Gotta Catch 'Em All™ - Pokémon!

Final Project – Team 213

IYSE 6740 – Fall 2024

Andrew Dean

Introduction & Background

Part of my motivation to join the OMSA program was to pursue a job in the video game industry incorporating data analytics. I have a deep passion for video games and know that analytics will continually be evolving in this industry. With that said, I decided to look at one of the most popular video game franchises to analyze for this project – Pokémon.

For those who are not familiar with the game, Pokémon are creatures that a person can capture and train to fight other Pokémon. You start as a novice Pokémon trainer, and as you go along in the game, you capture more and more Pokémon, continually leveling them as they win battles, and then try to beat all the other trainers in the "Poke-verse" to become a Pokémon Master. The franchise has been around since 1996[1] and is one of the first video game franchises I fell in love with as a kid.

Each Pokémon can train up to 4 moves to fight with, are associated with a certain element called a "Type" (Water, Fire, Lightning, etc.), and have different combination of stats (health, strength, dexterity, etc.). A trainer can only hold up to 6 Pokémon at a time which comprises their Pokémon team.

Problem Statement

From my original proposal, my plan was to address two main questions / problems in this project:

1. **Is the game balanced?**

   In any video game, game balance is imperative to make it fun. Every video game company struggles with making a game as balanced as possible. You don't want the game to be too easy and not challenging. Conversely, you don't want a game to be too hard and frustrate your consumer. Therefore, we want to answer, is the game balanced? I attempted to answer that with the following analysis:

   a. What groups of Pokémon are the strongest? What groups of Pokémon are the weakest? Is the distribution well balanced? (Classification).

b. Originally, I was going to also look at power moves by Pokémon, but I could not develop a comprehensive dataset that would allow me to do this. I did find quite a few, but nothing that covered my entire dataset. However, I will touch on this to help solve my question from Part A here and it will still be considered in some light.

2. **What is the optimal 6-Pokémon Team?**

   Once we know what Pokémon and moves are the strongest, the goal will be to put together the optimal Pokémon team that we can.

In general, my assumption going into this was that I expected the game would show some imbalance particularly around 'Type'. On the other hand, I believe we can optimally put together a 6 Pokémon Team.

Data Overview

I pulled my data from a few sources here and combined them together. Below is a quick outline.

Pokémon Data – Pulled two data files that have a list of all Pokémon[2] from Kaggle.com. The combined data had the following variables for each Pokémon: *[Name, Type1, Type2, hp, attack, defense, special attack, special defense, speed, generation, legendary]*

Power Moves Data – Also pulled from Kaggle.com which had a list of Pokémon, and their associated power moves [3]. I then combined this with another file which has a list of all power moves along with each moves' stats[4]. This was pulled from bulbapedia.net. The power move stats were as follows: *[Name, Type, Power, Accuracy, PP].* I will note here that I ran into a snag as mentioned in my problem statement. I could not match power moves to specific Pokémon as there was a lot missing from one dataset or the other and the game has changed drastically over the years. With that said, I was not able to quite line up power moves for each Pokémon with their respective power move stats. However, I really wanted to still incorporate power moves, and I ended up taking a more overarching approach I will detail later on.

Methodology & Approach

My general methodology to solve this problem was going to put Pokémon in a tier list based on their stats. The best Pokémon would be considered 'A Tier' and the worst Pokémon would be 'C Tier' with everything else landing in 'B Tier'. I decided I was going to do this by conducting Principal Component Analysis to reduce my feature variables and then using k-means clustering to see if I could generate this tier list through machine learning. I also planned on doing one-hot

encoding for my categorical variables as necessary. I then planned on doing a similar methodology for power moves.

My assumption was, if the game was balanced, I would assume the clusters would be very similar in stats and that the distribution of the number of Pokémon would be consistent. In addition, I had an assumption going into this that Pokémon 'Type' would show some imbalance. Meaning that certain 'Types' were generally going to be much stronger than others. So once I have my tiers, we can look at the difference in number of 'Types' in each cluster to see if there is a material difference in their distributions.

My final assumption was that Pokémon with, for example, higher hp would likely not be as fast or maybe not attack as hard. Conversely, strong damage Pokémon I would have assumed likely give up some speed or defense for the tradeoff of higher damage.

Once I had my tiers, I would have certain criteria to build my 6-Pokemon team and optimize as best as I could.

Evaluation

Pokémon Tier List

I first pulled in the formula from bulbapedia.net on damage output from a Pokémon when it conducts an attack. The formula [5] is as follows:

$$Damage = \left(\frac{\left(\frac{2 \times Level \times Critical}{5} + 2 \right) \times Power \times A/D}{50} + 2 \right) \times STAB \times Type1 \times Type2 \times random$$

Snippet courtesy of bulbapedia.net

Here is the overview of the variables in this formula:

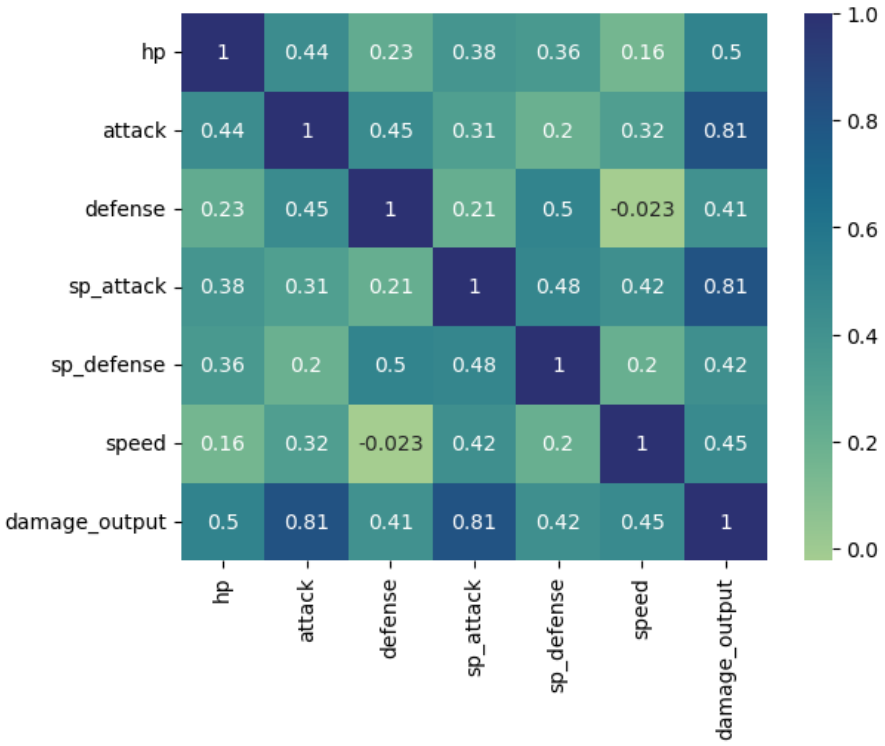
- **Damage** = The total amount of damage a Pokémon will do when they execute a move.
- **Level** = The level of the Pokémon. For this exercise, we will assume the level stays consistent at level 100.
- **Critical** = This number is 2 when it is a critical strike, 1 otherwise. For this analysis, we will assume all moves are non-critical strikes.
- **Power** = is the power of the power move. I ended up using a mean here due to my dataset not being able to tie to specific Pokémon.
- **A** = This is the "attack" of a move. Special Attack is also sometimes used depending on the move. For this analysis, we will just assume a 50/50 split between the two.

