

POKEMON DATASET ANALYSIS



Portofolio
Muhammad Fajar Jati Permana



UNDERSTANDING DATA



Dataset Pokemon ini berisi informasi tentang berbagai karakter Pokemon dari berbagai generasi. Data ini mencakup berbagai atribut yang menggambarkan karakteristik setiap Pokemon, seperti jenis, statistik dasar, dan apakah Pokemon tersebut adalah Legendary atau tidak. Dataset bisa diambil melalui link berikut bit.ly/data-pokemon-dsf





UNDERSTANDING DATA (TIAP KOLOM)

Name	Nama Pokemon.
Type 1	Jenis utama Pokemon.
Type 2	Jenis sekunder Pokemon (jika ada).
Total	Jumlah total dari semua kekuatan dasar.
HP (Hit Points)	Jumlah poin kesehatan Pokemon.
Attack	Nilai serangan fisik Pokemon.
Sp. Atk (Special Attack)	Nilai serangan khusus Pokemon.
Sp. Def (Special Defense)	Nilai pertahanan khusus Pokemon.
Speed	Kecepatan Pokemon.

UNDERSTANDING DATA (TIAP KOLOM)

Generation	Generasi Pokemon (1 sampai 6).
Legendary	Apakah Pokemon tersebut termasuk kategori Legendary (True atau False).

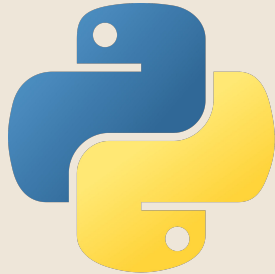
TUJUAN

- **Mengidentifikasi Distribusi Statistik**, Memahami bagaimana distribusi berbagai statistik dasar seperti HP, Attack, Defense, dll.
- **Membandingkan Jenis Pokemon**, Melihat bagaimana tipe Pokemon mempengaruhi statistik dasar.
- **Mengetahui Pengaruh Generasi**, Melihat apakah ada perubahan atau tren dalam statistik Pokemon dari generasi ke generasi.
- **Identifikasi Pokemon Legendary**, Memahami karakteristik khusus dari Pokemon Legendary dibandingkan dengan yang non-Legendary.

TOOLS

Bahasa
Pemrograman

PYTHON



IDE (Integrated
Development Environment)

**GOOGLE
COLABORATORY**



DATASET AWAL

Gambaran awal dari
dataset Pokemon

```
import pandas as pd
df = pd.read_csv("Pokemon.csv")
df
```

	#	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary
0	1	Bulbasaur	Grass	Poison	318	45	49	49	65	65	45	1	False
1	2	Ivysaur	Grass	Poison	405	60	62	63	80	80	60	1	False
2	3	Venusaur	Grass	Poison	525	80	82	83	100	100	80	1	False
3	3	VenusaurMega Venusaur	Grass	Poison	625	80	100	123	122	120	80	1	False
4	4	Charmander	Fire	NaN	309	39	52	43	60	50	65	1	False
...
795	719	Diancie	Rock	Fairy	600	50	100	150	100	150	50	6	True
796	719	DiancieMega Diancie	Rock	Fairy	700	50	160	110	160	110	110	6	True
797	720	HoopaHoopa Confined	Psychic	Ghost	600	80	110	60	150	130	70	6	True
798	720	HoopaHoopa Unbound	Psychic	Dark	680	80	160	60	170	130	80	6	True
799	721	Volcanion	Fire	Water	600	80	110	120	130	90	70	6	True

800 rows x 13 columns

DATASET DESCRIBE

Penjelasan dari dataset Pokemon

```
# gambaran singkat dari dataset  
df.describe()
```

