Analisis Data

Pandas dan Seaborn

26 Februari 2024

DataFrame Berkas CSV

Mengecek *header* dan lima baris teratas DataFrame.

```
# Read CSV
import pandas as pd

df = pd.read_csv("Pokemon.csv")
df.head()
```

	#	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def
0	1	Bulbasaur	Grass	Poison	318	45	49	49	65	65
1	2	Ivysaur	Grass	Poison	405	60	62	63	80	80
2	3	Venusaur	Grass	Poison	525	80	82	83	100	100
3	3	VenusaurMega Venusaur	Grass	Poison	625	80	100	123	122	120
4	4	Charmander	Fire	NaN	309	39	52	43	60	50

Informasi DataFrame

Mengecek tipe data objek dan kolom-kolomnya.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 800 entries, 0 to 799
Data columns (total 13 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   #           800 non-null    int64
 1   Name        800 non-null    object
 2   Type 1      800 non-null    object
 3   Type 2      414 non-null    object
 4   Total       800 non-null    int64
 5   HP          800 non-null    int64
 6   Attack      800 non-null    int64
 7   Defense     800 non-null    int64
 8   Sp. Atk     800 non-null    int64
 9   Sp. Def     800 non-null    int64
 10  Speed       800 non-null    int64
 11  Generation  800 non-null    int64
 12  Legendary   800 non-null    bool
dtypes: bool(1), int64(9), object(3)
memory usage: 75.9+ KB
```

Deskripsi Kolom

Mengecek agregasi data-data numerik pada kolom.

```
df.describe()
```

	#	Total	HP	Attack	Defense	
count	800.000000	800.00000	800.000000	800.000000	800.000000	8
mean	362.813750	435.10250	69.258750	79.001250	73.842500	
std	208.343798	119.96304	25.534669	32.457366	31.183501	
min	1.000000	180.00000	1.000000	5.000000	5.000000	
25%	184.750000	330.00000	50.000000	55.000000	50.000000	
50%	364.500000	450.00000	65.000000	75.000000	70.000000	
75%	539.250000	515.00000	80.000000	100.000000	90.000000	
max	721.000000	780.00000	255.000000	190.000000	230.000000	1

Filter Data

Menampilkan beberapa kolom pilihan.

```
# Filtering DataFrame
df.filter(["Name","HP","Speed"])
```

	Name	HP	Speed
0	Bulbasaur	45	45
1	Ivysaur	60	60
2	Venusaur	80	80
3	VenusaurMega Venusaur	80	80
4	Charmander	39	65
...	...	...	...
795	Diancie	50	50
796	DiancieMega Diancie	50	110
797	HoopaHoopa Confined	80	70

Sortir Data

Menampilkan data secara urut berdasarkan kolom tertentu.

```
# Sorting
df.sort_values("Total", ascending=True)
```

	#	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk
206	191	Sunkern	Grass	NaN	180	30	30	30	30
322	298	Azurill	Normal	Fairy	190	50	20	40	20
446	401	Kricketot	Bug	NaN	194	37	25	41	25
288	265	Wurmple	Bug	NaN	195	45	45	35	20
16	13	Weedle	Bug	Poison	195	40	35	30	20
...	...	...	...	...	...	...	...	...	...
424	383	GroudonPrimal Groudon	Ground	Fire	770	100	180	160	150
422	382	KyogrePrimal Kyogre	Water	NaN	770	100	150	90	180

Grup Data

Mengelompokkan data berdasarkan nilai kolom tertentu.

