Alex Perera

29th, April 2022

Data Wrangling & Exploratory Analysis with Dr. Shi

Pokémon with Stats Summary

1. Introduction to Dataset

This dataset allowed me to deep-dive into the competitive E-Sport side of the Pokémon Video Game Tournaments (VGTs), and break down the Pokémon used in their teams at the highest level in the Master’s Division. The dataset encompasses over 700 different Pokémon, however the dataset row index and the Pokedex ID index do not align with each other since some Pokémon have alternative forms (Mega Evolutions, the Pokémon Rotom has different appliance-forms, and some legendary Pokémon). All questions began with how certain Pokémon can be utilized and further question if there are other Pokémon available that allows for further flexibility in teams through the analysis of their in-game stats for their collective Total of Stats, HP (Health Points), Attack, Defense, Speed, and Special Attack and Defense. The Pokémon with Stats dataset allowed to explore not just their core stats which affect the game, but gave insight to each Generation/Series, their Pokédex ID, whether or not they’re legendary Pokémon, and their type (fire, water, poison, fairy, dragon). Utilizing this dataset did initially lead to many questions to be asked, and eventually the four prioritized questions formed based off the E-Sport results for the Pokémon Master’s Division VGTs in 2014 and 2015.

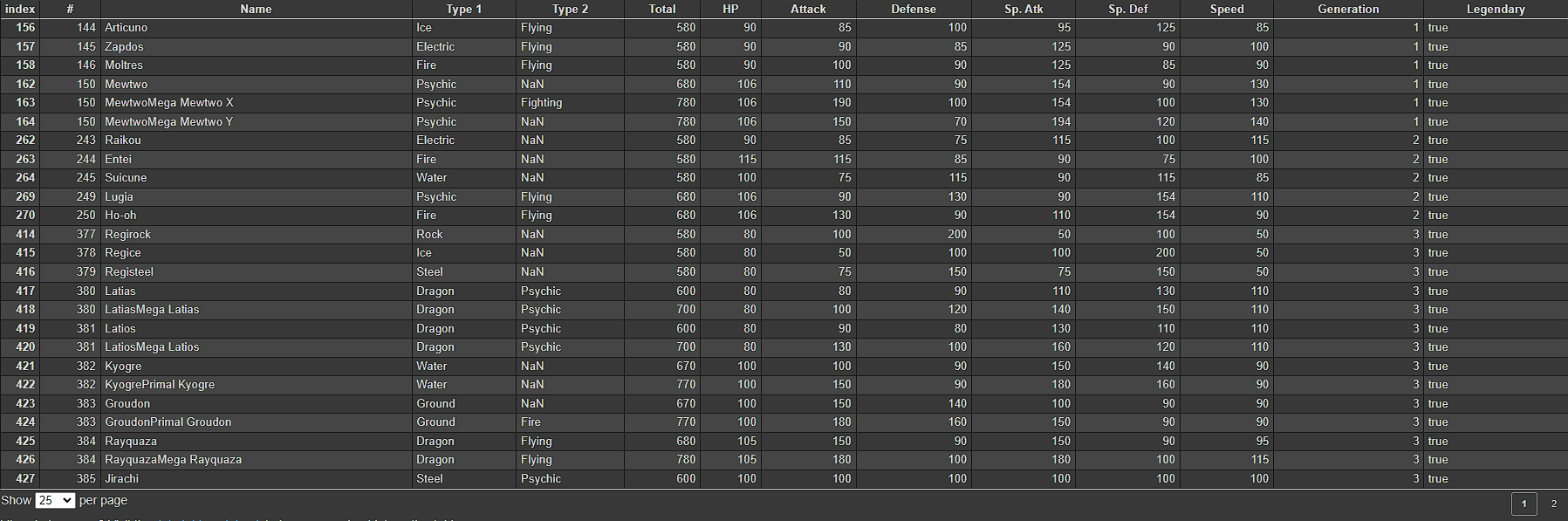


Figure - Legendary Pokémon Dataframe

1. Issues Encountered