- **D** = Defense of the target. For this analysis, we will use the mean defense of all Pokémon and use this as a static number. Similar to “Attack”, there is also Special Defense, and we will again split as 50/50.
- **STAB** = This is 1.5 if the move type matches the Pokémon type, 1.0 otherwise. For this analysis we will assume it is always 1.5 and that the Pokémon uses a move of its type.
- **Type 1 & 2** = These determine the effectiveness of the move against a certain type of Pokémon. It can range from 0.5 - 2.0. For this analysis, we will assume these are both 1.0 and will not take into account how effective a move is. However, we will consider this when optimizing a team composition.
- **random** = An integer between 217 to 255 and then divided by 255. We will just assume this also just stays at 1.0 for this analysis.

With the above formula defined, I was able to calculate ‘Damage_output’ for each Pokémon and create that variable.

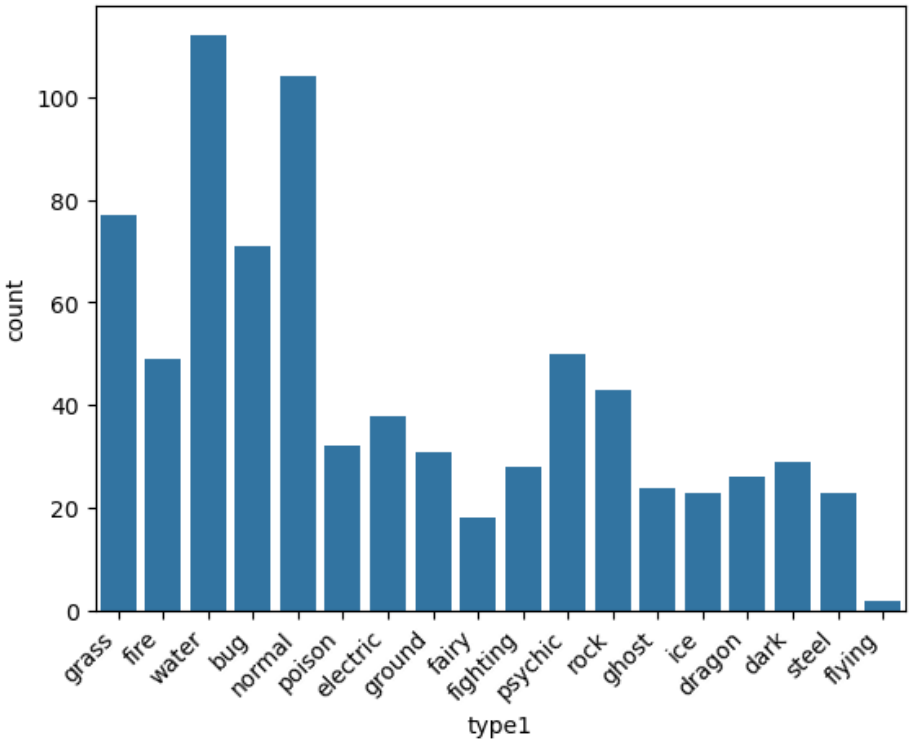
However, it is also important to consider a Pokémon's other stats like its hp, speed, and defense overall. While the damage formula is highly being driven by “Attack” and “Special Attack” these other variables should still affect a Pokémon's tier.

I then conducted some simple exploratory data analysis to see correlation and number of Pokémon by type. Below is the correlation matrix for the non-categorical variables.



Looking at the correlation, we obviously see a big relationship between Attack/Sp_attack and Damage_output. There is no surprise there. The rest of the variables, however, seem relatively uncorrelated, which is great for our dataset. I was a little surprised to see that we didn't have a more negative correlation in relation to some variables. For example, a Pokémon that has less damage I would've expected to have higher defense or higher hp and that just doesn't seem to be the trend based on this correlation matrix. I would argue this may point to some imbalance between Pokémon; supporting my original hypothesis. There theoretically should be tradeoffs in running a higher damage Pokémon, like less speed or defense, and that doesn't seem to be the case necessarily.

Now let's turn to Pokémon by type.

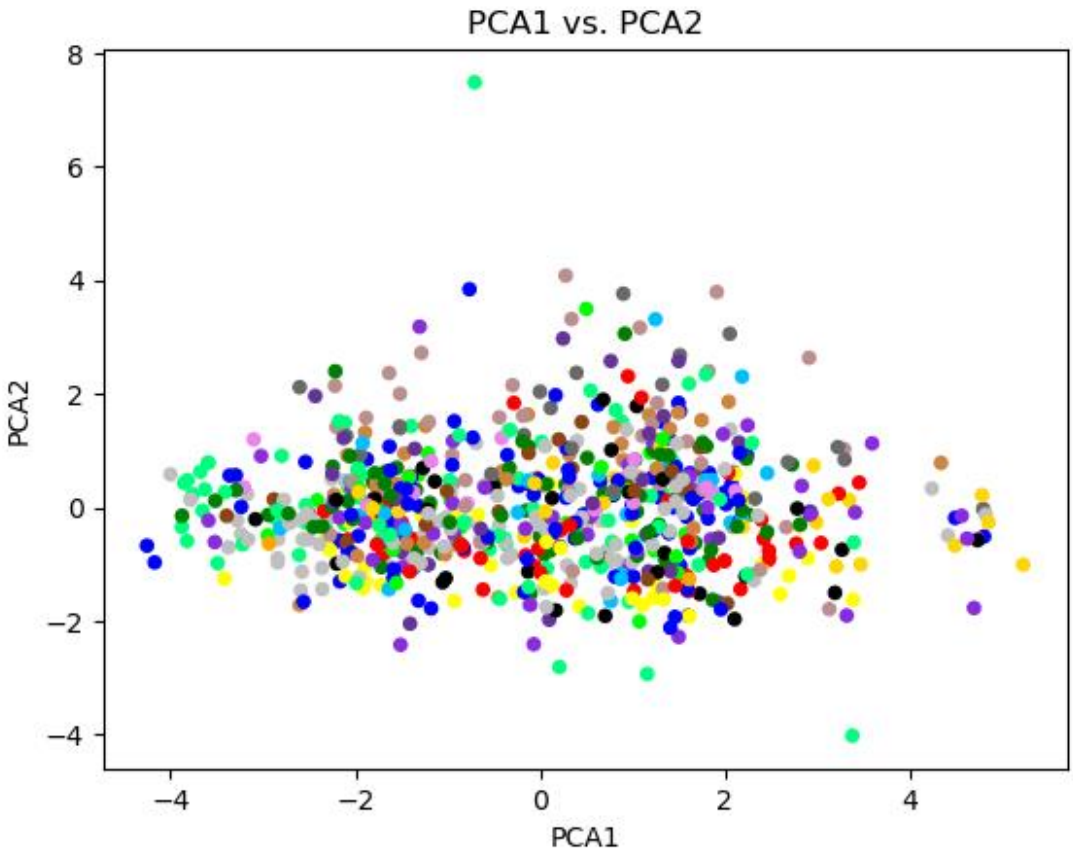


Looking at the data, we see that the distribution of Type is actually quite skewed. For example, there are many more water and normal type Pokémon compared to other types. We will dive into this later to determine if this matters, but something to note on the analysis as a whole.

Next I split my data column *'Type'* into a categorical variable by one-hot encoding them into their respective columns. If it was a "*Fire Type*" I made a column called *'Fire'* and then have 1's and 0's. I did this for each type. I also changed 'legendary' into a binary column instead of 'Yes' or 'No'. I then standardized my non-categorical feature variables and conducted PCA to reduce my feature variables into 2-dimensions. Below were my final feature variables I used:

['hp', 'attack', 'defense', 'sp_attack', 'sp_defense', 'speed','damage_output', 'legendary','ground', 'psychic', 'steel', 'fairy', 'grass', 'dark', 'normal','poison', 'fighting', 'water', 'flying', 'electric', 'ghost', 'rock', 'ice', 'dragon', 'bug', 'fire']

After conducting PCA, I then pulled my top two eigenvectors and plotted them against each other and obtained the graph below. I also color coded each Pokémon type to see if we could see any patterns.



