Pokemon Analyst Project

Presented by Nana CAW

START

Project Overview

This project **was conducted** to **analyze** the Pokemon dataset to **identify trends** and insights that can **help players** make the **good team composition**.

Project Goals

1. **Identify** and **understand** the **distribution** of Pokémon of **different** generations and **types** to **provide insight** into the **trends** and **popularity** of each category.

2. **Identify** Pokémon with the **highest** and **lowest** stats to **assist players** in **building** more **effective** and **strategic teams**.

Data Overview

	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary
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
3	VenusaurMega Venu	Grass	Poison	625	80	100	123	122	120	80	1	FALSE
4	Charmander	Fire		309	39	52	43	60	50	65	1	FALSE
5	Charmeleon	Fire		405	58	64	58	80	65	80	1	FALSE
6	Charizard	Fire	Flying	534	78	84	78	109	85	100	1	FALSE
6	CharizardMega Char	Fire	Dragon	634	78	130	111	130	85	100	1	FALSE
6	CharizardMega Char	Fire	Flying	634	78	104	78	159	115	100	1	FALSE
7	Squirtle	Water		314	44	48	65	50	64	43	1	FALSE
8	Wartortle	Water		405	59	63	80	65	80	58	1	FALSE
9	Blastoise	Water		530	79	83	100	85	105	78	1	FALSE
9	BlastoiseMega Blast	Water		630	79	103	120	135	115	78	1	FALSE
10	Caterpie	Bug		195	45	30	35	20	20	45	1	FALSE
11	Metapod	Bug		205	50	20	55	25	25	30	1	FALSE
12	Butterfree	Bug	Flying	395	60	45	50	90	80	70	1	FALSE
13	Weedle	Bug	Poison	195	40	35	30	20	20	50	1	FALSE
14	Kakuna	Bug	Poison	205	45	25	50	25	25	35	1	FALSE
15	Beedrill	Bug	Poison	395	65	90	40	45	80	75	1	FALSE
15	BeedrillMega Beedri	Bug	Poison	495	65	150	40	15	80	145	1	FALSE
16	Pidgey	Normal	Flying	251	40	45	40	35	35	56	1	FALSE
17	Pidgeotto	Normal	Flying	349	63	60	55	50	50	71	1	FALSE
18	Pidgeot	Normal	Flying	479	83	80	75	70	70	101	1	FALSE
18	PidgeotMega Pidge	Normal	Flying	579	83	80	80	135	80	121	1	FALSE
19	Rattata	Normal		253	30	56	35	25	35	72	1	FALSE

Data Overview

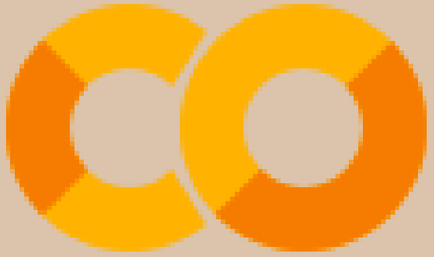
This dataset contains information about various types of Pokémon with the following attributes:

- **#**: ID for each pokemon
- **Name**: Name of each pokemon
- **Type 1**: Each pokemon has a type, this determines weakness/resistance to attacks
- **Type 2**: Some pokemon are dual type and have 2
- **Total**: sum of all stats that come after this, a general guide to how strong a pokemon is
- **HP**: hit points, or health, defines how much damage a pokemon can withstand before fainting

Data Overview

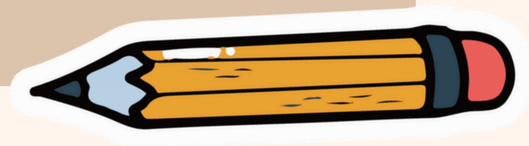
- **Attack**: the base modifier for normal attacks (eg. Scratch, Punch)
- **Defense**: the base damage resistance against normal attacks
- **SP Atk**: special attack, the base modifier for special attacks (e.g. fire blast, bubble beam)
- **SP Def**: the base damage resistance against special attacks
- **Speed**: determines which pokemon attacks first each round

Tools



Google colab

Spreadsheet



Import Library and Load Dataset

	#	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

```
# Import all libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans

# Load dataset
df = pd.read_csv('Pokemon.csv')
df.head()
```

First step, **import** the **necessary libraries** and **load** the '**Pokemon.csv**' dataset to be used. Next, display the **first 5 rows** of the dataset to give an **initial** overview.

Summary Dataset

```
# Summary Dataset

# Display informatic
print(df.info())

# Display the dimen
print(df.shape)
```

```
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
None
(800, 13)
```

From the **information**, the pokemon dataset **has 800 rows** of data and **13 columns**.

Cleaning Data

```
# Check Missing Values and Duplicated Data
print('Check missing values: ')
print(df.isnull().sum())

print('\nDuplicated data:')
print(df.duplicated().sum())
```

```
Check missing values:
#             0
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
```

```
Duplicated data:
0
```

From the result, there are **386 missing values** in the **'Type 2'** column, but it **doesn't affect** our **future** analysis **result**. Therefore, **I ignore** the **missing values** in **'Type 2'** column. And also from the dataset there is **no duplicate data.**

Drop Column

```
# Drop column
df.drop(columns = ['#'], inplace = True)
```

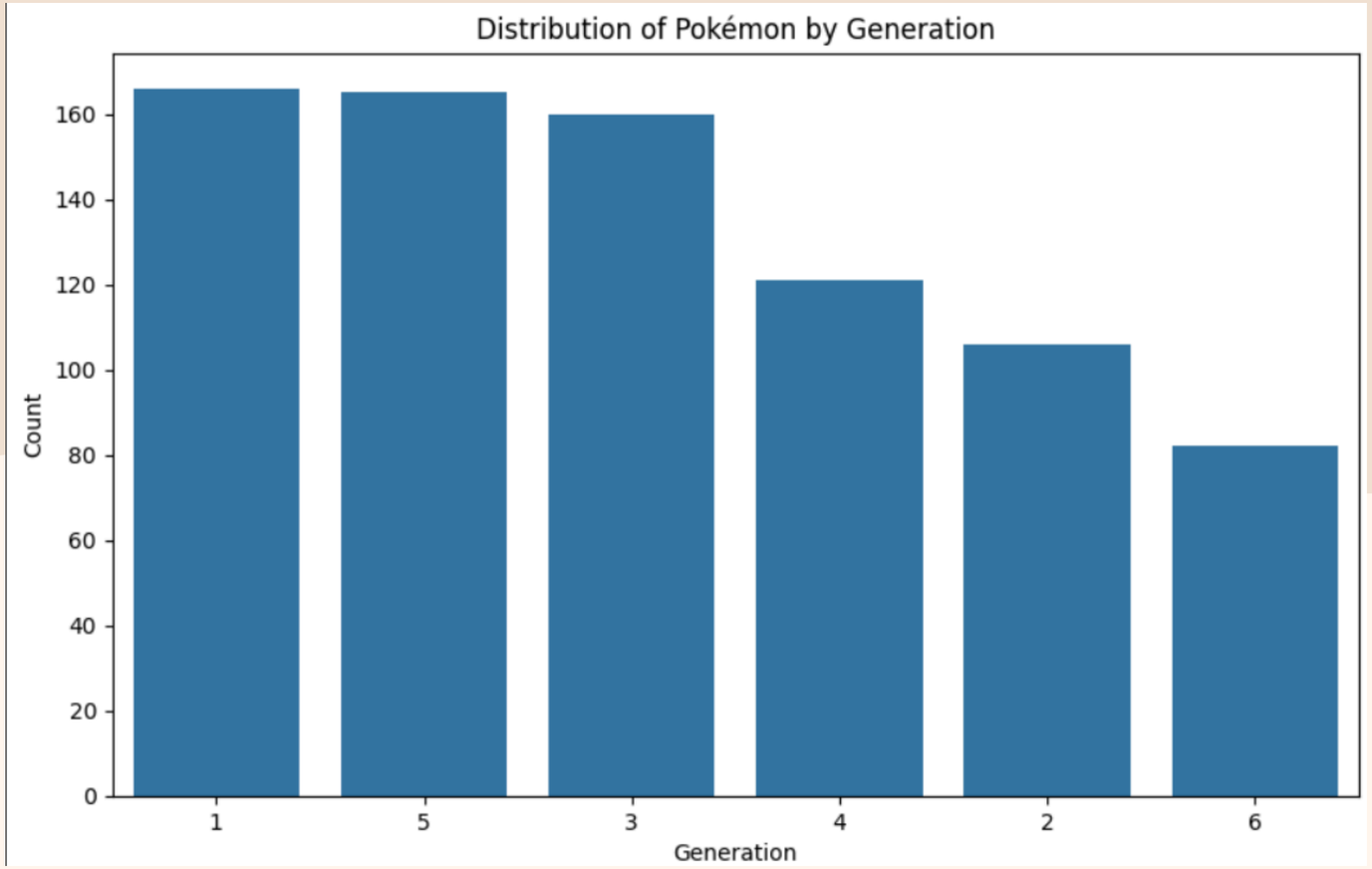
after column '#' **removed.**

I also **remove** the **unused** column ('#') because the column **doesn't affect** the **results** of the data analysis **later**.

