



ANALISIS DATA POKEMON DATASET

Background

Nama : Quraini Septy Wardhani

Pendidikan : S1-Statistika

Pengalaman Kerja:

- Data Analyst - Freelance (Februari 2023-Sekarang)
- Staf Keuangan - Arti Bakery (Mei 2022- Desember 2023)
- Customer Care - PT. Telkom Indonesia (Januari 2020-Februari 2020)

DATASET

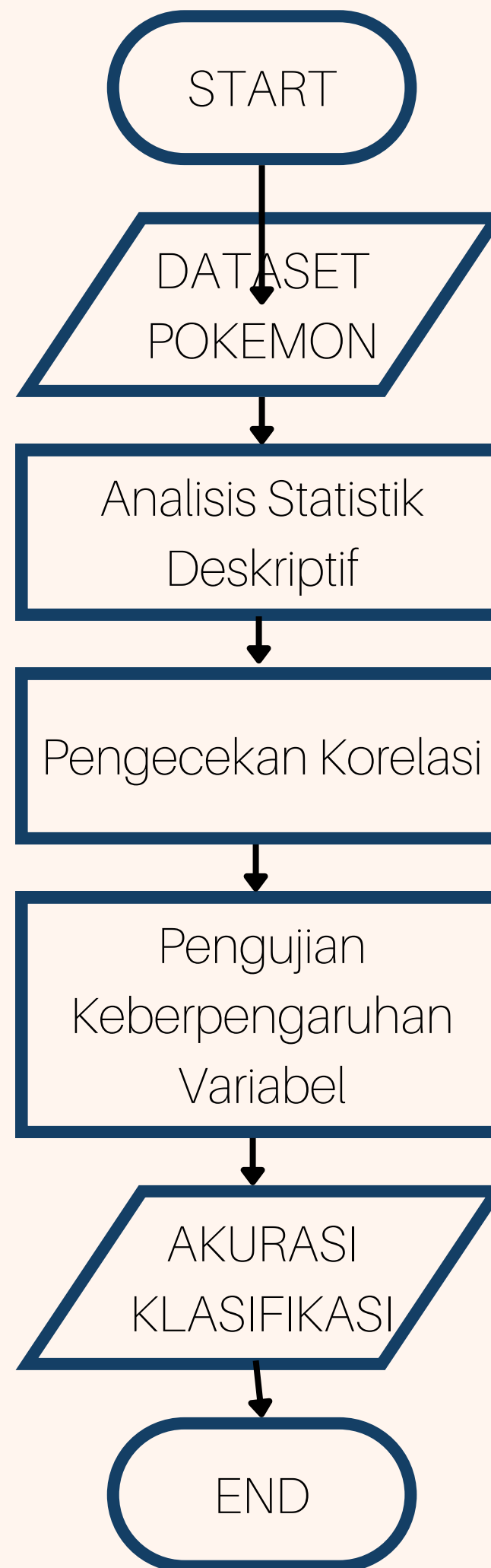
Dataset POKEMON terdiri dari 12 variabel data dengan 800 data

1. Name : Nama Pokemon
2. Type 1 : Tipe primer pokemon
3. Type 2 : Tipe sekunder pokemon
4. Total : Total skor dari attack, defense, sp. atk, sp. def, dan speed
5. HP : HP dasar
6. Attack : Skor serangan dasar pokemon
7. Defense : Skor pertahanan dasar pokemon
8. Sp. Atk : Skor serangan khusus pokemon
9. Sp. Def : Skor pertahanan khusus pokemon
10. Speed : Kecepatan
11. Generation : Generasi pokemon diperkenalkan
12. Legendary : Legend tidaknya pokemon

Tujuan

1. Melakukan analisis statistik deskriptif,
2. Pengujian keberpengaruhan masing-masing variabel terhadap legend tidaknya pokemon,
3. Pengklasifikasian dari dataset.

Alur Kerja (FLOWCHART)

