Pokémon! - Gotta Predict 'Em All!

In this project we had several questions up front that we tried to answer. These are:

- Does Pokemon with two types separate themselves from those who only have one?
- Do Pokemon types relate differently to each other in the newer generations?
- Can we predict if a Pokemon is legendary?
- Can we predict the type of a pokemon?
- Can we predict the outcome of a Pokemon battle, based on the features of the Pokemon that are fighting?

Alongside these questions, we performed exploratory data analysis in order to gather more insight and build upon the existing set of questions.

Exploratory Data Analysis

We explored the following areas in the EDA:

- Correlation between the average statistics distribution over the generations.
- A heatmap showing the correlation between all features in the dataset.
- The number of normal and legendary Pokemon in each generation.
- The number of each type in each generation.
- The fluctuation of the average pokemons BMI over the generations.

The correlation between the average statistics distribution over the generations shows the average Pokemon health points, attack, defense, special defense and special attack from each generation. This showed that pokemons in general seems to be stronger in generation 4 and 7. The graph reveals that the balacing of the Pokemon statistics have gone through quite a bit over the span of the 7 generations. for example we see that between generation 1 and 2, all the attack statistics have gone down while the defense statistics have gone up. This balances out a bit in generation 3 and then spikes on across all statistics in generation 4.

The correlation heatmap revealed what features seemed to be more important for prediction. We test this later on.

The number of each type in each generation raised the question of how the types actually relate to each other, and if this has changed over the generations.

The calculation of the average BMI in each generation did not raise any new questions, but it did reveal some interesting insights about the design of the pokemons in generation 7. In this generation, the average BMI is much larger than all other generations. This is due to the introduction of Pokemon such as Cosmeom, who weighs about 1 ton, but is only 10cm tall. Due to these extreme data points, we calculated the median instead. When doing so, we see that median BMI have barely changed over the generations.

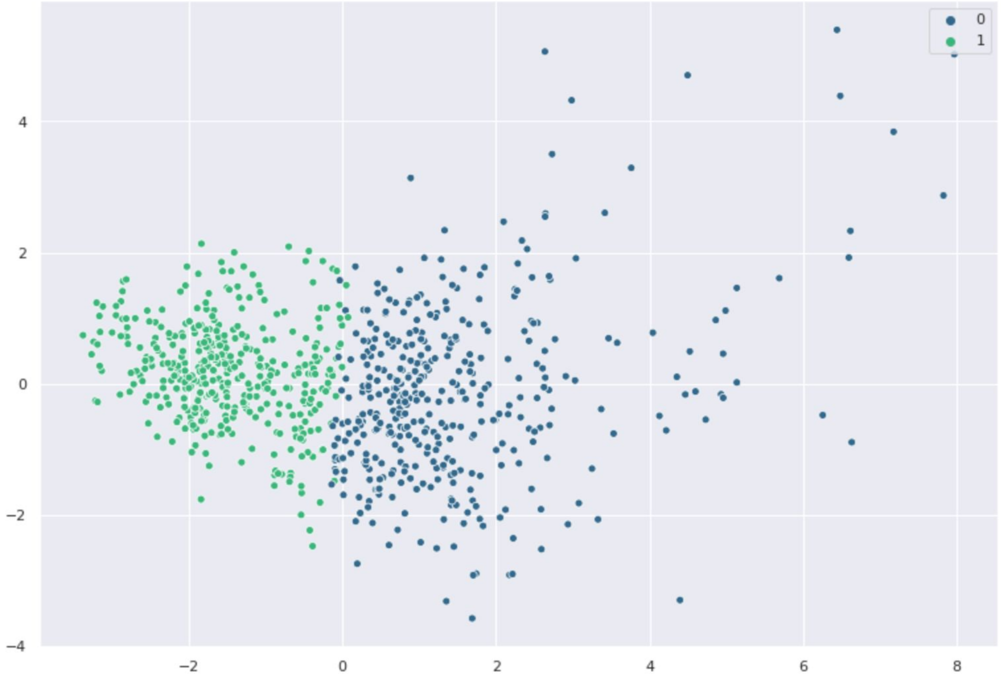
Unsupervised machine learning

In the unsupervised machine learning part we delved into the following questions:

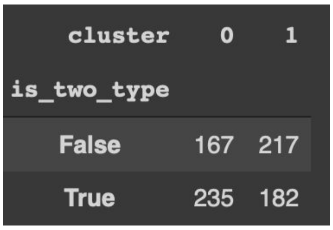
- Does it matters if a Pokemon has two types or not
- Are Pokemons types correlated?

Does it matter if a Pokemon has two types or not?

We made a scatterplot and found that there is in fact some kind of clustering that is very close, and some that are more dispersed in the upper right corner.



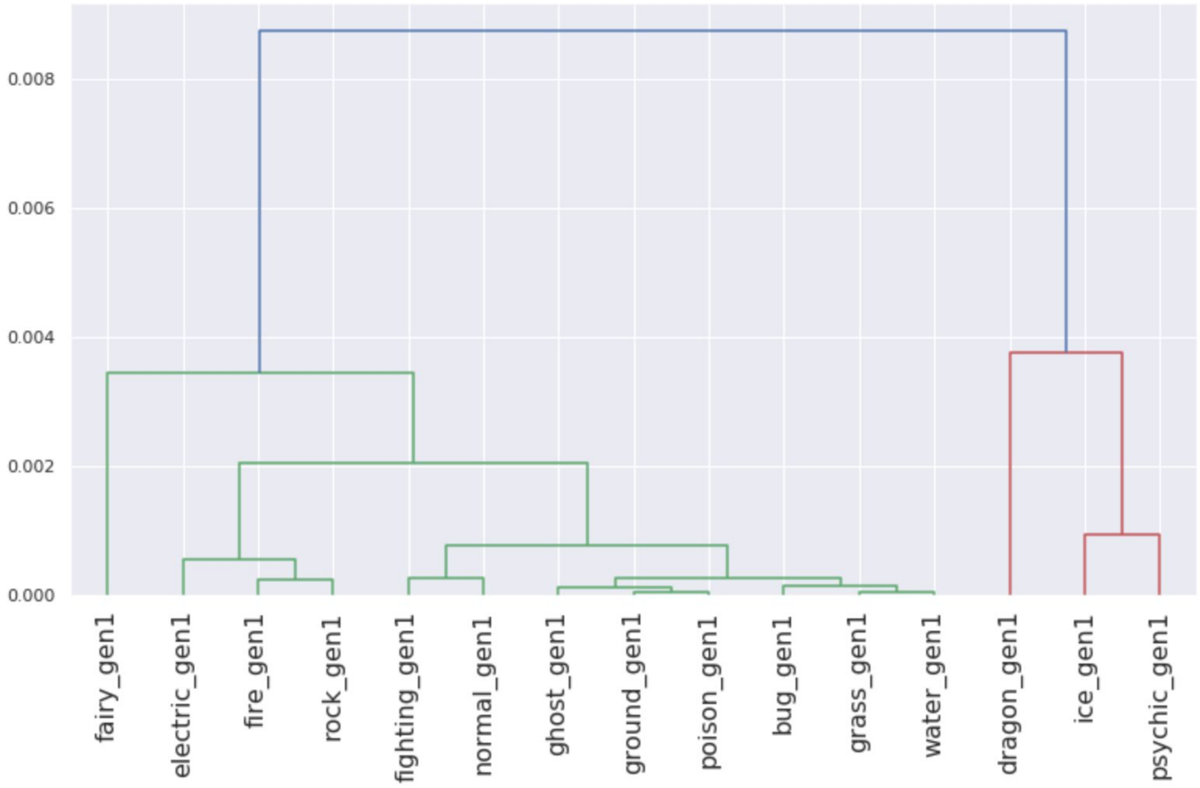
Though there is a clustering, we found that it was not whether they had two types or not,, as seen in the crosstab below. Note: If it was a factor, the numbers would have been more to one side, for example instead of “False: 167, 217”, it would be more like “False: 370, 14”, and reversed in the True row.



cluster	0	1
is_two_type		
False	167	217
True	235	182

Are Pokemon types correlated?