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

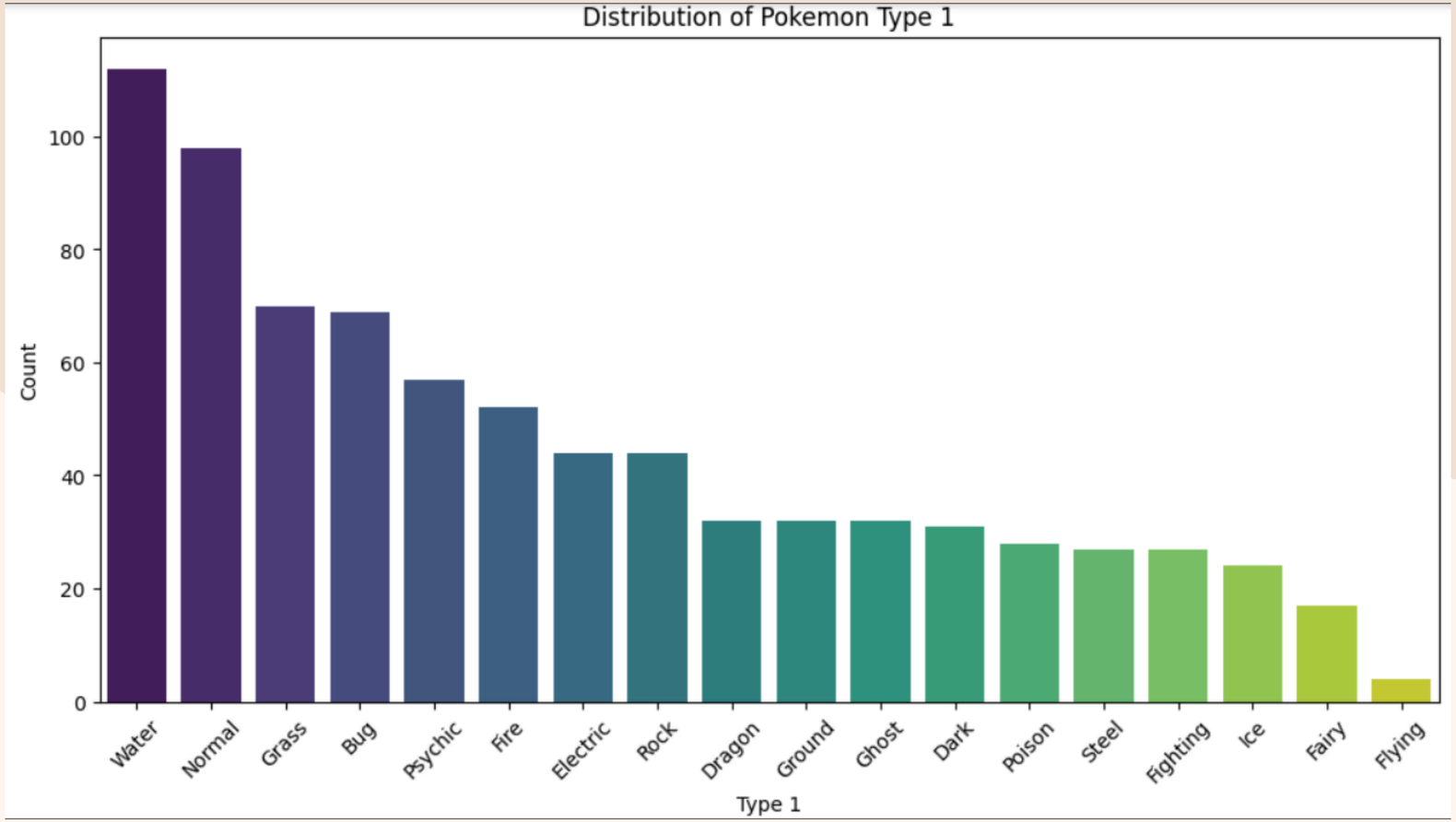
Insight

Distribution of Pokemon by Generation



Generation 1 has the **largest** distribution of pokemon, with **total 166** Pokemons, **followed** by generation **5** and **3** with **165** and **160** Pokemon.

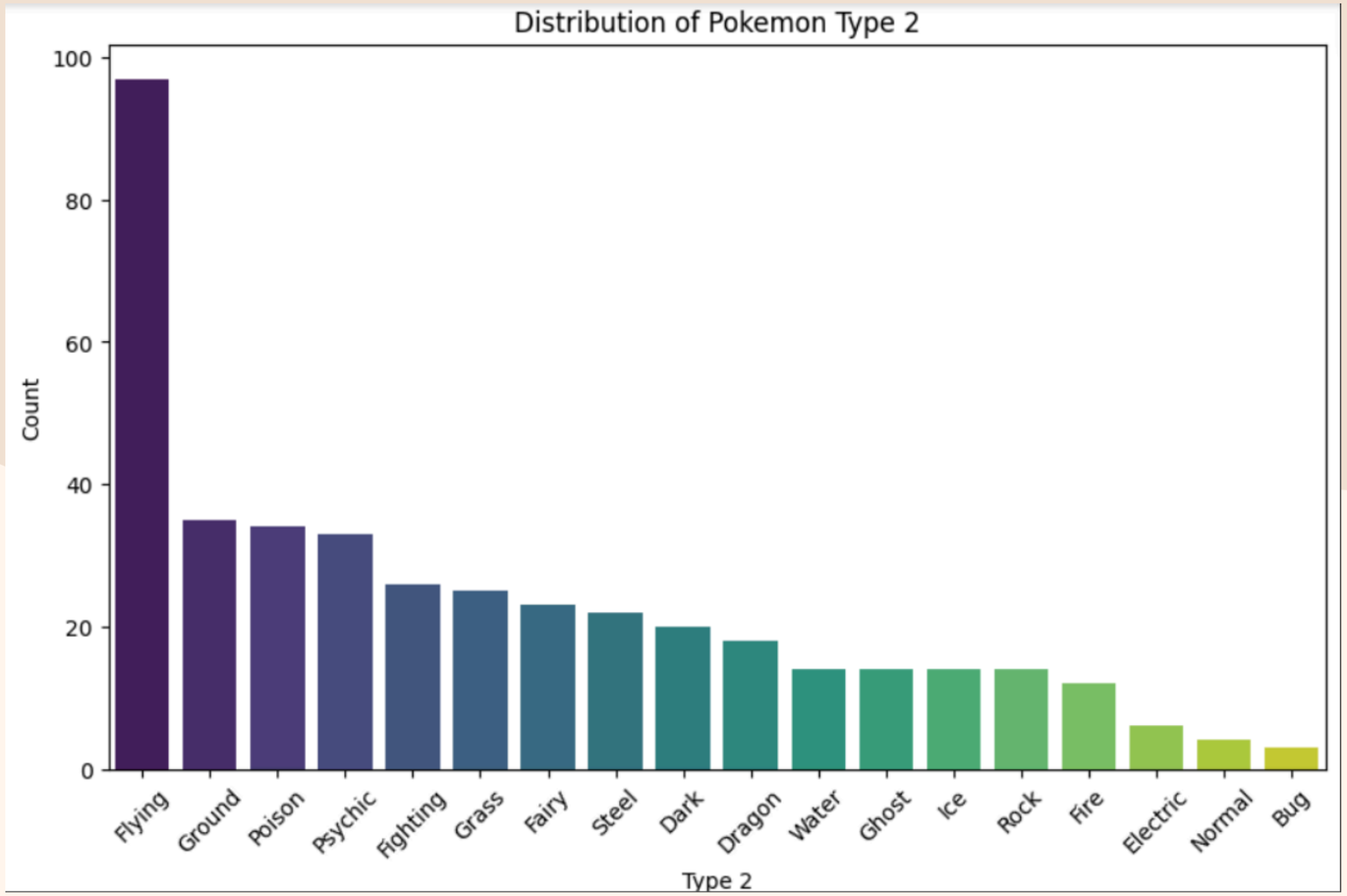
Distribution of Pokemon Type 1



Type 1	
Water	112
Normal	98
Grass	70
Bug	69
Psychic	57
Fire	52
Electric	44
Rock	44
Dragon	32
Ground	32
Ghost	32
Dark	31
Poison	28
Steel	27
Fighting	27
Ice	24
Fairy	17
Flying	4

The figure shows the **distribution** of '**Type 1**' pokemon, where the **highest element** in type 1 is **water** (112), **followed** by **normal** (98) and **grass** (70).