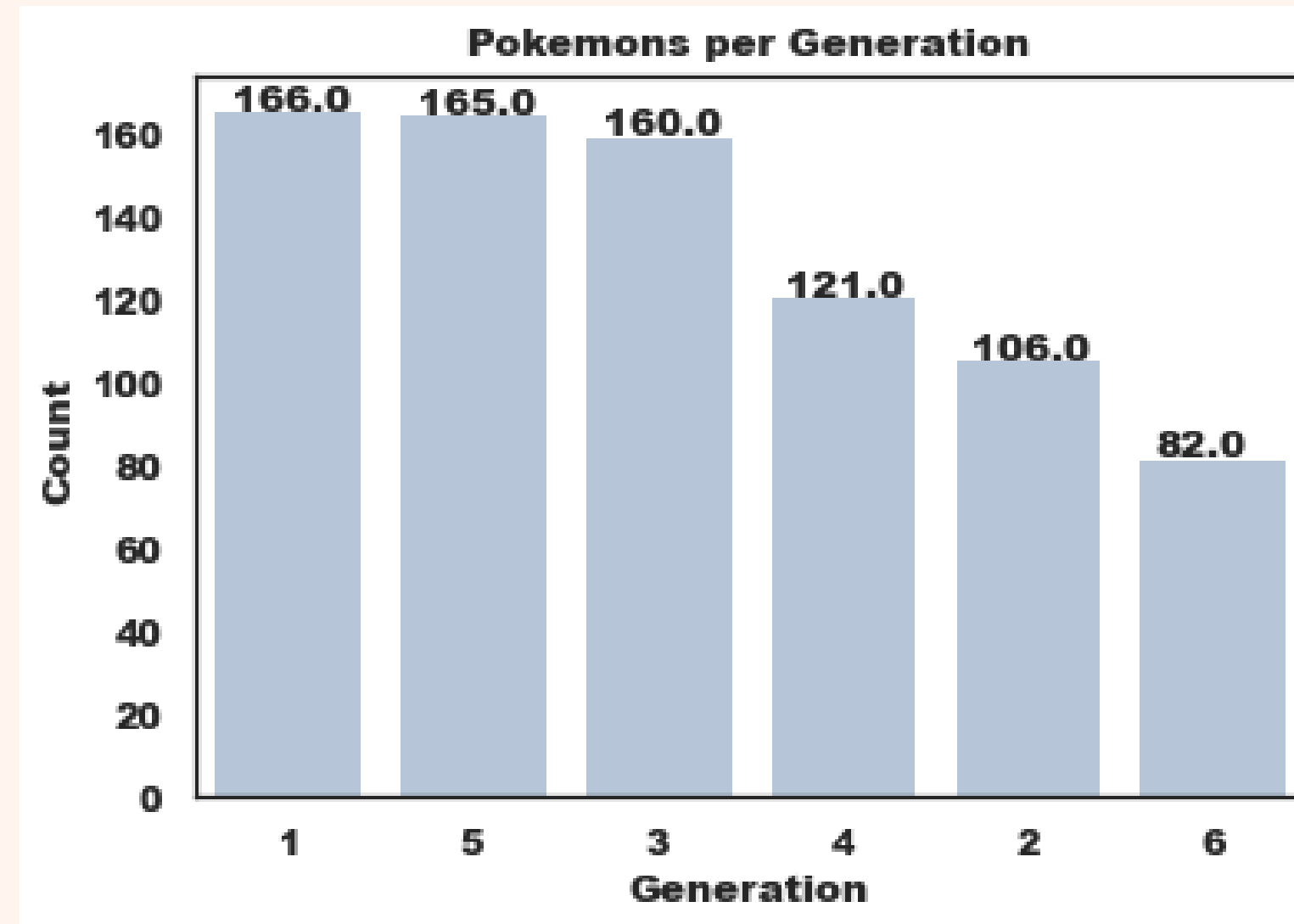


Statistik Deskriptif

	count	mean	std	min	25%	50%	75%	max
Total	800.0	435.10250	119.963040	180.0	330.00	450.0	515.0	780.0
HP	800.0	69.25875	25.534669	1.0	50.00	65.0	80.0	255.0
Attack	800.0	79.00125	32.457366	5.0	55.00	75.0	100.0	190.0
Defense	800.0	73.84250	31.183501	5.0	50.00	70.0	90.0	230.0
Sp. Atk	800.0	72.82000	32.722294	10.0	49.75	65.0	95.0	194.0
Sp. Def	800.0	71.90250	27.828916	20.0	50.00	70.0	90.0	230.0
Speed	800.0	68.27750	29.060474	5.0	45.00	65.0	90.0	180.0
Generation	800.0	3.32375	1.661290	1.0	2.00	3.0	5.0	6.0

