Final Project: Pokemon Data

John Ozuna

Introduction:

The dataset used in this project is called "Pokemon.csv" which contains information on 800 different Pokemon across many versions of the franchise. The data was collected by a group of data analysts who compiled the information from various sources including official Pokemon websites and licensed Pokemon games. Each row of the dataset represents a different Pokemon, while each column represents a different feature/value of each Pokemon.

Dataset features:

Column	Feature	Description of Feature
1	*Name*	The name of the Pokemon
2	*Type 1*	The primary type of the Pokemon
3	*Type 2*	The secondary type of the Pokemon (if applicable)
4	*Total*	The total value of stats of the Pokemon (HP + Attack + ...)
5	*HP*	The hit points value (health points) of the Pokemon
6	*Attack*	The attack strength value of the Pokemon
7	*Defence*	The defense strength value of the Pokemon
8	*Sp. Attack*	The special attack strength value of the Pokemon
9	*Sp. Defense*	The special defense strength value of the Pokemon
10	*Speed*	The speed value of the Pokemon
11	*Generation*	The generation the Pokemon belongs to
12	*Legendary*	Whether the Pokemon is a legendary Pokemon or not

Possible values for the features vary depending on the feature.

- The *“Name”* of the Pokemon is a qualitative feature and is unique to each Pokemon.
- The *"Type 1"* and *“Type 2”* features can have 18 possible qualitative values, representing the different types of Pokemon such as grass, fire, water, etc.
- The *"Legendary"* feature is a qualitative boolean value that is either True or False.
- The *"Total"* feature is the sum of all of the Pokemon's other stats (quantitative)
- The *“Generation”* feature is a qualitative value of when the Pokemon was introduced
- The rest of the features are quantitative (number) values of each Pokemon.

Final Project: Pokemon Data

John Ozuna

Work Description:

The first step in the analysis was to clean the data. The dataset contained some missing values, so those rows were dropped using the "dropna" function in pandas. After cleaning the data, the length of the dataset was 799.

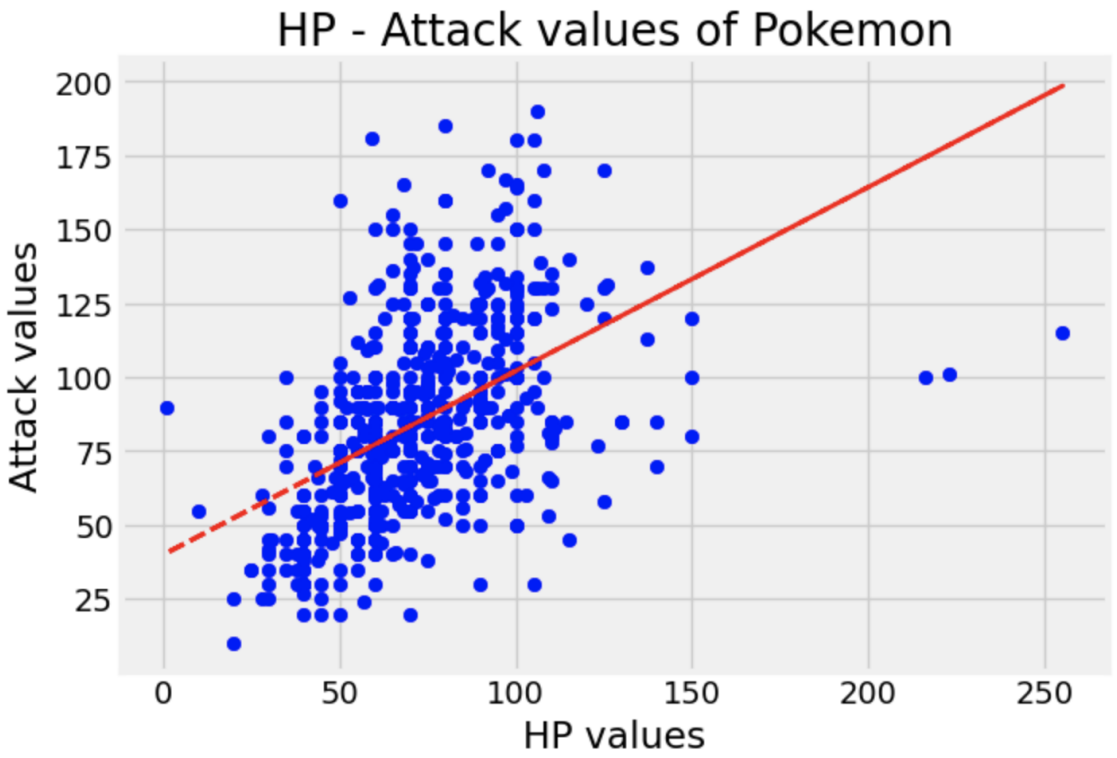
```
HP Stats:
Mean: 72.67
Median: 70.00
Variance: 664.85
Standard deviation: 25.78

Attack Stats:
Mean: 85.05
Median: 84.00
Variance: 1088.14
Standard deviation: 32.99
```

Next, the HP and Attack features were selected, and their statistics were calculated using numpy. The mean, median, variance, and standard deviation were calculated for both features. As determined, the HP values have a smaller mean and median, as well as less standard deviation and variance than attack stats.