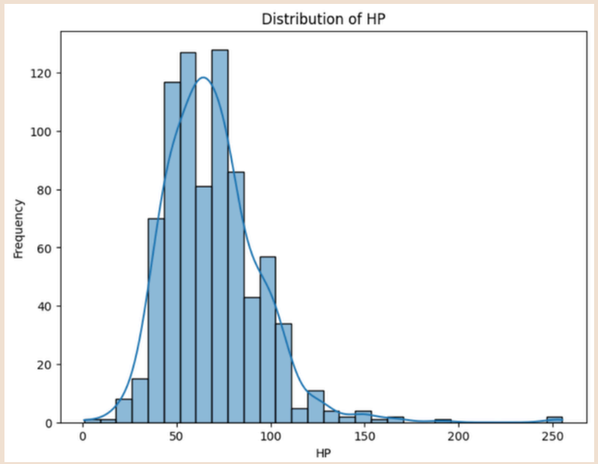
Distribution of Pokemon Type 2

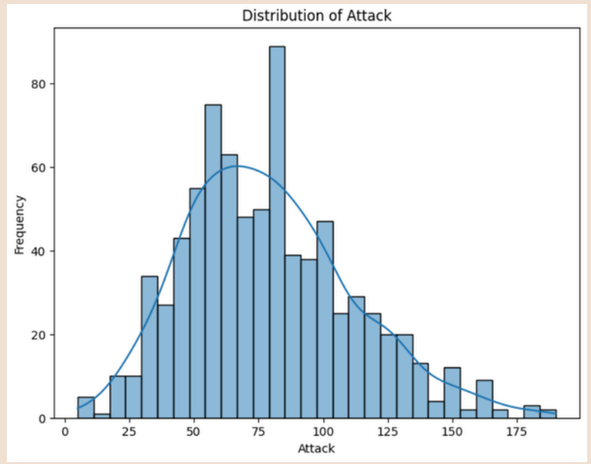


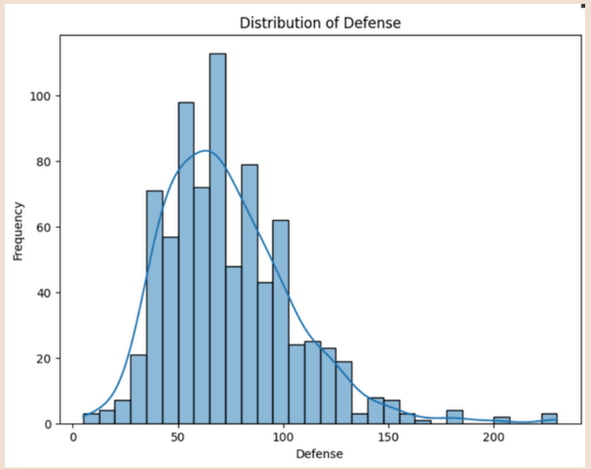
Type 2	
Flying	97
Ground	35
Poison	34
Psychic	33
Fighting	26
Grass	25
Fairy	23
Steel	22
Dark	20
Dragon	18
Water	14
Ghost	14
Ice	14
Rock	14
Fire	12
Electric	6
Normal	4
Bug	3

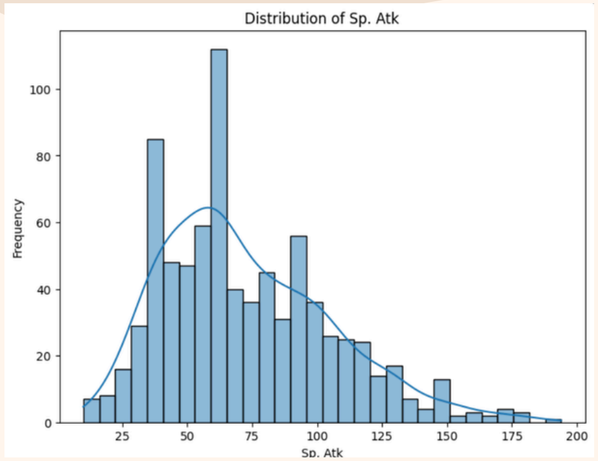
The figure shows the distribution of pokemon in '**Type 2**', where the **highest distribution** is in the **Flying** element with **97** pokemon, **followed** by **Ground** (35) and **Poison** (34).

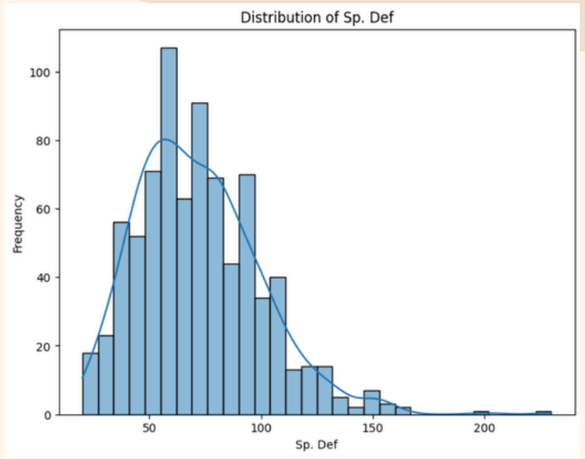
Histogram of Distribution by Attributes

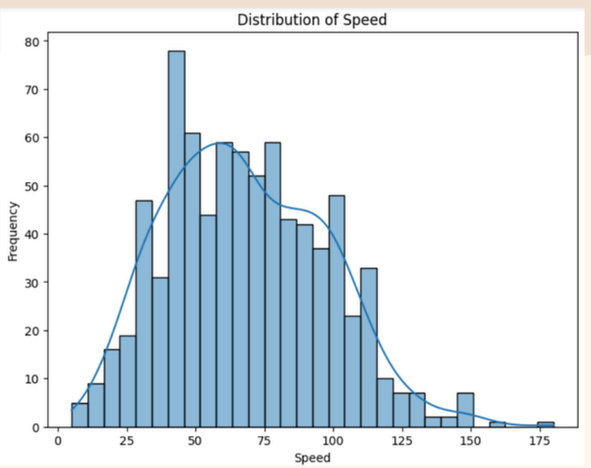






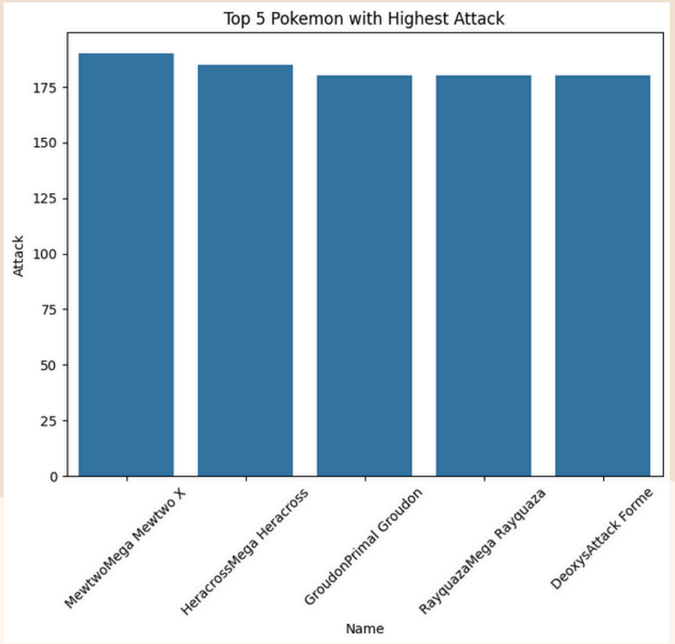


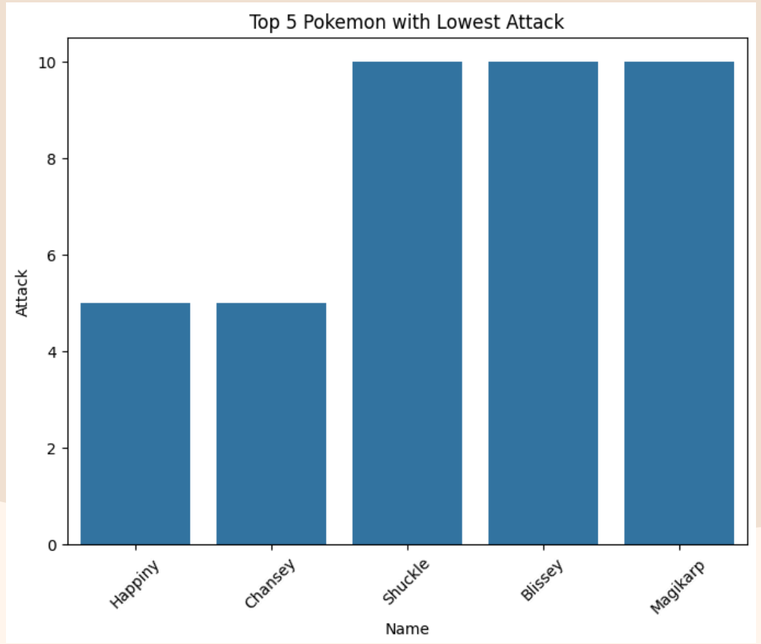




From **all** the **figures** above, all the distributions are **skewed** towards the **right** (it called **positive skewness**).