	#	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation
count	800.000000	800.000000	800.000000	800.000000	800.000000	800.000000	800.000000	800.000000	800.000000
mean	362.813750	435.10250	69.258750	79.001250	73.842500	72.820000	71.902500	68.277500	3.32375
std	208.343798	119.96304	25.534669	32.457366	31.183501	32.722294	27.828916	29.060474	1.66129
min	1.000000	180.00000	1.000000	5.000000	5.000000	10.000000	20.000000	5.000000	1.00000
25%	184.750000	330.00000	50.000000	55.000000	50.000000	49.750000	50.000000	45.000000	2.00000
50%	364.500000	450.00000	65.000000	75.000000	70.000000	65.000000	70.000000	65.000000	3.00000
75%	539.250000	515.00000	80.000000	100.000000	90.000000	95.000000	90.000000	90.000000	5.00000
max	721.000000	780.00000	255.000000	190.000000	230.000000	194.000000	230.000000	180.000000	6.00000

DATASET INFO

```
# info tentang dataset  
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
```

Berikut adalah penjelasan singkat mengenai dataset Pokemon dari `df.info()`:

- Jumlah Baris dan Kolom: Dataset ini terdiri dari 800 entri (baris) dan 13 kolom.
- Dataset ini mengandung tiga tipe data: integer (`int64`), objek (`object`), dan boolean (`bool`).
- Kolom Type 2 memiliki nilai yang tidak lengkap, hanya terisi sebanyak 414 dari 800 entri. Kolom-kolom lainnya lengkap terisi tanpa nilai kosong.

CLeaning DATA



1. Mengisi nilai 'unknown' pada kolom 'Type 2' yang terlihat kosong

```
# cek kekosongan (null) dari dataset
df.isnull().sum()
```

```
#
Name                0
Type 1              0
Type 2             386
Total              0
HP                 0
Attack             0
Defense            0
Sp. Atk            0
Sp. Def            0
Speed              0
Generation          0
Legendary           0
dtype: int64
```

```
# Mengisi nilai yang hilang pada kolom 'Type 2' dengan 'Unknown'
df['Type 2'].fillna('Unknown', inplace=True)
df
```



CLeaning DATA



	#	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary
0	1	Bulbasaur	Grass	Poison	318	45	49	49	65	65	45	1	False
1	2	Ivysaur	Grass	Poison	405	60	62	63	80	80	60	1	False
2	3	Venusaur	Grass	Poison	525	80	82	83	100	100	80	1	False
3	3	VenusaurMega Venusaur	Grass	Poison	625	80	100	123	122	120	80	1	False
4	4	Charmander	Fire	Unknown	309	39	52	43	60	50	65	1	False
...
795	719	Diancie	Rock	Fairy	600	50	100	150	100	150	50	6	True
796	719	DiancieMega Diancie	Rock	Fairy	700	50	160	110	160	110	110	6	True
797	720	HoopaHoop Confined	Psychic	Ghost	600	80	110	60	150	130	70	6	True
798	720	HoopaHoop Unbound	Psychic	Dark	680	80	160	60	170	130	80	6	True
799	721	Volcanion	Fire	Water	600	80	110	120	130	90	70	6	True

800 rows x 13 columns



CLeaning DATA



2. Menghapus kolom '#' dan menjadikan index sebagai kolom dengan nama 'num' dan diawali dari satu

```
# Menghapus kolom '#'
df = df.drop('#', axis=1)

# Mereset index dan menjadikannya kolom 'num'
df = df.reset_index()

# Mengganti nama kolom 'index' menjadi 'num'
df = df.rename(columns={'index': 'num'})

# Menambahkan 1 ke kolom 'num' agar dimulai dari 1
df['num'] = df['num'] + 1

# Mengatur kolom 'num' sebagai index
df.set_index('num', inplace=True)

df
```



