ReadMe

December 10, 2020

1 Pokemon Data Project

For this project, I was inspired by the Data Visualization lesson and decided to do a full analysis on a Pokemon dataset. There are just shy of 900 Pokemon now as of 2020, making it just the right size to base this project around.

To gather this data, I downloaded two datasets in CSV form from Kaggle (Main Data: https://www.kaggle.com/mariotormo/complete-pokemon-dataset-updated-090420?select=pokedex_%28Update_05.20%29.csv / Supplemental Data: https://www.kaggle.com/edgaro/pokedex-gen8) which I then wrangled and prepared for the exploratory data analysis detailed in the Jupyter notebook and Report.

2 Exploratory Data Analysis Findings Summary

In general, it seems that Pokemon is a very well balanced game - most of our findings in comparing various Pokemon variables support this conclusion as follows:

- The first analysis we did was to look at Pokemon type and their respective total stat distributions. Here, we actually see the biggest bias throughout our analysis in that Dragon type does have a noticeable edge over any other type.
- Next, we took a look at each individual stat (still grouped by Pokemon type), and noticed that while some types did have slight edges in certain categories (just like Dragon type did for stat totals), this was usually balanced by the fact that this same type tended to be trailing in the counterpart stat (Note: Special Attack is a counterpart to Attack and Special Defense is a counterpart to Defense). We verify this observation through the "Attack vs Special Attack" and "Defense vs Special Defense" data visualization.
- This result actually supports our overall story that Pokemon is well balanced, because types that have higher than average metrics for certain individual stats were balanced by having lower than average metrics for their counterpart stats. No type shows dominance across the board, and Dragon type's edge in stat totals comes from their consistently high (but not the best/dominant) performance in every stat.
- Next, we turned our interest to Pokemon Status. Here, we see that status does play a significant impact on Pokemon stat totals, but this is to be expected because Sub Legendary, Legendary, and Mythical Pokemon are by nature extremely rare, and are designed to be some of the strongest Pokemon in the game. The rarity of these special Pokemon is the balancing component that limits these special Pokemon from dominating the game.

- So after looking at each of these different variable interactions, we then turn to Generation, each of which represents about a 5 year gap in releasing new Pokemon (i.e. Generation 1 to Generation 2 is about a 5 year time difference). Our interest in looking at this time variable stems from the fact that in some video game series, game critics will point out how as time progresses, new characters (or Pokemon in this case) will sometimes steadily increase in power level which renders older characters/Pokemon obsolete (a phenomenon known as "powercreeping"). In Pokemon, we can sort of see a small trend like this in Generations 1-4, noting that in Generation 4, there is a higher proportion of Pokemon resting in the higher peak range of stat totals as opposed to Generation 1 which was a more even bimodal distribution. However, you can immediately notice that this actually rebalanced in Generation 5, and then when a similar slight creep towards the higher peak occurred again in Generation 6 and 7, this was rebalanced again in Generation 8.
- Now we begin to juxtapose some of these variables to see if they may infer some sort of correlation/causation between themselves. For Generation and Pokemon Status, we can see that there were more special Pokemon added in the progression from Generation 1 to Generation 4, but then this stabilized going into Generation 5 and 6, peaking in Generation 7, before balancing back down to average amounts for Generation 8. This trend is similar to the stat total upwards creep described in the previous section.
- Then looking at Pokemon Type and Pokemon Status together, as explained in our Exploratory Data Analysis report: Psychic type has the most "special" Pokemon, with the highest number of Sub Legendary (tied with Electric for 5), Legendary (7), and Mythical (5) Pokemon overall. This is interesting since when we first looked at overall stat distributions by type, Dragon type stood out as the slight winner above other types, and since we observed in the status analysis portion that being a "special" Pokemon tends to mean boasting much higher stats than "Normal" Pokemon, we would've expected Dragon types to have a much larger amount of special status Pokemon types than is the reality shown here.
- Finally, we end our analysis by looking at some final remaining variables, like the comparison between mono and dual type Pokemon (which we found to be relatively balanced, only a slight leaning towards higher stats for dual types) and weight and height (both of which had relatively unsurprising results - as a Pokemon gets bigger/heavier, it tends to be stronger).

In creating the Explanatory Analysis, I wanted to focus on the data points that would help support the claim that Pokemon is a very well balanced game.

3 Key Insights for Presentation:

Pokemon typically balances stat totals through the following key patterns:

For variables that have a tangible correlation to higher stat totals (Dragon Type and Special Statuses), Pokemon balances these competitive advantages by scarcity and spreading out these innately stronger Pokemon across different generations and types.

For individual stats (as components of the stat total), Pokemon balances advantages in individual stats by typically having disadvantages in related individual stats.

No Generation is meant to be significantly stronger than any other.

Stat totals are multimodal, allowing for flexible ranges of weaker and stronger Pokemon across

types, generations, and even statuses instead of having pocketed concentrations of totals around either strong or weak zones.