DSC 530 Final Project

Kevin Paulovici

<https://github.com/kevinpau/Bellevue_University_DSC_530/tree/master/Final_Project>

Pokémon Data Science

Bellevue University

2019

# Introduction

For this project my goal was to take a dataset of Pokémon information and answer the statistical/hypothetical questions summarized below. As part of this project, a simple battle simulation was created which ended up being closed tied to answering those questions. Through the analysis I discovered new insights that helped me understand some flaws in my initial questions and overall assessment of the dataset. Future work could include revisiting and adjusting the assumptions and improving the battle simulation to provide a more accurate assessment. For additional details, please visit the DSC530\_Paulovici\_Final\_Project.ipynb file in the provided GitHub link.

# Statistical/Hypothetical Question

1. Is the strongest lineup the perfect lineup?
2. Is one type of Pokémon clearly stronger than others?
3. What stat is most important when considering overall strength?
4. Where do the elite four Pokémon stand against other Pokémon?
5. Can you simulate a mock battle to determine outcome?
6. What limitations are there?

# Outcome of your EDA

Through the analysis explored in DSC530\_Paulovici\_Final\_Project.ipynb I was able to answer the above questions and will summarize the results here.

The assumption of strongest lineup (6 Pokémon) came from a variable called base\_total which is a summation of various other variables (e.g., hp, attack, defense, etc.). By picking out the six Pokémon with the highest values of base\_total I assumed this would be the perfect lineup. However, that is not true because the perfect lineup would be relative to another lineup. While my lineup is strong, that could also just be a coincidence and would not necessarily hold up against a lineup that I did not test. Additionally, the mechanics of the battle simulation would greatly influence this. The EDA process revealed that no one particular type was dominate over the others. This is good, considering it allows for a more even playing field when fighting various Pokémon. The assumption here, was that there could be an unfair advantage. Overall strength was assumed by base\_total. The EDA process did not reveal any one particular parameter that was more important. By comparing the elite four Pokémon to all generation 1 Pokémon, it was determined that they generally fall on the high end of the strongest Pokémon (based on mean to max values). A simulated battle was performed but the mechanics had some flaws that would impact the analysis performed. Finally, the limitations will be discussed in the sections below.

# What do you feel was missed during the analysis?

The overall connection between the battle simulation and overall strength was an oversight in my process. By adjusting these two items I could have ended up with different outcomes in the perfect lineup and overall results to check them.

# Were there any variables you felt could have helped in the analysis?

The dataset provides variables called against\_XXX, where XXX represents a type. As I mentioned earlier types are not equally impacted. For example, water types are weak against fire, so a water attack would be more deadly than what was accounted for. However, going this route would go beyond the scope of this analysis, so a compromise was made to exclude it. Again, these could drastically alter results determined.

# Were there any assumptions made you felt were incorrect?

In an effort to simplify the dataset and battle simulation I overlooked how nested overall strength and the battle simulation would work. Since my battle simulation only accounted for a few variables, I would have wanted to readjust strength based on those specific variables and exclude others.

# What challenges did you face, what did you not fully understand?

Overall challenges came from applying statistical methods and formulas to real data. It wasn’t always clear how to relate them or if they truly made sense.

# Submit a link to your repository to the assignment link during the final week of class.

All files associated with project can be found at: <https://github.com/kevinpau/Bellevue_University_DSC_530/tree/master/Final_Project>

DSC530\_Paulovici\_Final\_Project.ipynb

DSC530\_Paulovici\_Final\_Project.docx

DSC530\_Paulovici\_Final\_Project.pptx

Pokemon.csv

Additional libraries needed for DSC530\_Paulovici\_Final\_Project.ipynb