu.

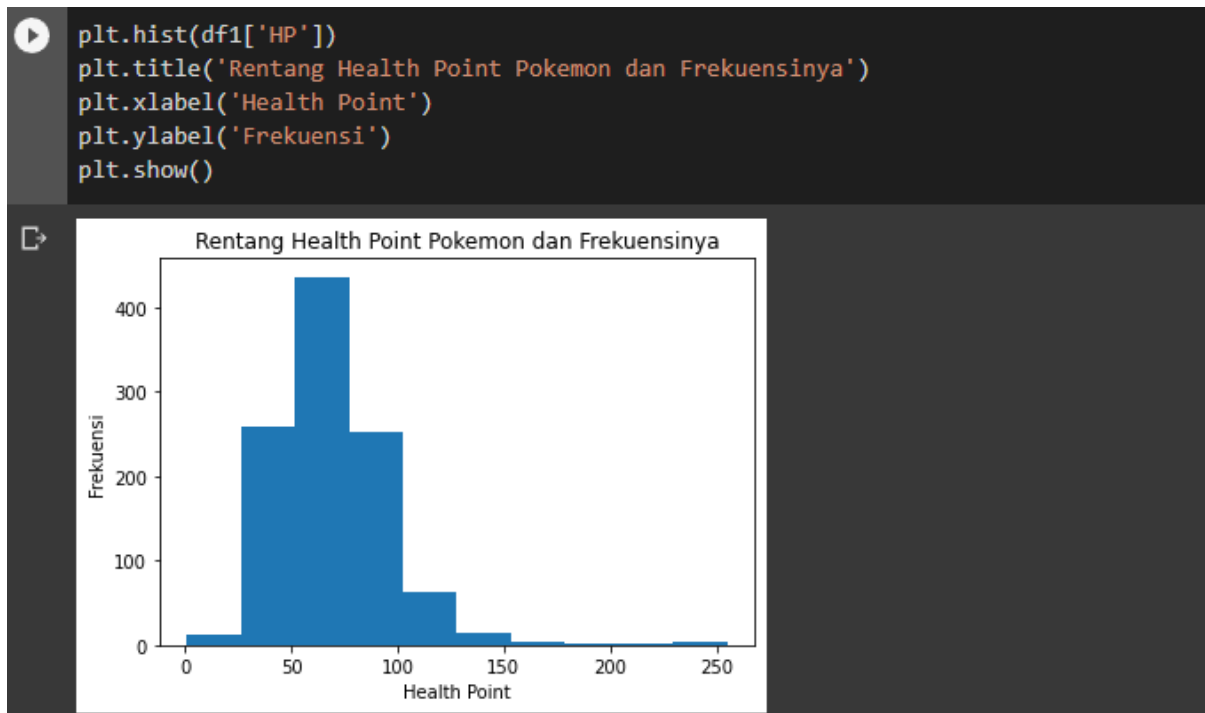
```
plt.figure(figsize=(10,4))
sns.countplot(df1['Generation'])
plt.title('Jumlah Pokemon berdasarkan Generasi')
plt.show()
```

