A Statistical Analysis of Pokémon Competitive Viability

STA210 Final Project - Jason Guan

Introduction

With over 480 million sales worldwide, Pokémon, short for Pocket Monsters, is one of the most successful game series in the world. Spanning 122 installations and roughly 28 years, the Pokémon games generally revolve around catching, training, and battling fictional creatures called "Pokémon" in a quest to become the Pokémon Champion, creating an entertaining gameplay experience for children and adults alike. Despite its simple premise and cartoon aesthetics, Pokémon is home to a rich competitive scene which has continued to grow and evolve since the establishment of the official Pokémon tournament circuit (Video Game Championships, or VGC) in 2009.

At the heart of this community is Smogon, an online hub which has been hosting competitions, forums, and strategy guides since the early 2000s. Above all, Smogon's greatest contribution to the competitive scene comes in the form of its widely respected tiering system, forming the basis on which most competitive online play is based. These tiers are based on Pokémon usage rates and dictate which Pokémon are allowed in different game modes, with more powerful Pokémon in the higher tiers and lesser used Pokémon inhabiting the lower tiers. At the top lies AG (anything goes) and Ubers, which include Pokémon like Arceus and Rayquaza that are typically banned from online play. Below this lies OU (OverUsed), UU (UnderUsed), RU (Rarely Used), NU (Never Used), PU (Perfectly Useless), and their respective Banlists, which contain Pokémon banned in certain tiers but allowed in higher tiers. At the bottom we see LC (Little Cup) and NFE (Not Fully Evolved), which consist of unevolved, statistically weak pokemon that are not typically used in normal online play.

This study seeks to answer the question of how factors such as stats, typing, and Mega/Legendary status correlate to a Pokémon's competitive rank while also determining the factors predicted to cause a Pokémon to be more viable in the Generation 6 competitive scene. In doing so, we will hopefully be able to recognize trends in Pokémon power-scaling and identify weaknesses in game balancing which could potentially be addressed in future generations.

Our Data

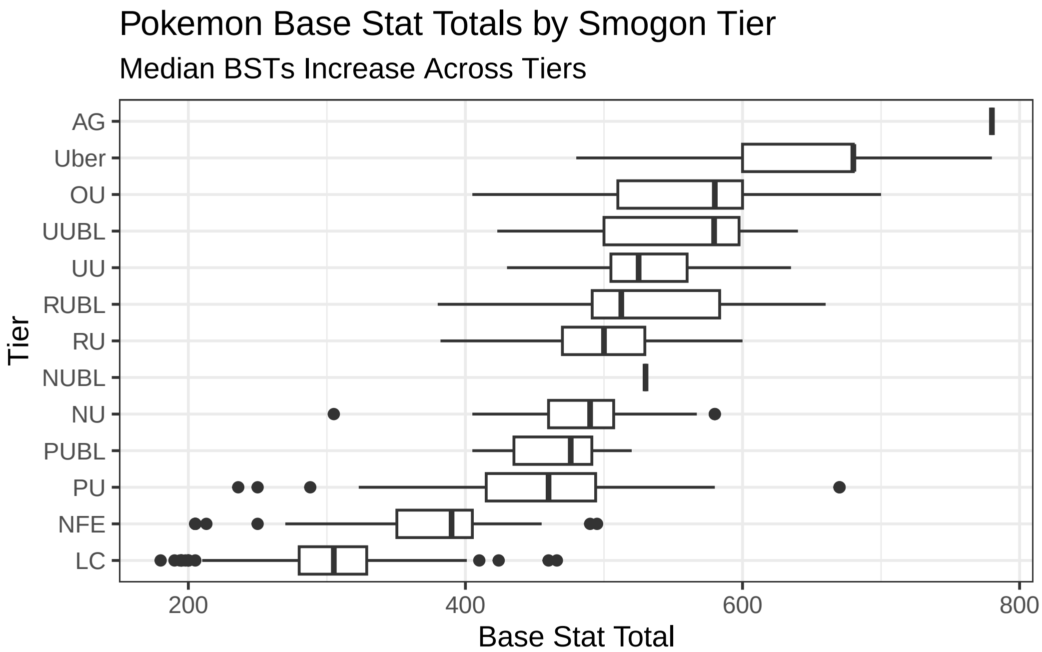
Due to a lack of available recent data, the two datasets used in this report come from Generation 6 of the Pokémon games, which includes the X/Y and Omega Ruby/Alpha Sapphire games. The smogon dataset features 499 observations and 21 columns which include all Pokémon in and above the PU (Perfectly Useless) tier. To include the Pokémon that were missing from this dataset (those below the PU tier), a subset of missing Pokémon was taken from the pokemon dataset, which features every Pokémon up to Generation 6 but does not store their competitive tiers. This new data was cleaned by changing all variable names to match those of the 'smogon' dataset and removing unnecessary variables such as catch rate and egg groups. It was then merged with the existing data to create the updated 'smogon' data frame used throughout this project. This final data frame was then further cleaned by giving the newly imported Pokémon their respective tiers (LC or NFE) and releveling the 'Tier' variable to reflect the actual ordering. Important variables in this data frame include:

- **HP, Attack, Sp. Atk, Defense, Sp. Def, and Speed**: 6 continuous variables representing a Pokémon's 6 in-game stats
- **Type.1, Type.2**: categorical variables which reflect the type(s) a Pokémon has (e.g. fire, water, grass, etc.)
- **Legendary/Mega**: boolean variables representing whether or not a Pokémon is Legendary (one of a kind in-game) or a Mega form
- **Tier**: a categorical variable storing the competitive tier (LC to AG) a Pokémon was placed in by Smogon;

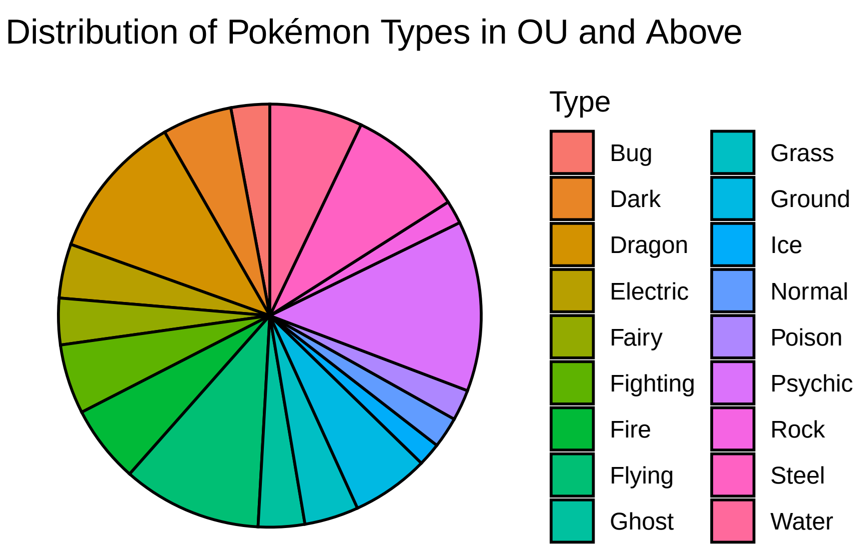
These tiers are ordered from lowest to highest as follows: LC, NFE, PU, PUBL, NU, NUBL, RU, RUBL, UU, UUBL, OU, Uber, AG

