Pokémon Strength Analysis

Trying to identify the strongest Pokémon through K-Means Clustering

Daniel Peslherbe - dpeslherbe.wixsite.com/website

Summer 2020

Introduction

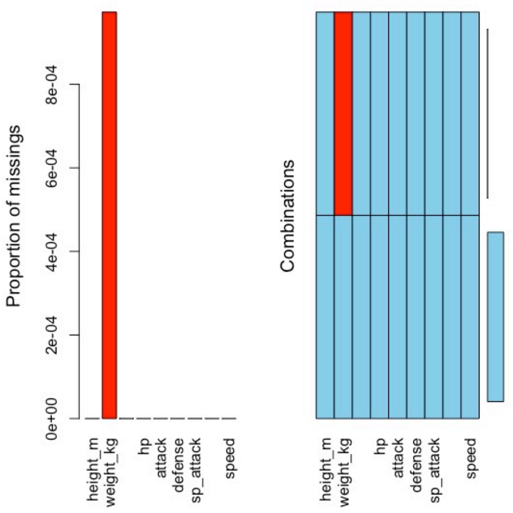
To most people belonging to Generation X, and Generation Y, Pokémon were a fun game playable on Nintendo gaming systems through the years where the goal was to acquire Pokémon creatures to complete a game's Pokédex, as well as use the captured Pokémon in trainer battles. While some may have outgrown this phase of gaming, for many others, Pokémon games and battles have reached a competitive status with the availability of online play. But the question remains; Which Pokémon are the strongest, and which are worth it to evolve or keep it to build a competitive 6 Pokémon lineup ?

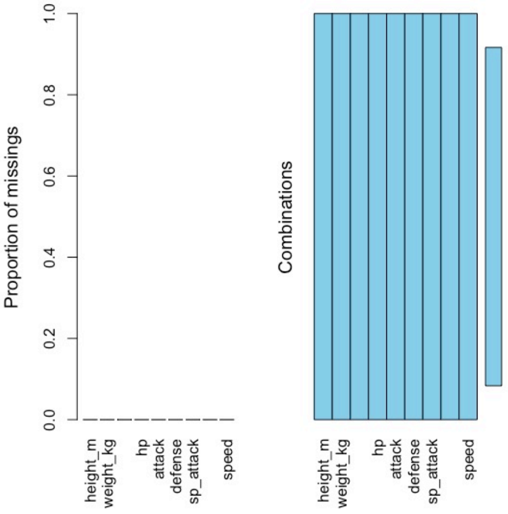
Methodology

To try and group Pokémon according to overall battle capacity, we will perform K-Means Clustering (a machine learning unsupervised algorithm that regroups data points into a given the number of clusters that are based on variable similarity). For the data used, we used the Complete Pokémon Dataset updated on 05.20 (available for download at https://www.kaggle.com/mariotormo/complete-pokemon-dataset-updated-090420). We used the R program in combination with the VIM, cluster, corrplot packages.

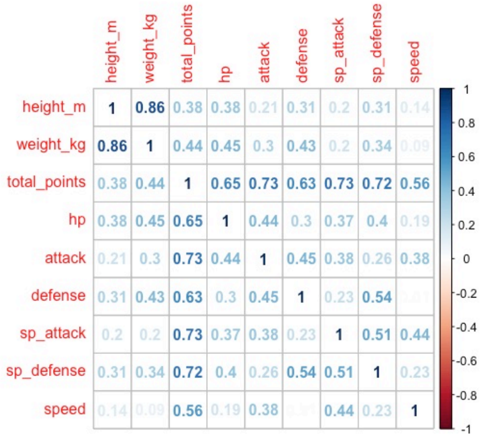
Note we retain only the important variables for our analysis and then scale them to reduce size effects. The variables used in our analysis are height (in m), weight (in kg), total points, hp, attack, defence, sp. attack, sp. defence, and speed. Note that for

Eternatus Eternamax, whose weight is not specified, we scale it in comparison to Eternatus (with the same weight-to-height ratio).