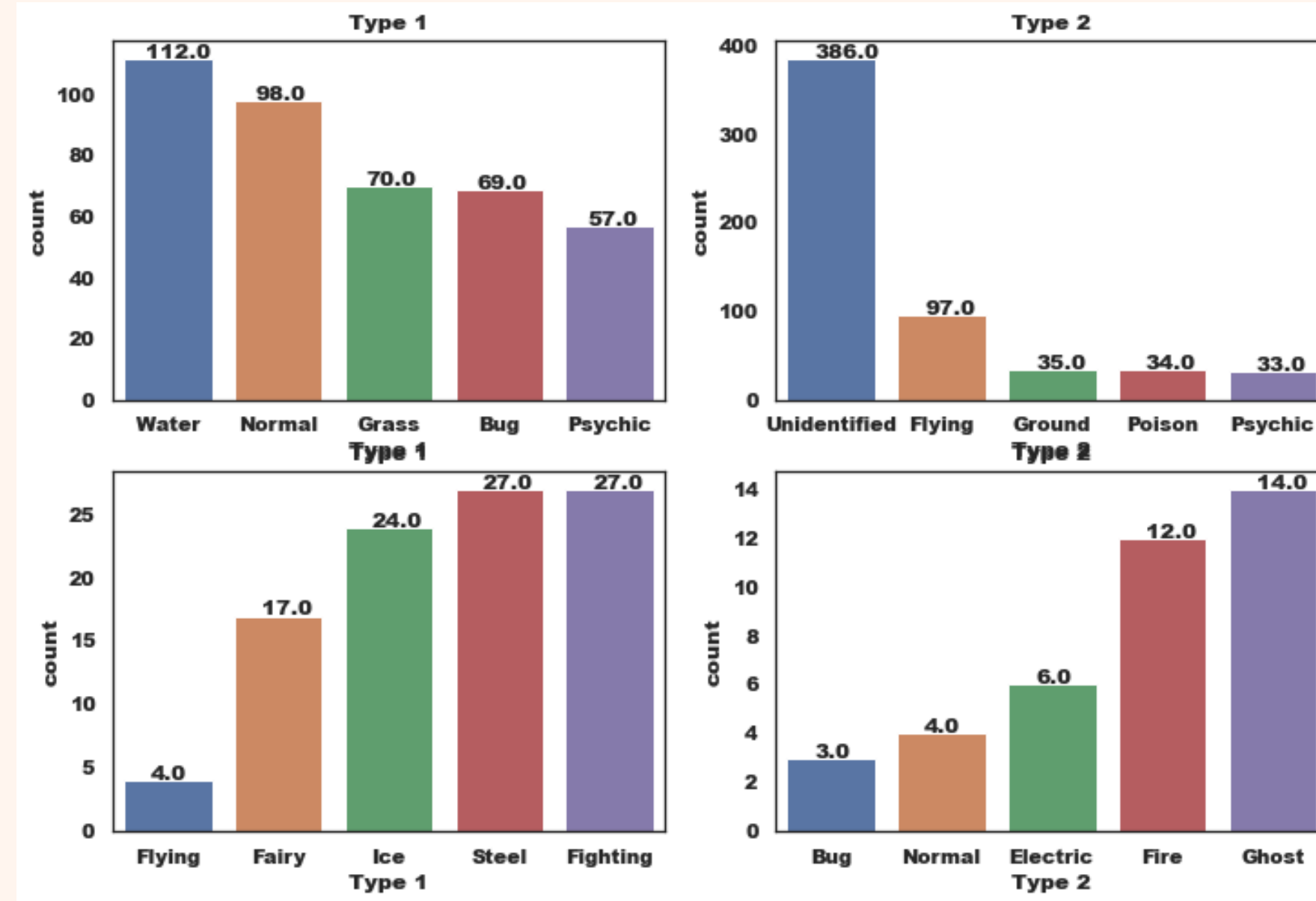
- Total data : 800

Pokemon Per Generasi



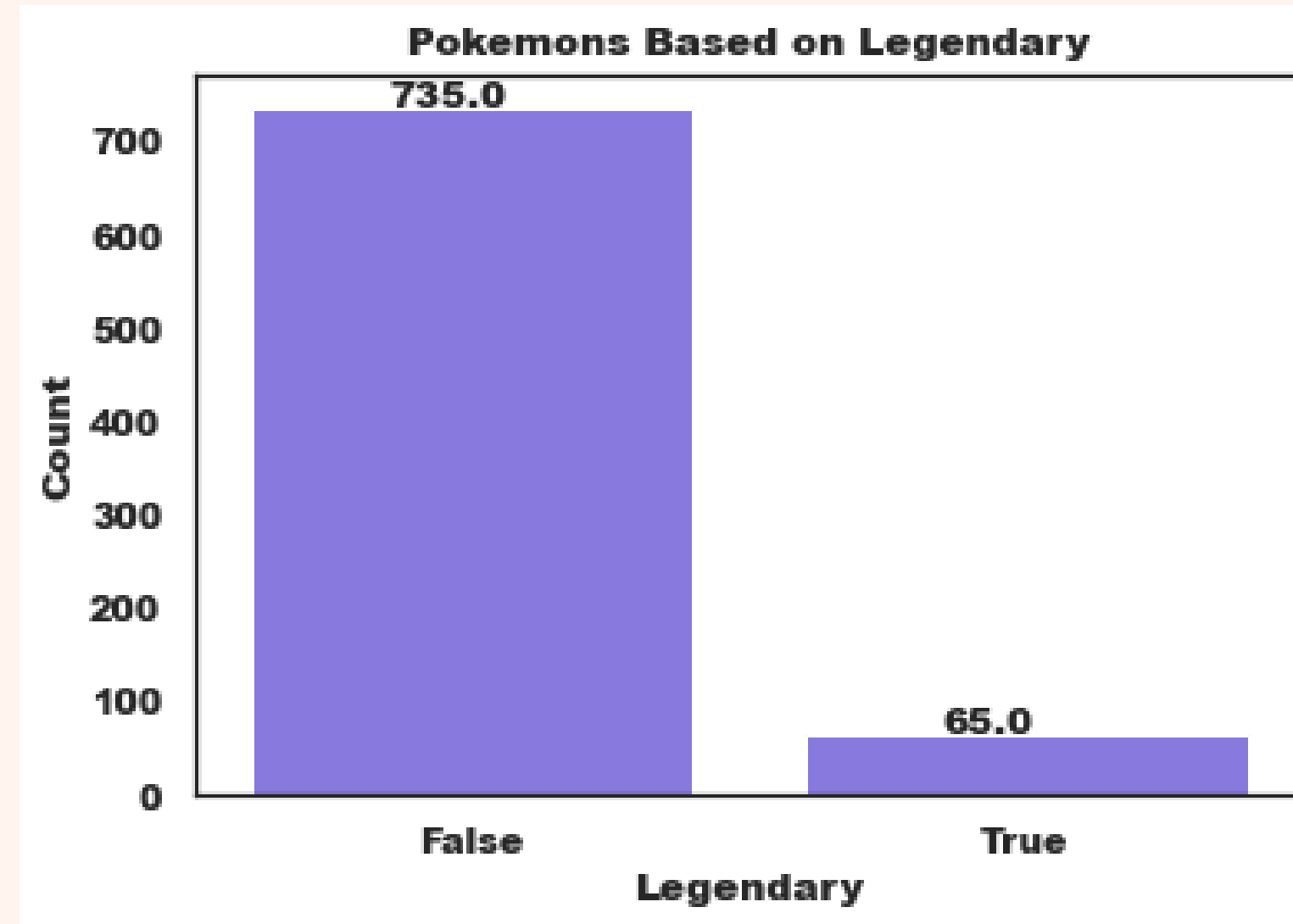
- Generasi 1 memiliki jumlah pokemon paling banyak.
- Generasi 6 memiliki jumlah pokemon paling sedikit, kurang dari setengah jumlah pokemon generasi 1.