Typing for Pokémon was an interesting an issue since some Pokémon are of a singular type, whereas others (especially as Pokémon evolve) may have dual-types or gain dual-typing. This issue this led to was not being able to use the function call: “drop.isna()” because it would remove all of the Pokémon who are of a single type and ultimately skew the dataset. Resolving this problem as to simply not utilize that function call, since the typing of Pokémon are sensitive and relevant to both the game and the competitive scene. Additionally, another problem I had encountered with the dataset, was its flexibility with different plotting functions (Seaborn and Matplotlib/Pyplot) because the data being passed into the arguments would sometimes call for error, seeking a Boolean syntax when nothing Boolean is being passed. Most of the plot issues encountered led either learning/recognizing a solution through the parameters and prior data preparation steps.

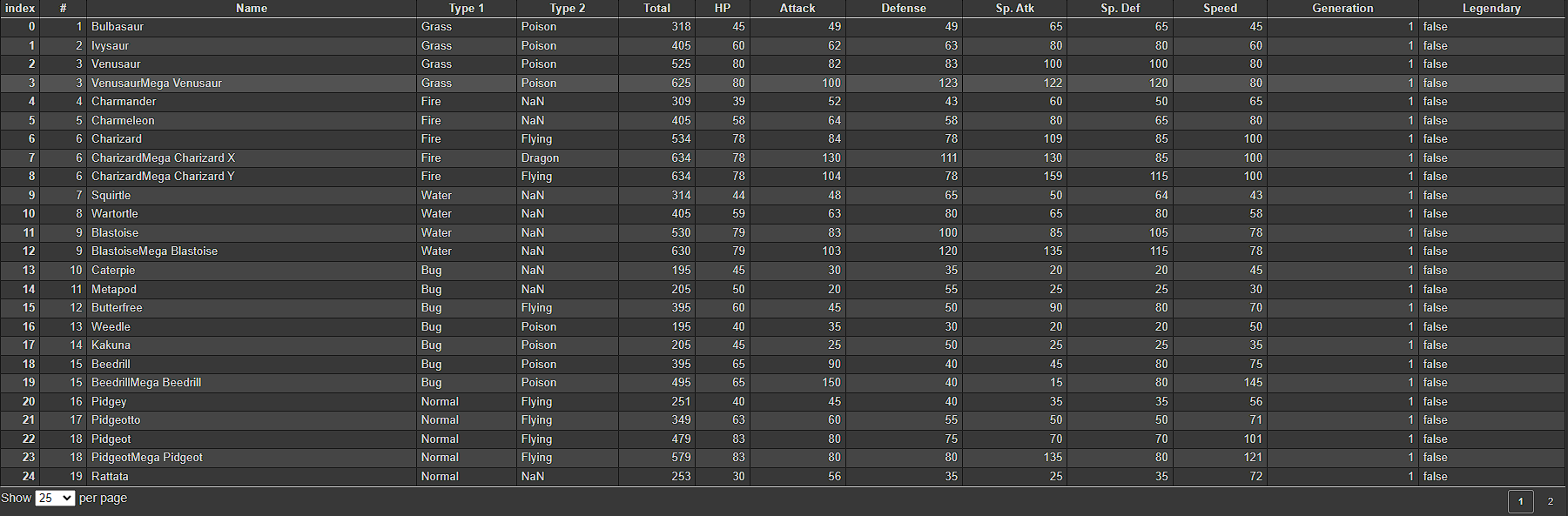


Figure - The Dataset expressing typing problem

1. What I learned/Conclusion

Exploring and Analyzing the dataset forced me to learn about the competitive Pokémon scene, and investigating the rules, rhythm, and discovering the previous champions of 2014 and 2015, which led to very interesting comparisons between the teams of the champions as well as the runner ups. The first question I explored in this data was with the reoccurring ‘Cute Electric Mouse’ Pokémon (AKA: Pikachu-like) and whether or not there are other stat-advantages within the similar Pokémon because the 2014 Champion had credited and utilized a Pokémon called Pachirisu to winning the championship. Exploring this I had discovered that, even though Pikachu can evolve into Raichu and receive a major stat-boost from evolving and generally had the highest overall stats; Pachirisu had higher Defensive (Defense and Sp. Defense) stats than all the other Pokémon, which ultimately provided an edge in their battle.