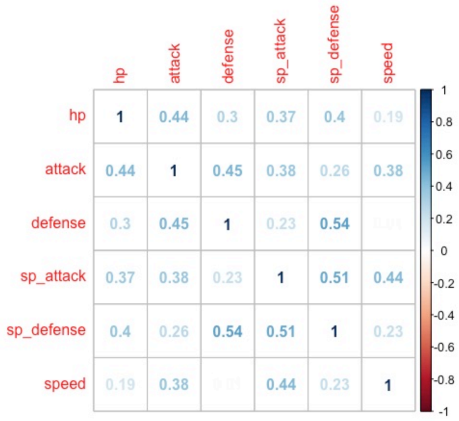




However, through our correlation plot, we notice that height and weight are highly correlated, and that total points are highly correlated with all other variables, so we shall drop them from our model analysis.

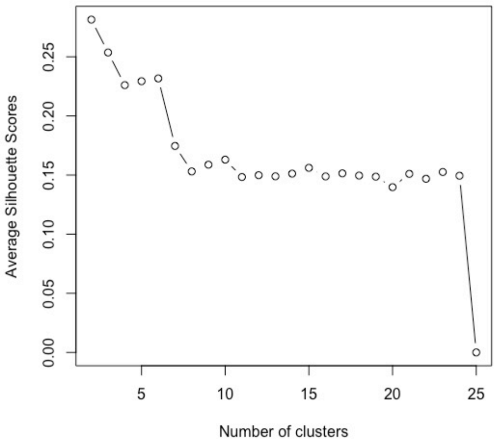


	height_m	weight_kg	total_points	hp	attack	defense	sp_attack	sp_defense	speed
height_m	1	0.86	0.38	0.38	0.21	0.31	0.2	0.31	0.14
weight_kg	0.86	1	0.44	0.45	0.3	0.43	0.2	0.34	0.09
total_points	0.38	0.44	1	0.65	0.73	0.63	0.73	0.72	0.56
hp	0.38	0.45	0.65	1	0.44	0.3	0.37	0.4	0.19
attack	0.21	0.3	0.73	0.44	1	0.45	0.38	0.26	0.38
defense	0.31	0.43	0.63	0.3	0.45	1	0.23	0.54	0.01
sp_attack	0.2	0.2	0.73	0.37	0.38	0.23	1	0.51	0.44
sp_defense	0.31	0.34	0.72	0.4	0.26	0.54	0.51	1	0.23
speed	0.14	0.09	0.56	0.19	0.38	0.01	0.44	0.23	1



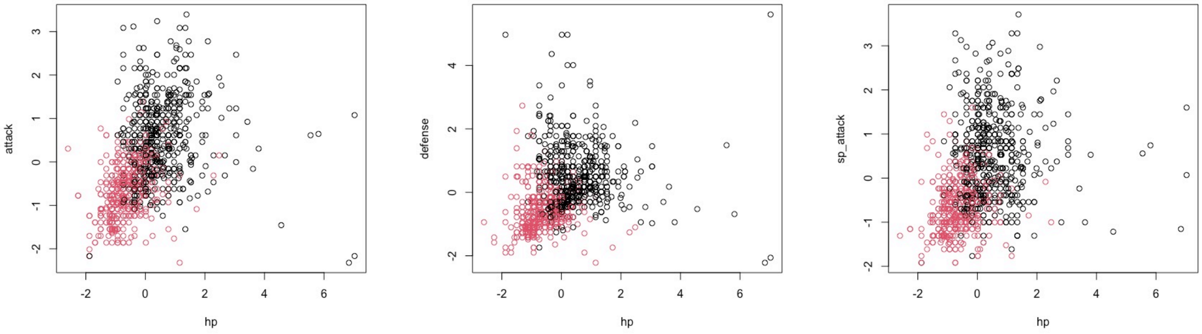
	hp	attack	defense	sp_attack	sp_defense	speed
hp	1	0.44	0.3	0.37	0.4	0.19
attack	0.44	1	0.45	0.38	0.26	0.38
defense	0.3	0.45	1	0.23	0.54	0.01
sp_attack	0.37	0.38	0.23	1	0.51	0.44
sp_defense	0.4	0.26	0.54	0.51	1	0.23
speed	0.19	0.38	0.01	0.44	0.23	1

