

STAT 527 Pokemon Analysis

Abstract

This project focuses on examining the influence of typing, base stats, and experience of Pokemon in order to find a strong team against each of the eight gym leaders in *Pokemon Platinum*. Each gym has its own typing specialization. Specifically, we wanted to find Pokemon that had strong type resistances for each gym, and would generally take less damage and avoid fainting. Only Pokemon in the Platinum Pokedex were considered, not including Legendary Pokemon. Typically, strategies utilizing special training, abilities, items and status moves can be employed to improve battle odds; however, we decided to leave these out of the analysis since they would add too much complexity. We decided to take a straightforward approach, focusing on the reality of obtaining and leveling different Pokemon, then finding which Pokemon should be chosen based on type resistances. Various data analysis and visualization techniques were used to ensure a thorough analysis was conducted. This research offers valuable insights into *Pokemon Platinum* gameplay. Despite our limitations, the project can serve as a good starting point for team-building with the goal of defeating gym leaders in *Pokemon Platinum*.

Introduction

Pokemon is a game based around catching and training pokemon to use in battle against other trainers. The foundation of the gameplay is built on using your pokemon to battle others to gain experience points, hereafter called EXP. Upon the defeat of every opposing Pokemon, you gain EXP for the Pokemon you used. Once reaching a certain EXP threshold, your pokemon will gain a level, where the higher the level, the stronger the Pokemon.

There are many factors that go into building a Pokemon team, such as the availability of Pokemon, their typing, stats, moves, experience, and more. This complexity makes the game exciting, and building the most effective team is the primary goal of many players. In this project, we are looking at the influence of typing, base statistics, and experience of the 201 Pokemon in the game *Pokemon Platinum* to try to find a strong team against each of the eight gym leaders in the game. The aim is to incorporate new Pokemon into the players team for each gym leader and find if it is reasonable to complete the game battling only trainers.

Data Description

To start collecting data, we looked at reputable websites to find existing datasets: namely kaggle.com, serebii.net, and pokemondb.net. We identified variables lacking information for our research goal and supplemented our datasets by web scraping and merging additional tables. All necessary data was found. The specifics will be discussed further.

First we had a dataset listing all the existing Pokemon, which we adapted as other datasets. We had a dataset that only includes the Pokemon found in the game Platinum, as well as a dataset that showed the different types of each Pokemon. There are 17 different types that are present in the game, which we included in our analysis. We also created a dataset with the locations each Pokemon could be found, so we

knew which Pokemon we could find before each gym. Next, we looked at the different moves gym Pokemon can perform. From this dataset, we were able to find which Pokemon would best resist the moves found in the gyms.

Secondly, there were three essential data sets for the analysis of EXP availability throughout the game. The first being a complete collection of all the trainers in *Pokemon Platinum*, this included the names and levels of all pokemon owned by the trainers. The second being a complete collection of all the pokemon available in the game, and their accompanying statistics, most importantly the base EXP yield of all pokemon, and their growth type. Where the growth type of a pokemon is one of six, this group determines the amount of EXP needed to reach the next level. Meaning that Pokemon gain Levels at a different rate. The third essential data set being the EXP requirements to reach individual levels for each of the growth types.

As a final note, we made sure to check for formatting issues or any other discrepancies in the datasets we chose to use. For example, we had to remove Fairy typing from our data, and replace it with the Pokemon's original typing as it appeared in the game. We often combined datasets to make them easier to use as well. Using all this data, we could perform our analysis to determine the best Pokemon team to defeat each gym, without repeating the same Pokemon within each team.