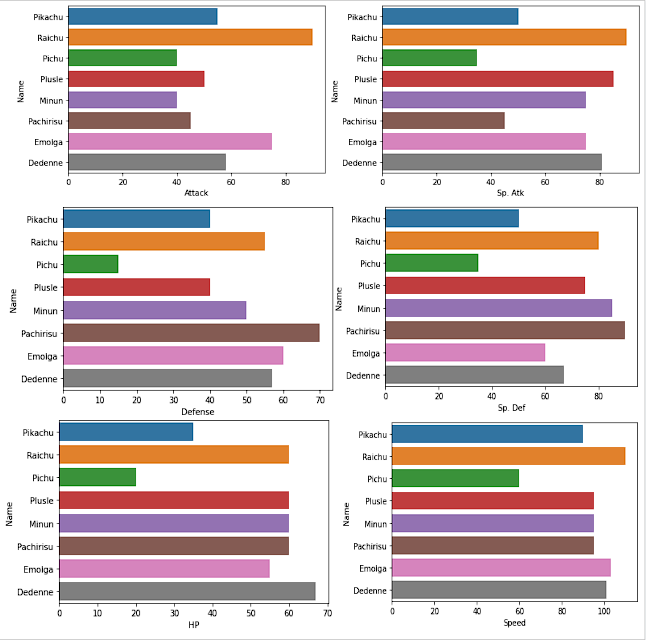


Figure - Pikas Dataset comparing Pokémon Stats

The second question begins with the starter-Pokémon which can heavily influence the start of the game, and commonly they find themselves in the teams of players in the VGTs. I created a data-frame of just the starter Pokémon and explored their Total stat distribution by their Generation. Exploring the data-frame generationally revealed that the first 4 Generations had very similar stat distributions amongst their 3 starter Pokémon. Generations 5 and 6 had Pokémon that had stats that were more similar to each other, creating more of a balance amongst the starting Pokémon.

