/01

Pokemon Data Analysis

Problem Statement

1.   How is the data distribution
2.   How is the amount of legendary data on each
      type of pokemon
3.   How is the legendary distribution in
      generation
4.   what are the best pokemon?
5.   how is the correlation between features?

/02

Import data


```
[ ]  df = pd.read_csv("/content/drive/MyDrive/KULIAH/Pribadi/Dibimbing/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 × 13 columns

Data Understanding

```
[ ] 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
```

The dataset consists of 800 data with 13 columns or features

Pokemon.csv

Name
Type 1
Type 2
Total
HP
Attack
Defense
Sp. Atk
Sp. Def
Speed
Generation
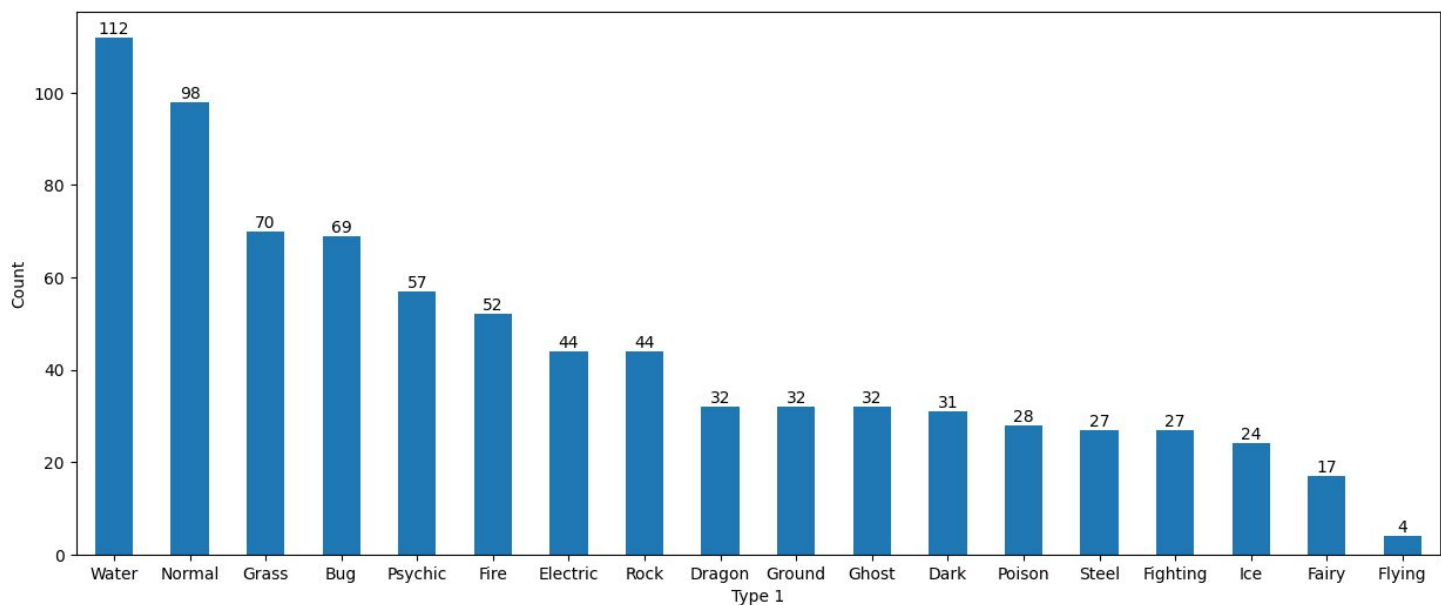Legendary

STATISTICAL ANALYSIS

```
[ ] df.describe()
```

	#	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation
count	800.000000	800.00000	800.000000	800.000000	800.000000	800.000000	800.000000	800.000000	800.00000
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

DISTRIBUTION
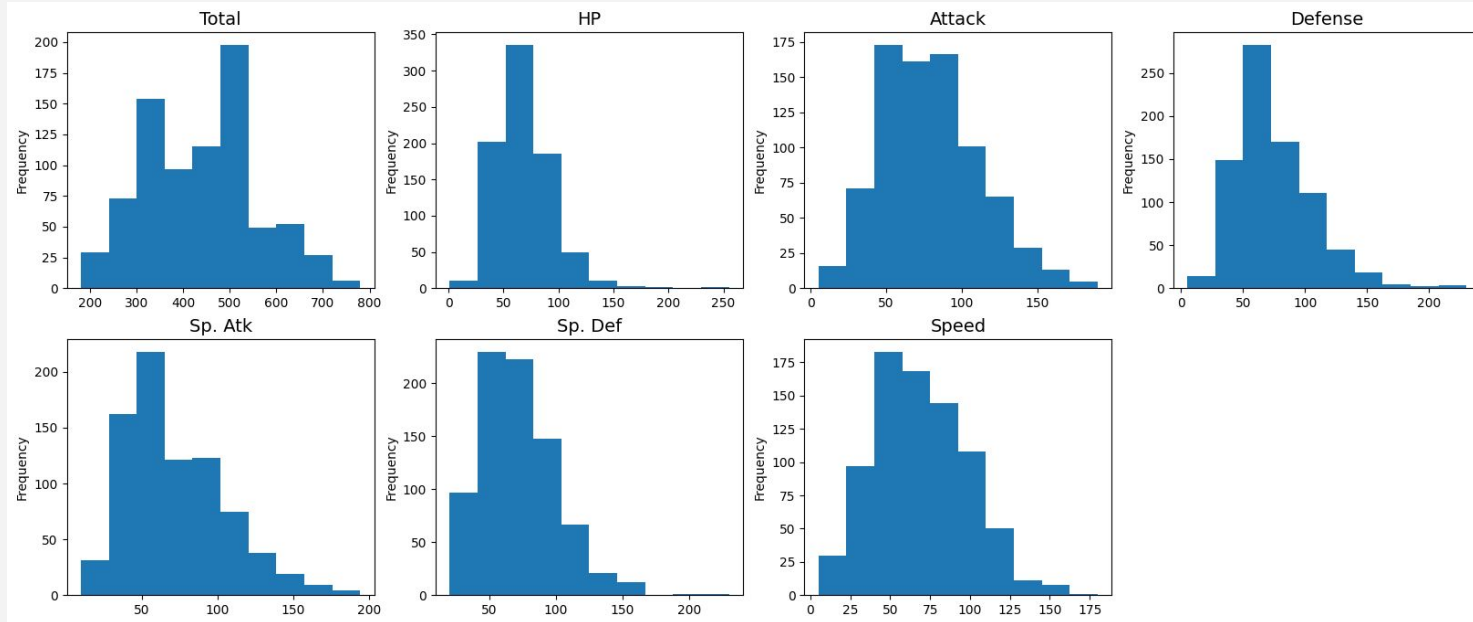
TYPE 1 OF POKEMON

The pokemon types with the highest number are Water, Normal, and Grass while the least number are Ice, Fairy, and Flying, respectively.

