To Be The Very Best

Richard Marks

4/27/2020

ABSRACT

Pokemon is an international mulitimedia phenominon spanning over twenty years. In this paper we looked to see what truly makes a pokemon strong. We found that being a legendary, having a balanced set of base stats, having a high egg hatch step rate, and being primarliy male has the best chance of predicting if a pokemon will be stong.

INTRODUCTION

On Februrary 27, 1996 a (at the time) small game developing studio named Game Freak partnered with Nintendo to release Pokemon Red and Green for the Nintendo Game Boy in Japan. The release would lead to a worldwide phenominon spanning videogames, trading cards, tv shows, etc. Through the 24 years since its inception Pokemon has skyrocketted to become the highest grossing media franchise of all time, grossing over 95 billion dollars in total revenue. To out this into perspective the second highest grossing franchise is Hello Kitty at 86 billion dollars in total revenue, with a 9 billion dollar gap between the two. Through its popularity Pokemon has grown to a household name, and with an ever expanding list of pokemon, with the first game introducing 151 to the Ultra Sun and Ultra Moon games having 801 unique and obtainable pokemon. With all of these added pokemon there is a question that most players come across, what makes a Pokemon strong? To answer this question we have to ask two more. What is the strongest pokemon type for each of the games 6 stats Hp(hit points), attack (physical attack), defense(physical defense), special attack, special defense, and speed? Also, which type has the highest base stat total? Alongside trying to figure out what makes a Pokemon strong there are a few other questions this report there will be focusing on. Which generation has the strongest pokemon? Can we create a model capable of predicting a pokemons base stat total based off of various factors? Finally is it possible to create a unsupervised clustering model able to tell if a pokemon is a legendary or not. This data was taken from Serebii.net, the most accurate and expansive pokemon database.

Methods/Results

Which Type Has the Highest Average For Each Base Stat?

To answer this question we will be taking the average of each stat for every type. Pokemon who have two types will be counted for both of their types. So for example Charizard will be counted in the averages for both the fire and flying types.

After averaging the base stats of each pokemon within each type we have found which Pokemon types have the highest average base statisitic for all six of the differnt base statistics. For HP the Dragon type had the highest average hit points with an average of 82.6. For the Attack stat the fighting type had the highest average at 99.4. For Defense the rock type had the highest average at 111.7. For Special Attack the fire type had the highest average with 92.9. For Special Defense the fairy type had the highest average with 84.4. The final stat is for speed with the flying type rocketing in with an average of 90.8.

Which Type Aside From Dragon Has The Highest Base Stat Total?

Base Stat totals are the sum of all 6 of a Pokemons stats. While most pokemon have a modest base stat total ranging from 300-400 there are some that excel, and some that fall far below their peers. For this analysis the Dragon type will be excluded, as the majority of Dragon type Pokemon are Legendary (rare Pokemon with extremely high base stat totals usually around or over 600) or PsuedoLegendary (rare Pokemon that, while not considered a legendary, have a base stat total of 600). This would make the Dragon Type have the highest base stat total by a wide margin. So to find which type has the highest average base stat total we will do the same thing as with each individual stat. We will average the base stat totals of all pokemon of each type, with pokemon that have Dual Types being considered as both of their types.

Of all 17 remaining types there was a clear winner with the Steel type having an average base stat total of 480.9 which is over 100 points higher than the type with the lowest base stat total the Bug type with an average of only 377.7 points. The Steel type having the highest average base stat total is interesting because it did not have the single highest average for any of the 6 stats, not even Defense which is what it is known for. So having the whole being greater than the sum of its parts is a very interesting observation when we take into account the question we are trying to answer, what makes a Pokemon strong. The question is answered by the Steel type, being well rounded. If a Pokemon were to excel in only 1 or 2 stats it could be strong in certain situations, but not all of the time. Lets use Alakasam as an example. Alakasam has an insane base special attack stat of 135, and a very fast speed of 120, so it should hit before any other Pokemon and very hard. Unless that Pokemon had a very high special defense stat such as Snorlax at 110, which would very easily survive a hit from an Alakasam and most likely defeat it in one hit, since Alakasam has an abysmal defensive stat. However a Pokemon such as Lucario who is very well rounded is extremely versitile in any situation having both high Attack and Special Attack stats while having very good Defense and Special Defense stats. A good Pokemon is one who is well rounded and viable in many different situations as opposed to the many Pokemon who have very specific roles to fill.