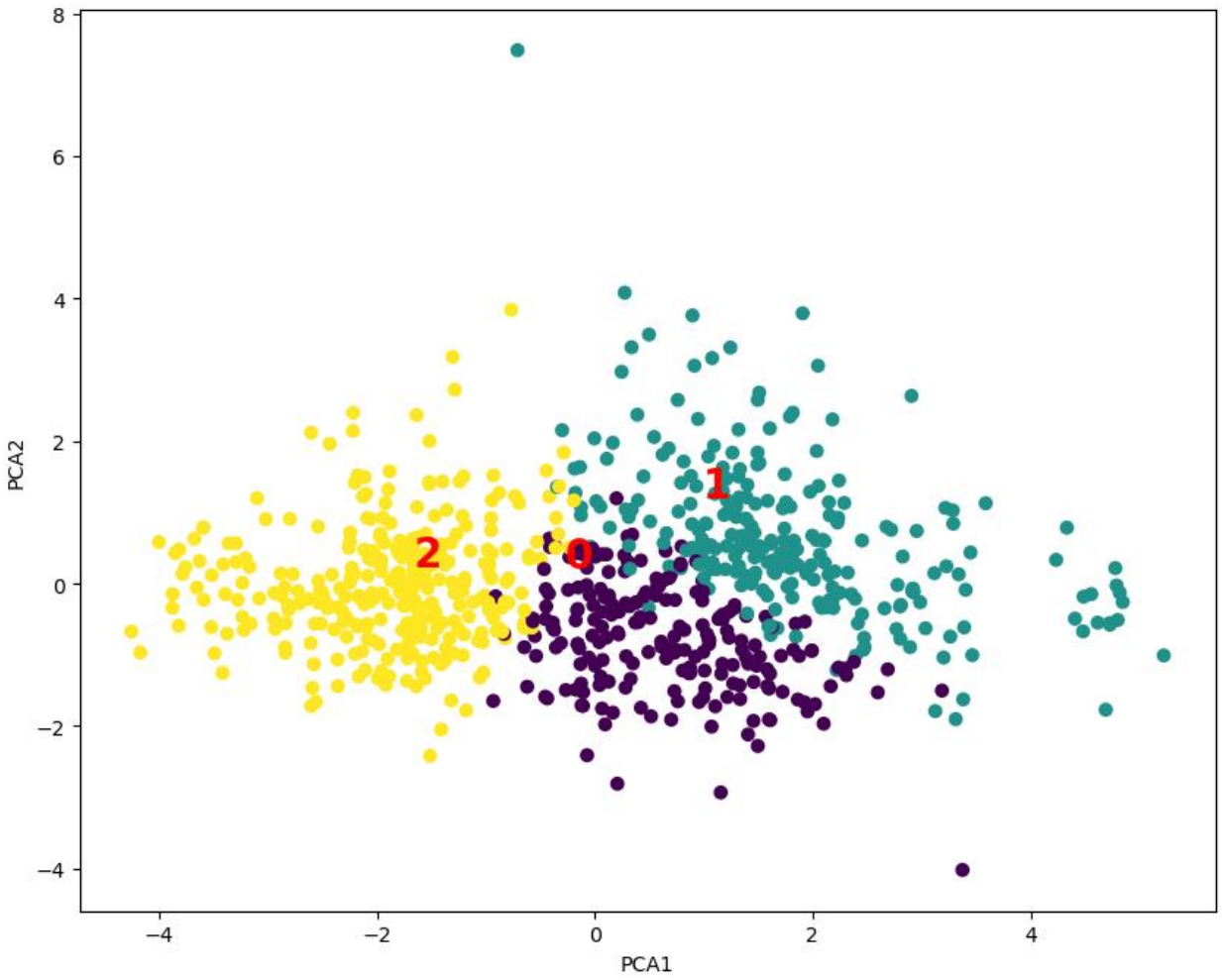
It is good to see that there aren't too many outliers. There are a couple, but seemingly within reason for a dataset this size. This would point to likely some sort of balance in the game and would be a consideration for rejecting my original hypothesis.

However, I then did a PCA chart by 'Type' to really see if there is a difference in the shape of each 'Type'. This can be found in the appendix in Figure 1.

The distribution by type is very different. So conversely while the universe of Pokémon seems to be distributed well, by 'Type', we see the opposite. This would point to some potential game imbalance as well; again, supporting my hypothesis.

After looking at this, I decided to go ahead with my k-means analysis to see if I could come up with a tier list.

After playing around with different clusters, I ended up using 3 as that seemed to be the best in terms of explaining variance. Below were the results of a linear k-means clustering with 3 clusters. The red numbers represent the cluster number that represent each other:



We see that the clusters are pretty well defined. I also attempted to run a Gaussian Mixture Model to see if the classification would look any better, but it ended up producing a worse result in terms of classifying. The result can be found in the Appendix in Figure 2.

Now that we have pretty distinct clusters, let's look at the descriptive statistics in each one to determine which cluster is A Tier, B Tier, and C Tier. For this, we are going to compare the means of each cluster.

Cluster 0	Pokemon Total = 215							
number	hp	attack	defense	sp_attack	sp_defense	speed	total_defense	damage_output
mean	69.5	79.0	66.1	80.0	73.0	89.5	298.2	132.9

Cluster 1	Pokemon Total = 252							
number	hp	attack	defense	sp_attack	sp_defense	speed	total_defense	damage_output
mean	88.4	99.3	97.1	85.0	89.3	63.9	338.7	153.5

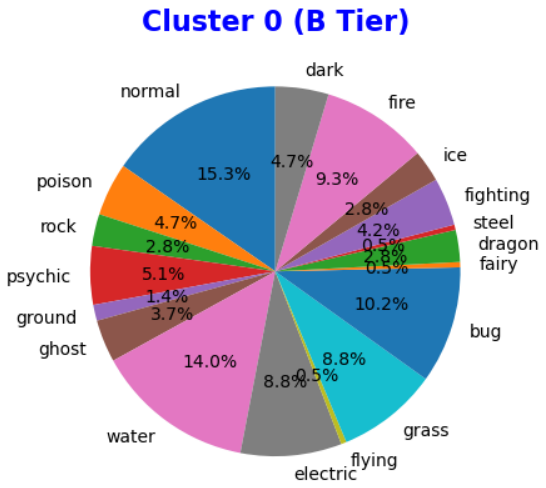
Cluster 2	Pokemon Total = 313							
number	hp	attack	defense	sp_attack	sp_defense	speed	total_defense	damage_output
mean	51.2	53.6	53.7	47.6	50.7	48.8	204.3	85.7

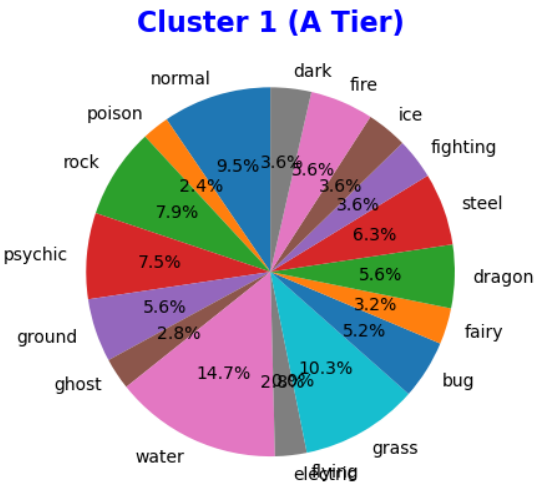
Overall Population	Pokemon Total = 780							
number	hp	attack	defense	sp_attack	sp_defense	speed	total_defense	damage_output
mean	68.3	75.4	71.1	68.6	69.3	64.9	273.6	120.6