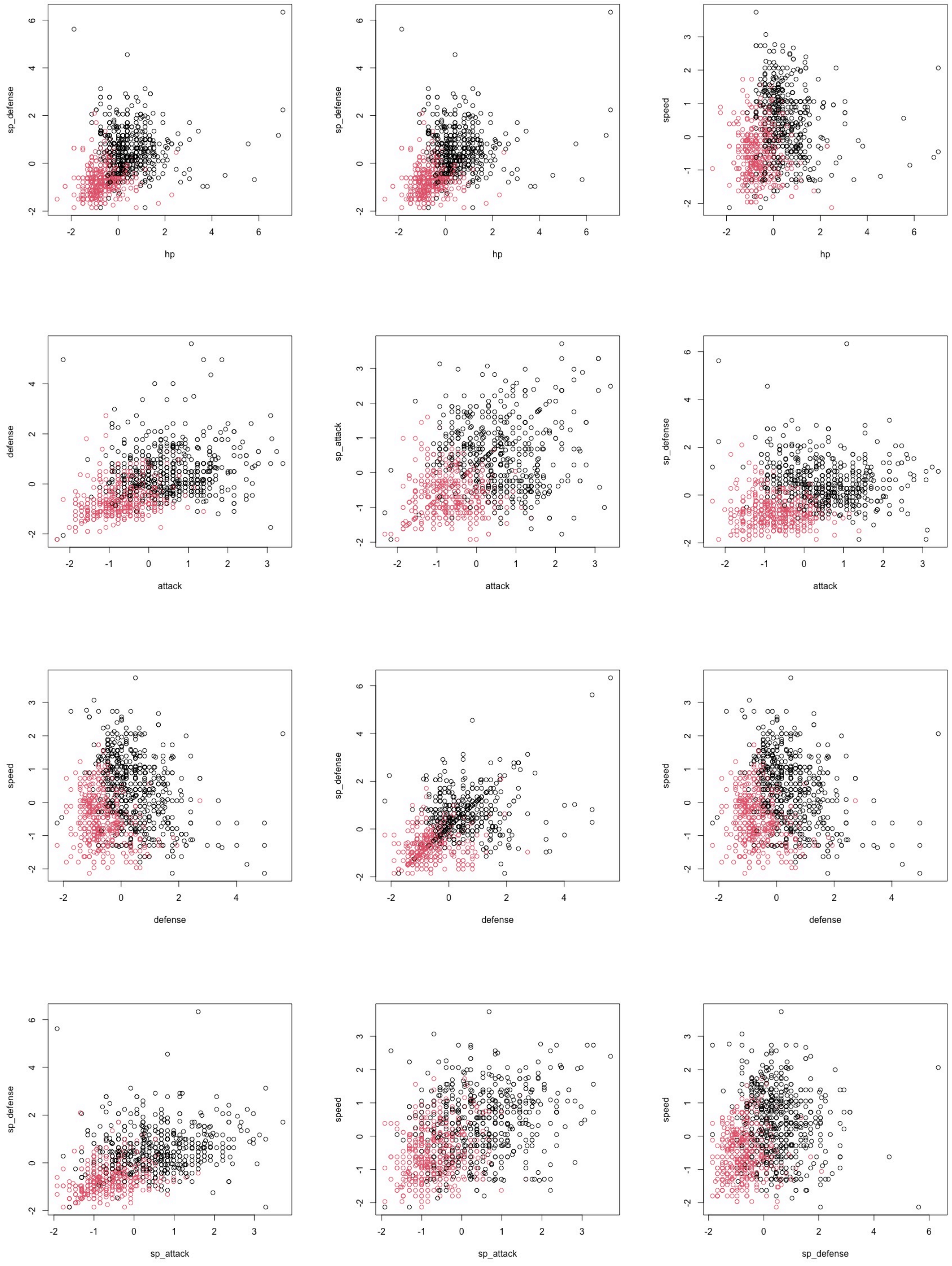
Then, using the kmeans( ) and silhouette( ) functions, we create a function silhouettrescore( ) which calculates the average silhouette score for a certain number of clusters on our dataset. Then we use this function to calculate the average silhouette score for *k* (number of clusters) ranging from 2 to 25. We then choose the number of clusters where the average silhouette score is highest, which is for 2 Clusters here.