In the exploration of the data, we saw that the number of Pokemon types created fluctuated greatly, which sparked the question of, if there is a correlation between types and if new types are added or existing ones are left out, does the correlation change?

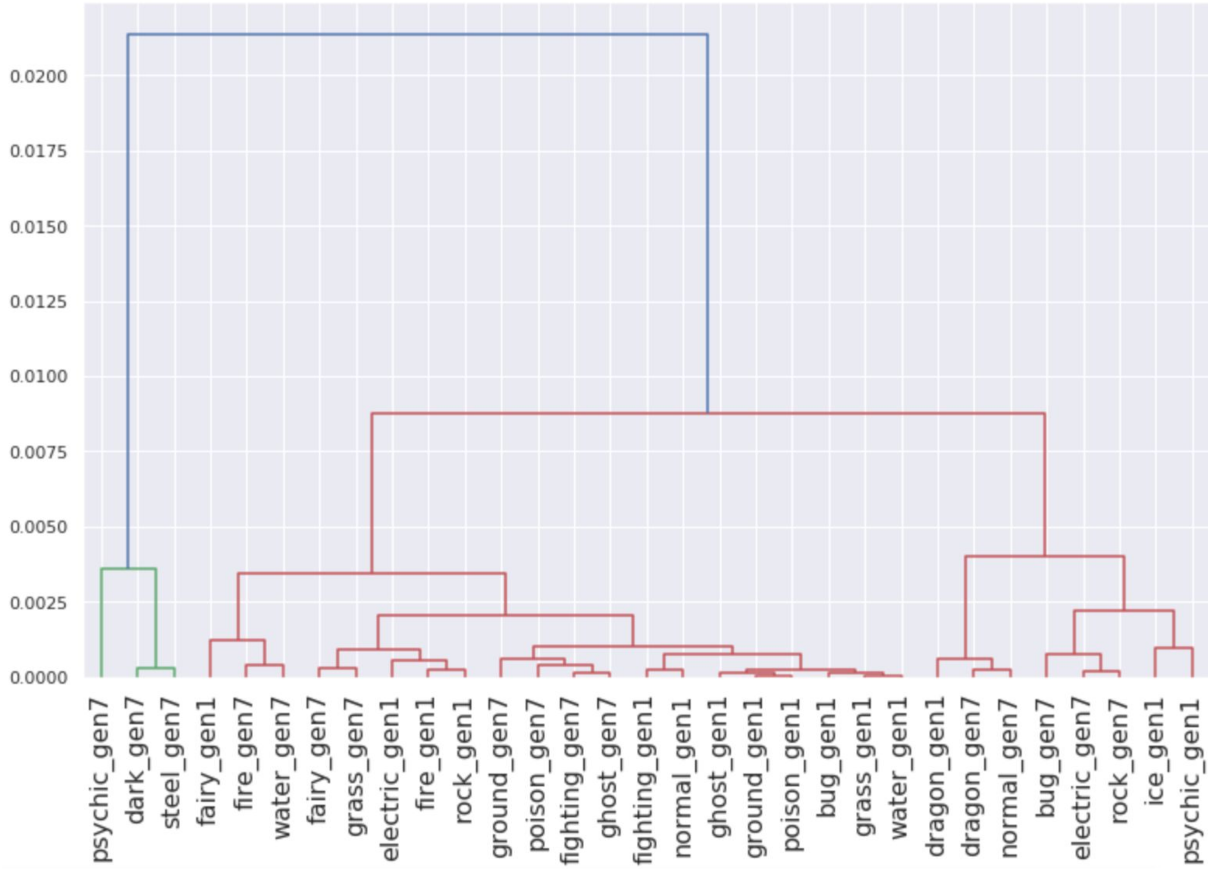


Just by glancing over the graph we see that the first clusters are created by ground/poison and grass/water. This might have something to do with their strength/weakness towards each other, since ground deals double damage to poison and grass deals double damage to water. The bug-type is added to the grass/water cluster afterwards, whereas bug deals double damage to grass.