TYPE 1 OF POKEMON

Type 1 Legendary

Psychic **14**
Dragon **12**
Fire **5**
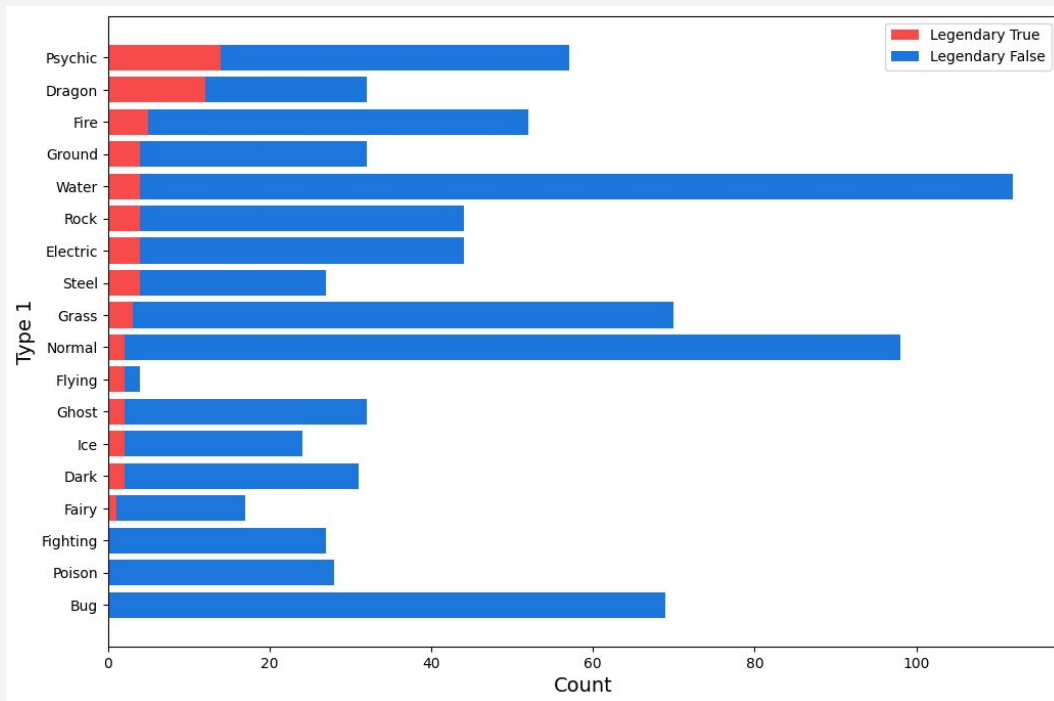Ground 4
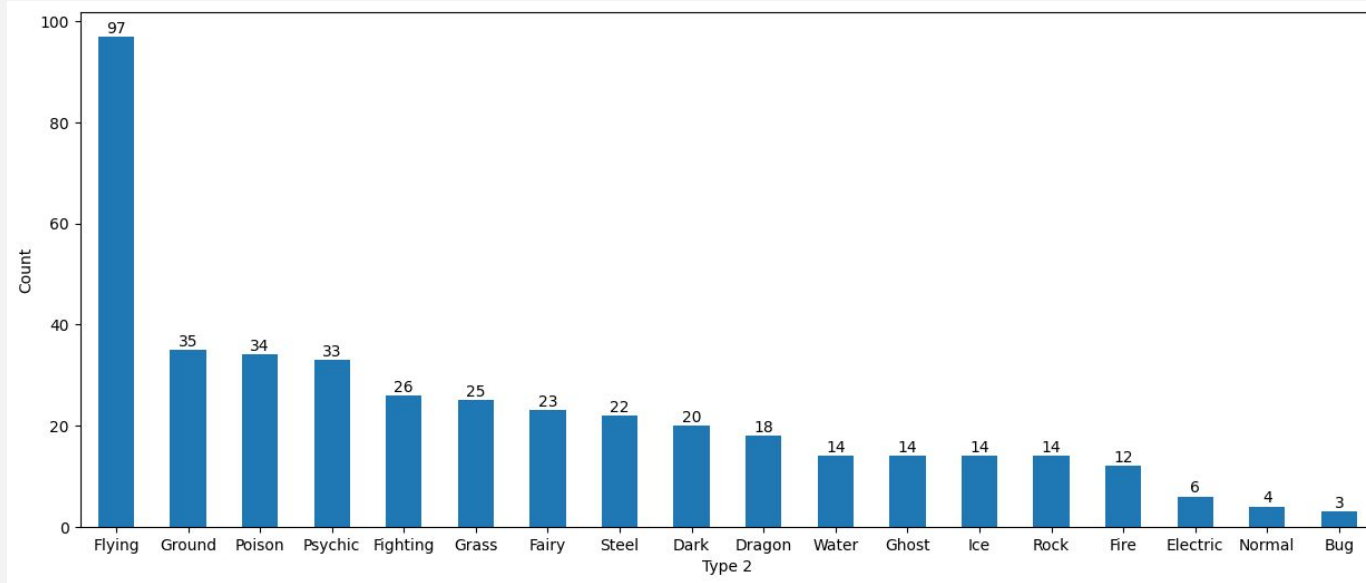Water 4
Rock 4
Electric 4
Steel 4
Grass 3
Normal 2
Flying 2