## Which Generation Has The Highest Average Base Stat Total

Older fans of the Pokemon series have recently been complaining about a sudden power creep, where the newer Pokemon introduced are significantly stronger than those of past generations, but is there any validity to this question, or is it just the older fans showing favoritism to the older generations? To answer this question we first must define what a generation is, every few years Pokemon comes out with a new pair of main series games that take place in an all new location. A new generation is created by a new game being created that takes place in a new location, and the Pokemon native to this new location are considered apart of that generation.

The Average Base Stat Totals for each generation are as follows: Gen 1= 408.3, Gen 2 = 407.2, Gen 3= 404.7, Gen 4= 447.5, Gen 5= 426.4, Gen 6= 436.2 , Gen 7= 448.2. This does confirm the older fans worries, after generation 3 the last game to be avaliable on the Gameboy Advanced, there is a significant power increase. This could very likely be due to the jump from the Gameboy Advanced to the Nintendo DS. To be sure that fans would stick around with the series and but a Nintendo DS to play the newest game GameFreak very well made stronger Pokemon in the hopes of having old players want to catch and use them. Whatever the explaination it is very much true that the most recent games pokemon have expirenced a major power boost compared to those of the very first games.

Building a Model to Predict a Pokemons Base Stat Total

Every Pokemon is unique, from its type combination, to the amount of steps it takes to hatch it out of an egg, every pokemon has a unique combination. So a question arises, is it possible to guess a Pokemons base stat total based off of its different properties? To test this we ran a multiple logistic regressions, a small regression using only a pokemons base statistics, a medium regression using any possibly useful information, a large regression using all of that information and all of its interactions, and an optimal regression using the optimal group of statistics, based off of a step function that was too large for our computer to put into the