Exploratory Data Analysis

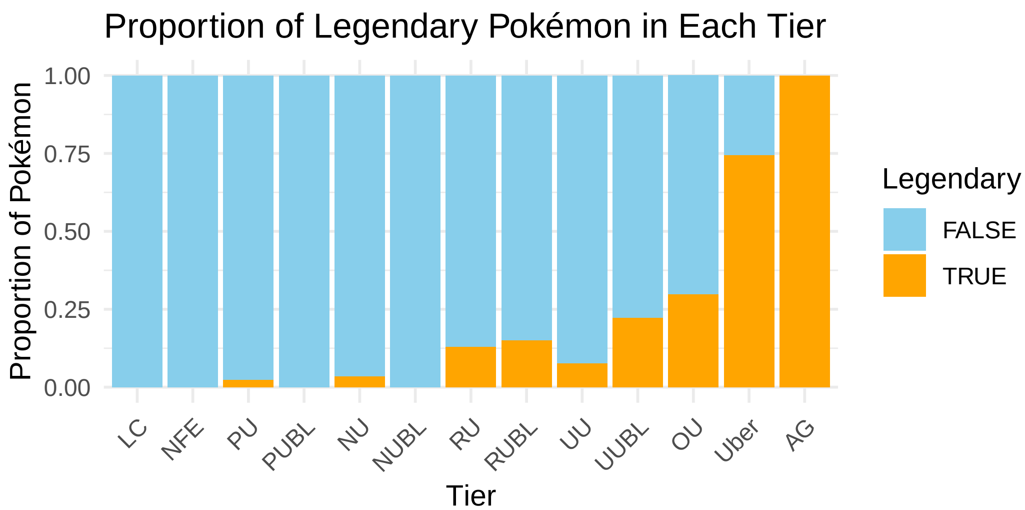
Given that the goal of this project is to analyze the factors behind Pokémons' competitive rank and viability, our exploratory data analysis will focus primarily on examining the distribution of various attributes among Pokémon across different tiers, such as stat totals, typings, and legendary or mega status. To start, the boxplot on the following page illustrates the distribution of Pokémon base stat totals across different competitive tiers. We can see in this plot that, with the exception of the NUBL tier, which only contains one Pokémon, the median base stat totals (BSTs) seem to **increase** as we move up the tiers, meaning that Pokémon in higher competitive tiers tend to have higher base stat totals. In the lower tiers, there are three outliers: Articuno (PU), Slaking (PU), and Regirock (NU). These Pokémon have high base stat totals, though all carry some competitive hindrance which keeps them in low competitive tiers. Regirock and Articuno boast weak/vulnerable typings, while Slaking's immense 670 BST is hindered severely by the "Truant" ability, which renders it useless every second turn.

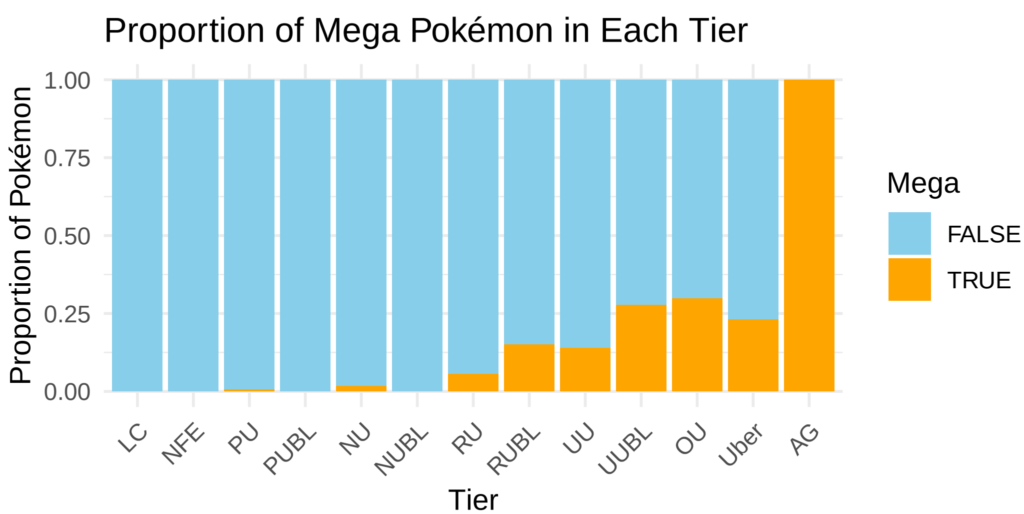


The figure below illustrates the distribution of typings across the OU, Uber, and AG tiers, the tiers where Pokémon are typically considered competitively viable. From this, we can see that Psychic, Flying, Dragon, and Steel type Pokémon are the most common in these tiers. However, note that Fairy, generally considered as one of the best typings in the game, was introduced in this generation, meaning there were significantly less Fairy Pokémon in the dataset compared to other typings.