CLeaning DATA



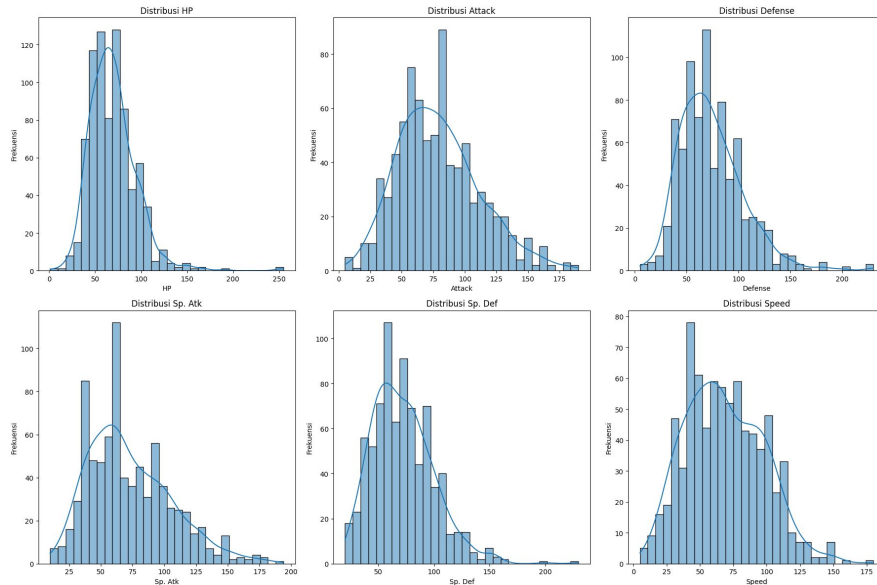
	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary
num												
1	Bulbasaur	Grass	Poison	318	45	49	49	65	65	45	1	False
2	Ivysaur	Grass	Poison	405	60	62	63	80	80	60	1	False
3	Venusaur	Grass	Poison	525	80	82	83	100	100	80	1	False
4	VenusaurMega Venusaur	Grass	Poison	625	80	100	123	122	120	80	1	False
5	Charmander	Fire	Unknown	309	39	52	43	60	50	65	1	False
...
796	Diancie	Rock	Fairy	600	50	100	150	100	150	50	6	True
797	DiancieMega Diancie	Rock	Fairy	700	50	160	110	160	110	110	6	True
798	HoopaaHoopaa Confined	Psychic	Ghost	600	80	110	60	150	130	70	6	True
799	HoopaaHoopaa Unbound	Psychic	Dark	680	80	160	60	170	130	80	6	True
800	Volcanion	Fire	Water	600	80	110	120	130	90	70	6	True

800 rows × 12 columns



DATA ANALYSIS

MENGIDENTIFIKASI DISTRIBUSI STATISTIK

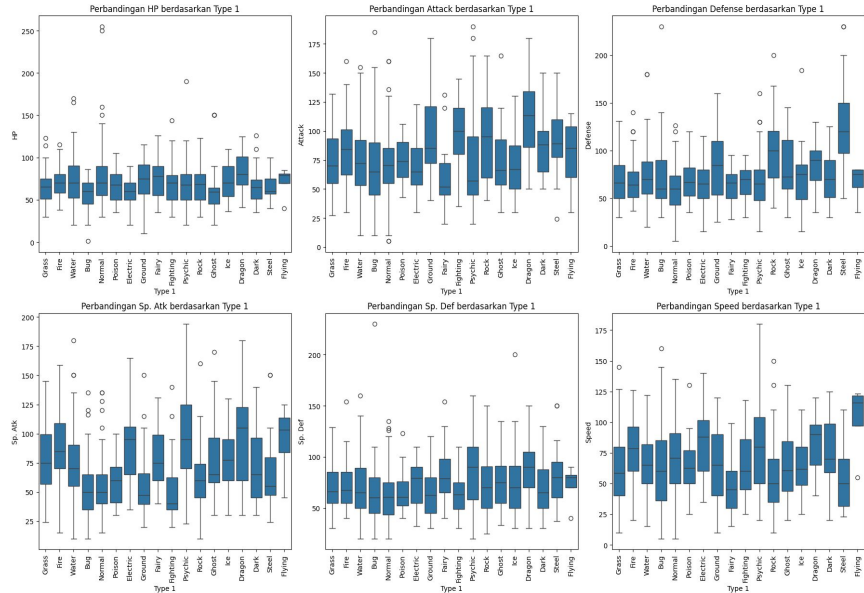


Sebagian besar atribut memiliki distribusi yang mirip, di mana sebagian besar nilai berada di sekitar rata-rata (antara 50 hingga 100), tetapi ada beberapa Pokemon dengan nilai yang sangat tinggi, menunjukkan mereka memiliki kemampuan luar biasa di area tersebut.