Ghost 2
Ice 2
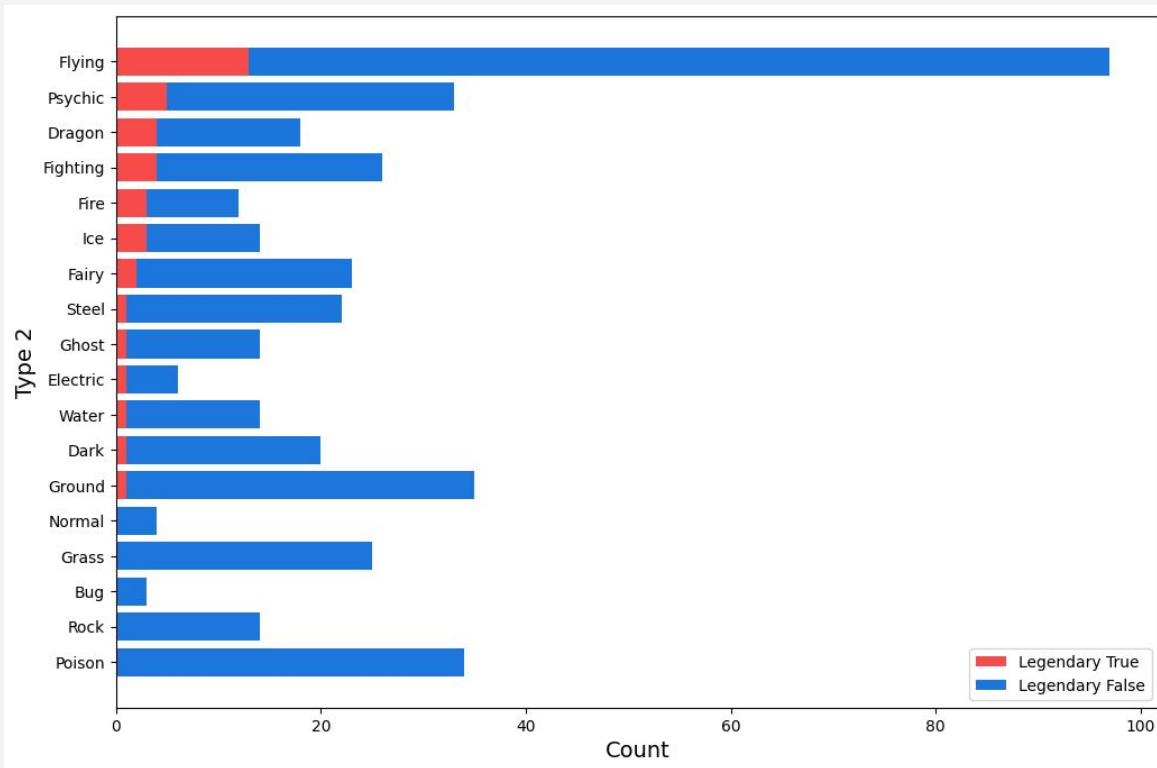Dark 2
Fairy 1

TYPE 2 OF POKEMON

The types of pokemon with the highest number are Flying, Ground, Poison while the least number is Electric, Normal, Bug, respectively.