Diagram Jumlah Pokemon per Tipe



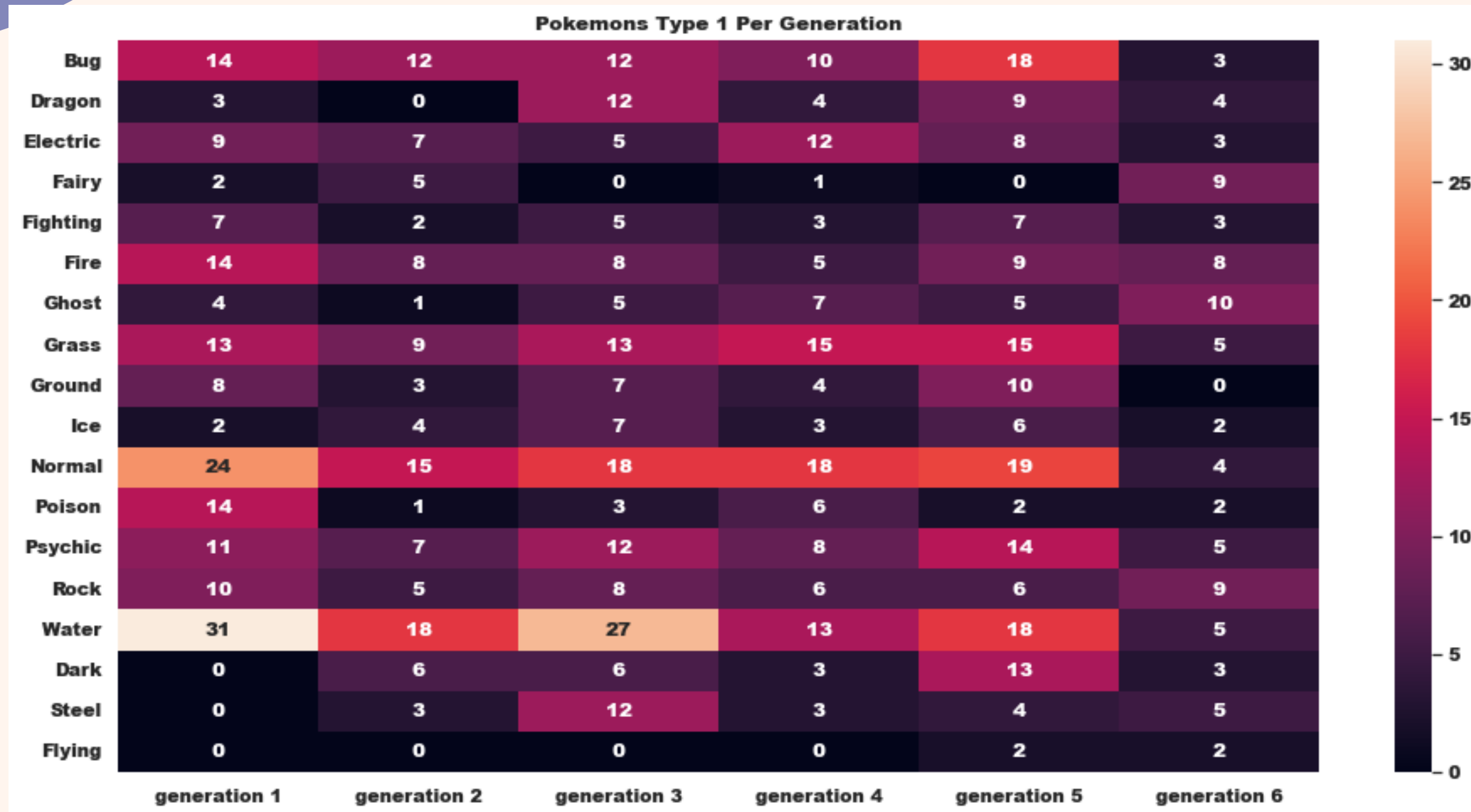
- Pokemon dengan nama water memiliki jumlah paling banyak di tipe 1.
- Pokemon yang tidak teridentifikasi namanya memiliki jumlah paling banyak di tipe 2.