Top 5 Highest vs Lowest Attack



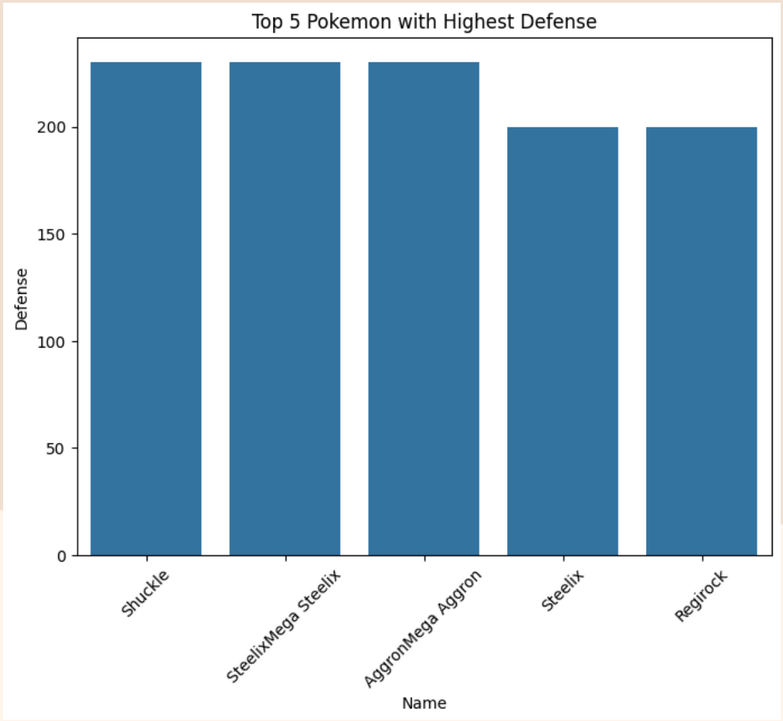


From the **first figure**, the **top 5** Pokémon with **the largest Attack** are obtained. One of them is **Mewtwo** (Mega Mewtwo X) from **generation 1** with an **Attack** of **190**. In addition, there is **Heracross** (Mega Heracross) from **generation 2** which is **not** a **legendary Pokémon** with an **Attack** of **185**. The **second** figure shows the **top 5** Pokemon with the **lowest attack**. There are 5 pokemon with one of them, **Happiny** and **Chansey** from generation **4** and **1** with an **attack** of **only 5**.

Top 5 Highest Defense

Top 5 Pokemon with Highest Defense:

	Name	Defense	Generation	Legendary
230	Shuckle	230	2	False
224	SteelixMega Steelix	230	2	False
333	AggronMega Aggron	230	3	False
223	Steelix	200	2	False
414	Regirock	200	3	True

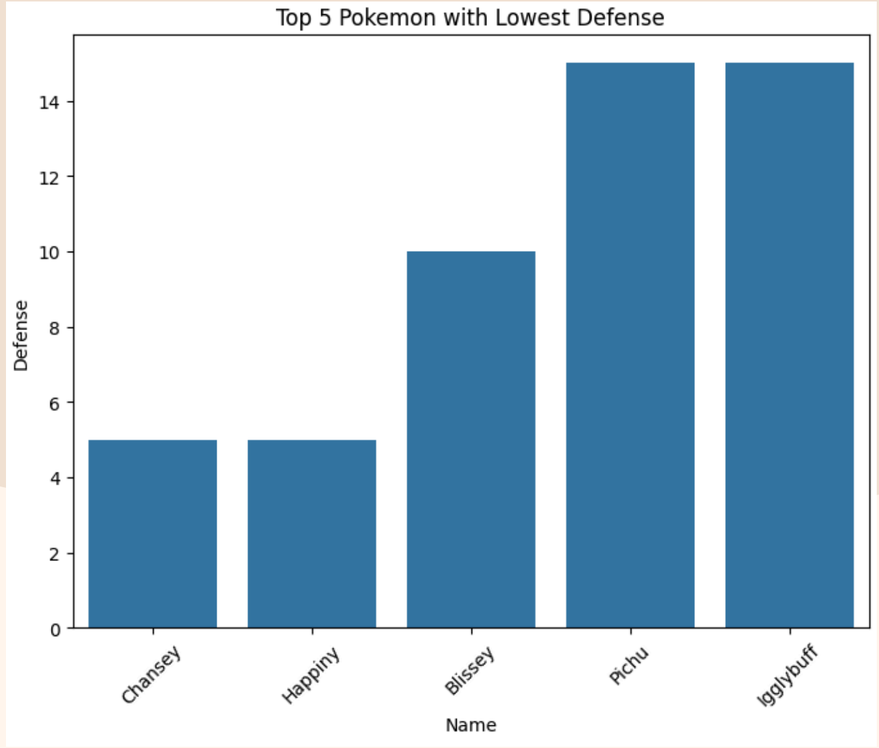


From the analysis and visualization results, it can be **seen** that **Shuckle**, **Steelix** (Mega Steelix), and **Aggron** (Mega Aggron) **have** the **highest Defense** among all Pokémon, with the **same value** of **230**. The **presence** of legendary Pokémon in this list is **only represented** by **Regirock**. The **majority** of Pokémon with the **highest Defense** are **from** Generation **2** and Generation **3**.

Top 5 Lowest Defense

Top 5 Pokemon with Lowest Defense:

	Name	Defense	Generation	Legendary
121	Chansey	5	1	False
488	Happiny	5	4	False
261	Blissey	10	2	False
186	Pichu	15	2	False
188	Igglybuff	15	2	False

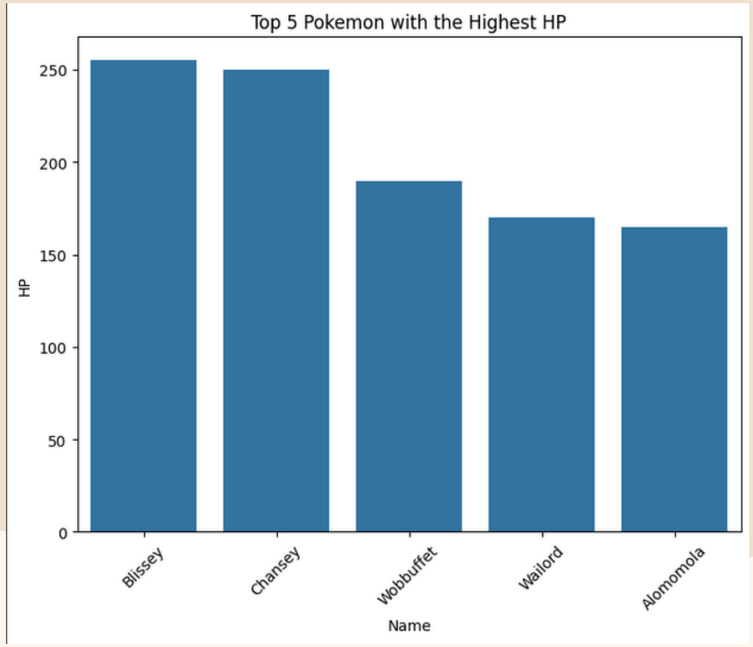


The bar graph created **shows** that the Pokémon with the **lowest Defense** is **Chansey**, Chansey **has** the lowest Defense **value**, at **5**. **All** of the Pokémon on this list **are not legendary Pokémon**, showing that the Pokémon with the lowest Defense are **usually non-legendary Pokémon**.

Top 5 Highest HP

Top 5 Pokemon with Highest HP:

	Name	HP	Generation	Legendary
261	Blissey	255	2	False
121	Chansey	250	1	False
217	Wobbuffet	190	2	False
351	Wailord	170	3	False
655	Alomomola	165	5	False



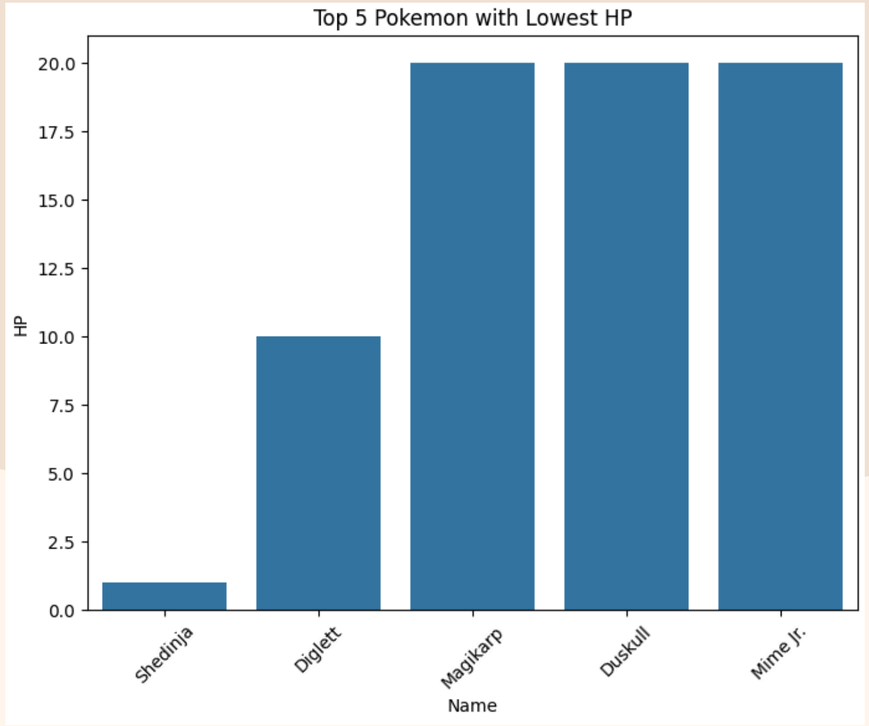
The bar graph created **shows** that the Pokémon with the **highest HP** is **Blissey**, followed by **Chansey**, **Wobbuffet**, **Wailord**, and **Alomomola**. Blissey has the highest HP **value** of **255**, **followed** by Chansey with **250 HP**. Then, Wobbuffet, Wailord, and Alomomola with HP values of 190, 170, and 165.

