Marco Noriega

Mario Banuelos

MATH 120

19th December 2025

Final Project Final Written Draft

Introduction:

Final Project GitHub Link: GitHub Link

My Final Project will be based on analyzing Pokemon.csv from Alberto Barradas. This dataset consists of Pokémon from Generations 1-6. This project matters to me as it helps me better understand the techniques required to extract information that will help me answer my questions. This project incorporates the use of visualizations, merging, classes, functions, and data cleaning.

Questions aimed to answer from my project:

- What are the most common primary, secondary, and overall typings? How are the average total base stats changed among the different Pokémon generations?
- How much higher, on average, are the total base stats of legendary Pokémon in comparison to non-legendary Pokémon?
- Which Pokémon typings are correlated to be more offensive (higher attack/special attack/speed) and more defensive (higher defense/special defense/lower speed)?

Methods:

- Used the built-in function drop_duplicates, which removes identical rows from the dataset. This is my method for cleaning my dataset.
- Built a Pokémon class where each stat is an attribute to the class with a method function “find” which returns the data of the Pokémon.
- Used concat from the Pandas library, as concat joins two different columns in the same dataset into one.
- Used groupby and its aggregation functions, such as “mean,” to find average base stats when generating visualizations.
- Used functions that call the Pokémon class to retrieve the typings and base stats. These functions were used to generate visualizations that answer the fourth question.
- Used a function that categorizes a typing on one of the following playstyles: Offensive, Defensive, and Balanced. The evaluation metric is based on whether the Attack or Defensive stat is the greater stat. The average speed of the Pokémon typing is also assessed determine playstyle.
- Created pie, line, and horizontal bar charts to display results. Imported plotly.express so that users can see the exact stat if they hover their mouse at a specific bar or dot of the visualization.

Results (Based on the questions I aimed to answer):

- Water is the most common primary type and overall type, accounting for 14.6% of primary typings and 11% of overall typings. Flying is the most common secondary type, making up 24.9% of the secondary typings.

- There are slight decreases in total base stats from the first three generations.
- Generation 4 is the peak of total base stats and also has the biggest differences when compared to any other generation in the dataset.
- The total base stats of the recent generations (Generations 4, 5, and 6) are substantially better than the older generations (Generations 1, 2, and 3).
- Throughout all Generations, the average total base stat for legendary Pokémon is greater than 600.
- Legendary Pokémon in five of the six generations maintain a 200 total base stat lead in comparison to non-legendary Pokémon. Generation 5 is the only generation where legendary Pokémon do not have a 200 total base stat lead (199.48 stat lead).
- Globally, legendary Pokémon have an average total base stat of 620.217, while non-legendaries have an average of 404.161, making legendaries substantially better than non-legendaries.
- Dragon, Fire, and Dark are the three typings that can be mostly associated with being offensive, as they have the three highest average attack stats out of any Pokémon, and are three of the five fastest typings. All three of these typings are also categorized as offensive by the notebook.
- Steel and Rock are the two typings that can be mostly associated with being defensive, as they are the two with the highest average defense. In addition, Steel and Rock are two of the three slowest Pokémon typings, suggesting that they are designed to resist damage. Both of these typings are categorized as defensive by the notebook.

Discussion:

- The use of the newer topics introduced in Weeks 14 and 15, such as Pandas and Groupby, definitely speeds up the process of generating visualizations. This method was the sole method to generate visualizations to answer three of the four questions.
- The use of functions and Pokémon class made the fourth question more time-consuming to get a meaningful analysis of than the first three questions combined.
- This project doesn't take into consideration the Type advantage/disadvantage system. The type advantage/disadvantage system may influence the visualizations for the fourth question, as some typings may not have impressive stats, but many type advantages or resistances. For example, Ground is in the lower half of the bar chart when it comes to pure attack stat, but has five type advantages. This is the most amount of type advantages any typing has.
- If I had more time, I would've added a Type advantage/disadvantage dataset and merged the dataset with the Pokémon dataset to get a better analysis on the fourth problem.

Conclusion:

The project gives me a better understanding of using the Pandas library and groupby. It also helped me understand how I can incorporate classes to retrieve data and create visualizations. Originally, I thought I could solely rely on the Pandas library to generate all my visualizations, but the fourth question made me realize I needed to use functions. This is because I needed to create a new dataset such that Type 1 and Type 2 are merged as one column, while being assigned a stat that is also a merge of two different columns. I also realized that bar charts were not a good enough assessment to categorize the playstyles of Pokémon's typings, so I made two more functions to evaluate each base stat.