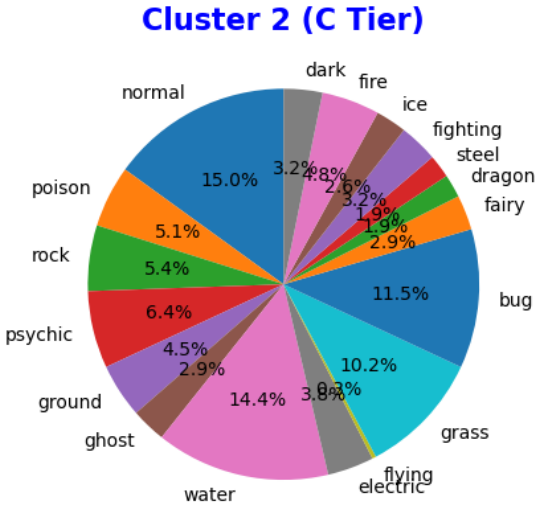
In general, we see that Cluster 1 by far has the most superior stats. It has the highest stats in essentially every category but 'Speed'. We also see that Cluster 2 really has no superior stats whatsoever and ranks dead last in every category. The number of Pokémon has some variation; however, I would say it is relatively even between the three. Closer than I had anticipated. In general, based on the above, I think it is safe to say that Cluster 1 is our 'A Tier', Cluster 0 is our 'B Tier', and Cluster 2 is our 'C Tier'.

It should also be noted at this point, that Pokémon can evolve in the game. For example, a Bulbasaur is a basic Pokémon, but as it levels, it can evolve into Ivysaur, and then eventually Venusaur. As it evolves, it gets more powerful. If we look at this evolution, Bulbasaur is in C Tier. Ivysaur is in B Tier, and finally, Venusaur happens to be in the A Tier. It is likely that a lot of our final evolution Pokémon can be found in A Tier which is good. You're typically not going to keep a Bulbasaur for the entire game. So, when looking at balance, one could argue you want to look at different stages of the game. Early game, you may have Pokémon only in C Tier, but as you progress, your team should start to get closer to A Tier naturally. Since we are looking at level 100 Pokémon only, my analysis will really only look at late-game team compositions. But overall, having a similar number of Pokémon in each tier is a good sign that this is more likely due to progression in the game versus balance or imbalance. Hence, going against my hypothesis.

However, let's turn to 'Types' and see if there is any skew in that within the tiers. Below are the distributions of each tier and the respective Type.







Based on the above, here is what we see and can conclude:

Differences	Types	Notes
Significant	Steel, Dragon, Rock Bug, Normal	A Tier is significantly stronger than B & C Mostly represented in B & C Tier
Medium	Psychic, Ground, Fairy Poison Electric	A & C Tier are more represented than B Tier A Tier is less represented than B & C Tier Mostly represented in B Tier
Normal	Dark, Ice, Fighting, Grass, Water, Ghost, Flying. Fire	No material differences between cluster

In general, most tiers are relatively balanced. There are a few significantly different (Steel, Dragon, Rock, Bug, & Normal), but overall, most seem within reason within a +/- 5% difference. Since we are specifically looking at late-game level 100 Pokémon, let's specifically look at the A Tier distribution against the Total distribution. The reason for doing this is we know from earlier, there are more 'Types' of certain Pokémon in the game versus others. By looking at the comparative distribution, I would expect A Tier to be similarly distributed as the Total population. Here is a table representing the two.

	Norm	dark	fire	ice	Fight	steel	Drag	fairy	bug
A Tier	9.50%	3.60%	5.60%	3.60%	3.60%	6.30%	5.60%	3.20%	5.20%
Total	13.30%	3.70%	6.30%	2.90%	3.60%	2.90%	3.30%	2.30%	9.10%

Difference From A Tier	-3.80%	-0.10%	-0.70%	0.70%	0.00%	3.40%	2.30%	0.90%	-3.90%

	grass	Fly	Elec	water	ghost	ground	Psy	rock	poison
A Tier	10.30%	0.00%	2.80%	14.70%	2.80%	5.60%	7.50%	7.90%	2.40%
Total	9.90%	0.30%	4.90%	14.40%	3.10%	4.00%	6.40%	5.50%	4.10%

Difference From A Tier	0.40%	-0.30%	-2.10%	0.30%	-0.30%	1.60%	1.10%	2.40%	1.70%

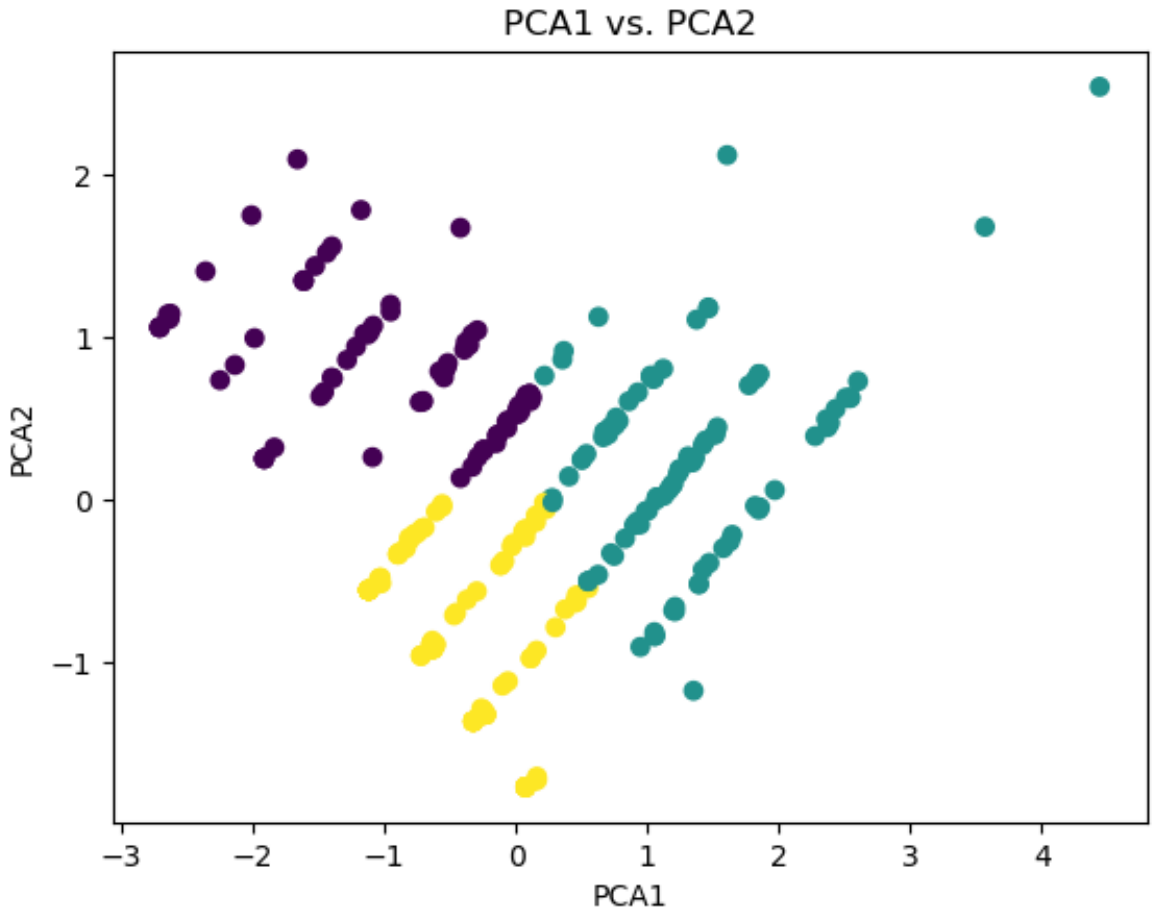
Again, we see very minimal differences actually. Some minimal imbalances particularly in Dragon, Electric, and Rock. Then a little more significant imbalances with Normal, Steel, and Bug. Again though, I would've expected that some tiers would be +/- 5% which is where I would've said that probably represented materially imbalanced 'Types', but we are not seeing that.