Description of Data Analysis

After we found all our data and organized it in a way that was easy to use, we could begin our analysis. When beginning the analysis process, we used tidyverse and many of its accompanying packages, such as ggplot2. We made sure all our data was imported in our session. As an overview, we wanted to look at generally which type in the game has the best average base stats, so we can then compare to see if these types appeared more often as parts of the best team for each gym leader. To do this, we had to take the average of each of the six statistics - HP, Attack, Defense, Special Attack, Special Defense, and Speed - and store these values in a new column. Then, since some Pokemon have two types and there were two columns for typing, we combined the column to make sure the code counted every Pokemon with a certain type, regardless of if it was in the Type 1 or Type 2 column. Then, we created a new data frame to store the types of each Pokemon. From there we could take the average of all the statistics for each type, and finally plot it in a bar chart, seen in Figure 1. Based on this analysis, we see that the type with the worst average base stats is Bug and the type with the best average stats is Dragon. This information can be useful to see how often a certain type appears in a team for each gym. It is important, however, to keep in mind that there are type weaknesses, so although one type may be better overall, it still has a weakness that could make it less effective against a certain gym.

To explore the type resistance of all Pokemon against gyms, two main factors were considered: type resistance against physical and special attack moves, and type resistance against the gym Pokemon themselves. For example, Cranidos is a Rock type, which gives it a 1.5x damage multiplier when using Rock moves; however, it could still use a non-Rock move like Pursuit (a Dark move). Both aspects needed to be accounted for. Then a "resistance score" was calculated as $0.5 * (\text{type_resist}) + (\text{movetype_resist})$, which was used to rank Pokemon. If resistance scores were identical, the pokemon with the higher total base stats was prioritized. Only Pokemon

obtainable before each gym battle were considered, as mentioned previously. This methodology allowed us to create “resistant teams” for each gym, based on strong defensive typing. Full teams of six were gathered in this step, although the feasibility of obtaining a full new team for each gym is studied later on.

For Gym One: Rock, and Gym Two: Grass, some special considerations were at play. The starter Pokemon gifted by Professor Rowan (as the player's first Pokemon) are unique to the playthrough: when you choose one you can no longer obtain the others, except by trading; they are generally strong Pokemon. Against the Pokemon in these two gyms, Turtwig is the higher ranked starter; however Piplup is only slightly weaker and becomes a consistently recommended Pokemon later on, so it gets the edge. It is also important to note that you can get strong Pokemon before the first gym by utilizing Honey Trees; Honey Trees can be checked every six hours with a low chance to encounter a strong pokemon. These Pokemon will be included in the rankings, but it may not be realistic or necessary to obtain all of them before Gym One. The recommendations (by resistance) were:

Gym One: Rock-- Leader Roark

Resistant Team: Onix, Heracross, Cherubi, Snorlax, Ambipom, Piplup
(in rank order, 1 → 6)

note: Heracross, Snorlax, and Ambipom are only obtained via Honey Trees

Gym Two: Grass-- Leader Gardenia

Resistant Team: Dustox, Skorupi, Golbat, Roserade, Vespiquen, Mothim

note: Bug types get worse later on; Flying types will be useful for next gym

The next gyms were fairly straightforward. Although the Pokemon Rotom is worth discussing. Considered by some as an unofficial Legendary, Rotom has high base stats and can only be caught once. Furthermore, its typing can be changed with the use of special items. These additional typings were not considered in the analysis due to our rules for items. An alternative recommendation was provided in case the player decides against using Rotom. Here are the recommendations for gyms three and four:

Gym Three: Ghost-- Leader Fantina

Resistant Team: Togekiss, Chatot, Staravia, Noctowl, Honchkrow, Stunky

Gym Four: Fighting-- Leader Maylene

Resistant Team: Spiritomb, Bronzor, Mismagius, Duskull, Scizor, Rotom

note: Rotom can be swapped for Gastrodon; Spiritomb is a must (see Figure 5)

For all remaining gyms there was only one special consideration. For Gym Five: Water, the highest resistance Pokemon were Water types. This also means that using Water type moves against the gym would not be effective; the results from our methodology should be reviewed with this in mind. If a Pokemon could learn/be taught moves that were effective against Water or specific typings of gym Pokemon, they might have performed better than our recommendations that could not. Regardless, here were the results:

Gym Five: Water-- Leader Crasher Wake

Resistant Team: Empoleon, Vaporeon, Golduck, Floatzel, Octillery, Lumineon

note: Gyrados is a strong honorable mention for its high base stats; Pokemon that can learn Grass and Lightning type moves are strong against this gym

Gym Six: Steel-- Leader Byron

Resistant Team: Gastrodon, Whiscash, Quagsire, Gliscor, Gabite, Empoleon

Gym Seven: Ice--Leader Candice

Resistant Team: Empoleon, Scizor, Lucario, Weavile, Bibarel, Magnezone

note: Bronzong honorable mention; Empoleon is a must (see Figure 5)

Gym Eight: Electric-- Leader Volkner

Resistant Team: Gastrodon, Whiscash, Quagsire, Steelix, Rhyperior, Golem

These teams were used as a basis to investigate Pokemon further. With respect to final Pokemon team building, the feasibility of EXP gain needed to be factored-in.

To investigate the EXP availability and feasibility of leveling a full team of Pokemon to the next gym leader cap. We work with the Pokemon and trainers data sets, cleaning, and analyzing as follows. As we have already sorted the Pokemon into gym sections, the same is done for trainer encounters. This was done by looking at the max level of each trainer's Pokemon, if its level was within a range of a gym leader's level cap, it was placed in that gym section. The trainers' data set needed far more organization and manipulation than other datasets. Not only was a gym section added to all the trainers, there was also a need to run the EXP gain formula on all of the pokemon for all of the trainers. This EXP value for individual Pokemon was summed and added to each trainer, this gave a total amount of EXP gained from defeating that trainer. From there the total amount of EXP available from trainers per gym section was calculated and visualized via plotly.

With the total EXP available and organized by the gym section, the next step was to find all of the pokemon available in each gym section, with the next gym section including all of the previous pokemon. From this we can find the total number of pokemon of each growth type available in each gym section. This was visualized with ggplot2. From here the final step was to find the total amount of EXP needed to reach the gym leader level cap from section to section (i.e EXP needed to reach gym leader 1 cap to gym leader 2 cap. Assuming after gym 1 your team is all at the level of the leader recently defeated). This was done for all growth types, and visualized with ggplot2 such that, you can see how many pokemon can reach the next level cap before running out of available EXP. Finally, taking the average amongst all growth types, you can find on average how many pokemon from your team you can reach the next gym leader cap battling only trainers.

Visualization of Results

Figure 1:

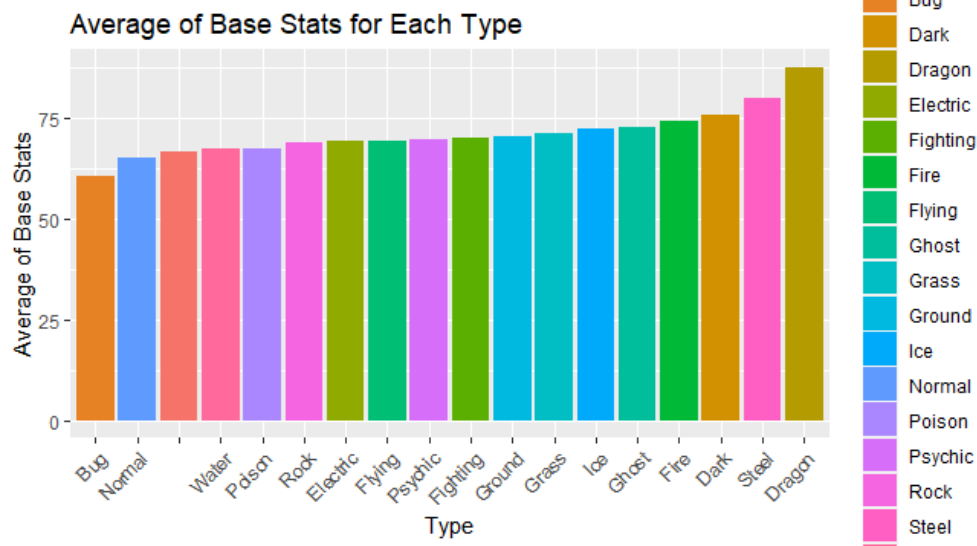


Figure 2:

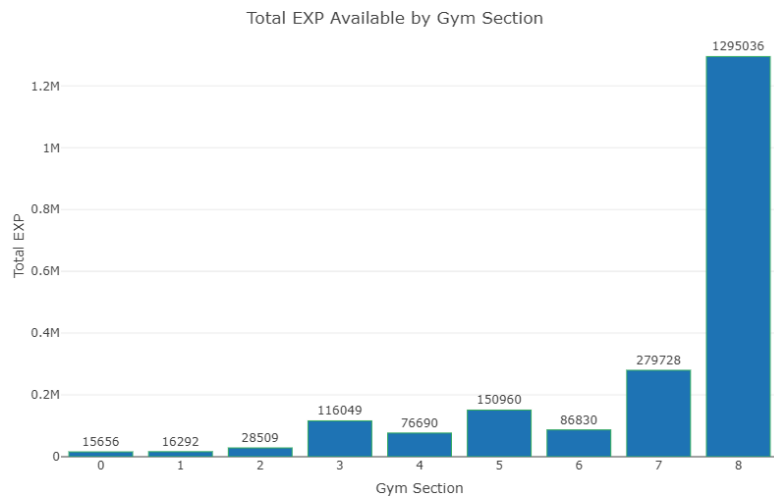


Figure 3:

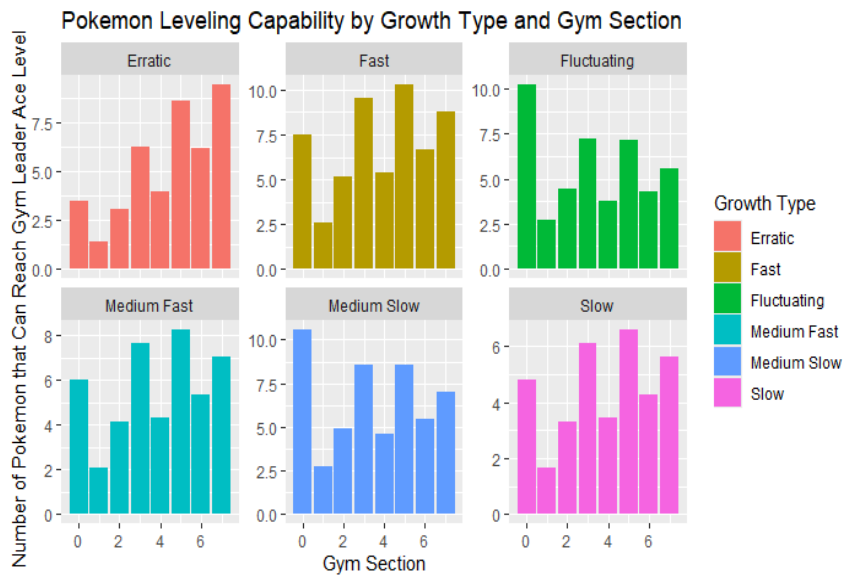


Figure 4:

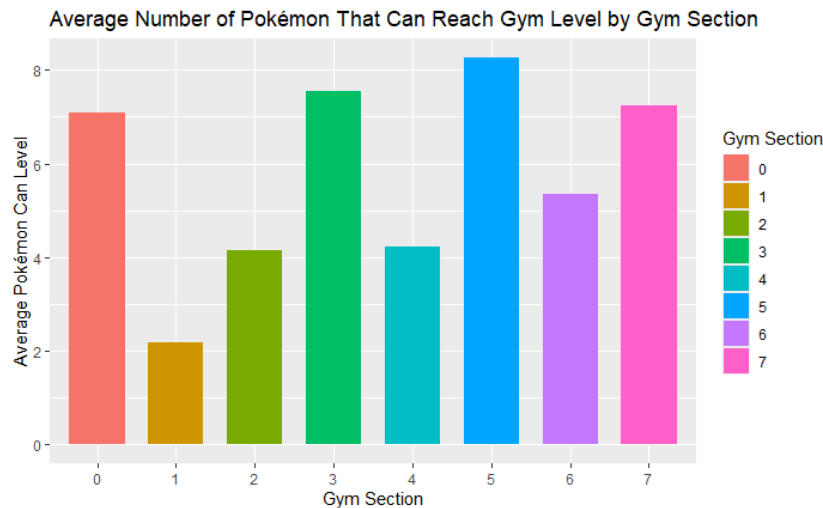


Figure 5:

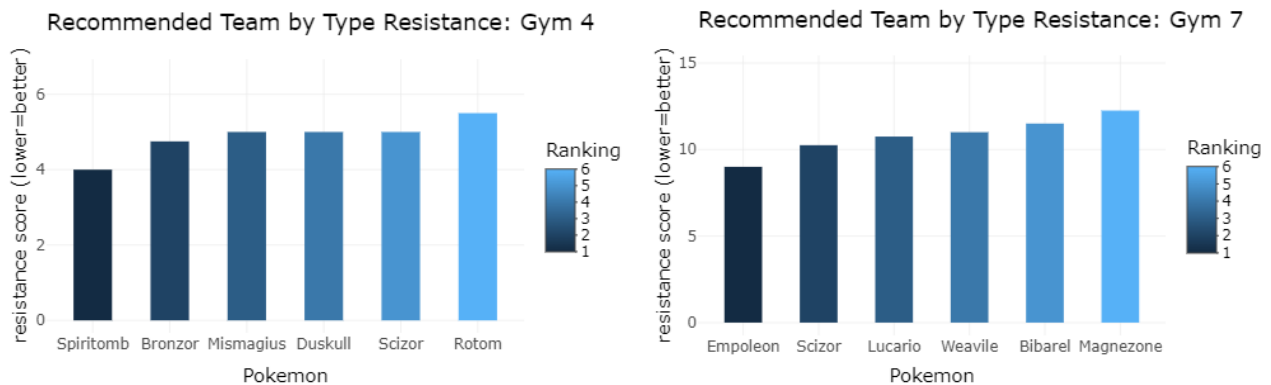


Figure 1: is a bar chart that shows the average base stats of each Pokemon in Platinum based on their typing. Although type weaknesses are important to keep in mind, this can be a good start at finding which Pokemon to use.

Figure 2: is a bar chart showing EXP available in each gym section.

Figure 3: is a series of charts showing how many pokemon from each growth type can reach the level cap for each gym section.

Figure 4: shows the average number of Pokemon that can reach the gym cap per gym section. Figure 4 gives us the most useful insight on how many Pokemon our team of six will be able to reach the max level for the next gym leader. The chart shows that for half of the gym battles, the player will not have a team of max level Pokemon. Instead they must rely on a portion of the team being weaker than the optimal maximum level.

Figure 5: shows a couple of cases where there was notable variance in resistance scores within the recommended teams. The graphs help show how much better the first

rank Pokemon was suited for the gym battle opposed to the least suited Pokemon in the team (at rank 6).

Individual Contribution to Project

I contributed the entire EXP analysis and visuals that accompanied it. Along with a fair portion of data acquisition and processing. We were able to find datasets on Kaggle that were useful as well. Most notably, I scraped and cleaned the data for the trainers, the EXP, the level of when a pokemon can be captured in the wild, and gym sections for all the data sets that used them. My group mates were great and I think we worked very well together.