All Pokémon on **this list** are also **not legendary Pokémon**, showing that **some non-legendary** Pokémon **have** very high HP.

Top 5 Lowest HP

Top 5 Pokemon with Lowest HP:

	Name	HP	Generation	Legendary
316	Shedinja	1	3	False
55	Diglett	10	1	False
139	Magikarp	20	1	False
388	Duskull	20	3	False
487	Mime Jr.	20	4	False

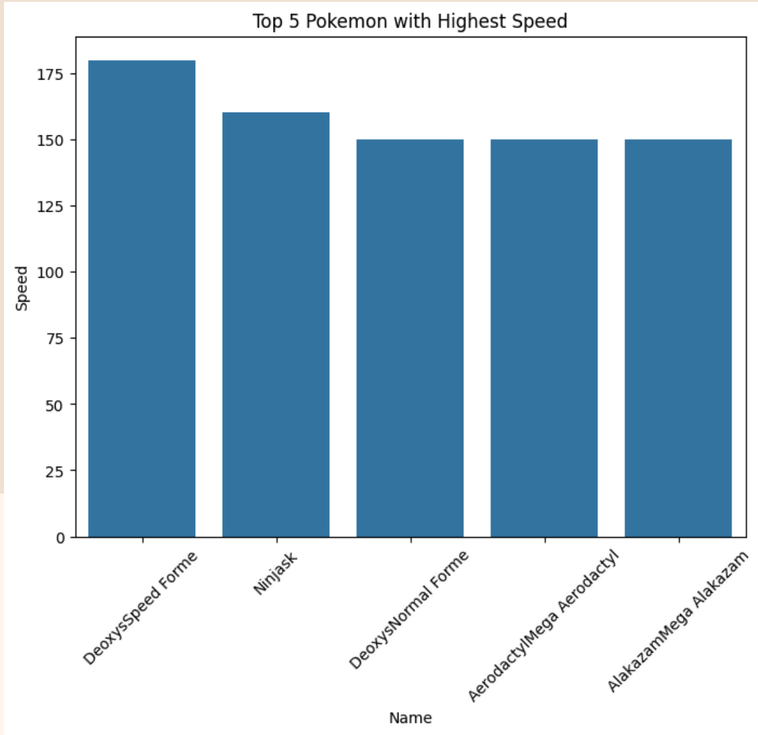


Shedinja has the **lowest HP** value of **1**, which is **very unique** and **far** below other Pokémon. **Diglett**, with an **HP** of **10**, also has a very low HP value. Magikarp, Duskull, and Mime Jr. each have an HP of 20. All the Pokémon on this list **come** from **multiple generations** and are **all non-legendary.**

Top 5 Highest Speed

Top 5 Pokemon with Highest Speed:

	Name	Speed	Generation	Legendary
431	DeoxysSpeed Forme	180	3	True
315	Ninjask	160	3	False
428	DeoxysNormal Forme	150	3	True
154	AerodactylMega Aerodactyl	150	1	False
71	AlakazamMega Alakazam	150	1	False

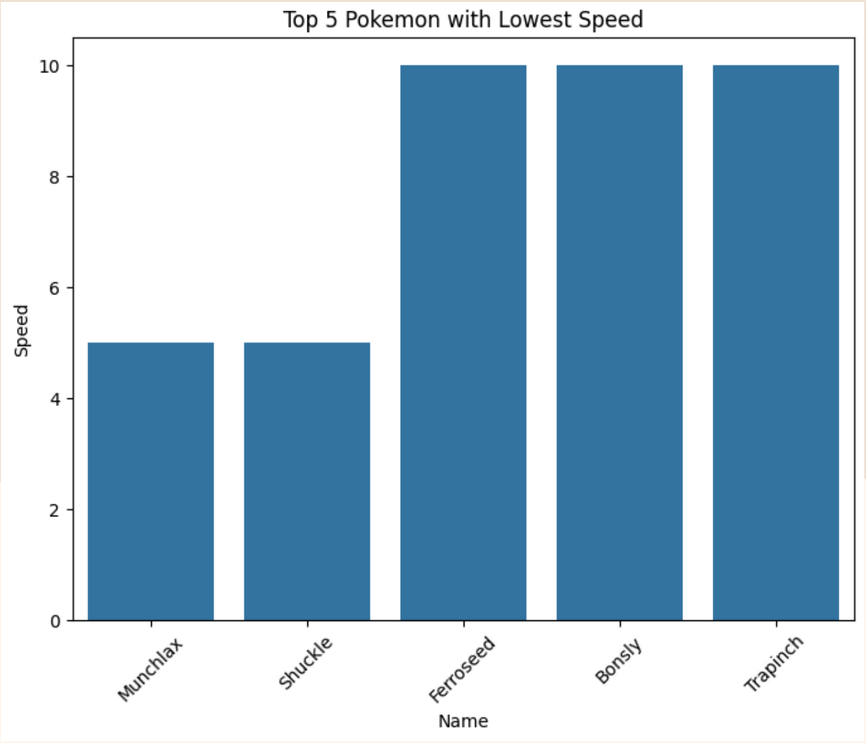


It shows **Deoxys** (Speed Forme) **has** the **highest Speed** value of **180**, far above other Pokémon. This list is **dominated** by Pokémon from **Generation 3** and **Generation 1**. In addition, **non-legendary** Pokémon such as **Ninjask**, **Aerodactyl** (Mega Aerodactyl), and **Alakazam** (Mega Alakazam) also have **very high Speed**.

Top 5 Lowest Speed

Top 5 Pokemon with Lowest Speed:

	Name	Speed	Generation	Legendary
495	Munchlax	5	4	False
230	Shuckle	5	2	False
658	Ferroseed	10	5	False
486	Bonsly	10	4	False
359	Trapinch	10	3	False

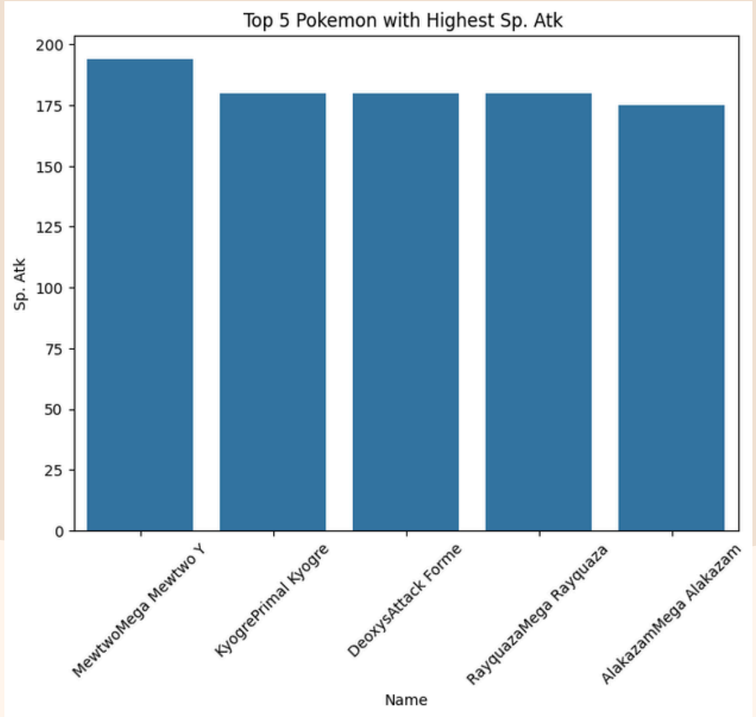


Munchlax and **Shuckle** have the **lowest Speed** value, which is **5**. The other Pokémon on the list, **Ferroseed**, **Bonsly**, and **Trapinch**, have a **slightly higher** Speed of 10. All Pokémon on this list **are non-legendary**, indicating that the Pokémon with the lowest Speed **generally don't fall** into the **legendary** category.