Now let's turn to Pokémon moves. This is the last component we will look at to see if we can get any more insight into whether the game is balanced before making our determination.

Power Moves Tier List

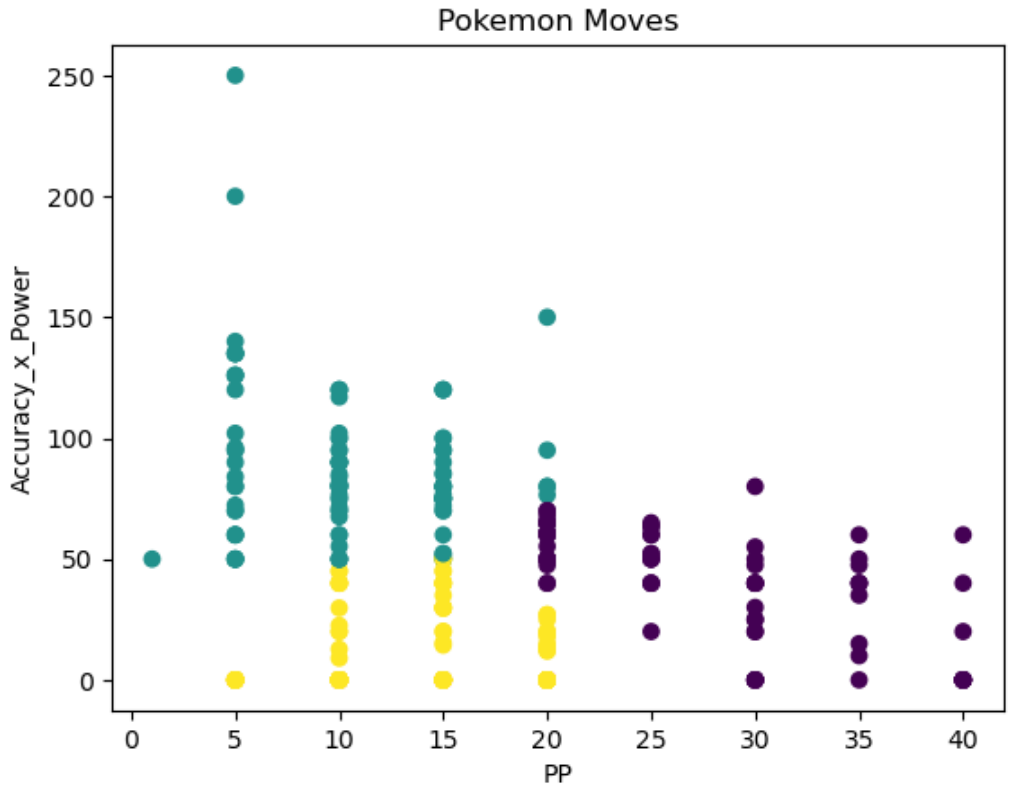
I took a very similar approach here. In my dataset, I had 'Type', 'Power', 'Accuracy', and 'PP' for each move. PP is how many times a Pokémon can use a move before it can't use it any longer. I ended up multiplying power and accuracy together as I felt that was an accurate way to

represent those two variables together. I one-hot encoded my types again and then I normalized all my variables. I then conducted PCA on all my feature variables to reduce into 2-dimension and then conducted a 3-cluster k-means and plotted the results which are follows. The one thing to note is 'PP' is in increments of 5, so that is why the chart below looks the way it is a bit.



Overall, k-means did a great job of classifying, but I did again try a Gaussian Mixture Model as well just out of curiosity. Again, those results were very poor, but you can see the output in the Appendix.

I then looked at the graph in terms of 'PP' and 'Power_x_Accuracy' and obtained the following:

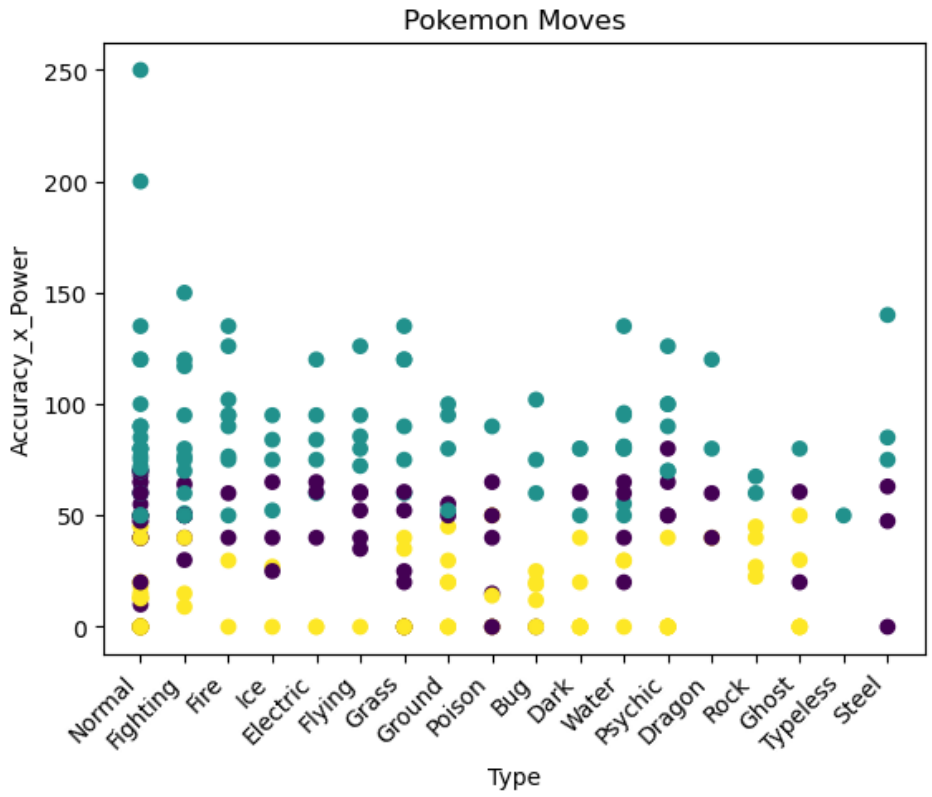


So overall we see that there is a subset of moves (in the yellow) that have low damage and low amounts of PP which would be our worst moves. Then you have moves that you can't use too often, but output a lot of damage (teal), and then finally you have moves that output low damage but can be used often (purple).

In general, you would argue that yellow are C-Tier, and then you can make a debate between purple and teal. Because a 6-pokemon team fight is relatively short, I am going to put a higher weight on damage output than 'PP'. Hence, I would say my purple cluster is B-Tier, and my Teal cluster is my A-Tier.