Results (2 Clusters)

Running our code separates our dataset into two clusters, separated by colours (red being the second cluster, and black being the first) on the following graphs which compare different attributes against each other (health vs attacks, health vs defence, and so on…)

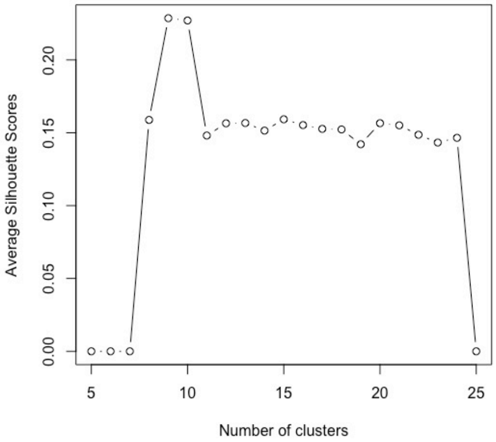



sp_defense
6
4
2
0
-2
-2
0
2
4
6
hp
sp_defense
6
4
2
0
-2
-2
0
2
4
6
hp
speed
3
2
1
0
-1
-2
-2
0
2
4
6
hp
defense
4
2
0
-2
-2
-1
0
1
2
3
attack
sp_attack
3
2
1
0
-1
-2
-2
-1
0
1
2
3
attack
sp_defense
6
4
2
0
-2
-2
-1
0
1
2
3
attack
speed
3
2
1
0
-1
-2
-2
0
2
4
defense
sp_defense
6
4
2
0
-2
-2
0
2
4
defense
speed
3
2
1
0
-1
-2
-2
0
2
4
defense
sp_defense
6
4
2
0
-2
-2
-1
0
1
2
3
sp_attack
speed
3
2
1
0
-1
-2
-2
-1
0
1
2
3
sp_attack
speed
3
2
1
0
-1
-2
-2
0
2
4
6
sp_defense

We notice from the cluster means that the first cluster correspond to Pokémon that are on the higher end of the variables (positive average cluster means for each of hp, attack, defense, sp_attack, sp_defense, and speed).

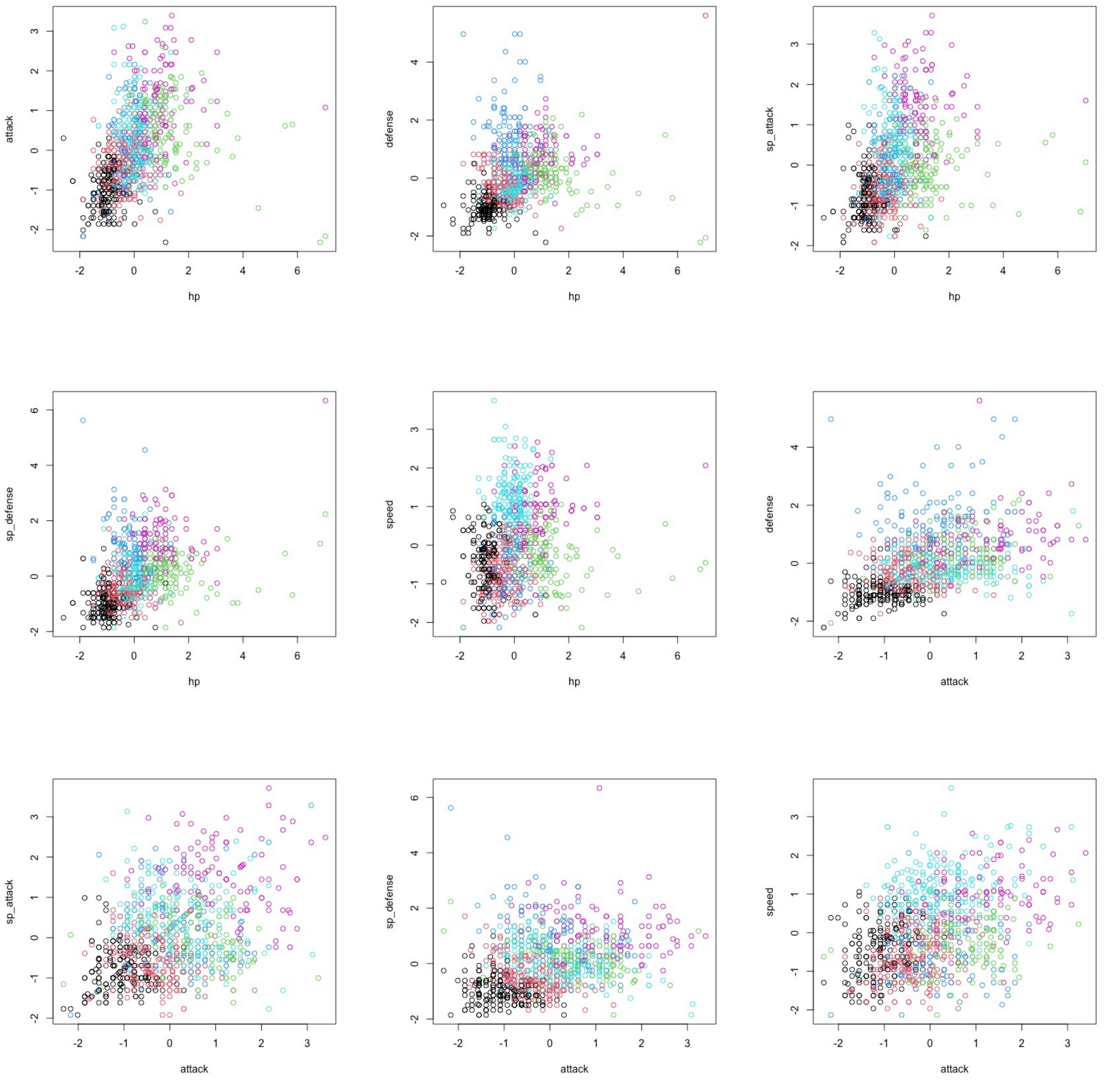
Cluster	Health	Attack	Defense	Sp. Attack	Sp. Defense	Speed
1	0.4952192	0.5300545	0.4840779	0.5136510	0.5449158	0.3930374
2	-0.6261141	-0.6701571	-0.6120280	-0.6494177	-0.6889465	-0.4969239