Then, 4 graphs were constructed to have a better understanding of the dataset.

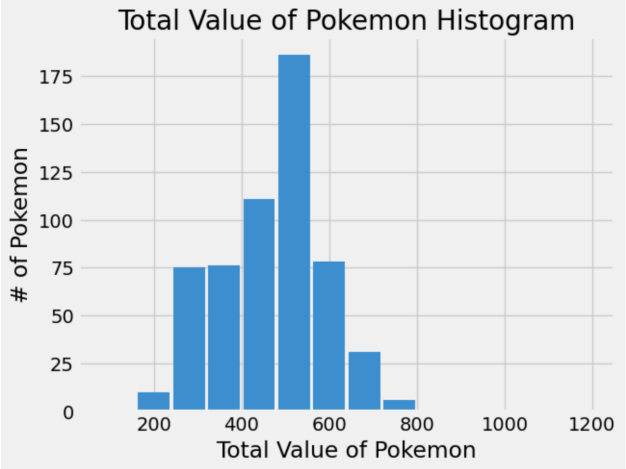
Firstly, a scatter plot with a line-of-best-fit was created for HP and Attack values. As seen below, the graph demonstrates a positive correlation between a Pokemon's HP and its Attack, meaning that as the HP value increased, the Attack value generally also increased. The graph also shows 3 outlier points with abnormally high HP values. This did not greatly affect the stats.

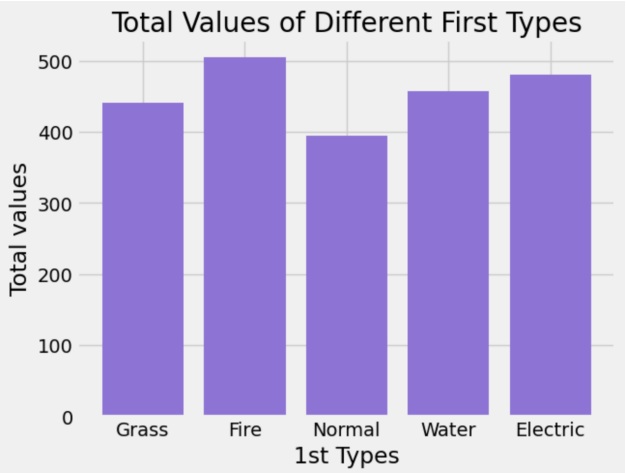


Final Project: Pokemon Data

John Ozuna

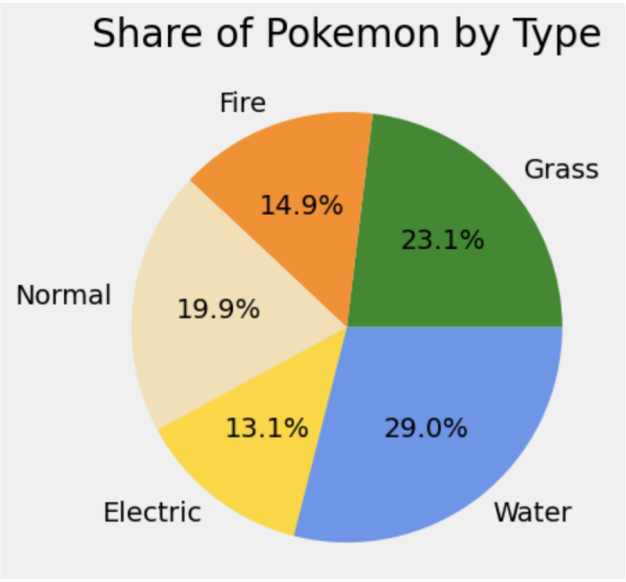
Secondly, the Total feature was analyzed by creating a histogram. The histogram shows that most Pokemon have a total value around 300 to 700, with a small outlier of around 1100.





Thirdly, a bar graph was constructed to compare the mean Total values of different Pokemon by Type 1. The dataset was grouped by Type 1 using pandas' "groupby" function, and the mean Total values for each type were calculated and compared between the different types. The picks 5 popular Pokemon types and shows that Fire types, on average, have greater Total values than Grass, Normal, Water, or Electric types. Electrip type has the second highest Total value, then comes water, grass, and normal in last place. This can be seen above.

Finally, a pie chart was constructed to show the share of 5 popular Pokemon by Type. The chart shows that out of the 5 popular types (Grass, Fire, Water, Normal, and Electric), Water types have a plurality of number of Pokemon. Of note, although fire and electric types have the two highest total values of the 5 as seen in chart 3, they also have the lowest shares of the 5 types.



Overall, the analysis of the Pokemon dataset showed some interesting insights into the correlations and differences between various Pokemon types and features.