Jodi Yip

April 22, 2019

NBA Players vs. Pokemon Comparison

Motivation:

My project compares Pokemon to NBA players. I always enjoyed watching basketball especially the NBA. I also play Pokemon Go and study the strengths and weaknesses of Pokemons. I wanted to combine my favorite interests from the real world and compare it to my interest in the virtual reality world. I wanted to analyze which Pokemon corresponds to which NBA player based on similar variables (offense and defense) and find whether they carry similar characteristics. I also wanted to see which group is stronger – which group would win in a fight if their worlds were joined together.

Data Sources:

The NBA dataset was found on the Fox Sports website. The dataset is titled '2018-2019 NBA Stats and Leaders'. The file is separated into 4 different links with 50 entries in the first 3 links and 33 entries in the last link. The first NBA dataset can be found here. The links to the next 3 files can be found on the bottom of the link.

The variables I used from the website was Advanced, Offense Rating (OFF RTG), Defense Rating (DEF RTG) and Minutes Per Game (MPG). These statistics are based on the regular season.

I found the Pokemon dataset on Kaggle. The file is titled 'Pokemon With Stats' by Alberto Barradas. The CSV file titled Pokemon.csv can be found here. The file size is 43 KB. It is

800x13. The data comes from the Pokemon games, not including Pokemon cards or Pokemon Go.

The variables I used in my project included name of Name, Attack and Defense. Name is the name of the pokemon. Attack is the base modifier for normal attacks. Defense is the base damage resistance against normal attacks. The data from the CSV file comes from pokemon.com, pokemondb and bulbapedia.

Data Manipulation Methods:

The specific needs were to find complementary variables in both datasets. The end goal was to compare Pokemon and NBA players based on 50% offense and 50% defense. Because the Pokemon and NBA dataset were based on different scales, I normalize the scores so I can compare using the same baseline variables.

NBA Data

Read HTML files and join into single dataframe: Because the entire dataset was found on 4 separate links, I read each url using pd.read_html() and saved it into a dataframe. I used concat to merge all 4 dataframes into a dataframe titled 'basketball'.

Sort Offense and Defense column to normalize scores: I used the sort function twice to find the max offense and defense rating. I found the normalized scores by dividing by the maxes. I multiplied each value by 0.5 so that offense and defense ratings are worth 50% each. I then added the two scores together. A perfect offense and defense NBA player would have a score of 1.

Sort MPG column to find normalize scores (error found and fixed in this step): When the results displayed, I realized the best NBA players did not display on top. The players with the top normalized scores were benchwarmers and unheard of. I looked through the data and noticed some NBA players had the possibility of a high offense or defense rating even if they don't get much playing time. I had to factor in the MPG (minutes per game). Using the same concept as offense and defense, I normalized the MPG by dividing all MPG by the max minutes per game.

Normalized NBA score = (normalized offense score + normalized defense score) * normalized minutes per game

Pokemon Data

Read CSV file into dataframe: I read in the file using pd.read_csv() and added into a dataframe titled pokemon_csv.

Filter out entries with keywords: There was no missing data but an excess of data in the pokemon dataset. I noticed some Pokemon were repeated because the dataset included the "Mega" or "Primal" forms of the pokemon. Mega and primal are stronger versions of the same pokemon. I decided to filter out those data entries because I do not want pokemon repeats. I converted the Pokemon name column to a string and filtered out words using the contain function.

Sort Attack and Defense column to normalize score: I used the same technique as the NBA dataset to find the normalized Attack and Defense scores.

Normalized Pokemon Score = normalized attack score + normalized defense score

Joining NBA and Pokemon dataset: Because I used the reset_index function on both the Pokemon and NBA datasets, I was able to concat them into a large dataframe titled data_joined.

Since there are 721 data entries for Pokemon and only 183 players in the NBA, many weaker Pokemon could not be matched with a NBA player. The sample size of the NBA dataset is a lot smaller because there are less players in the NBA. I could not create or find new data to add to the NBA dataset because there are a specific number of players in the NBA.

The biggest challenge was analyzing the data. It was difficult to compare NBA players and Pokemon because they come from very different backgrounds. Pokemons exist in a imaginary world while NBA players are real life. The Pokemon attack/offense is used to defeat Pokemon while the defense is used to prevent incoming attacks. The NBA player offense is used to score baskets while the defense is for preventing the other team from scoring. Even though the two variables are the same, they serve different purposes.

Analysis and Visualization:
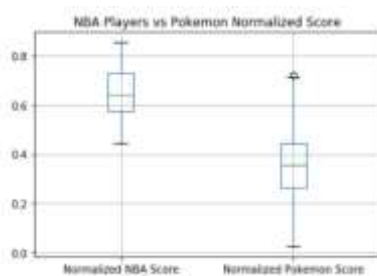
The normalized score is found based on 50% offense and 50% defense. The results of my data manipulation showed that normalized NBA score has a higher max and min and a smaller range compared to the Pokemon normalized score. In a hypothetical situation where the NBA and Pokemon existed in the same world, the average NBA player would be stronger offensively and

defensively than the average Pokemon. The Pokemon and NBA normalize score range is significantly different because in order to become a NBA player, you have to excel at the sport. Therefore, the range is small. However, pokemons are the general population within the virtual world. If we compared legendary Pokemons to NBA players, the average normalized score would be more similar.

	ADVANCED	Normalized NBA Score	Normalized Pokemon Score	Name
0	Beal, Bradley Beal, X. WAS	0.854561	0.713014	Dreodon
1	58 Harden, James Harden, J. HOU	0.848242	0.712260	Regirock
2	Lillard, Damian Lillard, D. POR	0.842331	0.688860	Aggron
3	13 George, Paul George, F. OKC	0.826604	0.662375	Regigigas
4	46 Capela, Clint Capela, C. HOU	0.822327	0.671498	Rhyperior
5	Holiday, Jrue Holiday, J. NOP	0.814108	0.670884	Steelix
6	Durant, Kevin Durant, K. GSW	0.811307	0.661898	Slaking
7	Booker, Devin Booker, D. PHX	0.809881	0.657828	Metagross
8	Walker, Kemba Walker, K. CHA	0.800478	0.655183	Cloyster
9	47 James, LeBron James, L. LAL	0.794811	0.621081	Wigybe
10	Harris, Tobias Harris, T. PHI	0.794267	0.619562	Golem

The data says James Harden is most similar to Regirock. I believe this statistic is incorrect because James Harden has a low defense rating while Regirock has a high defense rating. However, they are paired because they both have high overall score. The comparison I found the most interesting and accurate is Paul George and Regigigas. They are both lethargic on defense but quick and strong on offense. In addition, Regigigas is a combination of the rock, steel and ice Regis and contains moves from each Regis. Paul George is a jack of all trades. He is a decent ball-handler, great mid-ranged and three-point shooter and solid athleticism.

My mistake in the project is limiting the comparing variables to only Offense and Defense. Other characteristics such as speed and size should also be considered for a better comparison between Pokemon and NBA player. However, I could not find a NBA dataset that included player speed, athleticism, etc.