It is also interesting to see there are only a few outliers out of the 297 moves, which again, would probably point to the game being relatively balanced in terms of Pokémon moves overall.

The last thing I wanted to check is to see if there are any interesting insights on Pokémon moves vs. Type.



We see that every type basically has at least 1 ‘A Tier’ move and some other moves in-between. This is a good indication that really any type of Pokémon is likely viable in a fight. Below is a count of A Tier moves per ‘Type’:

Norm	Fight	Fire	Water	Gra	Psy	Fly	Elec	Grou	Ice	Bug	Dark	Steel	Rock	Drag	Pois	Gho
26	12	9	9	6	6	5	5	4	4	3	3	3	2	2	1	1

There are a ton of Normal moves based on the above. However, it is important to note that many Pokémon can learn Normal Type moves so this makes sense and probably shouldn’t be really used to determine evaluate the distribution.

It is worthy to note that Fighting, Fire, and Water have many moves that are in A Tier than the rest. For the rest of our comparison, let’s recall the distribution of A Tier Pokémon by Type as well (by %):

Norm	Fight	Fire	Water	Gra	Psy	Fly	Elec	Grou	Ice	Bug	Dark	Steel	Rock	Drag	Pois	Gho
9.5%	3.6%	5.6%	14.7%	10.3%	7.5%	0%	2.8%	5.6%	3.6%	5.2%	3.6%	6.3%	7.9%	5.6%	2.4%	2.8%

Comparing these two tables, here are some notable observations:

- Water makes up 14.7% of the A Tier Pokémon and has 9 Power Moves that are also in the A Tier. That is extremely powerful and could be used to make a claim that Water may be one of the best Pokémon 'Types' in the game.

- Steel, Dragon, and Rock had the strongest Pokémon, however, they also have the least amount of Power Moves that are in the A Tier. While their stats are great, their moves are not as stellar to choose from. Hence adding some balance back into the game.

- Bug was not represented well in A Tier compared to the total overall Pokémon Types. And yet, their moves are also on the lower end of the number in A Tier. Meaning that this is probably one of our worst types of Pokémon.

- Grass and Fire are also notable mentions here. Grass had a large part of the Pokémon Type A Tier distribution, and they are at the higher end of the Power Moves table.

- Since Normal Moves can be learned by many Pokémon, regardless of Tier status, this could help even out a Pokémon who might be not as powerful but could still learn a A Tier Power Move; offsetting the imbalance.

Final Results

Key takeaways on Game Balance

Now that we have taken a deep dive into both Pokémon statistics and the Power Moves in the game, we can revisit our ultimate question of, "is the game balanced?" Here are some of the key observations pointing towards positive game balance:

- The distribution of Types in our A Tier compared to the overall number of Pokémon in the game was very similar for most of our subgroups.

- The number of Pokémon in each tier list was very similar and with not many outliers. This points to likely a natural progression of starting with not as powerful Pokémon earlier in the game and getting stronger Pokémon later.

- Some subgroups of Types (Dragon, Steel, and Rock) are significantly represented in our A Tier list, however, they have not as many Power Moves to choose from offsetting some of their powerfulness.

- Power Moves that are very strong tend to have lower PP which makes it so the player cannot just keep using the same powerful move over and over again.

- Each subgroup of Type had A Tier Power Moves that can be chosen. And being able to also learn Normal moves helps offset some of the imbalance, if it exists.

On the flipside, there are lots of elements of the game that do support some imbalance:

- When a Pokémon's stats offensive stats (attack, sp_attack, etc.) are higher, there doesn't seem to be a tradeoff in defensive stats (hp, speed, etc.). I would've expected to see some negative correlation here to balance that tradeoff.

- There are some subtypes, like Water, where there are a lot of represented Pokémon in A Tier and Power Moves in A Tier. Conversely, there are subtypes, like Bug, who have the opposite pattern.

- Some subtypes of Pokémon have a higher or lower distribution compared to the overall population which could lead to some game imbalance. However, all differences were less than 5%.

All in all, I was surprised to see how much balance was represented in the game. Based on my findings, I would say my original hypothesis was incorrect. I do believe that the game seemingly is relatively balanced. There are aspects that could be argued that are not, but overall, the game makers seemed to have done a good job of trying to offer the player a lot of options, flexibility, and ways to play the game without really losing too much effectiveness. In addition, because Pokémon evolve and can learn other types of moves, if there is some imbalance, it allows for some of it to be reduced and create an environment where almost anything can succeed.

What about an Optimized Team?

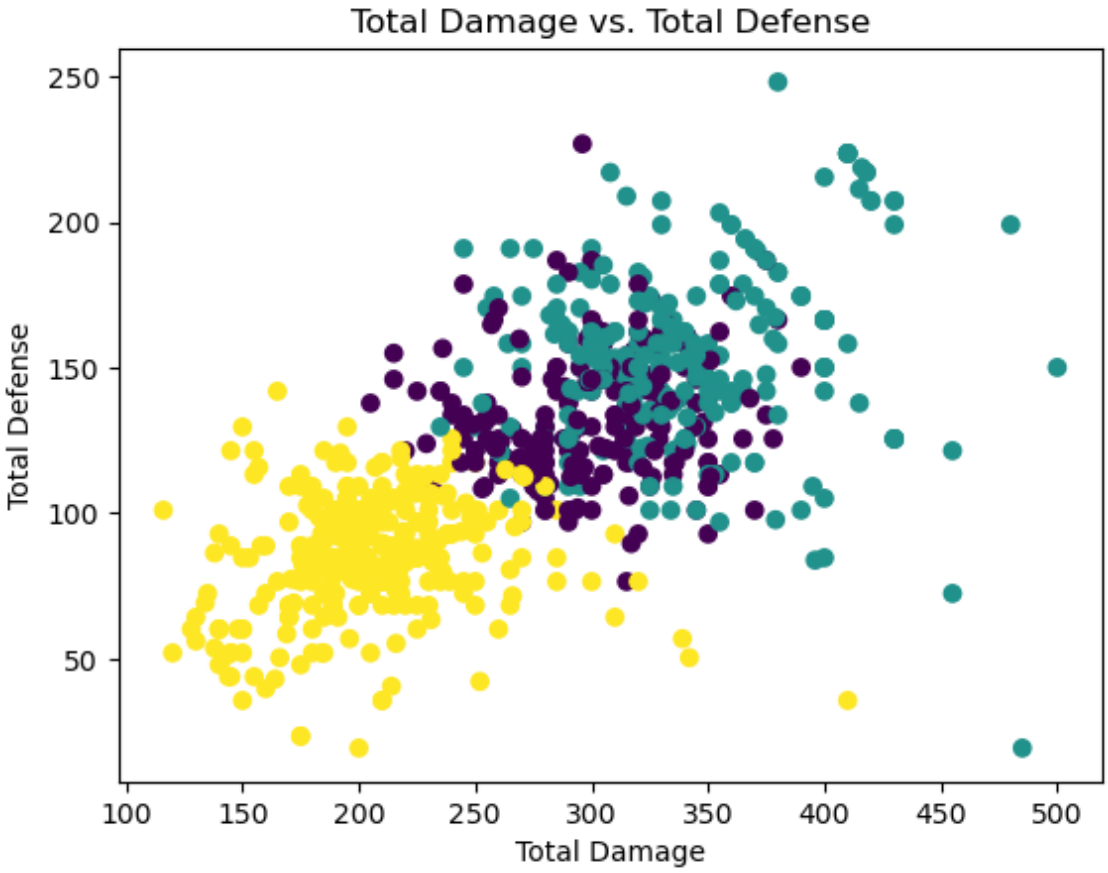
Going into this, I thought that I would probably have to run a complicated optimization problem to figure out if I could make an optimal team. In general, I do think this is a possibility and likely a whole other project if you wanted to get really complex with opponents and matchups.

