

Pokemon with stats

This data set includes 800 Pokemon, including their number, name, first and second type, and basic stats: HP, Attack, Defense, Special Attack, Special Defense, and Speed. It has been of great use when teaching statistics to kids. With certain types you can also give a geeky introduction to machine learning.

These are the raw attributes that are used for calculating how much damage an attack will do in the games. This dataset is about pokemon games (NOT pokemon cards or Pokemon Go).

The data as described:

#: 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

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

Decide the project objective(s): What do you want to tell using the data?

1. Find the top 10 Type 1 that yielded the highest average "total" in Pokemon and the number of each type 1.
2. Find type 1 and type 2 pair recorded the top 10 highest average "total" and each type 1 and 2 pair recorded.
3. Find the avg total of Legendary and non-Legendary of Pokemon and the number of them.
4. Find the name and number that the Legendary column is False have the total greater than the avg total of Legendary Pokemons.
5. Find the relationship between attack and the avg total.

Collect metadata of the dataset(s)

The number of entries: 800

Features: all data info of Pokemon

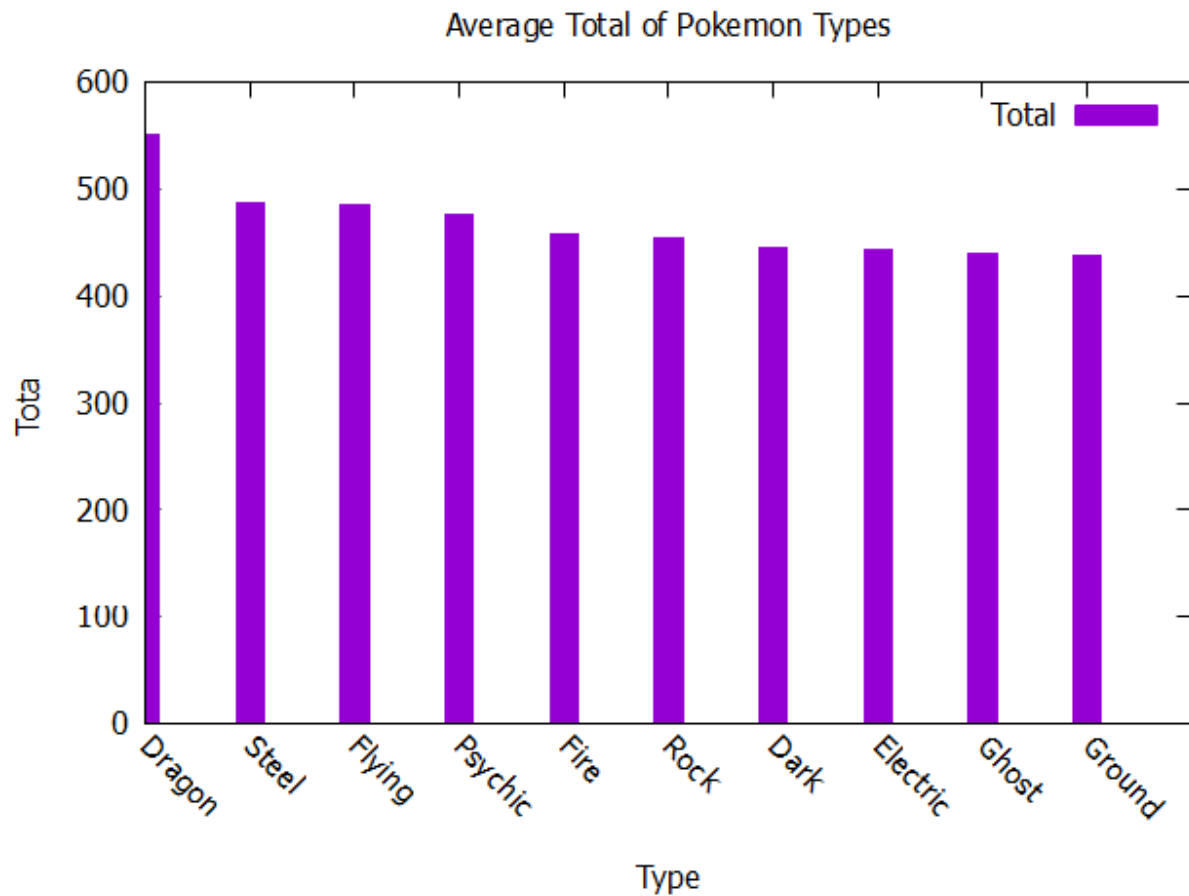
Value ranges: 5-780

Extract "nontrivial" information from the dataset.

What can you say from the data?

1. Find the top 10 Type 1 that yielded the highest average "total" in Pokemon and the number of each type 1.

Dragon type has the most avg total points.



2. Find type 1 and type 2 pair recorded the top 10 highest average "total" and each type 1 and 2 pair recorded.

Ground-Fire has the most avg total points with 2 different types of Pokemon.

3. Find the avg total of Legendary and non-Legendary of Pokemon and the number of them.

The difference of avg total points between Legendary and non-Legendary Pokemon is around 200.

637.385 True 65

417.214 False 735

4. Find the name and number that the Legendary column is False have the total greater than the avg total of Legendary Pokemons.

Those 6 Pokemons have higher total points than the avg total points of Legendary Pokemons.