Diagram Pokemon Berdasarkan Legendary



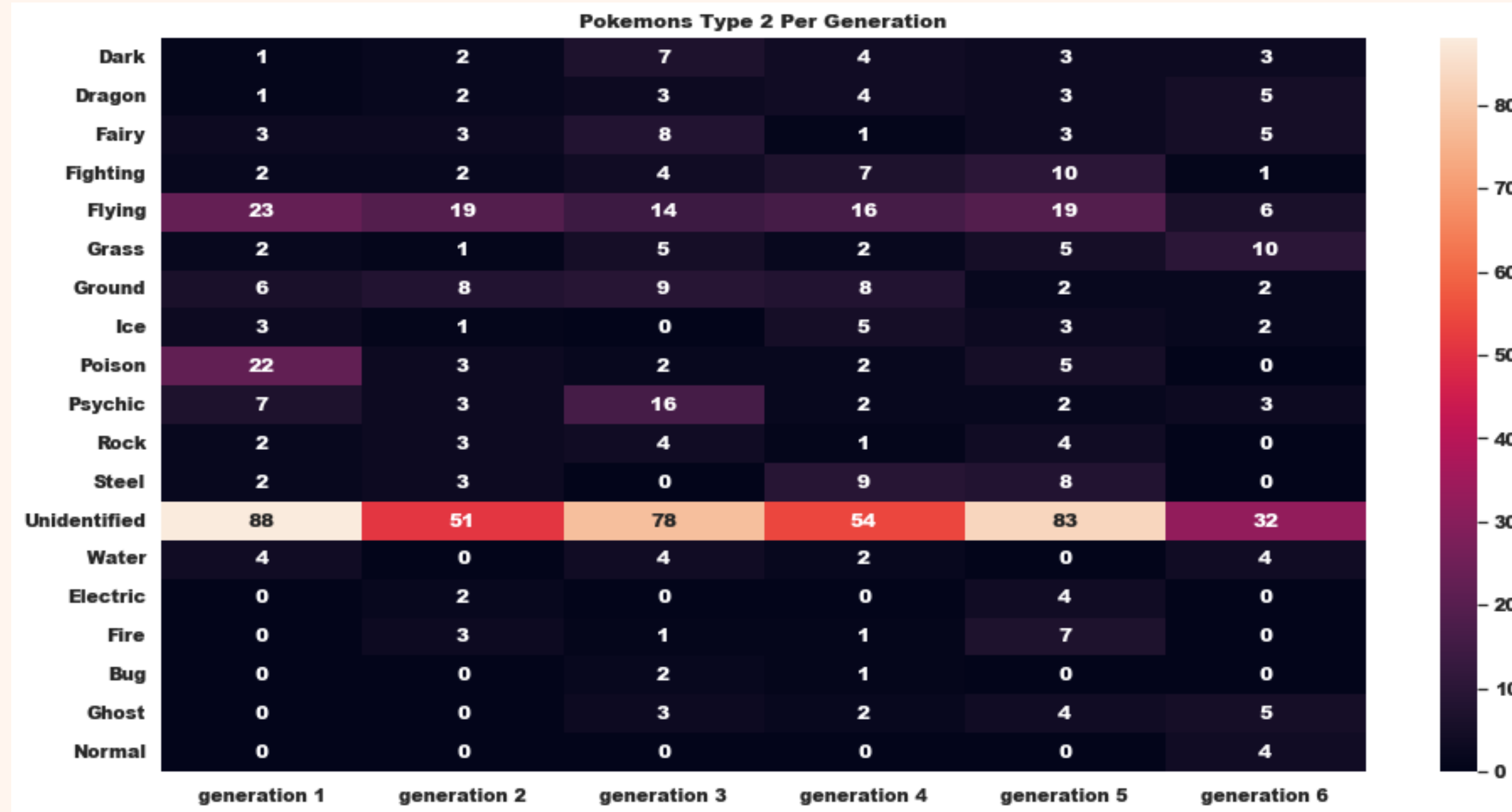
Sebagian besar pokemon tidak Legend

Pokemon Tipe 1 Per Generasi



- Tidak ada generasi yang memiliki semua tipe 1 dari pokemon.
- Pokemon tipe 1 yang paling banyak adalah tipe water pada generasi 1.

Pokemon Tipe 2 Per Generasi



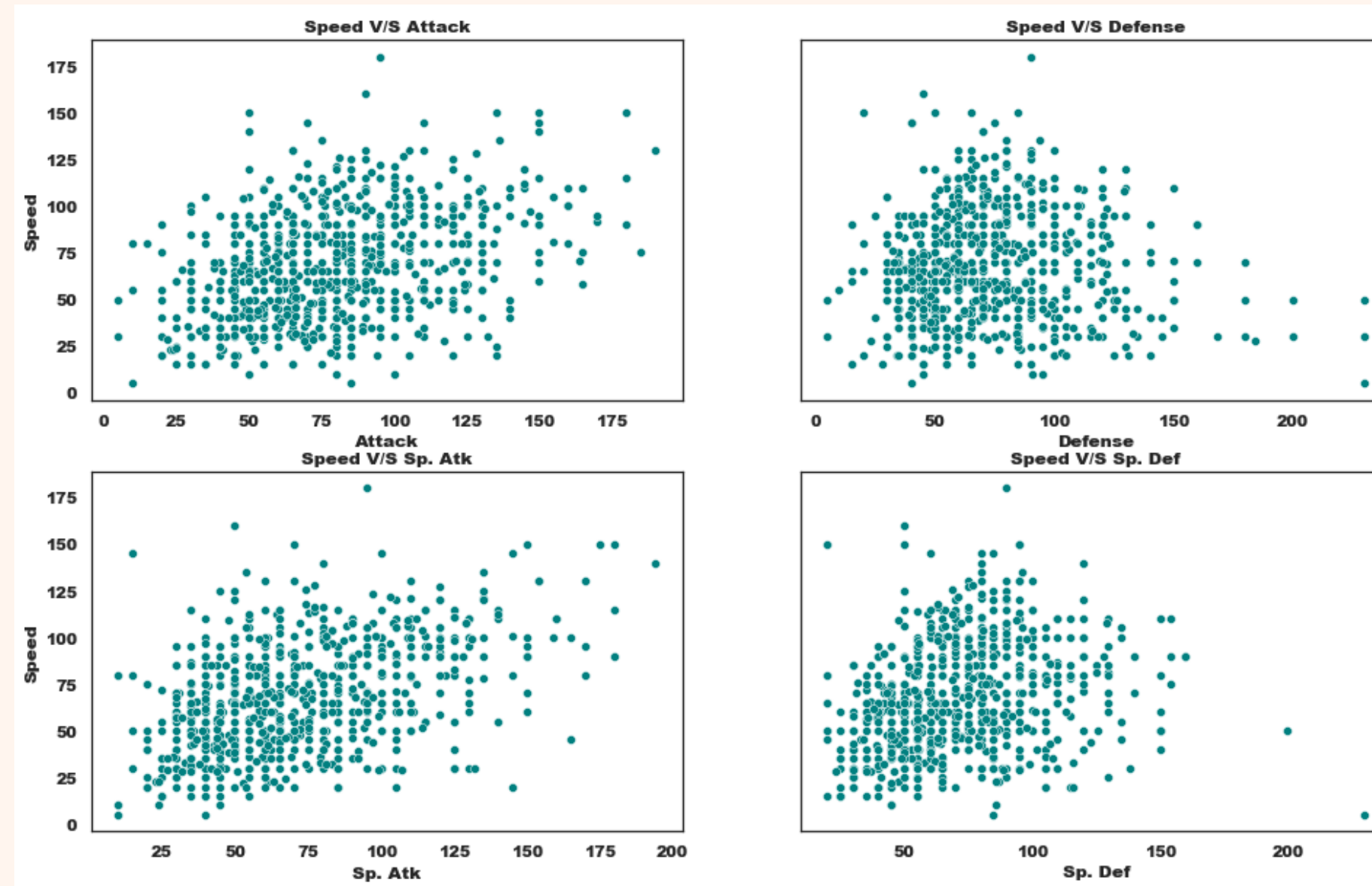
- Tidak ada generasi yang memiliki semua tipe 2 dari pokemon.
- Pokemon tipe 2 paling banyak yaitu pokemon unidentified (tidak teridentifikasi)