However, if you use the simplicity of what we found out here, I think coming up with a team is straightforward.

Here are some thoughts on what would constitute the most well-balanced team:

- We have not talked about much of this yet, but most Pokémon have a weakness. They will take double damage if they are versing a Pokémon that they are weak against. Hence, our team should be very balanced and have a variety of Pokémon types to combat this.
- We can focus just on our A Tier Pokémon since they are the strongest. No reason to take Pokémon from other tiers really.
- Since Power Moves seemingly show up in every A Tier subtype, we can mainly focus on just Pokémon Types statistics like overall attack and defense.

So, let's look at a graph of overall Total Offense vs. Total Defense stats.



From this, we are going to want to pull the Pokémon as far to the top right as possible. These are Pokémon with high defense and damage output. So let's look at the top 20 from my code:

	name	type1	total_defense	damage_output	defense_add_damage	weakness
487	Arceus	normal	480	199.047188	679.047188	fighting
248	Lugia	psychic	500	150.035391	650.035391	bug
770	Lunala	psychic	430	207.215821	637.215821	bug
769	Solgaleo	psychic	430	207.215821	637.215821	bug
701	Yveltal	dark	418	217.018181	635.018181	fighting
700	Xerneas	fairy	418	217.018181	635.018181	steel
149	Mewtwo	psychic	416	218.651907	634.651907	bug
634	Reshiram	dragon	410	223.553087	633.553087	fairy
635	Zekrom	dragon	410	223.553087	633.553087	fairy
479	Dialga	steel	410	223.553087	633.553087	fire
480	Palkia	water	410	223.553087	633.553087	electric
482	Regigigas	normal	430	199.047188	629.047188	fighting
382	Rayquaza	dragon	380	248.058986	628.058986	fairy
381	Groudon	ground	420	207.215821	627.215821	water
380	Kyogre	water	420	207.215821	627.215821	electric
287	Slaking	normal	415	211.300138	626.300138	fighting
636	Kyurem	dragon	400	215.384454	615.384454	fairy
483	Cresselia	psychic	455	121.445176	576.445176	bug
378	Latias	dragon	410	158.204024	568.204024	fairy
383	Jirachi	steel	400	166.372657	566.372657	fire

Overall, we see that the following Pokémon would make up our 6-Pokemon team if we are going by top stats and diversifying our type of **both *Type & Weakness:***

1. Arceus *(Type = Normal / Weakness = Fighting)*
2. Lugia *(psychic / bug)*
3. Xerneas *(fairy / steel)*
4. Reshiram *(dragon / fairy)*
5. Dialga *(steel / fire)*
6. Palkia *(water / electric)*

However, one could argue that this isn't full proof by any means. For example, we know that there are a ton of Water Pokémon in the game. So, it might be worthwhile to dig further and find an Electric Type Pokémon to bring along for that counter. We also know that there is a decent amount of Grass / Bug types which typically have a Fire weakness. So again, we

may want to find a Fire Pokémon to bring along to have that counter as well. There is a lot to consider here and part of it is going to be very situational based on what the player is versing. However, if we are just primarily going with a diversified best stats team, without knowing anything about your opponent, the one above is likely one of the best outcomes.

With that said, I do believe this is a pretty good, optimized team with some optionality to switch out a couple of Pokémon. I believe this part of my hypothesis was correct that we do have the ability to create an optimal team.

Other Considerations

I will say that while I believe my analysis did a good job of trying to dig deep into the world of Pokémon, that there are further things that probably should be explored to continue debating this topic. Here are some elements:

- I made a lot of assumptions on the damage formula, and I only used the basic Generation I formula. The game has much evolved, and the formula is much more complex than I even made it out to be in this paper.

- I generalized power moves and defense in my damage output formula. These should probably be tweaked to be more customized based on each individual Pokémon.

- Many power moves were not included in this analysis. I only included ones that do damage. There are some moves that, for example, do no damage but lower a target's defense. Or put a negative condition on an opposing Pokémon (like sleep for example) and don't do damage. I would call these utility moves, and they were not considered in this analysis.

- I did not take Pokémon weaknesses into account prior to the optimization part of this paper. That makes the gameplay much more complex than I laid it out to be.

- Pokémon also have subtypes and can learn various types of different moves outside of its same 'Type'. It is likely that it would be worthwhile to dig into the Power Moves associated with each Pokémon to see if there are any other imbalances.

- There likely should be some assumptions on your opponent's Pokémon as well.

- Every year or two, a new Pokémon game comes out. The entire universe of Pokémon may or may not be in a specific version of a game.

Conclusion

Overall, I am very happy with my results. I think I was able to dig deep into this analysis while still realizing there could be a much deeper dive if someone wanted to take this a step further. I believe the game makers have done a pretty good job of trying to keep things balanced and feel like my original hypothesis was debunked.

Game balance is such an important part of game design, and I feel like the makers have done a great job of trying to keep that in mind. While it isn't the only element of a video game, a lot of care has to be done to make sure the game is fun for the target audience. Pokémon seemingly has done a great job of this, and probably one of the reasons why they have been such a successful franchise over the years and continues to grow.

Appendix

[1] https://en.wikipedia.org/wiki/Pok%C3%A9mon

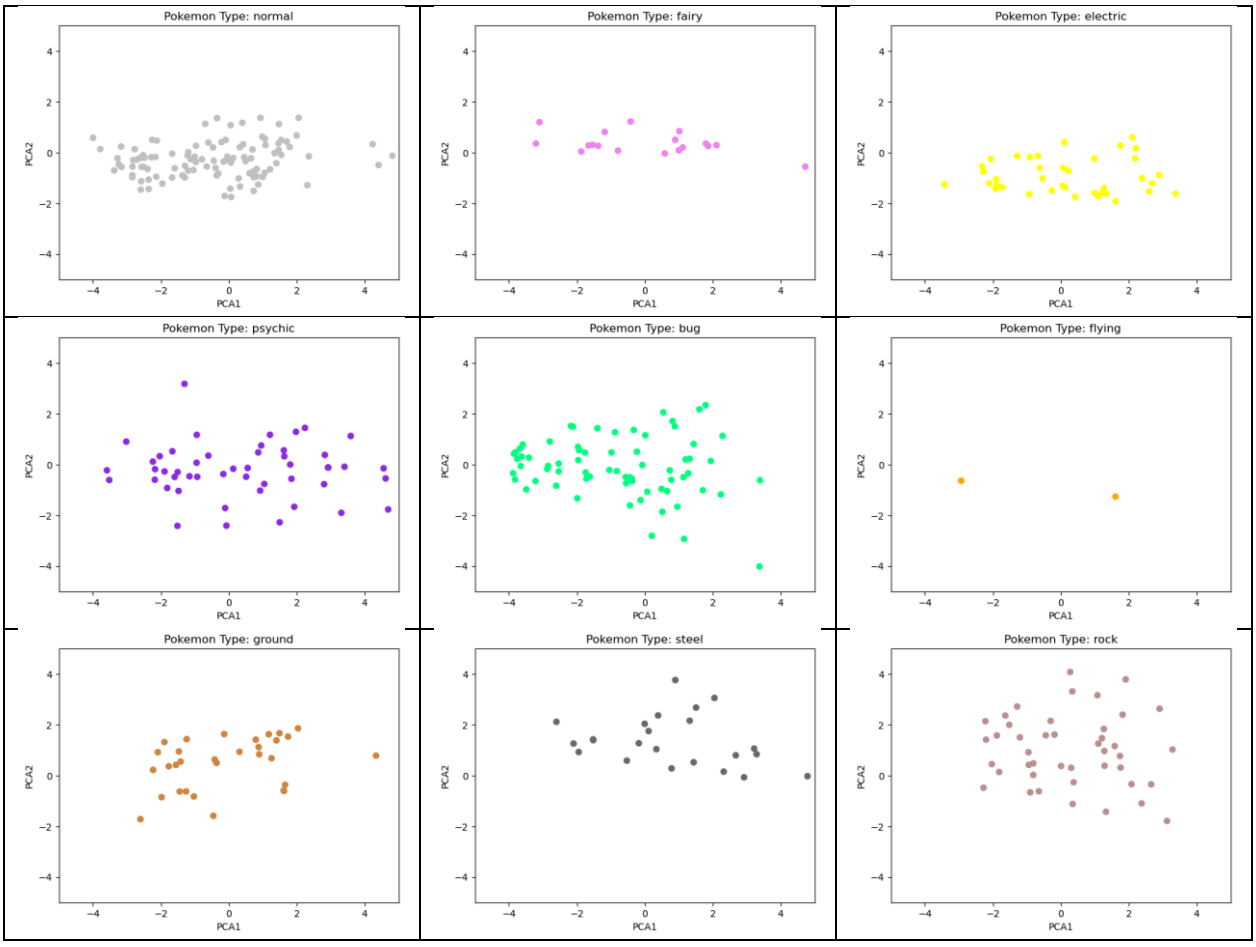
[2] https://www.kaggle.com/datasets/rounakbanik/pokemon