report. Important to note, for the type1 category the flying type was removed from the data set, because there are only three pokemon with the primary flying type (Noibat, Tornadus, and Novivern), with the multiple train set splits created there was no way to ensure that both the train and test sets had a primary flying type.
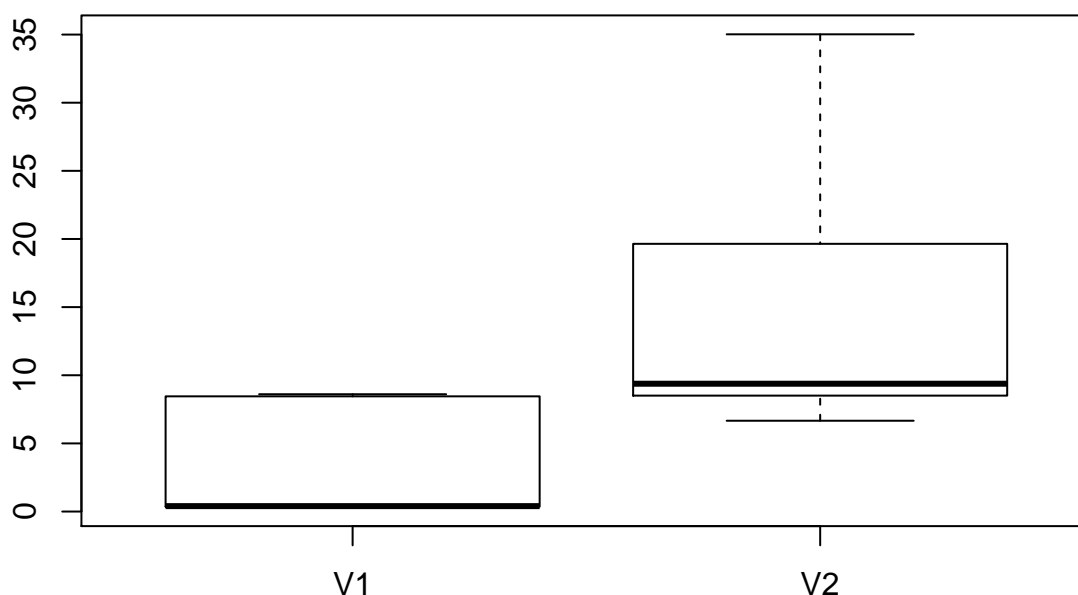
Shown by our above boxplot is the average root mean squared values for the two best performing models. V1 is our smallest model which predicted base stat totals from the pokemons 6 base stats. V2 is our Optimized model which was creaated using many different factors and interactions. Our smallest model is the most accurate, with an average RMSE of .3, so insanely close to the actual responses, however this makes alot of sense, since the base stat total is just a summation of the 6 core statistics, so of course we could predict the base stat total accurately with those. However what is interesting is for our optimal model we had an average RMSE of about 10. This is still a very good prediction of pokemons base stat total, predicting an average of only 10 points away from the actual total. An interesting question within this regression is, what factors had the biggest effect on predicting base stat total, aside from a pokemons core stats. To answer this question we looked at which factors showed up the most in the model, which lead to the three biggest factors being,how many steps it takes to hatch an egg , what percentage of the pokemon are male, and what is the primary typing. Pokemon are able to breed with each other in these games to create stronger pokemon, and these pokemon are hatched out of eggs. The stronger the pokemon is the longer it generally takes to hatch out of an egg, so it makes sense that the egg steps is a good predictor of base stat total. Interestingly a pokemons primary type has a large effect on predicting its base stat total, earlier we looked into the average base stat total for each type, but that was for both primary and secondary typings. Having a pokemons primary type being a strong predictor of base stat total shows how important a primary type is than a secondary. The last strong predictor of base stat total is percentage male, which is very interesting. Most pokemon have an even 50/50 split of female to male, however there are certain pokemon such as Gallade who can only be male, and Miltank who can only be female. This being a strong predictor of base stat total shows a dispairty between the strengths of single gendered pokemon.