The next two visualizations show the proportion of Pokémon in each tier that are legendary or mega-forms. Legendary Pokémon are characterized by their rarity, high BSTs, and prominent roles in the stories of each Pokémon game, while mega evolutions are new Pokémon forms introduced in Generation 6 that give existing Pokémon a new "mega" evolution, often with very high BSTs and new abilities. We see in the first plot that as we move up to more competitive tiers, the proportion of legendary Pokémon in each tier seems to increase, though not in a discernibly linear pattern. The proportions of mega Pokémon, on the other hand, do not seem to follow a pattern and are more randomly distributed between tiers. Finally, we can also see that both legendary and mega proportion peak in the AG tier. This is because it only contains one Pokémon, Mega Rayquaza, which is both legendary and a mega.





Methodology & Results

As previously mentioned, Smogon's tiers follow an ordered progression from weaker, less used Pokémon to more powerful, widely used Pokémon. This means that our corresponding 'Tier' variable is both categorical and ordered, making it a suitable response variable for an ordinal regression model. Therefore, an ordinal model was used to analyze possible factors influencing the placement of a Pokémon within Smogon's tiers.

The predictors used in fitting this model were chosen through exploratory data analysis. They include the six battle statistics (HP, Attack, Special Attack, Defense, Special Defense, and Speed) which act as continuous predictors, all 18 Pokémon types (Fire, Water, Grass, etc.) which act as binary variables storing '1' if a Pokémon is a certain type, as well as Legendary and Mega, two boolean variables which represent whether a Pokémon is legendary or a mega form. Below, we can see the exponentiated coefficients (odds ratios) of each predictor, while the full model output is shown on the next page.

```
           HP       Attack      Defense       Sp..Atk       Sp..Def
    1.0443814    1.0310772    1.0384013     1.0403195     1.0250162
        Speed       Normal         Fire         Water      Electric
    1.0478742    2.0579361    0.8158548     1.6681801     1.6532029
        Grass          Ice     Fighting        Poison        Ground
    0.8714908    0.7191876    6.8906596     3.3884788     2.8340858
       Flying      Psychic          Bug          Rock         Ghost
    1.9652854    1.7501212    5.3341063     0.4489678     3.3720892
       Dragon         Dark        Steel         Fairy      MegaTRUE
    1.1306751    4.3643223    4.1563215     9.2316649     0.3399962
LegendaryTRUE
    0.7797634
```

Firstly, we can see from this table that Speed boasts the highest odds ratio of the six battle statistics, followed by HP and Special Attack. This means that when we increment the speed of a Pokémon by 1 point, the odds that it is in the next highest Smogon tier are predicted to increase by **a factor of** 1.0478742 when holding all other variables constant. This makes sense since Speed is the stat which determines which Pokémon moves first, giving faster Pokémon a massive advantage by allowing them to potentially knock out opposing Pokémon before they can even move. This makes fast, fragile Pokémon such as Weavile and Gengar very effective.

Next, we see that the Fairy type has by far the highest odds ratio of any type, with an odds ratio of 9.2316649. This means that, holding all other variables constant, if a Pokémon is Fairy type, the odds that it is in the next highest Smogon Tier are predicted to be 9.2316649 **times** greater than if it were not a Fairy type. Fighting, Bug, Dark, and Steel boast the next highest odds ratios, with values ranging from 5.3341063 to 4.1563215. This also makes sense as Fairy types have strong offensive and defensive coverage on top of being very strong against Dragon types, which were formerly the strongest in the competitive scene.