Individual Takeaways of Project

First and foremost the biggest learning opportunity and takeaway was the importance of, when working with multiple data frames, normalizing naming conventions and formats wherever possible. It felt as if every step of the way my work was hindered by misspelled column names, different naming conventions, and poor planning/commenting. For example, many of the data sets I used for the EXP analysis had a column 'gym section'. I, however, named them 'gym_section', 'gym.section', 'gym.Section' ... This in conjunction with not enough comments and notes, made it very difficult to join data frames, and pull the information that I needed. Therefore, working on this over the course of weeks, I would regularly forget my place, train of thought, and what I was going to do next. Another takeaway was how much of the work was processing and cleaning the data to get it in a state that was easily worked on. This was at least 80% of the work that I did. Thus, I am becoming more familiar with the importance of having clean, workable, and normalized data, before trying to do any kind of analysis. Pokemon is much more dense than we previously believed. There are so many variables and aspects to the game, it was very difficult to find any kind of meaningful information without making many simplifying assumptions. That is not to say our investigation was not useful, it is however largely significant only when playing the game with self imposed rules. Our results are meaningful and immediately understandable to someone who has prior domain knowledge and experience. As an avid pokemon player and fan, I felt I had learned useful information. Especially because I often play the game with many of the self imposed rules our simplifying assumptions are based on.

Discussion & Conclusion

Through our thorough investigation of pokemon typing, base stats, experience, and gym leaders in *Pokemon Platinum*, we were able to create eight different teams that would be well suited to defeat each gym leader. The primary objective was to create diverse teams that could effectively resist gym pokemon attacks and stay in the fight to avoid fainting. Our project successfully identified and assembled strong pokemon teams for each gym leader in *Pokemon Platinum*, along with showing that it is not feasible to have a fully max level team for every gym leader battle. This means that the lower ranked Pokemon will be a lower level than the highest ranked Pokemon for many of the gyms. Overall, our project provides valuable information on the strategy of team building with the goal of defeating gym leaders. By understanding the importance of typing, base stats, experience, and strategic considerations, trainers can create

teams that are well-equipped to take on all the gym leaders in *Pokemon Platinum* and emerge victorious.

In the future, it may be interesting to look into some of the strategies that we decided to exclude from our analysis, due to the complexity. For example, further research could focus primarily on the role of items when facing gym leaders. Would including these items make a significant difference in our findings? We could also examine which item would be the most effective against a certain gym. As items often need to be purchased in the game, this research could help players justify items to buy. Since we decided to utilize a more simplistic approach for our analysis, further research including more variables could be beneficial to finding the true “best” team to use when facing each gym leader.

References

“Experience- Bulbapedia.” *Bulbapedia, the Community-Driven Pokémon Encyclopedia*, bulbapedia.bulbagarden.net/wiki/Experience#Erratic. Accessed 3 May 2024.

“List of Pokemon by Effort Value Yield (Gen IV).” *Bulbapedia, the Community-Driven Pokémon Encyclopedia*, bulbapedia.bulbagarden.net/wiki/Experience#Erratic. Accessed 3 May 2024.

“Pokemon Pokedex: list of Pokemon with stats.” *Bulbapedia, the Community-Driven Pokémon Encyclopedia*, bulbapedia.bulbagarden.net/wiki/List_of_Pokémon_by_base_stats_(Generation_II-V). Accessed 3 May 2024.

“Pokemon Platinum Trainer Data.” *Smogon University - Competitive Pokémon Community*, smogon.com/forums/threads/pokemon-platinum-trainer-data.3665884/. Accessed 3 May 2024.

“Pokémon Evolution Charts.” *Pokémon Database -- the Fastest Way to Get Your Pokémon Information*, pokemondb.net/evolution. Accessed 3 May 2024.

“Pokémon Moves: lists of attacks.” *Pokémon Database -- the Fastest Way to Get Your Pokémon Information*, pokemondb.net/move/all. Accessed 3 May 2024.

“Pokémon Platinum- Sinnoh Pokedex.” *Pokémon Database -- the Fastest Way to Get Your Pokémon Information*, pokemondb.net/pokedex/game/platinum. Accessed 3 May 2024.

“The Complete Pokemon Dataset.” *Kaggle*, www.kaggle.com/datasets/rounakbanik/pokemon. Accessed 3 May 2024.

EXP gain formula: $((b \cdot L) / 7) \cdot 1.5$

b=base exp of a pokemon. L = level of the defeated pokemon