Legendary Pokemon are extremely rare Pokemon that are very strong, while legendaries are rare is there anything else that really connects them toghether. To answer this question we will be running a Principle Component Analysis to try to accurately group pokemon into either a regular pokemon group or legendary group. You may very well be asking why does this matter though? It matters because for years legendary pokemon have always had an air of wonder around them. Sure they are stronger than average pokemon, but is that truly the only real difference? To create these clusters we will be using two main grouping features, a pokemons base stats and the types that are super effective against it. In pokemon every pokemon has one or two typings , each of these types is weak to and super effective against. This combination is part of what allows pokemon to be so unique. Type combinations are intergral to the core gameplay of Pokemon, so being able to group legendary Pokemon (the strongest in the game) by which types they are weak to can show which typings are the weakest in competitive Pokemon, and which would be the strongest.
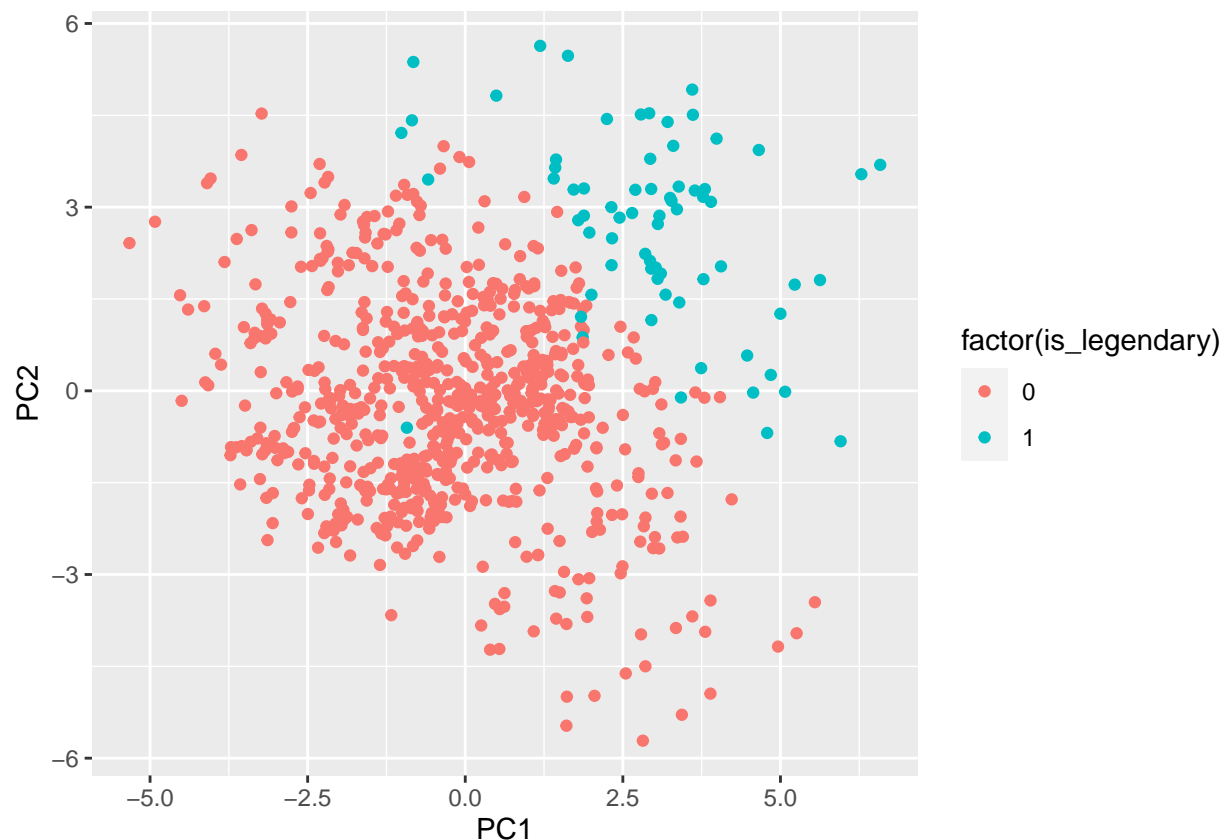
Seen above is the results of our Principle Component Analysis, as you can see there are two significantly different groups, the large red group is made up of the normal pokemon and the smaller blue group in the top right corner is comprised of the Legendary Pokemon. This shows that legendary pokemon do have similar characteristics and their Legendary status is not just arbitrary. They all have high base stat totals and most of them are weak and resistant to the same typings, with having a weakness to Ghost being the strongest predictor of being a legendary, and being resistant to Psychic being the strongest predictor of being a legendary.

Conclusion

Throughout this analysis there has been one thing made abundantly clear, Pokemon are all unique and overall relatively balanced. This is shown by the fact that for each of the six main stats the type with the

highest average stat is never repeated (HP: Dragon, Attack: Fighting, Defense: Rock, Special Attack: Fire, Special Defense: Fairy, speed: Flying), and on top of that the type with the highest average stat total, Steel, does not have the highest single stat, showing that each type is meant to be balanced or better at one specific stat. This shows that the best pokemon are those that are well rounded and balanced, where the highest base stat totals are with the typing that is not the best in anything specific just overall. Also shown through base stat totals was that there has been a significant power creep as the Pokemon generations go up. With the newer pokemon having a significantly higher average base stat total than those of the first few generations. Also important to note is the model created to predict a pokemons base stat total, with the most important predictors being a pokemons primary typing, how many steps it takes to hatch an egg, and the proability of a pokemon to be male. Finally we showed that a pokemon being classified as a legendary is not just an arbitrary concept, it is possible to classify legendary pokemon based off od their base stats and their weaknesses to every type, with legendary pokemon being more likely to be weak to ghost types. This all goes to show that Pokemon is an extremely complex game, where the stongest pokemon are most likely to be legendary. But to find the strongest non-legendary pokemon it is clear that it needs to be a pokemon with well rounded stats, a high amount of steps to hatch its egg, and have a higher chance to be a male pokemon rather than female.