term	estimate	std.error	statistic	coef.type
HP	0.0434248	0.0040025	10.8493326	coefficient
Attack	0.0306041	0.0036570	8.3686825	coefficient
Defense	0.0376823	0.0037812	9.9657987	coefficient
Sp..Atk	0.0395279	0.0037017	10.6784159	coefficient
Sp..Def	0.0247084	0.0038922	6.3482315	coefficient
Speed	0.0467636	0.0037098	12.6055096	coefficient
Normal	0.7217036	0.3148500	2.2922144	coefficient
Fire	-0.2035189	0.3210385	-0.6339394	coefficient
Water	0.5117333	0.2656796	1.9261295	coefficient
Electric	0.5027146	0.3501012	1.4359123	coefficient
Grass	-0.1375500	0.2764959	-0.4974757	coefficient
Ice	-0.3296330	0.3978871	-0.8284586	coefficient
Fighting	1.9301668	0.3385724	5.7008985	coefficient
Poison	1.2203811	0.3025610	4.0335042	coefficient
Ground	1.0417194	0.2914424	3.5743584	coefficient
Flying	0.6756375	0.2397446	2.8181547	coefficient
Psychic	0.5596850	0.2965606	1.8872535	coefficient
Bug	1.6741213	0.3068314	5.4561613	coefficient
Rock	-0.8008042	0.3219824	-2.4871057	coefficient
Ghost	1.2155325	0.3612850	3.3644696	coefficient
Dragon	0.1228148	0.3358923	0.3656376	coefficient
Dark	1.4734629	0.3168871	4.6498044	coefficient
Steel	1.4246304	0.3543638	4.0202486	coefficient
Fairy	2.2226394	0.3891619	5.7113486	coefficient
MegaTRUE	-1.0788209	0.3509835	-3.0737080	coefficient
LegendaryTRUE	-0.2487648	0.3319842	-0.7493271	coefficient
LC|NFE	14.1872229	0.6998987	20.2703959	scale
NFE|PU	15.6909592	0.7593159	20.6645996	scale
PU|PUBL	18.5774939	0.8680888	21.4004543	scale
PUBL|NU	18.6947912	0.8706721	21.4716793	scale
NU|NUBL	19.5622984	0.8903915	21.9704460	scale
NUBL|RU	19.5780898	0.8907300	21.9798251	scale
RU|RUBL	20.4791496	0.9117596	22.4611277	scale
RUBL|UU	20.8363113	0.9200681	22.6464882	scale
UU|UUBL	22.1192106	0.9478058	23.3372814	scale
UUBL|OU	22.5374163	0.9556882	23.5823946	scale
OU|Uber	24.5850214	1.0122671	24.2870887	scale
Uber|AG	30.6729419	1.5913466	19.2748348	scale

Another goal of this study is to determine the factors that make a Pokémon competitively viable. To do this, we can characterize "viability" as the state of being in any of the OU, Uber, or AG tiers since these Pokémon are either the most used in online play (OU) or deemed too powerful for OU competitive play (Uber/AG). We can model this using a logistical regression, where our response variable, 'Viable', is a new binary variable created that stores '1' if a Pokémon is in any of the aforementioned tiers. The predictors used to fit this model include base stat total, a continuous predictor representing the sum of the 6 battle statistics, legendary and mega status, which are both binary predictors, and the previously defined 18 binary "type" variables which each represent whether or not a Pokémon has a specific type. The complete model output can be found below:

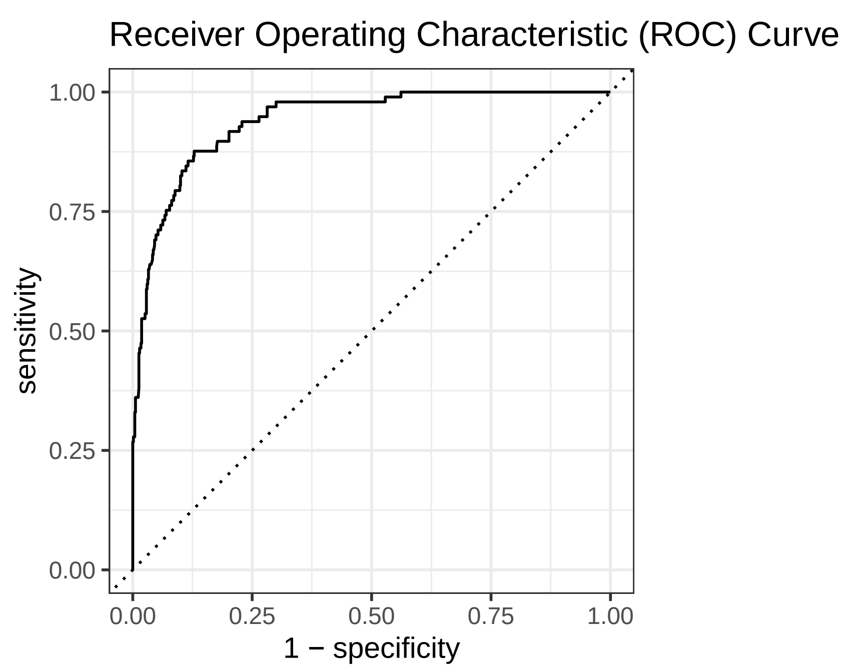
term	estimate	std.error	statistic	p.value
(Intercept)	-13.8591315	1.8784912	-7.3777995	0.0000000
Total	0.0202499	0.0034539	5.8628355	0.0000000
MegaTRUE	0.3652744	0.5084158	0.7184560	0.4724762
LegendaryTRUE	0.9009912	0.5045504	1.7857309	0.0741428
Normal	-0.3129475	0.7891866	-0.3965443	0.6917035
Fire	0.3319178	0.6074808	0.5463841	0.5848020
Water	0.7892390	0.5722356	1.3792205	0.1678268
Electric	1.0481576	0.6715366	1.5608347	0.1185628
Grass	0.5592544	0.6036886	0.9263955	0.3542405
Ice	-0.6977699	0.8746278	-0.7977907	0.4249920
Fighting	0.9661625	0.5915732	1.6332086	0.1024251
Poison	0.9761627	0.7119868	1.3710404	0.1703624
Ground	0.9502310	0.6008827	1.5813920	0.1137884
Flying	0.9645798	0.4874955	1.9786436	0.0478561
Psychic	0.8633685	0.5589087	1.5447396	0.1224093
Bug	0.6946905	0.7220632	0.9620910	0.3360039
Rock	-1.2550038	0.8790030	-1.4277583	0.1533614
Ghost	1.1271990	0.7005842	1.6089416	0.1076291
Dragon	0.2702855	0.5795304	0.4663871	0.6409384
Dark	1.2111897	0.6074089	1.9940269	0.0461491
Steel	1.7664456	0.5520412	3.1998437	0.0013750
Fairy	1.4197424	0.7193905	1.9735350	0.0484346

To determine which, if any, of these predictors are statistically significant, we can perform hypothesis tests using a significance level of 0.05. In doing so, we find that five predictors have p-values below this significance level: Total (which represents the base stat total), Flying, Dark, Steel, and Fairy, which carry associated p-values of 4.55e-09, 0.04786, 0.04615, 0.00138, and 0.04843 respectively. This means that for each of these variables, we can reject the respective null hypothesis (which states that there is no relationship between the predictor and the outcome variable) and conclude that there is sufficient evidence to suggest a relationship between each of these variables and the outcome variable.

We can interpret these results as follows. Total has an associated coefficient of 0.020250, meaning that when we increment a Pokémon's base stat total by 1 point while holding all other variables constant, the odds that a Pokémon is viable are predicted to increase by **a factor of** e^0.020250 = 1.02045642225. This makes sense given the earlier visualization which shows a tendency for Pokémons' median BST to increase across tiers.

Of the four type variables that are considered statistically significant, Steel carries the highest coefficient of 1.766446, followed by Fairy (1.419742), Dark (1.211190), and Flying (0.964580). This means that, holding all other variables constant, the odds that a Steel-Type Pokémon is viable are predicted to be e^1.766446 = 5.85002538178 **times** greater than if that Pokémon were not Steel-Type. This makes sense given that Steel is resistant to and strong against the very popular Fairy type while also being defensively resistant against many other powerful offensive types such as Ice, Flying, Psychic, and Dragon.