```
# Groupby
group_type1 = df.groupby("Type 1")
group_type1.first()
```

Type 1	#	Name	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Sp
Bug	10	Caterpie	Flying	195	45	30	35	20	20	
Dark	197	Umbreon	Flying	525	95	65	110	60	130	
Dragon	147	Dratini	Flying	300	41	64	45	50	50	
Electric	25	Pikachu	Steel	320	35	55	40	50	50	
Fairy	35	Clefairy	Flying	323	70	45	48	60	65	
Fighting	56	Mankey	Psychic	305	40	80	35	35	45	
Fire	4	Charmander	Flying	309	39	52	43	60	50	
Flying	641	TornadusIncarnate Forme	Dragon	580	79	115	70	125	80	

Agregasi Data

Menampilkan agregasi data numerik berdasarkan kelompok data.

```
# Aggregation of Group
df.groupby("Type 1").agg(['mean','median'])
```

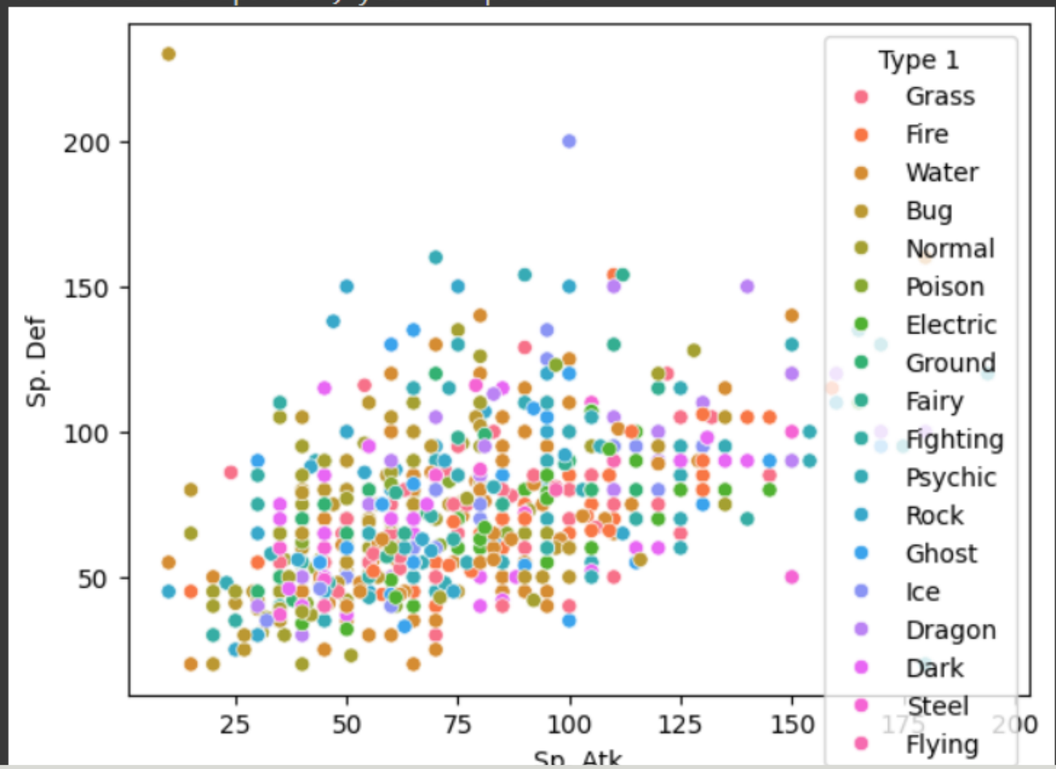
```