Korelasi Statistik



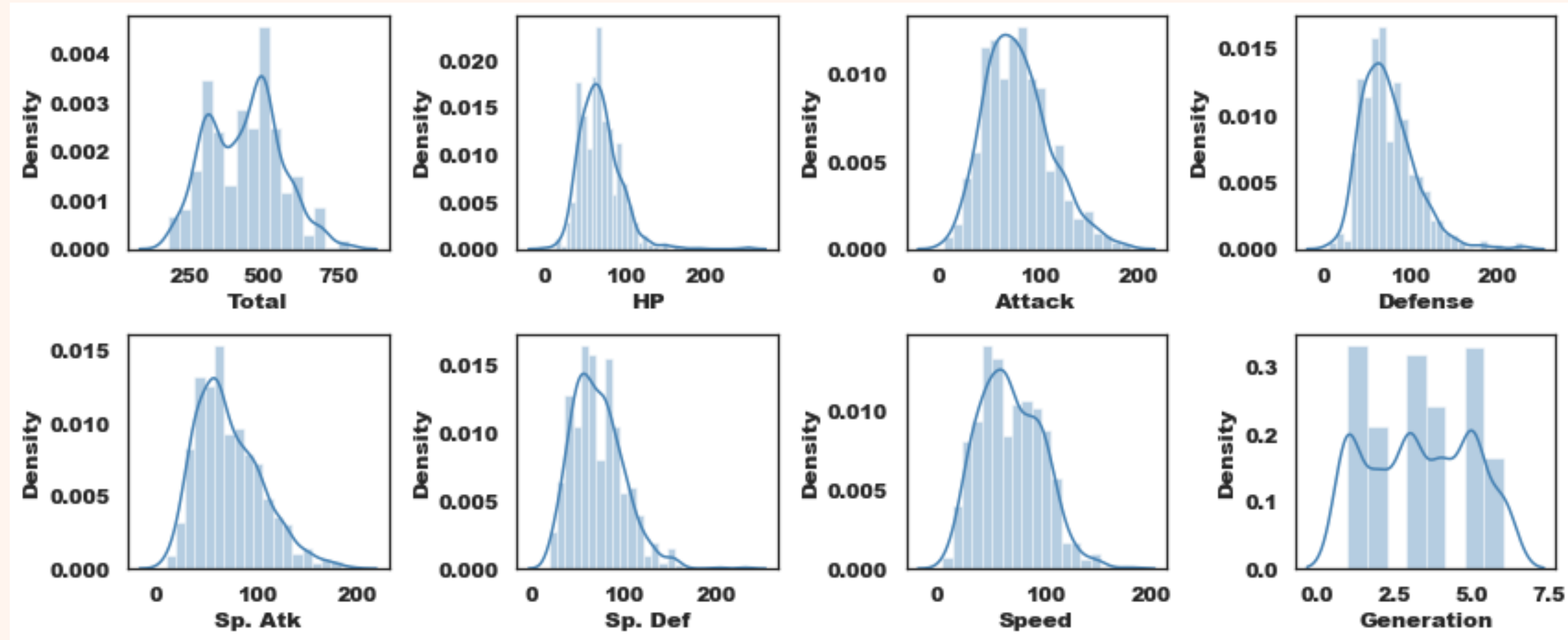
- Nilai korelasi tertinggi : Sp. Atk - Total
- Nilai korelasi terendah : Speed - Defense
- Korelasi negatif : Generation - Speed

Plot Korelasi Speed dengan Variabel lain



- Skor penyerangan sedikit bergantung pada kecepatan, sebagian pokemon dengan kecepatan rata-rata memiliki skor penyerangan yang tinggi.
- Skor pertahanan tidak terlalu bergantung pada kecepatan.
- Skor penyerangan khusus berbanding lurus dengan kecepatan, semakin cepat pokemon maka skor pertahanan khususnya semakin tinggi.
- Pokemon dengan kecepatan rendah memiliki skor pertahanan khusus rendah.

Grafik Distribusi Data



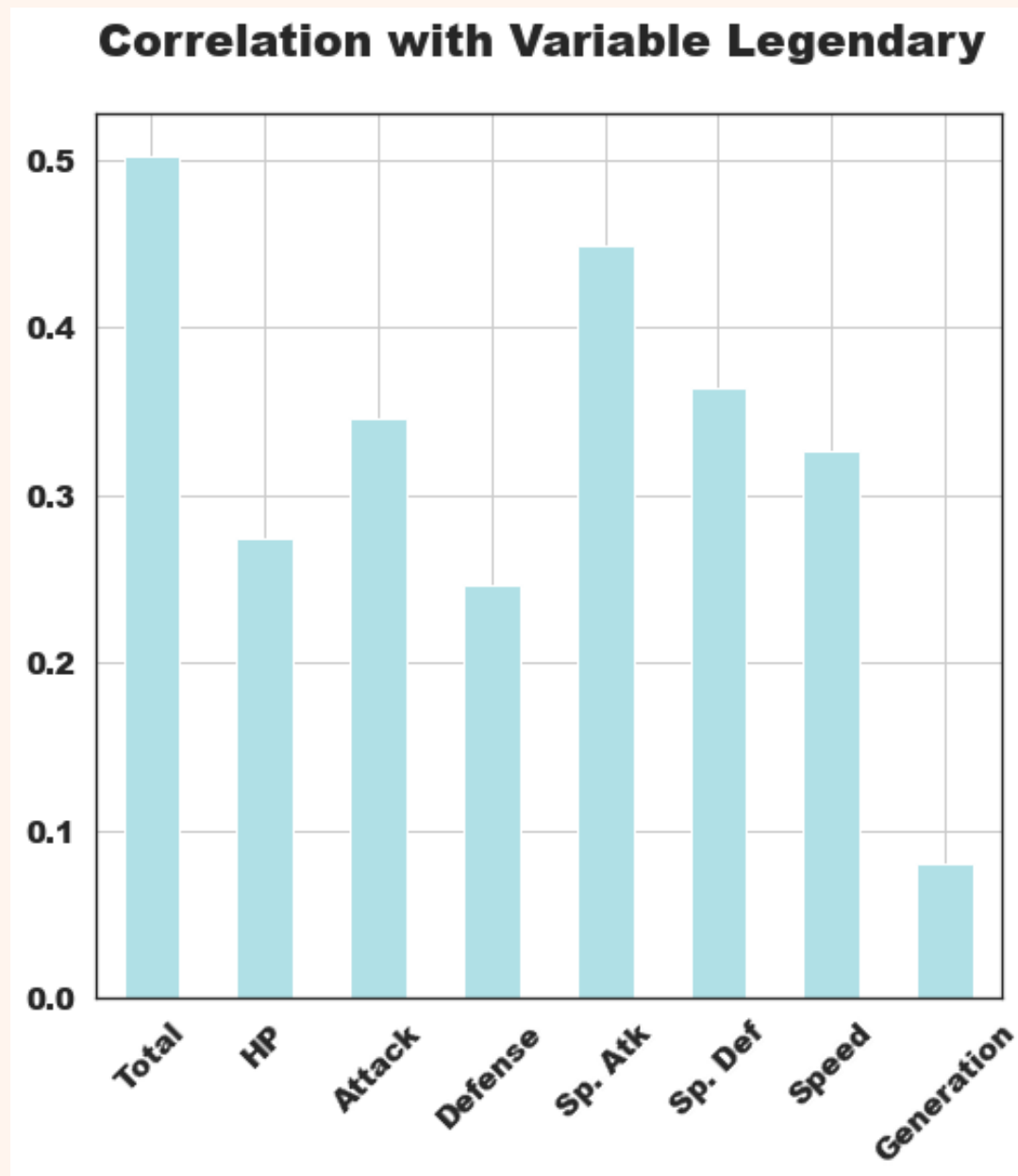
- Hampir semua variabel memiliki distribusi yang mendekati normal.
- Semua variabel memiliki ekor disebelah kanan (positive skewness) yang artinya sebagian besar data ada di nilai rendah.