[3] https://www.kaggle.com/datasets/jaidalmotra/pokemon-dataset

[4] https://bulbapedia.bulbagarden.net/wiki/List_of_moves

[5] https://bulbapedia.bulbagarden.net/wiki/Damage

Figure 1 – PCA1 vs. PCA2 by type



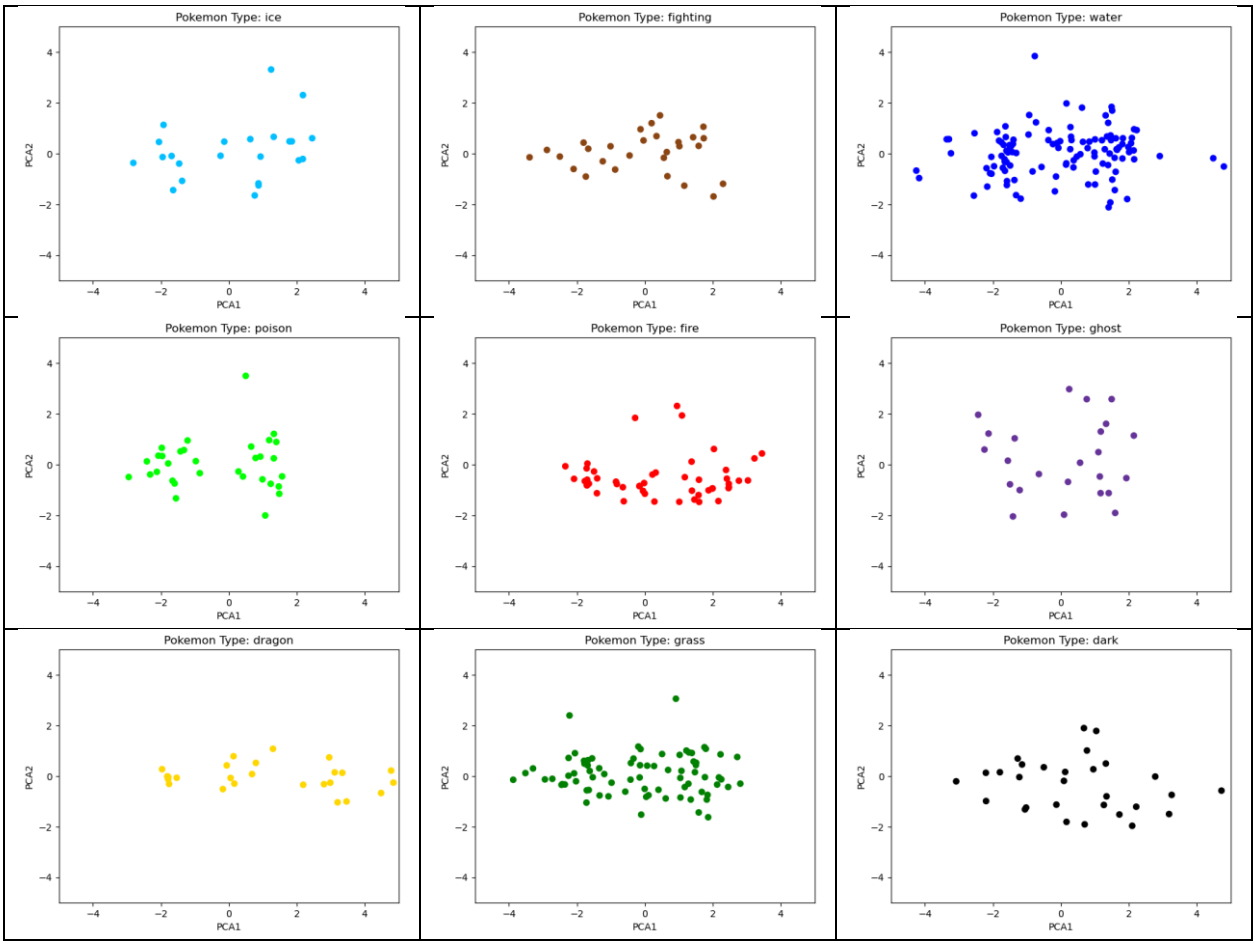

Pokemon Type: ice
Pokemon Type: fighting
Pokemon Type: water
Pokemon Type: poison
Pokemon Type: fire
Pokemon Type: ghost
Pokemon Type: dragon
Pokemon Type: grass
Pokemon Type: dark
PCA1
PCA2

Figure 2 – Gaussian Mixture Model Clustering - Type

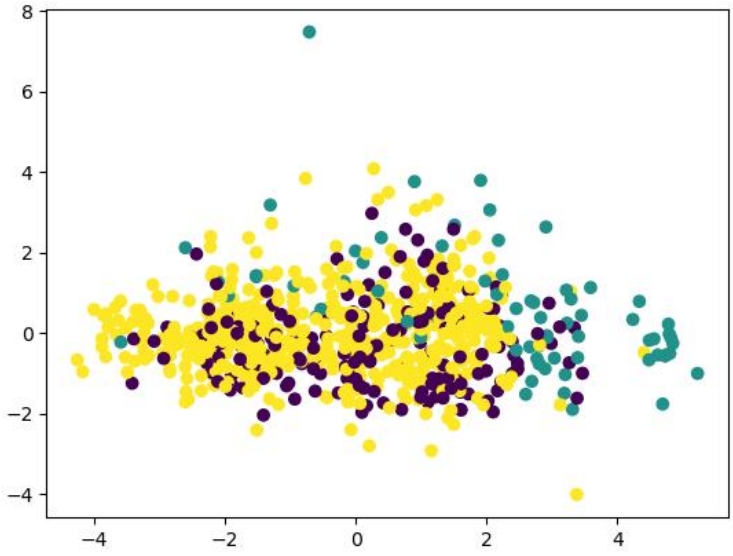


Figure 3 – Gaussian Mixture Model Clustering – Power Moves