Top 5 Highest Sp. Atk

Top 5 Pokemon with Highest Sp. Atk:

	Name	Sp. Atk	Generation	Legendary
164	MewtwoMega Mewtwo Y	194	1	True
422	KyogrePrimal Kyogre	180	3	True
429	DeoxysAttack Forme	180	3	True
426	RayquazaMega Rayquaza	180	3	True
71	AlakazamMega Alakazam	175	1	False

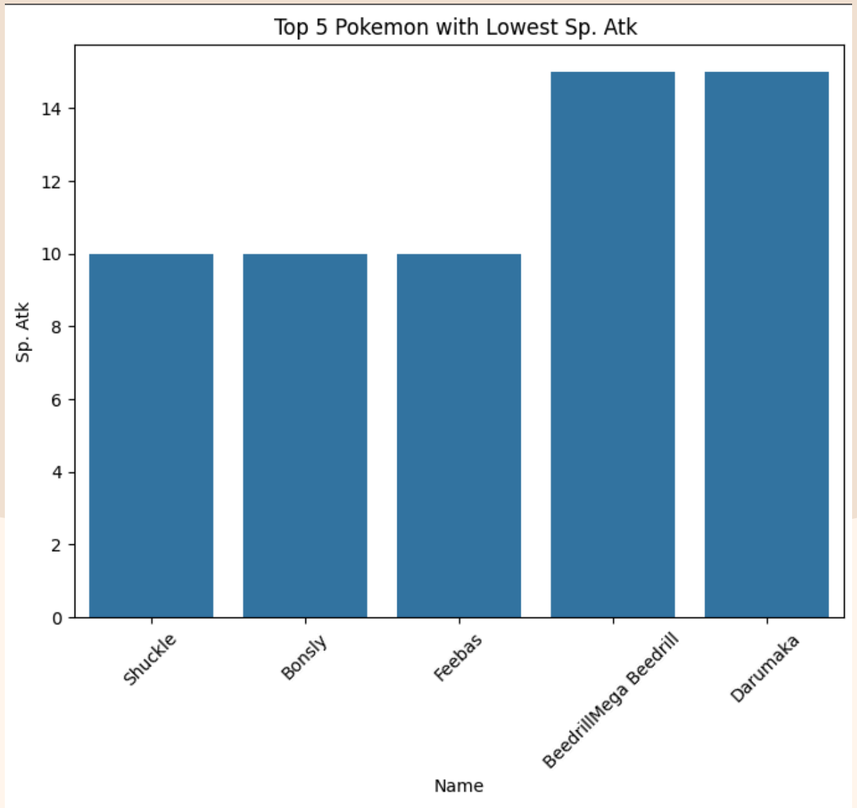


Mewtwo (Mega Mewtwo Y) **has the highest Sp. Atk**, at **194**, which is **much** higher **compared** to other Pokémon. This list also is **dominated** by **legendary Pokémon** from Generation **3**, with Kyogre, Deoxys, and Rayquaza having the same Sp. Atk of 180. **Alakazam** (Mega Alakazam), **although non-legendary**, has a very high **Sp. Atk** of **175**. This shows that many legendary Pokémon have very powerful **special attack abilities**.

Top 5 Lowest Sp. Atk

Top 5 Pokemon with Lowest Sp. Atk:

	Name	Sp. Atk	Generation	Legendary
230	Shuckle	10	2	False
486	Bonsly	10	4	False
381	Feebas	10	3	False
19	BeedrillMega Beedrill	15	1	False
614	Darumaka	15	5	False

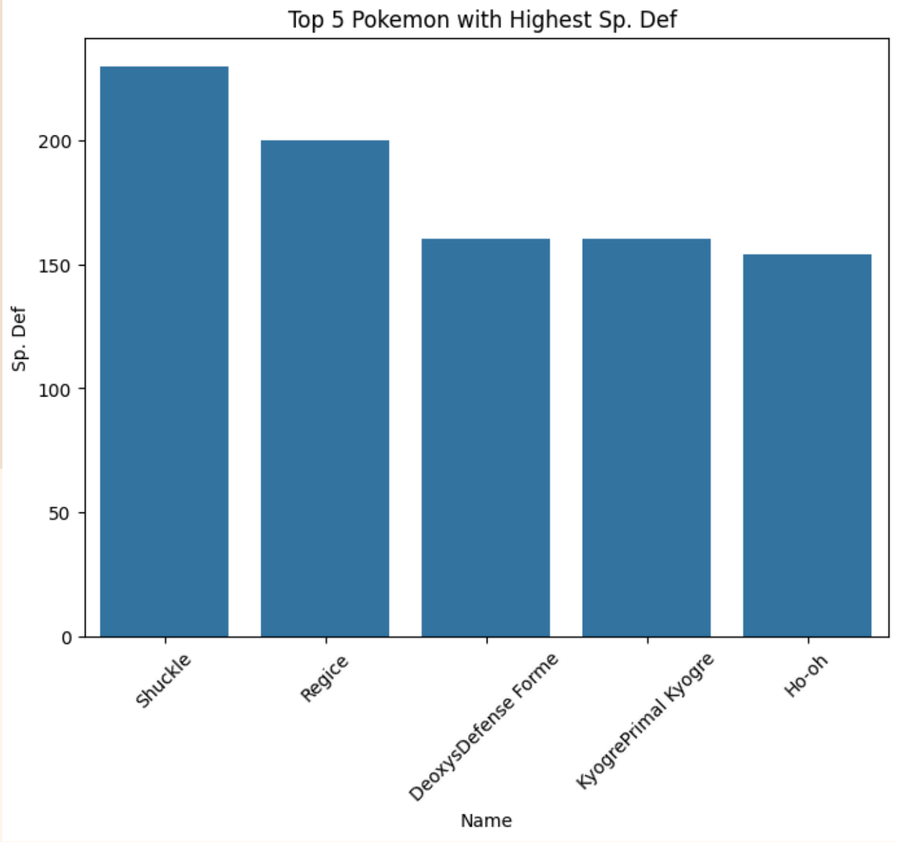


Shuckle, **Bonsly**, and **Feebas** have **the lowest Sp. Atk** value of **10**, indicating **very limited** special attack abilities. All Pokémon on this list **are non-legendary**, indicating that Pokémon with the lowest Sp. Atk are **generally not** legendary Pokémon.

Top 5 Highest Sp. Def

Top 5 Pokemon with Highest Sp. Def:

	Name	Sp. Def	Generation	Legendary
230	Shuckle	230	2	False
415	Regice	200	3	True
430	DeoxysDefense Forme	160	3	True
422	KyogrePrimal Kyogre	160	3	True
270	Ho-oh	154	2	True

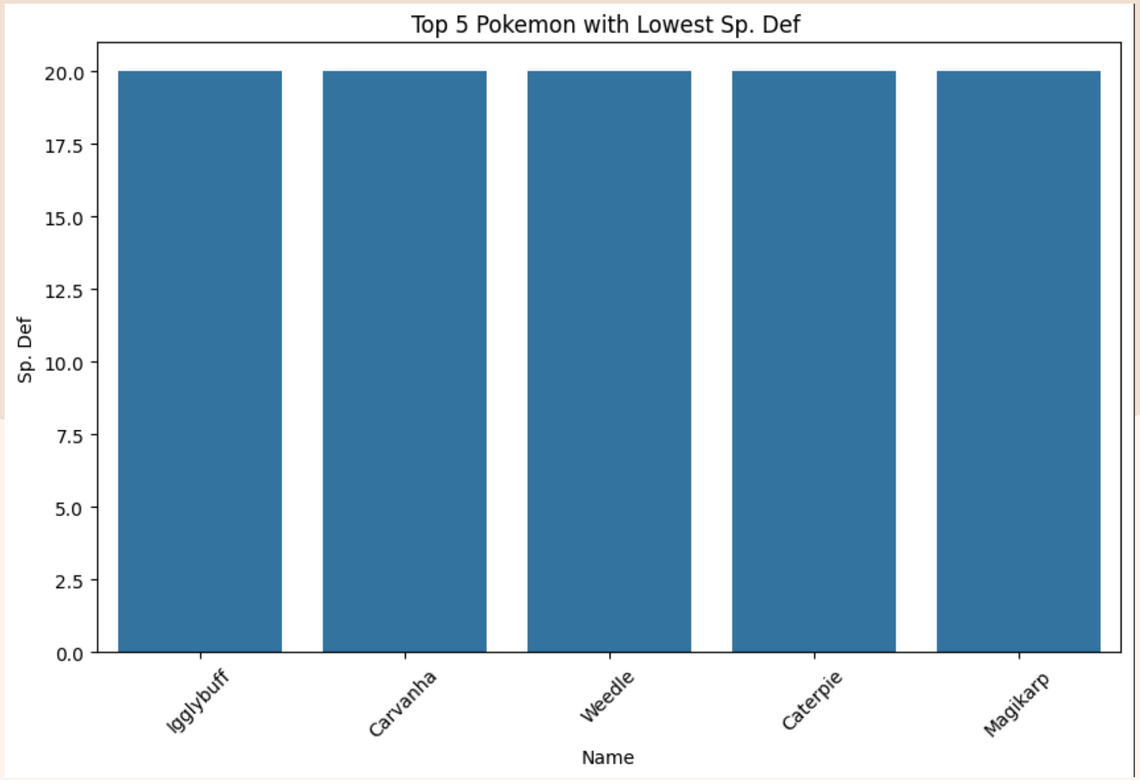


The bar graph **shows** that **Shuckle** has **the highest Sp. Def**, which is **230** and shucke is **not** a legendary pokemon. Then, **followed** by **Regice** with **Sp. Def 200**.