Since this is a binary classification model, we can evaluate this model's performance by analyzing the ROC curve, which can be seen below:



```
# A tibble: 1 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 roc_auc binary         0.940
```

We can see from the data above that the AUC (Area Under Curve) is 0.9399485. Since this value is quite close to 1, this means that the model has strong discriminatory power and is able to correctly identify Viable and Non-Viable Pokémon most of the time.

```
                          0   1
  Not Viable (Under OU) 679  40
  Viable (OU or Above)   21  57
```

The confusion matrix above shows us all of the true/false positives and negatives. From this, we find that we have a sensitivity of ~0.588, a specificity of 0.97, a positive predicted value of ~0.731, and a negative predicted value of ~0.944. Overall, this means that the model is better at making negative predictions than positive predictions, having a higher specificity and NPV with a lower sensitivity and PPV.

Conclusion & Discussion

In the end, we find that of a Pokémon's six battle statistics, Speed is the stat which is predicted to cause the greatest multiplicative increase in the odds of being in a higher competitive tier. Typing-wise, the very strong Fairy type is predicted to cause the greatest multiplicative increase in the odds of being in a higher competitive tier by a very large margin, followed not-so-closely behind by Fighting, Bug, Dark, and Steel. This tells us that, despite its novelty at the time and limited representation, the Fairy type has been a competitively strong type since its inception. In terms of viability, we see that there is a statistically significant positive correlation between a Pokémon base stat total and the odds it is competitively viable. Similarly, we also observe a statistically significant positive correlation between a Pokémon being either Steel, Flying, Dark, or Fairy type and the odds it is viable. This not only reinforces the fact that Fairy is a strong type, but also emphasizes the viability of types like Steel which directly counter it.

There are several limitations to this project. For one, the data used in this project comes from a past generation of the Pokémon games, and much has changed since then. With the release of the Generation 9 games, the competitive scene has witnessed the rise of even stronger Pokémon, new entries in the "AG" tier, and a new battle mechanic which allows Pokémon to change types. Specifically, the emergence of new Fairy Pokémon, which were very limited in the Generation 6 games, and the rise of other new strong Pokémon may change the distribution of stats, types, and other factors, making these results potentially outdated. As such, one piece of future work that we can consider performing once the necessary data is available is to fit these same models using Generation 9 data and seeing how the results yielded are different. Still, despite these limitations in data gathering, this data offers an interesting snapshot into a pivotal period in the development of competitive Pokémon.

Another limitation of this project is the exclusion of two important factors in competitive viability. Though efforts were made to include every relevant predictor in the models we fitted, there are a couple notable aspects of competitive battling which were not included.

Firstly, this study does not analyze dual-typings or different type combinations. Though we have looked into which single types seem to be stronger than others, there are certain type combinations can make Pokémon with weak individual types much more powerful. For instance, the Bug type is widely considered one of the worst types due to a plethora of weaknesses. However, when we pair the Bug type with the Steel type, which perfectly complements its strengths and weaknesses, we create a strong dual-typing with only one weakness in Fire. Though this data is available, it would be extremely time-consuming and difficult to fairly assess whether each dual-typing has any statistical significance and if it is predicted to cause any changes in a Pokémon's odds of viability since this would involve making new variables to represent each of the 171 unique type combinations. However, given that this data is available, there is potential for future work to be conducted regarding this subject, such as creating a variable which simply returns if a Pokémon is dual-typed at all and including this in future regressions.

Additionally, this study completely neglects a very important aspect of competitive Pokémon: abilities. Abilities, which are unique passive effects that automatically activate during battle, can drastically change the viability of certain Pokémon. For instance, Mega-Blaziken is the mega form of Blaziken that gains the ability "Speed Boost" which automatically raises its Speed stat every turn it is on the field, making it extremely powerful and causing it to be placed in the Uber tier. On the other hand, the aforementioned Slaking boasts an incredibly high base stat total, only to be hindered by its "Truant" ability, which effectively renders it useless every other turn. Despite their competitive significance, it is not possible to account for these abilities within our models since they can have complex interactions with other game mechanics as well as subtle effects that are difficult to quantify or accurately predict.