We then have two more clusters that merge, which is rock/fire and fighting/normal, which also seems to be grouped by their strength/weakness correlation. Rock deals double damage to fire and fighting double to normal.

Dragons merge to ice/psychic, which is weird, considering ice doesn't have a strength/weakness correlation with psychic, but rather deals double damage to dragon.

This shows that the strength/weakness correlation between the Pokemon is a big factor in the dendrogram, but not the only one. It also shows the mindset of the design of Pokemon, not only types or the number of Pokemon of each type, but the correlation between the types.



The dendrogram still separates the generations in the smaller clusters. The first generation 7 cluster is made from fighting/ghost, which is different from generation 1, where the correlation of fighting was with normal. This might be due to the newer type2, that made for a wider combination of pokemons, an example is the pokemon Marshadow, who is of type fighting/ghost in generation 7.

In generation 7 there were no ice-types, and therefore no correlation with psychic. The correlation for psychic is completely different and is now with two new types, that didn't exist in generation 1, steel/dark.

In this graph we see the difference of thinking in the design of Pokémon and how new types and types left out, influence different correlations between Pokemon types.

Supervised machine learning

In the supervised machine learning part, we delved into the following questions:

- Predicting if a Pokemon is legendary.
- Predicting the victor of a fight.
- Predicting types.

For the first question where we predict if a Pokemon is legendary, we simply use the legendary feature as our target and use the rest of the features to train the model. We used 3 different models and found that all of them did well when predicting the legendary status of a pokemon.

From the correlation heatmap we created in the EDA, we saw that a specific feature called "base_egg_steps" had a high correlation with the legendary status. Removing this feature from the training also revealed a decrease in performance with about 20%. This makes sense as the feature describes how many steps the player has to take in the game, to hatch an egg containing the pokemon. Legendary pokemons of course take longer time, and thereby more steps, to hatch. This also explains why there was a bad performance when predicting whether a Pokemon is legendary in the mini assignment, where we were working with a much smaller dataset of the Pokemon, as it simply lacked the proper features of the pokemons to make accurate predictions.

Predicting the victor of fight were done by generating a combined dataset that combined all the pokemon features of 2 Pokemon with an true/false outcome of the battle. Once again we tried 3 different models and found a model that had about 75% accuracy when predicting the outcome, when hyperparameter tuning our model.

The fights datasets consisted of 50.000 battles and their outcomes, which even though it sounds like a lot, probably still isn't enough to capture the likeliness of a given pokemon winning, given it and its opponents statistics. Many things factor into a Pokemon battle besides the statistics of the Pokemon fighting. Weather effect, randomness factor, attack sequence and level are just some of the factors, that we do not know when predicting the outcome of a battle.

Here it could be interesting to create an algorithm that would simulate Pokemon battles and generate a hopefully more accurate dataset of battles when would create a better foundation for training a model in predicting the outcome of a given battle.

When we predict the types of Pokemon, we isolate the type feature as our target and apply 3 different models with hyperparameter tuning. We found that our models in general had a high accuracy when predicting the types, which lead us to try and predict the types of generation 7, based on the previous 6 generations of Pokemon. We managed to do so with an accuracy of about 80%, which is pretty good, but also indicates that there has been a change in the way that types are assigned to Pokemon, or a change in their design philosophy. This fits with our observation of the hierarchical clustering of the types got generation 1 and 7.