Top 5 Lowest Sp. Def

Top 5 Pokemon with Lowest Sp. Def:

	Name	Sp. Def	Generation	Legendary
188	Igglybuff	20	2	False
347	Carvanha	20	3	False
16	Weedle	20	1	False
13	Caterpie	20	1	False
139	Magikarp	20	1	False

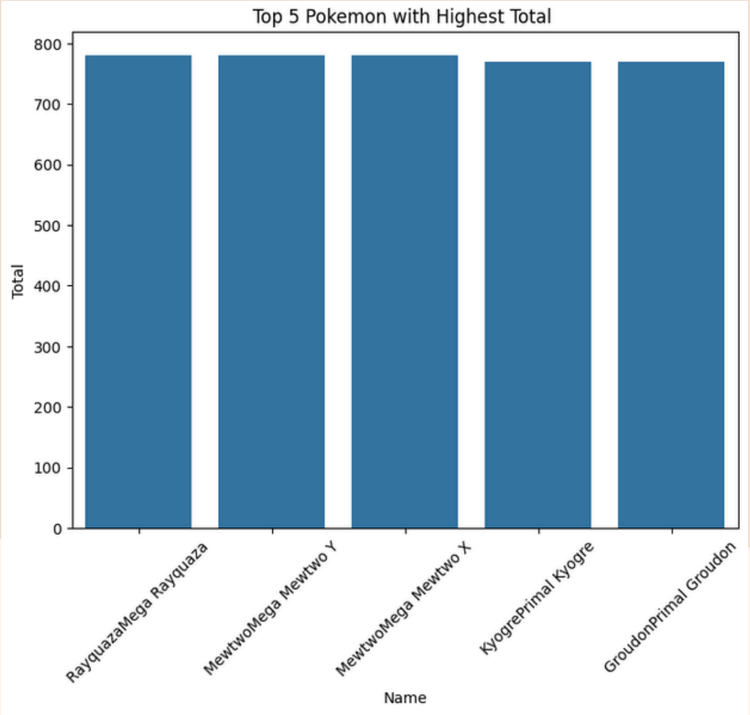


The bar graph **shows** that **Igglybuff**, **Carvanha**, **Weedle**, **Caterpie**, and **Magikarp** all have the **lowest Sp. Def**, which is **20**. This list is **dominated** by **non-legendary** Pokémon from various generations, **especially** Generation **1** and **3**.

Top 5 Highest Total

Top 5 Pokemon with Highest Total:

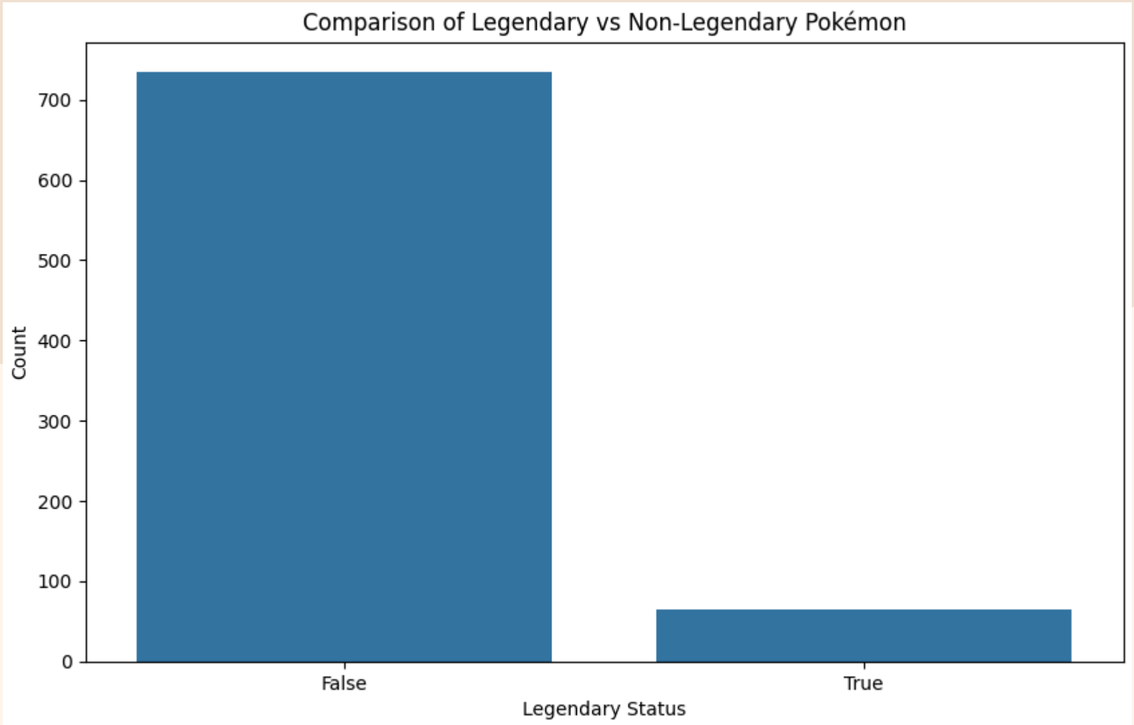
	Name	Total	Generation	Legendary
426	RayquazaMega Rayquaza	780	3	True
164	MewtwoMega Mewtwo Y	780	1	True
163	MewtwoMega Mewtwo X	780	1	True
422	KyogrePrimal Kyogre	770	3	True
424	GroudonPrimal Groudon	770	3	True



Rayquaza (Mega Rayquaza), **Mewtwo** (Mega Mewtwo **Y**), and **Mewtwo** (Mega Mewtwo **X**) have the **highest Total** values of **780**, indicating that **Mega Pokémon** with this **form** are very **strong** overall. **Kyogre** (Primal Kyogre) and **Groudon** (Primal Groudon) also have very high Totals of 770. All of the Pokémon on this **list** are **legendary**, showing that legendary Pokémon **often** have **very high** Totals **compared** to **non-legendary** Pokémon.

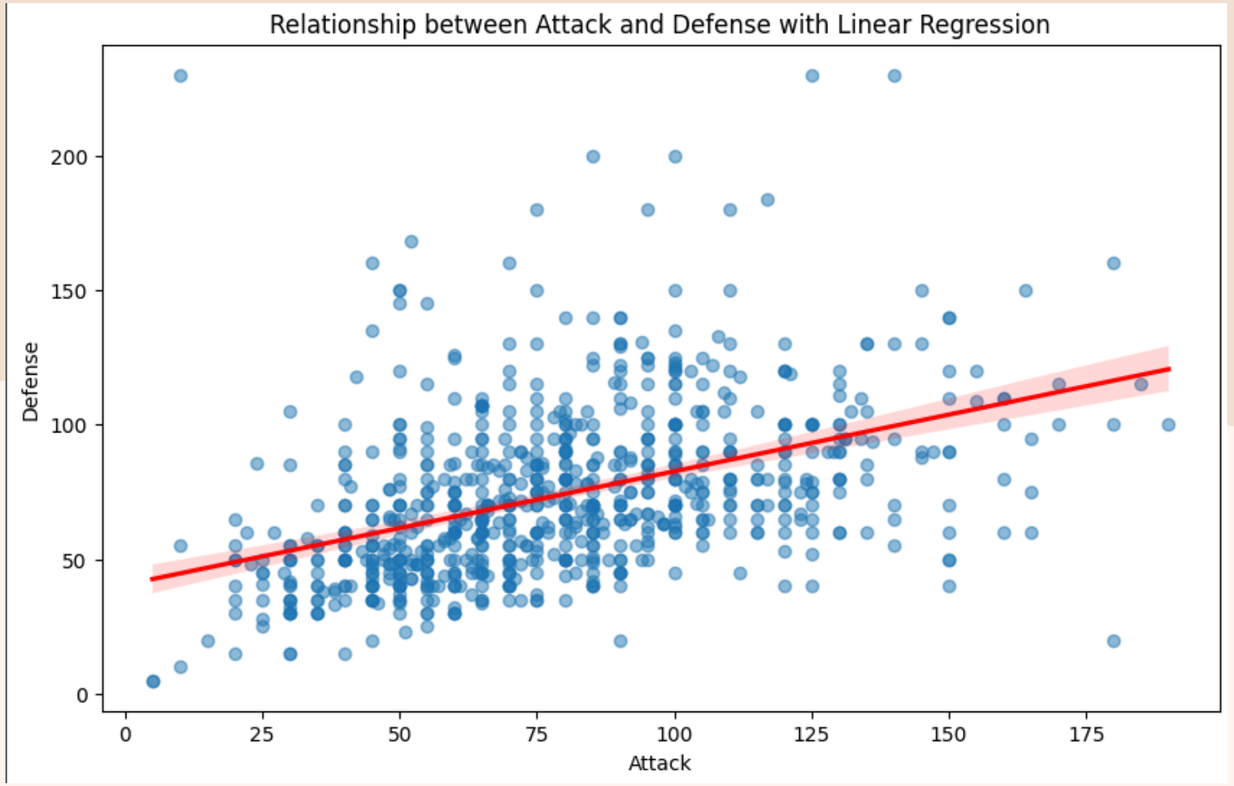
Legendary vs Non-Legendary

```
Comparison of Legendary vs Non-Legendary:
Legendary
False    735
True      65
```



There are **more non-legendary** Pokémon with a **total** of **735** Pokémon **compared** to **65 legendary** Pokémon.