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following FutureWarning



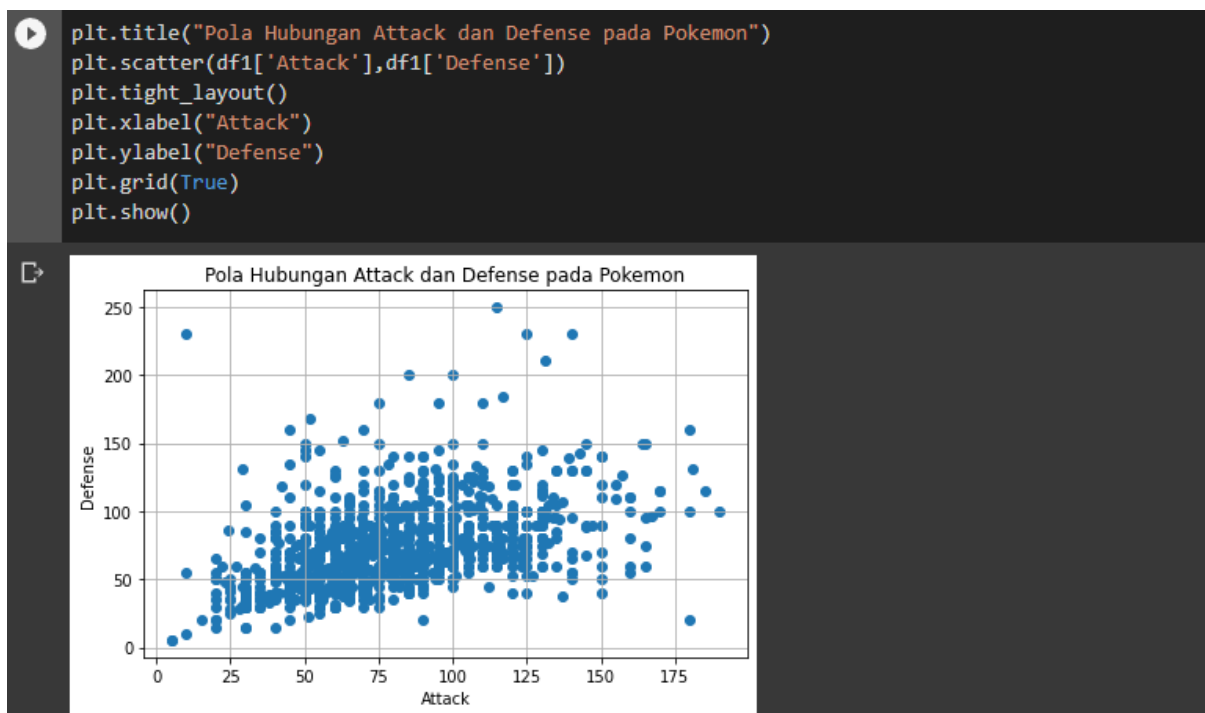
Grafik diagram batang di atas menggambarkan pengelompokan jumlah pokemon yang terdapat pada tiap generasi dari 1-8. Berdasarkan data yang dimiliki, jumlah pokemon paling banyak dirilis pada generasi pertama yaitu hampir mencapai 200 dan paling sedikit pada generasi keenam.

- **Satu label numerical data**, yaitu memvisualisasikan rentang data dengan frekuensinya.



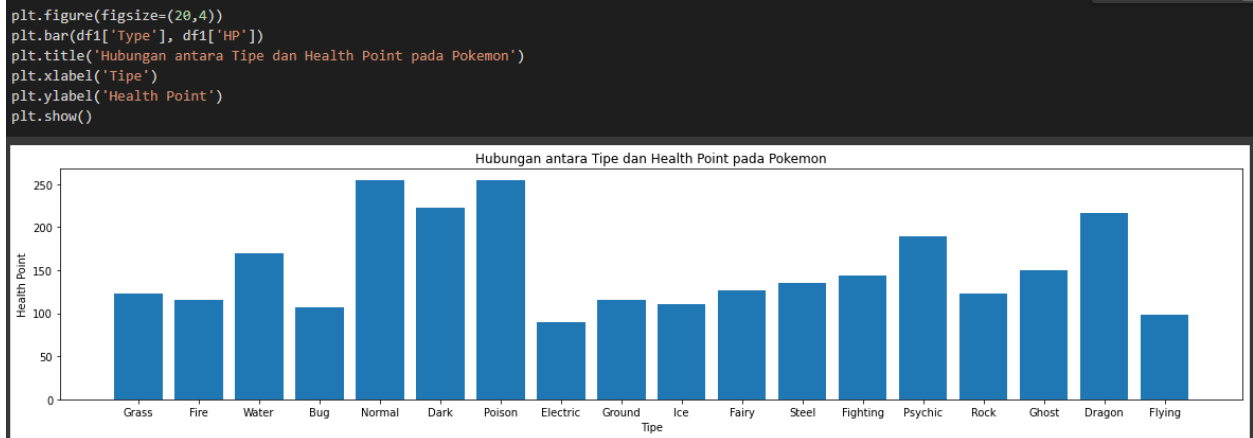
Grafik histogram diatas menggambarkan rentang HP (Health Point) dari pokemon dan menunjukkan jumlah pokemon yang memiliki HP tertentu. Jumlah terbanyak terdapat dalam rentang HP 50-100.

- **Dua label data numerical dan numerical.**



Grafik scatter plot diatas memvisualisasikan pola huungan antara Attack dan Defense pada pokemon. Dapat dilihat bahwa data sangat bervariasi sehingga sulit untuk bisa langsung menentukan bagaimana pola Attack dan Defense yang terdapat pada pokemon.

- **Dua label data numerical dan categorical.**



Grafik diagram batang di atas menunjukkan hubungan jumlah HP yang dimiliki oleh pokemon dengan tipe yang dimilikinya, yang mana dapat disimpulkan bahwa tipe normal dan poison umumnya memiliki jumlah HP yang cukup banyak dibanding tipe lainnya.