GyaradosMega Gyarados

TyranitarMega Tyranitar

Slaking

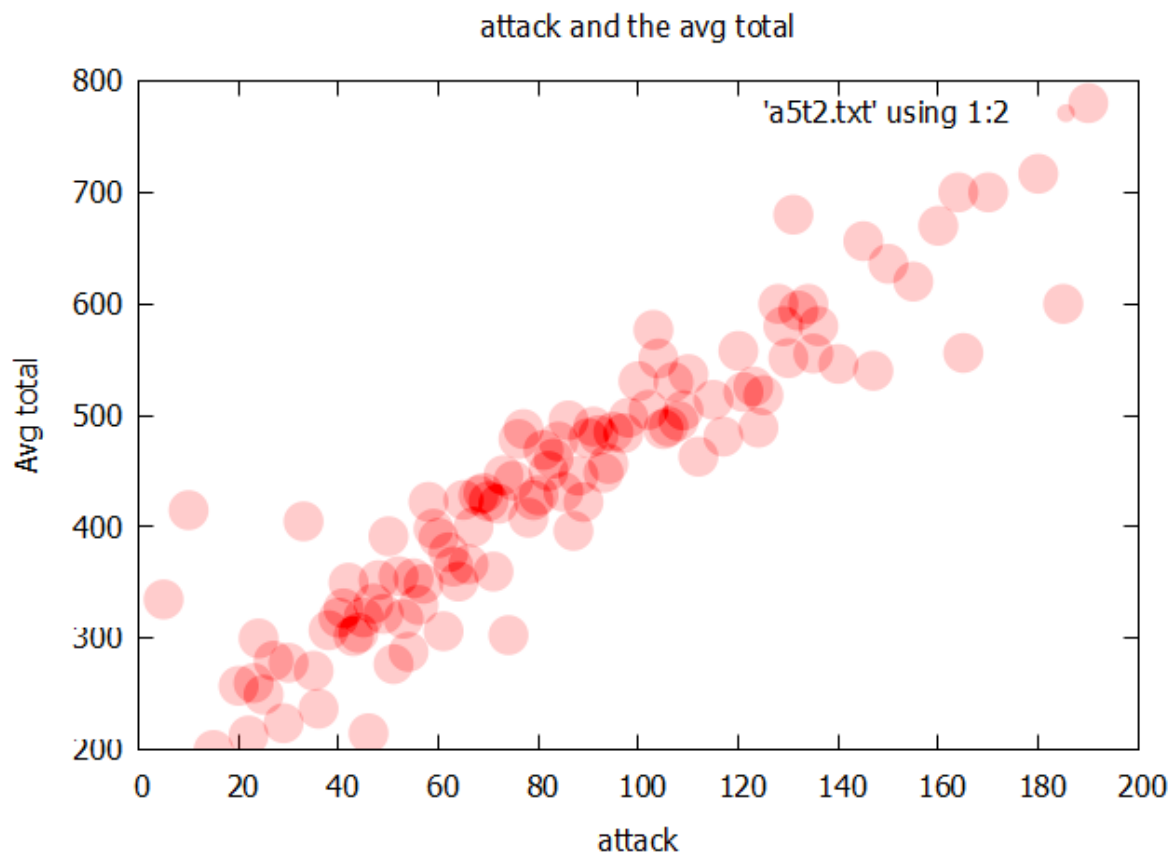
SalamenceMega Salamence

MetagrossMega Metagross

GarchompMega Garchomp

5. Find the relationship between attack and the avg total.

The relationship between attack and the avg total is linear.



Any interesting data trends that cannot be easily guessed or can be used to justify some hypothesis?

How many Non- Legendary Pokemons have great talent?

Those 6 Pokemons have higher total points than the avg total points of Legendary Pokemons.

GyaradosMega Gyarados

TyranitarMega Tyranitar

Slaking

SalamenceMega Salamence

MetagrossMega Metagross

GarchompMega Garchomp

Commands I used for this project:

1. find the top 10 Type 1 that yielded the highest average "total" in Pokemon and the number of each type 1.

```
cut -d, -f3,5 Pokemon.csv > temp.txt
```

```
awk -F, 'NR>1 {sum[$1]+=$2; count[$1]++} END {for (i in sum) print sum[i]/count[i], i, count[i]}'  
temp.txt > temp1.txt
```

```
sort -nr temp1.txt | head -10 > a5.txt
```

2. type 1 and type 2 pair recorded the top 10 highest average "total" and each type 1-2 pair recorded.

```
cut -d, -f3,4,5 Pokemon.csv > temp.txt
```

```
awk -F, -v OFS=' ' '{print $1"-"$2,$3}' temp.txt > temp1.txt
```

```
awk -F, 'NR>1 {sum[$1]+=$2; count[$1]++} END {for (i in sum) print sum[i]/count[i], i, count[i] }'  
temp1.txt > temp2.txt
```

```
sort -nr temp2.txt | head -10 >> a5.txt
```

3. The avg total of Legendary and none-Legendary of Pokemon and the number of them.

```
cut -d, -f13,5 Pokemon.csv > temp.txt
```

```
awk -F, 'NR>1 {sum[$2]+=$1; count[$2]++} END {for (i in sum) print sum[i]/count[i], i, count[i]}'  
temp.txt >> a5.txt
```

637.385 True 65

417.214 False 735

4. Find the name and number that the Legendary column is False have the total greater than the avg total of Legendary Pokemons.

```
awk -F, '$13 == "False" && $5 > 637.385 {count++; print $2} END {print "Number of none-Legendary:"  
count}' Pokemon.csv >> a5.txt
```

5. Find the relationship between attack and the avg total

```
cut -d, -f5,7 Pokemon.csv > temp.txt
```

```
awk -F, 'NR>1 {sum[$2]+=$1; count[$2]++} END {for (i in sum) print i, sum[i]/count[i] }' temp.txt >> a5.txt
```

Data reference:

[Pokemon with stats \(kaggle.com\)](https://www.kaggle.com/datasets/pschmitt/pokemon-with-stats)