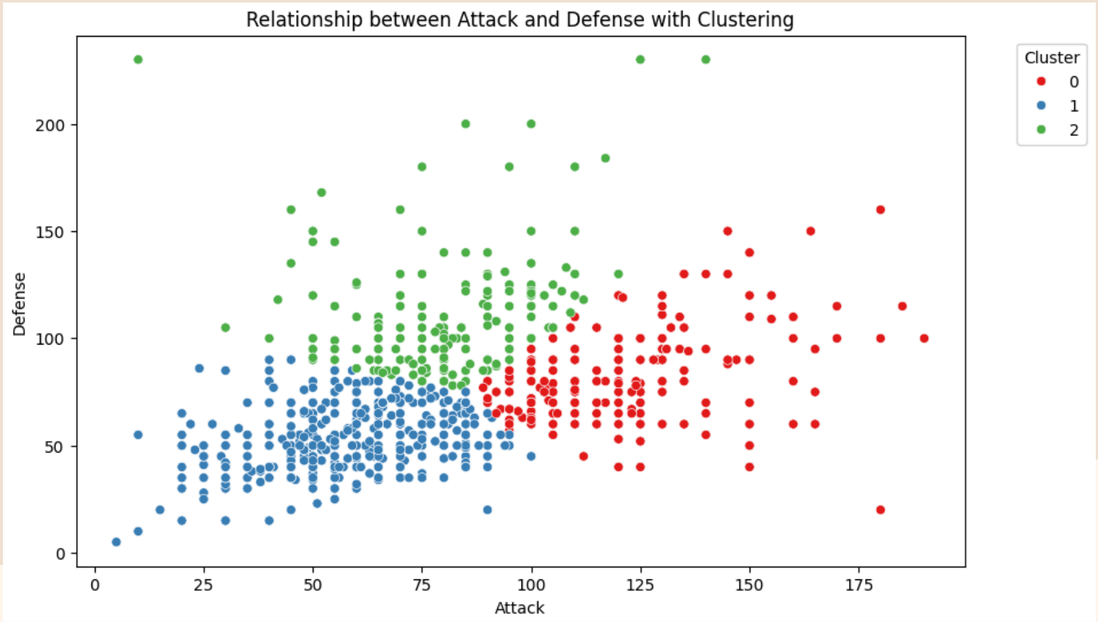
Linear Regression Attack vs Defense



The **regression line gives** a general idea that as '**Attack**' **increases**, '**Defense**' **tends** to **increase** as well, but the **prediction may** be **less accurate** due to the **wide spread** of the **data**.

Clustering K-Means

This **scatter plot** with **clustering** shows **how** the data **can** be **grouped based** on **attack** and **defense characteristics**.



Pokémon di Cluster 0:

	Name	Type 1	Type 2	Total	HP	Attack	Defense
7	CharizardMega Charizard X	Fire	Dragon	634	78	130	111
8	CharizardMega Charizard Y	Fire	Flying	634	78	104	78
19	BeedrillMega Beedrill	Bug	Poison	495	65	150	40
39	Nidoking	Poison	Ground	505	81	102	77
52	Parasect	Bug	Grass	405	60	95	80
..	...	...	...	...	...	...	...
792	Xerneas	Fairy	NaN	680	126	131	95
793	Yveltal	Dark	Flying	680	126	131	95
796	DiancieMega Diancie	Rock	Fairy	700	50	160	110
797	HoopaHoopa Confined	Psychic	Ghost	600	80	110	60
798	HoopaHoopa Unbound	Psychic	Dark	680	80	160	60

Cluster 0 (Red):
Contains data with **high** '**Attack**' **values** (around 75-175) and '**Defense**' **values varying** from **low** to **high** (around 0-150). This distribution **indicates** that this cluster **contains** Pokemon with high attack but with varying defense.

Pokémon di Cluster 1:

	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk
0	Bulbasaur	Grass	Poison	318	45	49	49	65
1	Ivysaur	Grass	Poison	405	60	62	63	80
4	Charmander	Fire	NaN	309	39	52	43	60
5	Charmeleon	Fire	NaN	405	58	64	58	80
9	Squirtle	Water	NaN	314	44	48	65	50
..	...	...	...	...	..	...	...	...
781	PumpkabooSmall Size	Ghost	Grass	335	44	66	70	44
782	PumpkabooLarge Size	Ghost	Grass	335	54	66	70	44
783	PumpkabooSuper Size	Ghost	Grass	335	59	66	70	44
790	Noibat	Flying	Dragon	245	40	30	35	45
791	Noivern	Flying	Dragon	535	85	70	80	97

Pokémon di Cluster 2:

	Name	Type 1	Type 2	Total	HP	Attack	Defense
2	Venusaur	Grass	Poison	525	80	82	83
3	VenusaurMega Venusaur	Grass	Poison	625	80	100	123
6	Charizard	Fire	Flying	534	78	84	78
11	Blastoise	Water	NaN	530	79	83	100
12	BlastoiseMega Blastoise	Water	NaN	630	79	103	120
..	...	...	...	...	...	...	...
788	Bergmite	Ice	NaN	304	55	69	85
789	Avalugg	Ice	NaN	514	95	117	184
794	Zygarde50% Forme	Dragon	Ground	600	108	100	121
795	Diancie	Rock	Fairy	600	50	100	150
799	Volcanion	Fire	Water	600	80	110	120

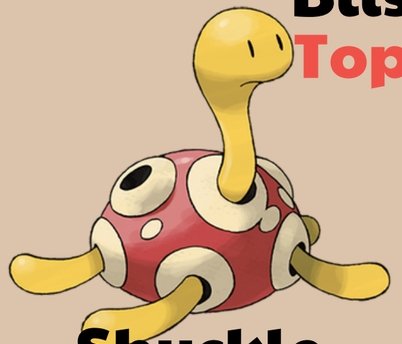
Cluster 1 (Blue):
Contains data with **low** to **medium** '**Attack**' values (around 0-75) and '**Defense**' values **tending** to be **low** to **medium** (around 0-75). This cluster **seems** to **contain** Pokemon with **lower** and **balanced** attack and defense **abilities**.

Cluster 2 (Green):
Contains data with **medium** to **high** '**Attack**' values (around 50-125) and **medium** to **high** '**Defense**' values (around 50-200). This cluster **indicates** that there are Pokemon with **medium** to **high attack** and **defense** that also t**end to be high**.

Top Highest

Mega Mewtwo X
Top Attack

Blissey
Top HP

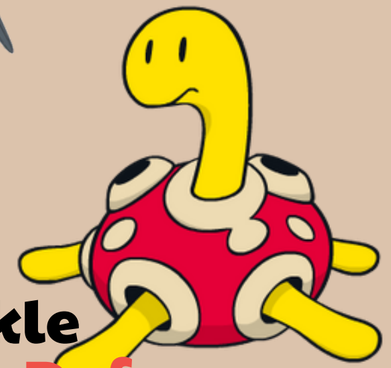


Shuckle
Top Defense

TOP TOTAL

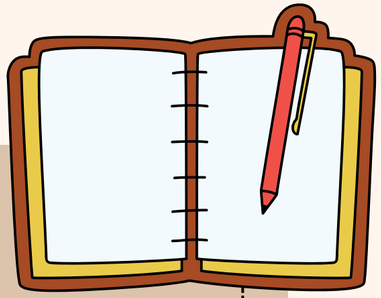
Mega Mewtwo Y
Top Sp. Atk

Deoxys
Top Speed



Shuckle
Top Sp. Def

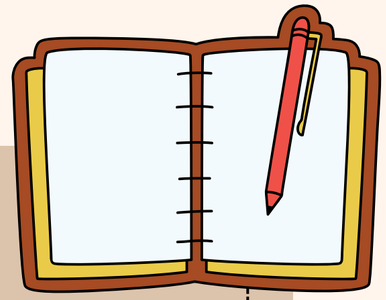
Conclusion



From the **overall analysis**, it can be concluded that **Generation 1** has the **highest number** of Pokémon. The element **'Water'** is the **most** commonly **found**, followed by '**Flying**' from '**Type 2**'. In addition, the number of **legendary Pokémon** is **much** less **compared** to **non-legendary**, indicating that legendary Pokémon are **very rare**.

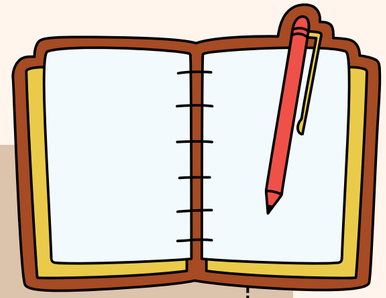
Then, the **clustering performed** can **help** group Pokémon with **similar** statistics. This can provide **insight** into the common characteristics among the group,,,, and Pokémon with the **highest scores** in certain statistics have **great** potential in game **strategy**.

Recommendation



1. **Use** Pokemon from '**Cluster 0**' for the **attack**.
2. **Utilize** Pokémon with the **highest stats** according to their **element**.
3. **Combine** Legendary Pokémon with Non-Legendary that have **certain advantages** for a **more effective strategy**.

Appendix



GOOGLE COLAB

GITHUB

Lets Work Together

: awnana123@gmail.com

: +62 812 1845 1055

: /in/nana-caw/

THANK YOU SO MUCH!