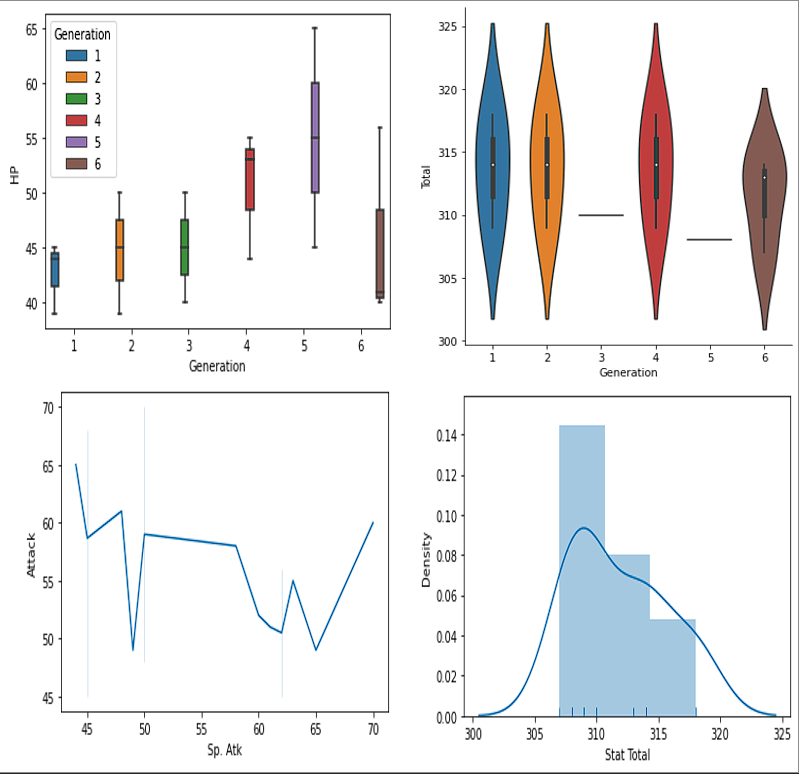


Figure - Starter Analysis

Lastly the third and fourth question explore the competitive teams used in the 2014 and 2015 championships. The third question compares the difference between a casual player who enjoys playing Pokémon and the Runner Up from the 2015 Championship (Wolfe Glick). The casual player had chosen Pokémon he had felt was a strong team and could handle just about any other team thrown it, while the Wolfe’s team was designed to be defensive. When examining the stat differences through a bar-plot, it was noticed that Wolfe’s team had rounded out high-stats, while the Casual player had stats that were varying. Higher level players tend to scrutinize and play the game to the smallest details, where every move made towards designing the team is relevant.

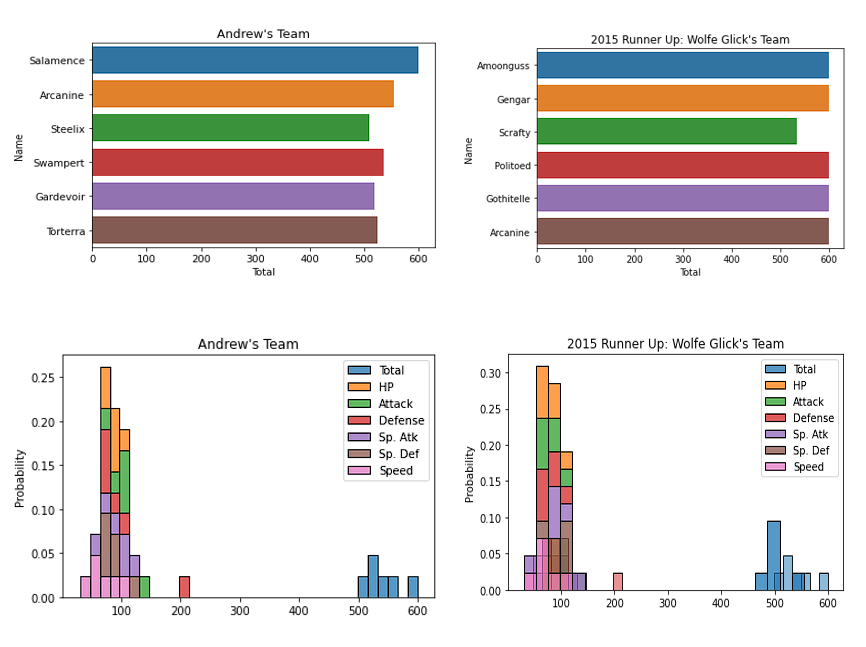


Figure - Andrew's Casual Player Stats

Lastly the competitive team differences were explored through histograms, violin plots, and a pie chart. Violin plots compared the stats of their Pokémon and gave a greater insight to how the team’s stats distributed in their respective fields. The 2014 championship showed different stat comparisons for their 4 Pokémon doubles battle, as it paced similarly to a normal Pokémon battle, where small combinations of certain stats for Se Jun Park’s Pachirisu (high defenses and speed) granted the victory. 2015’s Championship was different, and involved a higher level of mind games. Both players fought hard, playing defensively and offensively, often time swapping their Pokémon out. The two teams were similar in stats, however where the battle made its differences lied in the move sets (list of attacks/options a Pokémon has) of the Pokémon they have chosen. Ashton Cox, the 2015 Champion played for a long game, and used a move called “Perish Song” which will 1-hit-KO your opponent’s Pokémon by a 3rd turn. There are a lot of other interesting insights that can be pulled from here, especially with using the data to predict a winner based of their stats and possibly their move set.