Uji Keberpengaruhan variabel Terhadap Legendary

	variable	chisquare	pvalue	Kesimpulan
0	Name	800.000000	4.833763e-01	Tidak Berpengaruh Signifikan
1	Type 1	90.420491	5.118547e-12	Berpengaruh Signifikan
2	Type 2	31.378530	2.601501e-02	Berpengaruh Signifikan
3	Total	592.094600	1.149036e-40	Berpengaruh Signifikan
4	HP	284.308225	2.959964e-21	Berpengaruh Signifikan
5	Attack	233.608788	6.544111e-11	Berpengaruh Signifikan
6	Defense	207.384549	3.567665e-09	Berpengaruh Signifikan
7	Sp. Atk	323.489764	2.385355e-24	Berpengaruh Signifikan
8	Sp. Def	261.117012	2.404296e-18	Berpengaruh Signifikan
9	Speed	227.446273	1.091919e-10	Berpengaruh Signifikan
10	Generation	9.876763	7.880338e-02	Tidak Berpengaruh Signifikan

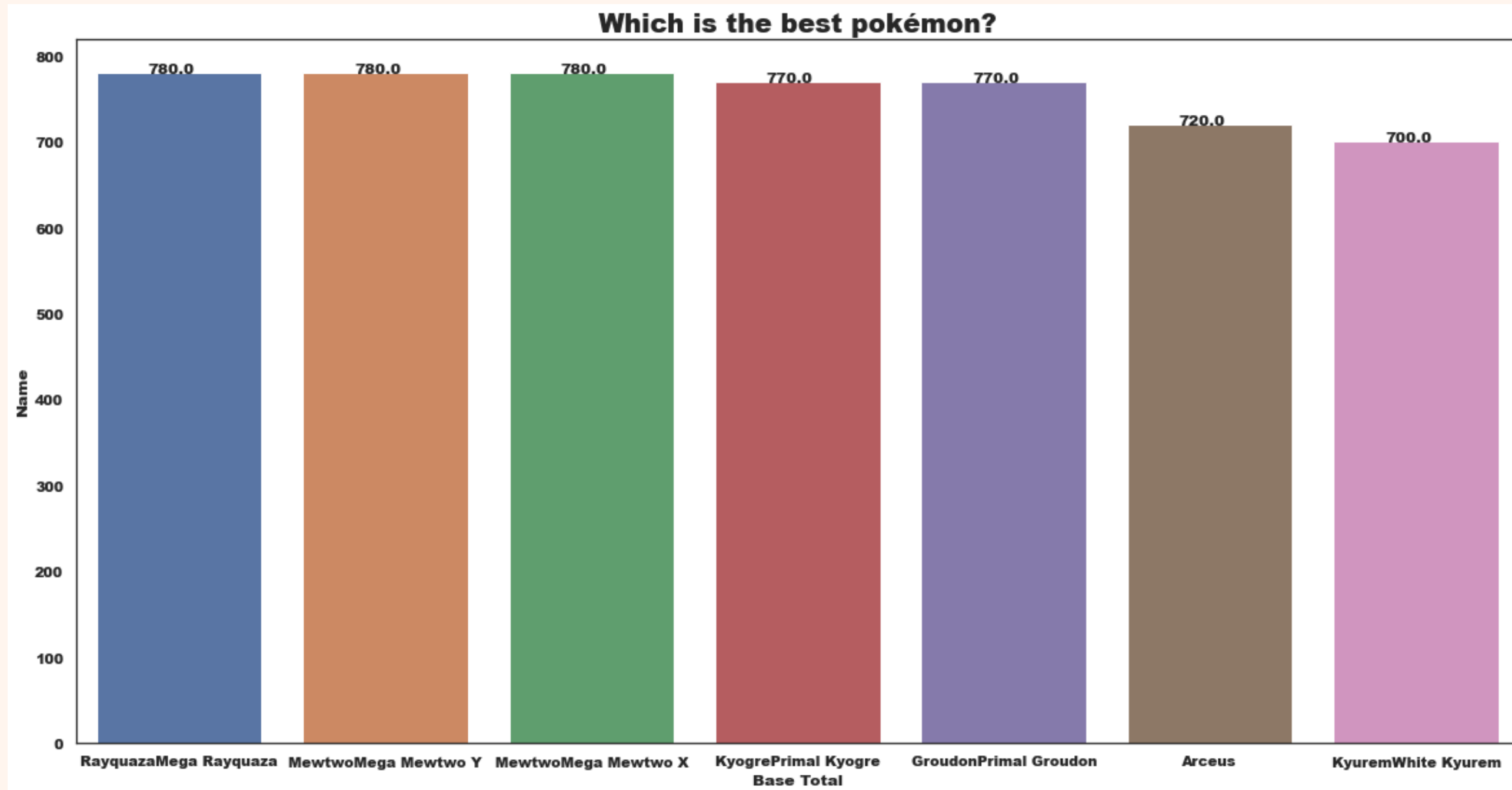
- Hampir semua variabel memiliki pengaruh signifikan terhadap legend tidaknya pokemon.
- Hanya variabel nama dan generasi yang tidak berpengaruh signifikan terhadap legend tidaknya pokemon.