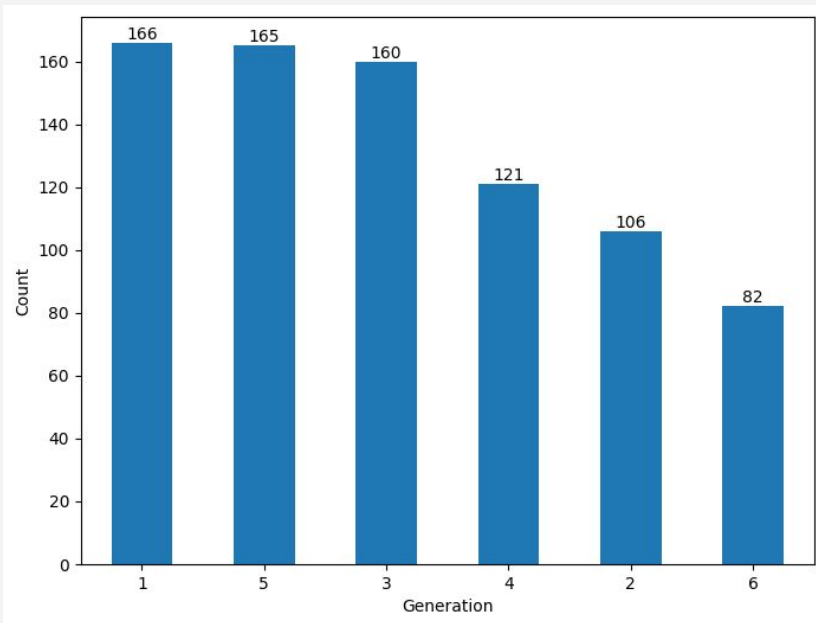
TYPE 2 OF POKEMON

Type 2 Legendary

Flying **13**
Psychic **5**
Dragon **4**
Fighting 4
Fire 3
Ice 3
Fairy 2
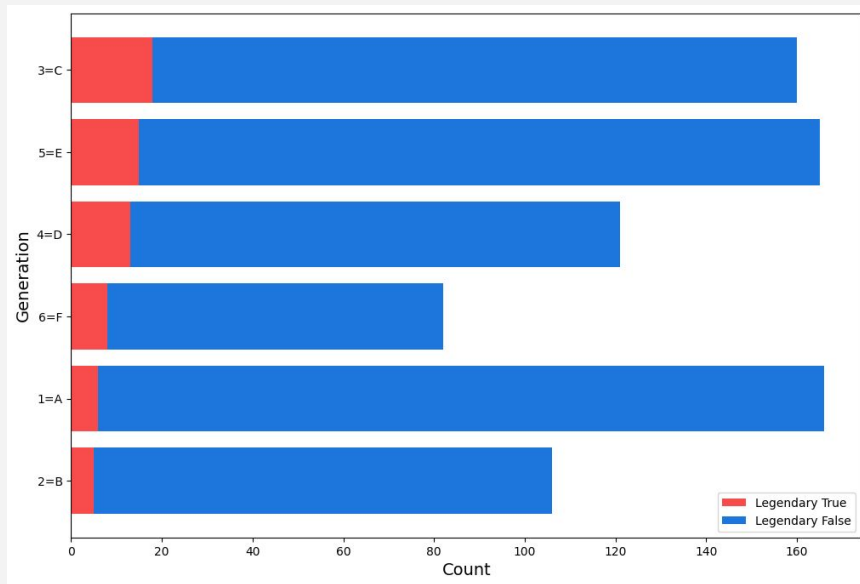Steel 1
Ghost 1
Electric 1
Water 1
Dark 1
Ground 1

GENERATION POKEMON

The most consecutive generations are generation 1, 5, 3 with values 166, 165, 160.