DATA ANALYSIS

MEMBANDINGKAN JENIS POKEMON

Secara umum, kita bisa melihat bahwa tipe-tipe tertentu memiliki kecenderungan statistik dasar yang berbeda. Tipe Dragon seringkali menonjol dengan nilai yang tinggi pada berbagai statistik, sementara tipe Bug cenderung memiliki nilai yang lebih rendah. Hal ini dapat membantu dalam strategi pertempuran dengan memilih tipe Pokemon yang sesuai dengan kebutuhan statistik tertentu.



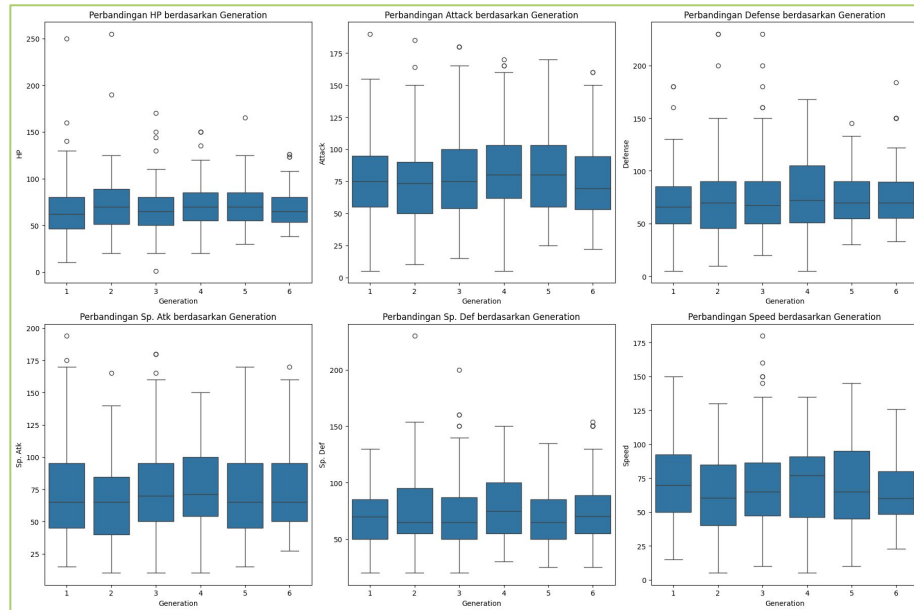
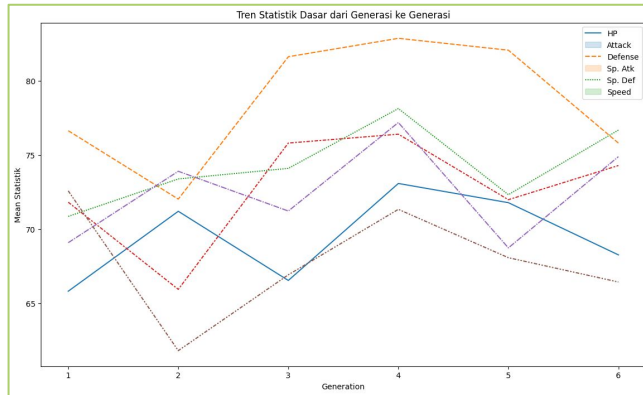
DATA ANALYSIS

Grafik tersebut merupakan perbandingan statistik dasar Pokemon berdasarkan Type 1. Berikut adalah penjelasan singkat dari masing-masing grafik Membandingkan Jenis Pokemon:

- **HP:** Grafik menunjukkan bahwa tipe Pokemon seperti Dragon, Steel, dan Flying memiliki rentang HP yang lebih tinggi dibandingkan dengan tipe lainnya. Sebaliknya, tipe Bug, Ghost, dan Grass cenderung memiliki rentang HP yang lebih rendah.
- **Attack:** Grafik ini menunjukkan bahwa tipe Pokemon Dragon, Fighting, dan Ground memiliki rentang Attack yang lebih tinggi. Tipe Fairy, Bug, dan Psychic memiliki rentang Attack yang lebih rendah.
- **Defense:** Tipe Pokemon seperti Steel, Rock, dan Dragon memiliki rentang Defense yang lebih tinggi. Tipe Fire, Electric, dan Normal memiliki rentang Defense yang lebih rendah.
- **Sp. Atk:** Tipe Pokemon Dragon, Psychic, dan Ghost memiliki rentang Sp. Atk yang lebih tinggi. Tipe Bug, Grass, dan Poison memiliki rentang Sp. Atk yang lebih rendah.
- **Sp. Def:** Tipe Pokemon Dragon, Steel, dan Flying memiliki rentang Sp. Def yang lebih tinggi. Tipe Bug, Fighting, dan Grass memiliki rentang Sp. Def yang lebih rendah.
- **Speed:** Tipe Pokemon Flying, Electric, dan Psychic memiliki rentang Speed yang lebih tinggi. Tipe Rock, Steel, dan Ghost memiliki rentang Speed yang lebih rendah.

DATA ANALYSIS

MENGETAHUI PENGARUH GENERASI



DATA ANALYSIS

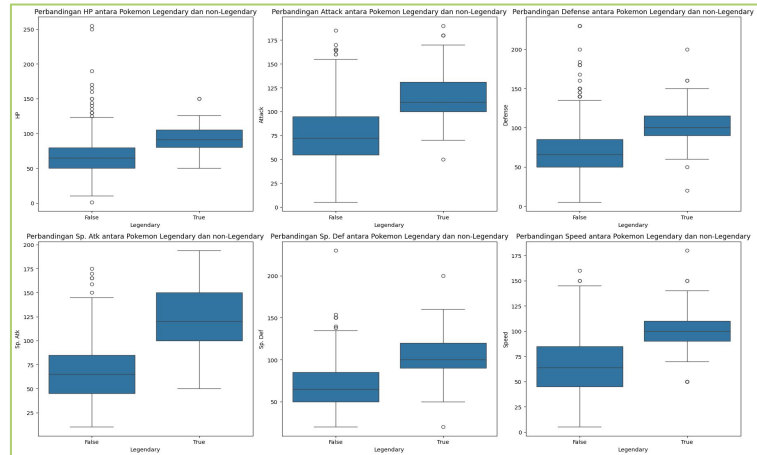
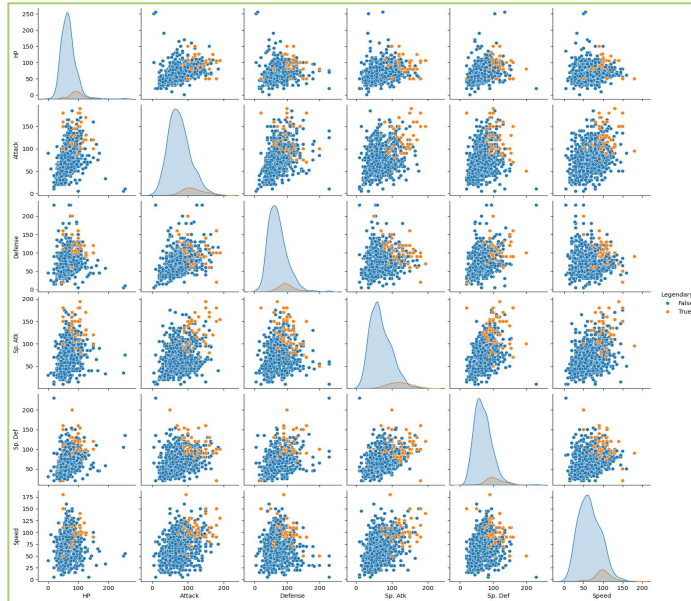
Grafik di atas menunjukkan tren statistik dasar Pokemon dari generasi ke generasi. Berikut penjelasannya:

- **HP:** Rata-rata HP cenderung meningkat pada Generasi 3 dan 4, kemudian menurun pada Generasi 5 dan 6.
- **Attack:** Rata-rata Attack meningkat secara signifikan dari Generasi 1 hingga 4, dengan puncak di Generasi 4, kemudian menurun pada Generasi 5 dan sedikit meningkat lagi pada Generasi 6.
- **Defense:** Rata-rata Defense meningkat dari Generasi 1 hingga 3, menurun pada Generasi 4, dan kembali meningkat pada Generasi 5 dan 6.
- **Sp. Atk:** Rata-rata Sp. Atk meningkat dari Generasi 1 hingga 4, dengan puncak di Generasi 4, kemudian menurun pada Generasi 5 dan sedikit meningkat lagi pada Generasi 6.
- **Sp. Def:** Rata-rata Sp. Def cenderung stabil dari Generasi 1 hingga 4, dengan sedikit penurunan pada Generasi 2, dan meningkat kembali pada Generasi 5 dan 6.
- **Speed:** Rata-rata Speed cenderung menurun dari Generasi 1 hingga 3, kemudian meningkat pada Generasi 4, menurun pada Generasi 5, dan sedikit meningkat lagi pada Generasi 6.

Secara umum, ada peningkatan yang signifikan pada beberapa statistik di Generasi 4, sementara Generasi 5 cenderung menunjukkan penurunan pada beberapa statistik. Tren ini menunjukkan variasi dalam desain dan keseimbangan Pokemon dari satu generasi ke generasi berikutnya.

DATA ANALYSIS

IDENTIFIKASI POKEMON LEGENDARY



DATA ANALYSIS

Box plot menunjukkan perbandingan statistik dasar (HP, Attack, Defense, Sp. Atk, Sp. Def, Speed) antara Pokemon Legendary dan non-Legendary.

- HP, Attack, Defense, Sp. Atk, Sp. Def, Speed: Pokemon Legendary umumnya memiliki nilai median yang lebih tinggi dibandingkan dengan Pokemon non-Legendary di semua statistik dasar. Ini menunjukkan bahwa Pokemon Legendary secara keseluruhan memiliki statistik yang lebih tinggi dibandingkan Pokemon non-Legendary.

Scatter plot dalam pair plot menunjukkan distribusi statistik dasar antara Pokemon Legendary dan non-Legendary.

- Distribusi: Pokemon Legendary (ditandai dengan warna oranye) cenderung berada di bagian atas atau kanan grafik untuk setiap pasangan statistik dasar, menandakan bahwa mereka memiliki statistik yang lebih tinggi dibandingkan Pokemon non-Legendary (ditandai dengan warna biru).
- Korelasi: Terdapat korelasi positif antara beberapa pasangan statistik dasar, seperti antara Attack dan Sp. Atk, Defense dan Sp. Def. Ini menunjukkan bahwa Pokemon dengan statistik dasar yang tinggi cenderung memiliki statistik tinggi di area lainnya juga.

Secara keseluruhan, grafik-grafik ini menunjukkan bahwa Pokemon Legendary memiliki statistik dasar yang lebih tinggi dibandingkan dengan Pokemon non-Legendary di semua aspek yang diukur.

ADDITIONAL

SOURCE CODE



github.com/mfajarjati/pokemon_analysis

HASIL CLEANING



bit.ly/pokemon_cleaned

THANK
YOU