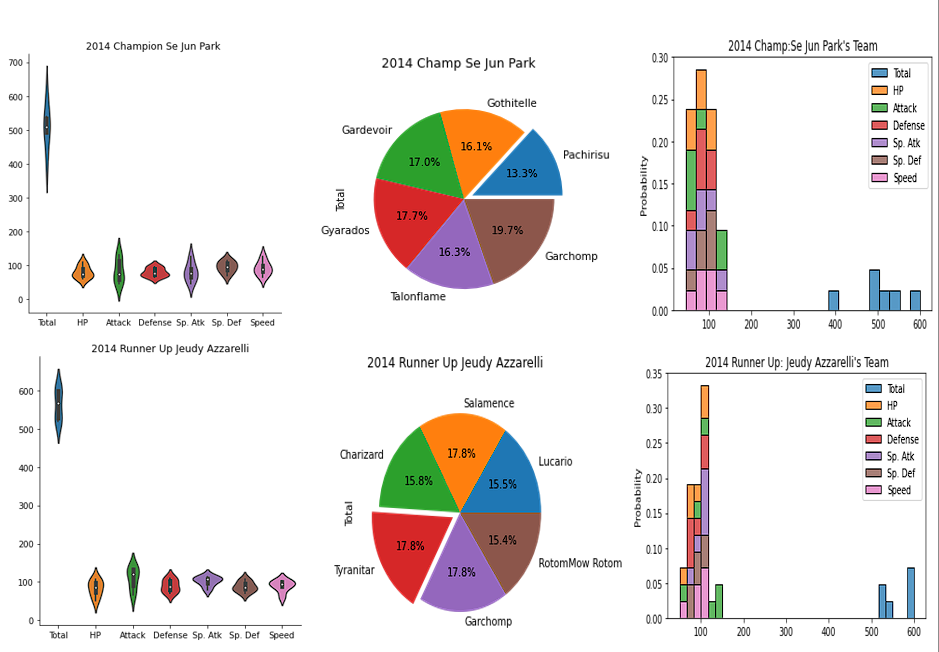


Figure - 2014 Championship Analysis

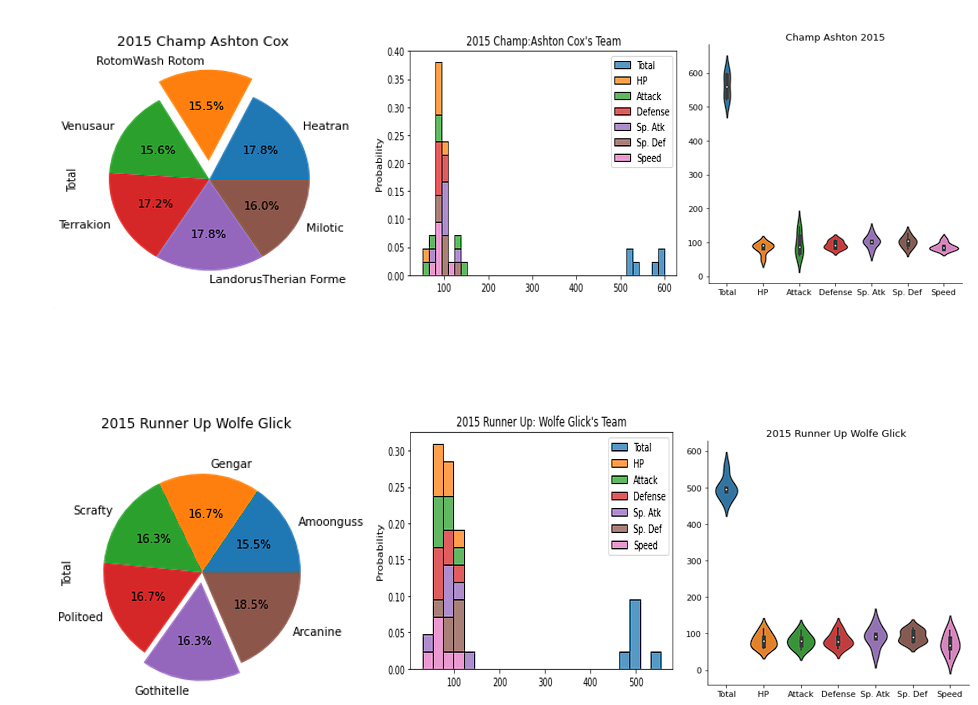


Figure - 2015 Championship Analysis