GENERATION POKEMON BY LEGENDARY

Gen 3 = **18**
Gen 5 = **15**
Gen 4 = **13**
Gen 6 = **8**
Gen 1 = **6**
Gen 2 = **5**

/12

LEGENDARY
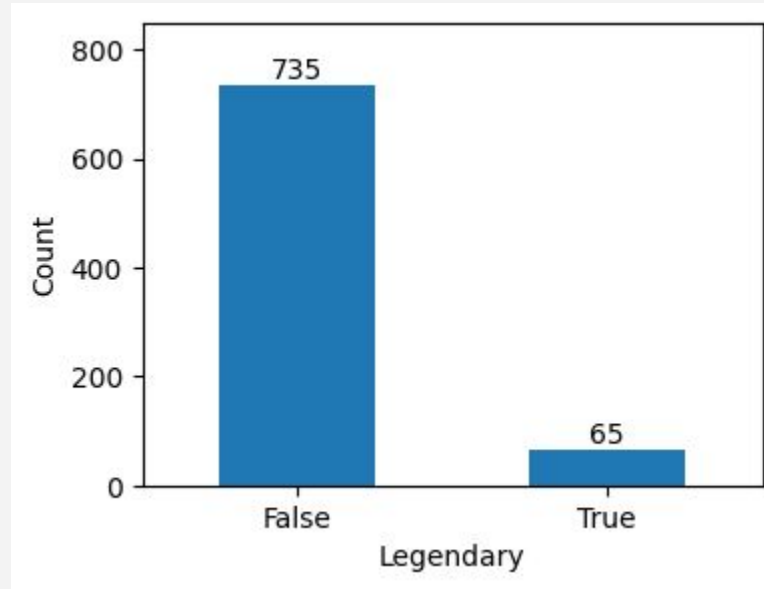
Legendary : **65**
Non Legendary : **735**

Data Science | Dibimbing.id

BEST POKEMON BASED ON FEATURE

HP	Attack	Sp. Atk	Sp. Def	Speed
Blissey	Mega Mewtwo X	Mega Mewtwo Y	Shuckle	Deoxys

/14

CORRELATION

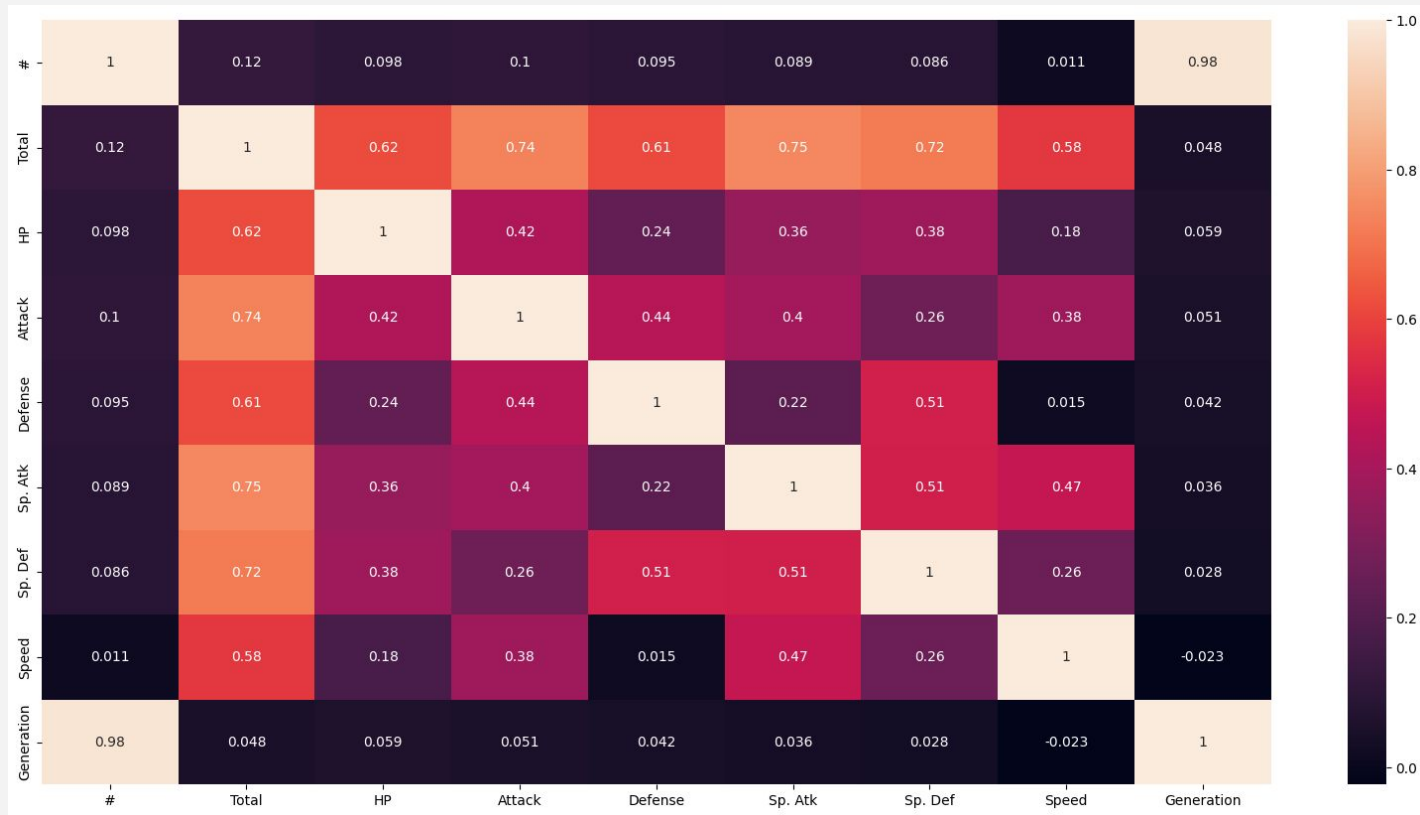|	#	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation
#	1	0.12	0.098	0.1	0.095	0.089	0.086	0.011	0.98
Total	0.12	1	0.62	0.74	0.61	0.75	0.72	0.58	0.048
HP	0.098	0.62	1	0.42	0.24	0.36	0.38	0.18	0.059
Attack	0.1	0.74	0.42	1	0.44	0.4	0.26	0.38	0.051
Defense	0.095	0.61	0.24	0.44	1	0.22	0.51	0.015	0.042
Sp. Atk	0.089	0.75	0.36	0.4	0.22	1	0.51	0.47	0.036
Sp. Def	0.086	0.72	0.38	0.26	0.51	0.51	1	0.26	0.028
Speed	0.011	0.58	0.18	0.38	0.015	0.47	0.26	1	-0.023
Generation	0.98	0.048	0.059	0.051	0.042	0.036	0.028	-0.023	1
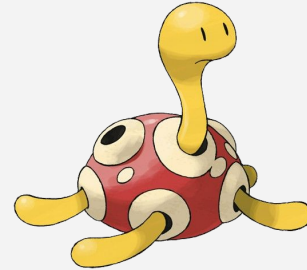
CORRELATION

1. **HP and Attack** (0.42) => **Medium Positive**, indicating that pokemon with high HP tend to have high attacks
2. **HP and Speed** (0.18) => **Low Positive**, indicating that pokemon with high HP tend to have high speed but not strength
3. **Attack and Defense** (0.44) => **Medium Positive**, indicating that pokemon with high Attack tend to have high defence
4. **Attack and Sp Def** (0.26) => **Low Positive**, indicating that pokemon with high attack have high defence specials, but are not strong.
5. **Speed and Defence** (0.015) => **Low Positive**, indicating that pokemon with high speed then have weak defence
6. **Defense and Sp. Def** (0.51) => **Medium Positive**, indicating that pokemon with high defence tend to have high defence specials
7. **Sp Attack and Speed** (0.47) => **Medium Positive**, indicating that pokemon with attack specials tend to have high speed

Thank you!

Access Code **s.id/23sXp**

✉ afanfanani03@gmail.com

in linkedin.com/in/afanfanani03

github.com/fanani03

Data Science | Dibimbing.id