Diagram Korelasi Variabel Terhadap Legendary



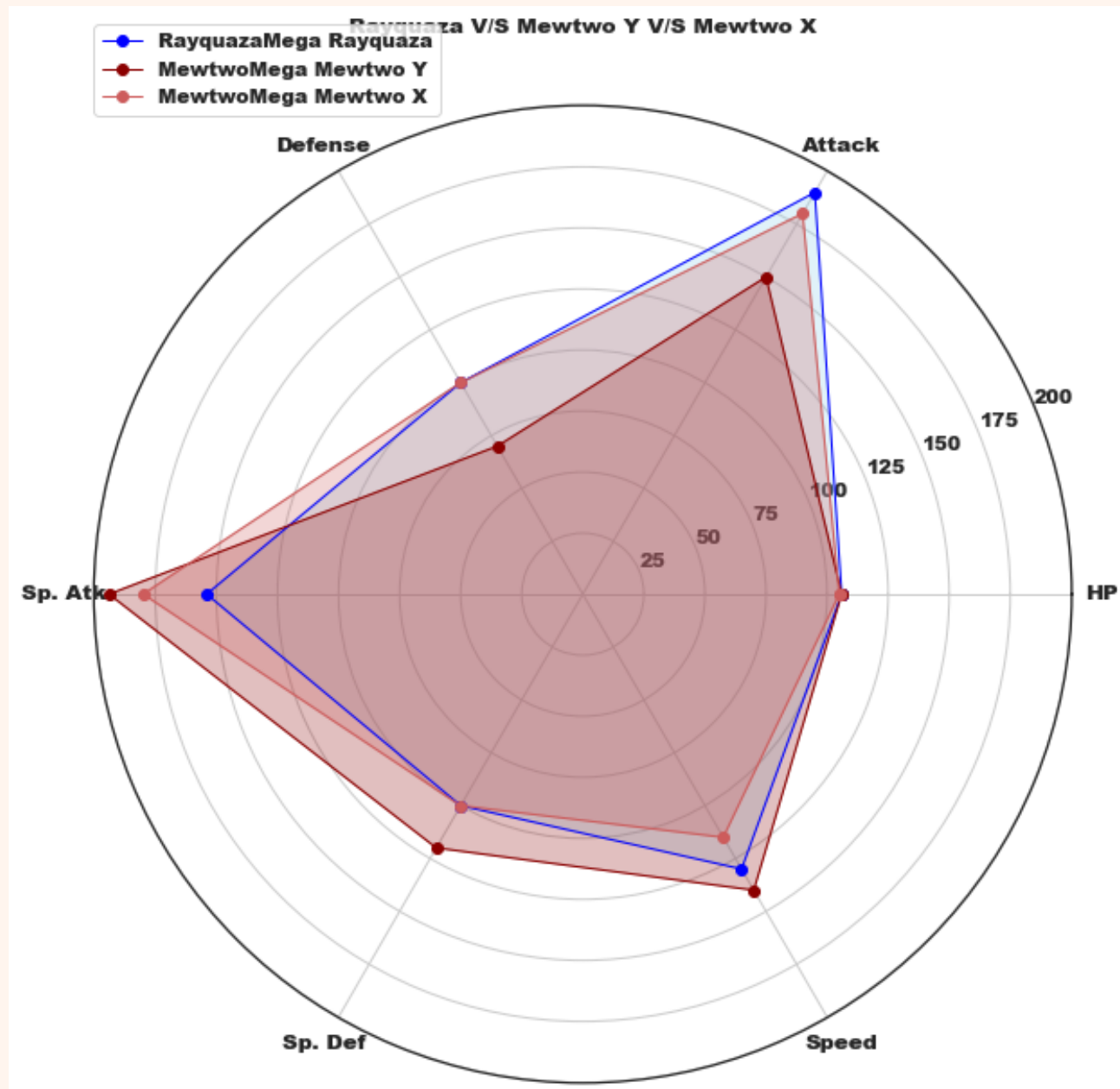
- Variabel Total memiliki korelasi paling tinggi terhadap Legendary
- Semua variabel berkorelasi positif dengan legendary.

Pokemon Terbaik



Pokemon dengan nama RayquazaMega Rayquaza, MewtwoMega Mewtwo Y, dan MewtwoMega Mewtwo X merupakan tiga pokemon terbaik dari dataset.

Perbandingan 3 Pokemon Terbaik



- Mewtwo Y memiliki kemampuan pertahanan paling rendah dari 2 lainnya.
- Kemampuan penyerangan terbaik dimiliki oleh pokemon Rayquaza.
- Dari segi kecepatan, Mewtwo Y lebih unggul dari lainnya.
- Meski kemampuan pertahanan Mewtwo Y lebih rendah, pokemon ini memiliki kemampuan penyerangan khusus dan pertahanan khusus paling baik.
- Mewtwo X memiliki kecepatan paling lemah.

Akurasi Klasifikasi

The accuracy score we achieved is: 91.25 %

Nilai akurasi dari klasifikasi pokemon berdasarkan legendary adalah sebesar 91.25%



Terima
kasih