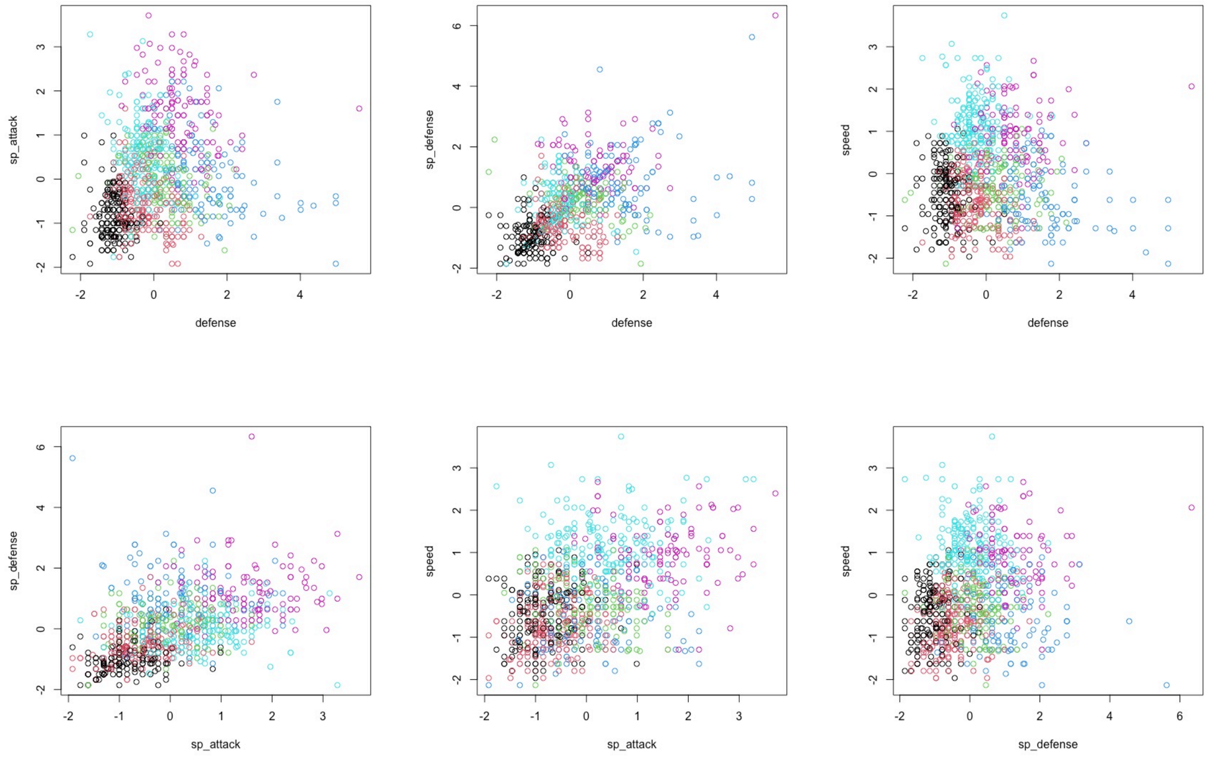
However, this only separates strong Pokémon from weak Pokémon. The issue this brings is that most of first cluster Pokémon are usually either the last Evolution or Mega-Evolution of weaker Pokémon; thus, while the average silhouette score is highest when there are 2 clusters, this is not enough to cluster Pokémon for decision making for competitive Pokémon Trainers. Then, to remedy this, we reuse our silhouettescore( ) function, but using *k* clusters ranging from 5 to 25 instead. The best value for cluster numbers is then 6.