<ipython-input-13-f88ebeadb52a>:2: FutureWarning: ['Name', 'Type 2'] did not
  df.groupby("Type 1").agg(['mean','median'])
```

	#		Total		HP		Attack
	mean	median	mean	median	mean	median	mean
Type 1							
Bug	334.492754	291.0	378.927536	395.0	56.884058	60.0	70.971014
Dark	461.354839	509.0	445.741935	465.0	66.806452	65.0	88.387097
Dragon	474.375000	443.5	550.531250	600.0	83.312500	80.0	112.125000
Electric	363.500000	403.5	443.409091	477.5	59.795455	60.0	69.090909
Fairy	449.529412	669.0	413.176471	405.0	74.117647	78.0	61.529412
Fighting	363.851852	308.0	416.444444	455.0	69.851852	70.0	96.777778
Fire	327.403846	289.5	458.076923	482.0	69.903846	70.0	84.769231

Visualisasi Scatterplot

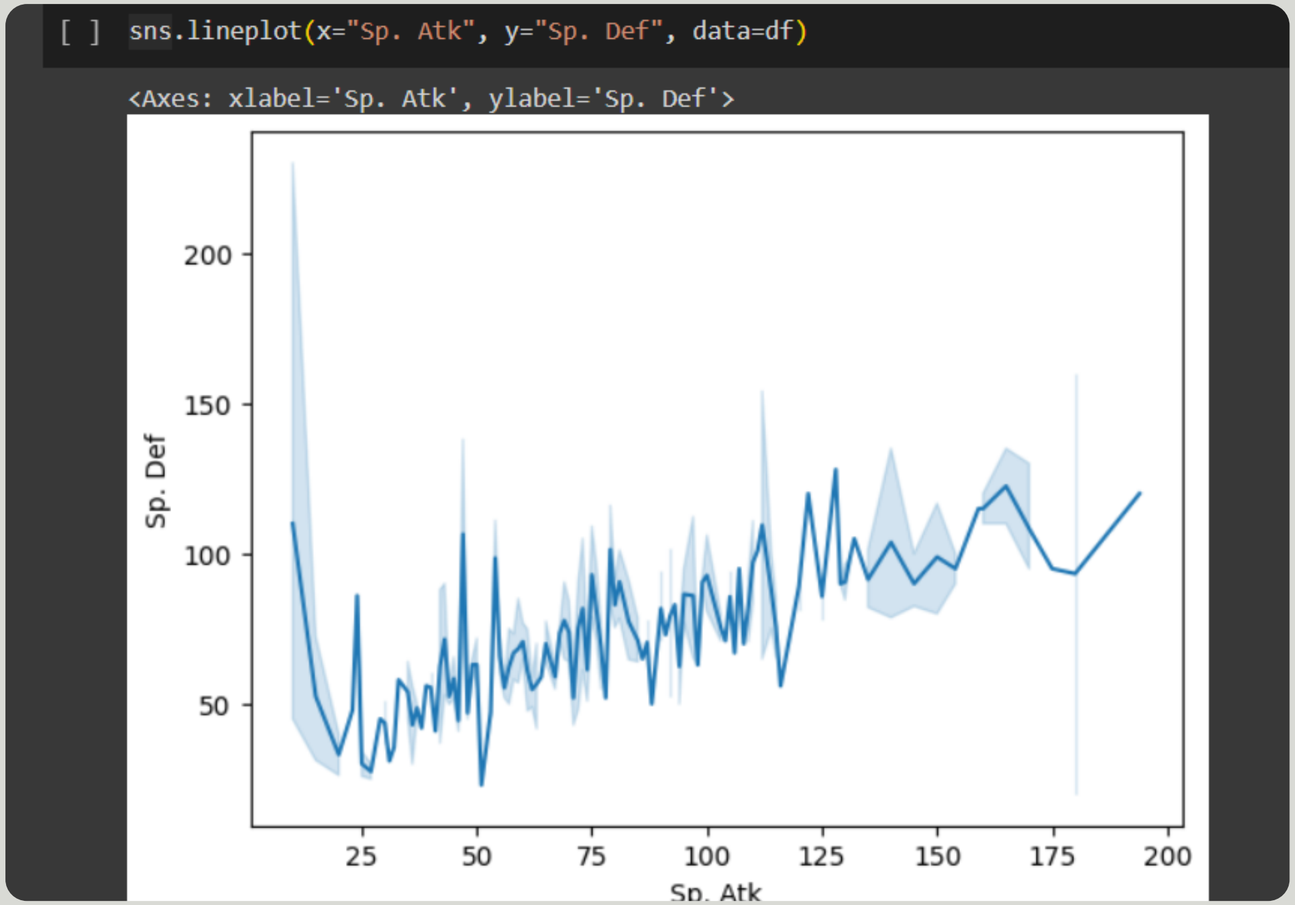
Menampilkan sebaran data dengan grafik scatter (sebar).

