

Project: Milestone 1

Author: Jordan, Andrew

Date: 06/15/2022

1. Project Subject Area

The three data sources that I will be using all originate from the strategy role-playing game (RPG) Pokemon. Within this video game the player captures, trains, evolves, and battles with creatures known as Pokemon. Each Pokemon has a type, statistics-such as attack, defense, and speed-for battling, a list of attacks, special ability, and a pokedex number. Pokemon evolution refers to the permanent transformation of a Pokemon into a new pre-determined species with increased statistics and a different pokedex number. The pokedex is a fictional tracker in the Pokemon video game, in which each Pokemon is assigned a unique pokedex number.

1. Data Sources

The first dataset I will be using is a table with ten columns from the website "Pokemon Database" that includes the Pokemon name, type, individual statistics, total combined statistics, and pokedex number. [Pokemon Database](#)

The second dataset I will be using is the api PokeApi. This api includes some of the same characteristics as the Pokemon Database table, but also includes many more details such as each Pokemon's learnable attacks, type advantage, special abilities, location in game, color, height, weight, and carried items. [PokeApi](#)

The final dataset I will be using is a csv file containing over 50,000 rows of battle simulation results. The first two columns contain the pokedex number of each pokemon in the simulated battle, and the third row contains the winner of said simulation. [Pokemon Battle Simulation Results](#)

1. Dataset Relationships

Within the Pokemon video game the tracker, called a pokedex, assigns each Pokemon a unique value known as their pokedex number. This number is consistent across all of the Pokemon universe, and creates a relationship between all three of my datasets via this built-in key. Using the pokedex number as my key ID, the api has a one to many relationship with both the csv and table data, while the csv and table data have a many to many relationship.

1. Project Proposal Plan

Using these three datasets I believe I can create a combined dataset that provides a one-stop source of data for Pokemon battling information and strategy formation. To do so I will need to first prepare my data, a step that will include homogenization of column names, separating types into two columns where necessary, identifying which Pokemon are legendary (a special category of exceptionally strong Pokemon), and coloring the data frame by type for ease of navigation. To reliably utilize the simulated battle data I must next remove any outliers from the csv dataset, which in this scenario will include battles between Pokemon of different evolutionary tiers.

Using visualizations covering, among others, type distribution, stats by type, and battle winners will supply a basis upon which the most effective Pokemon for battling can be determined. Of particular note, I believe that a radar chart of each type's average stat distribution per evolutionary tier can lay out the most effective Pokemon to use at each stage of the video game.

If successful there are indeed ethical implications of my research, most notably in the professional Pokemon player circuit. Professional Pokemon players battle one another with teams of Pokemon in tournaments where the tournament champion receives a monetary prize. My goal of creating a one-stop source of data that objectively identifies the optimal combination of Pokemon for a battling team could create a scenario in which professional Pokemon players' income is negatively impacted. As the public will now have the ability to quickly and easily identify the optimal Pokemon battling team, professionals will potentially be met with an influx of new players entering tournaments and face steeper odds of winning a monetary prize.