Results (6 Clusters)

This time around, the graphs are as follows; (with black representing the first cluster, red representing the second, green representing the third, dark blue representing the fourth, light blue representing the fifth, and lilac representing the sixth).





We see that the clusters seem to correspond to Pokémon with specific stat differences. From the cluster means, we have the following average for each variable according to clusters:

Cluster	Health	Attack	Defense	Sp. Attack	Sp. Defense	Speed
1	-0.99301457	-1.04617375	-1.0547421	-0.826741073	-0.99695878	-0.4721187
2	-0.36316077	-0.38754276	-0.2727654	-0.577229611	-0.50481763	-0.6724617
3	1.31017450	0.60454396	0.2477921	-0.008192179	0.18887565	-0.3118777
4	-0.16812378	0.08094902	1.4814636	0.081438126	0.91156089	-0.5497746
5	-0.04680911	0.24415377	-0.2440637	0.437618752	0.04878462	1.1131722
6	0.92175938	1.09495521	0.7294563	1.447424997	1.18608246	0.9190101

Thus, we notice for example that the first cluster regroups only the weakest Pokémon with very low averages in all statistical categories (trainers will want to avoid using these Pokémon in a competitive setting, unless they have an Evolution that falls into a better cluster and are battling their way into the Evolution). The second cluster regroups another set of weak Pokémon; however the Pokémon in this cluster have higher average stats in every category than the first cluster, with speed being the exception. We can then infer that the first cluster consists of smaller weak Pokémon, whereas the second regroups larger weak Pokémon. Thus we name the first cluster the Small Weak, and the second cluster the Big Weak.

The third cluster regroups Pokémon with the highest average health statistics as well as above average attack, defence, and sp_defense, while being below average in sp_attack, and speed; we can infer that these are slow large Pokémon with high physical attributes, and shall name this cluster the Health Tanks (based on the usual meaning in gaming of tanks, who are large characters with high health and physical attributes, but also usually not as quick as other player types, also sometimes referred to as meat sponges for their ability to receive a fair amount of damage).

The fourth cluster consists of Pokémon with the highest defence and sp_defense stats, while their health and speed points are below average. The attack and sp_attack points are just barely above average and do not seem to be huge factor in the cluster. Given these characteristics, we rename this cluster the Defensive Specialists (given their propensity for defensive attributes).

The fifth cluster corresponds to Pokémon with the highest average speed points, as well as above average attack, sp_attack, and sp_defense attributes, with below average health and demesne points. Thus, this cluster is named as Fast Attackers (since it contains speed focused Pokémon with above average attack power).

The sixth and last cluster corresponds to Pokémon who are above average in all point categories, where they have the highest average attack, sp_attack, and sp_defense, and the second highest health, defence, and speed points. In essence, these Pokémon have very little to no observable weaknesses stats wise, and are ideal for Competitive Pokémon Battles.

It is important to note however that this cluster analysis as revealed who are the strongest Pokémon when it comes to stat points, but to build the ultimate Competitive Pokémon, one must also consider the best combination of Pokémon types and abilities to create the best possible squad.

Note that to help trainers with decisions, the list of Pokémons including Name, Type(s), Pokédex Number, Abilities, and Cluster Classification, are available and exported in two .csv files named results and results2 (for the 2 cluster, and the 6 cluster results respectively) on the same Google Drive Link as this report, as well as all files used for this analysis and this report. Gotta Catch'Em All !