```
[ ] import seaborn as sns
    sns.scatterplot(data=df, x="Sp. Atk", y="Sp. Def", hue="Type 1")
```



Visualisasi Lineplot

Menampilkan hubungan data dengan diagram garis.

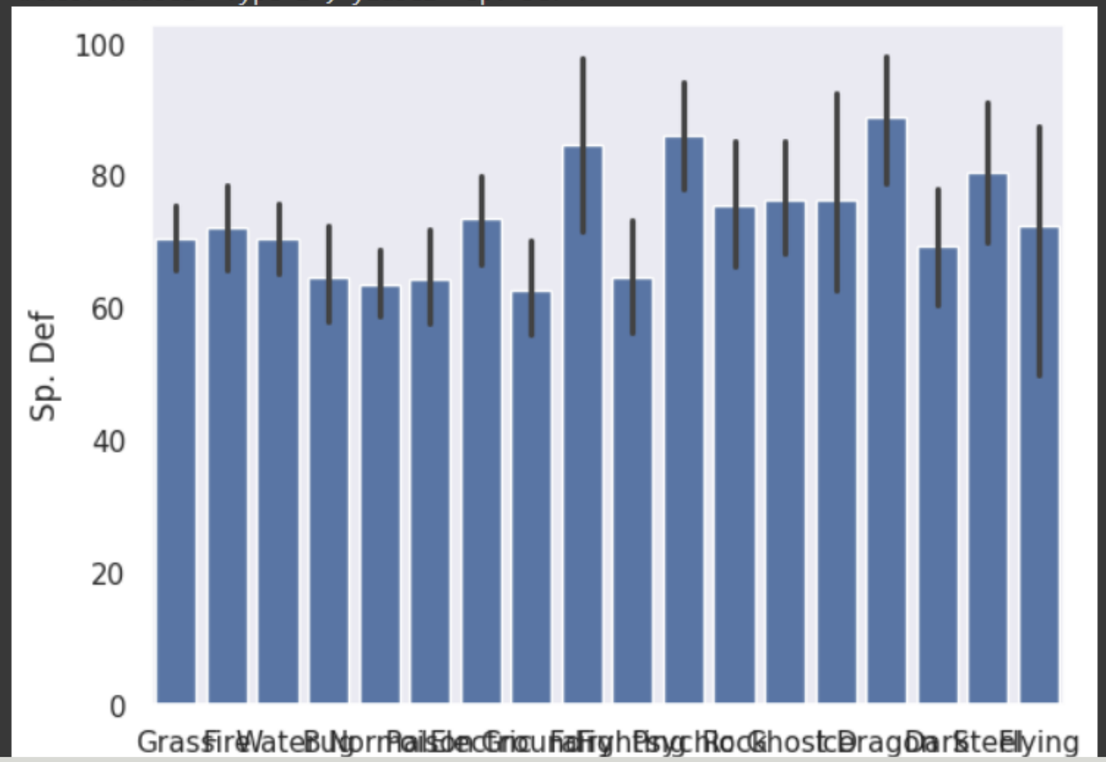


Visualisasi Barplot

Menampilkan hubungan data dengan diagram batang.

```
[ ]  sns.set(style="dark ")
     sns.barplot(data=df, x="Type 1", y="Sp. Def")
```

```
<Axes: xlabel='Type 1', ylabel='Sp. Def'>
```



Dukungan

Berkas CSV: bit.ly/data-pokemon-dsf

Referensi: https://www.geeksforgeeks.org/data-analysis-with-python/