**Exploratory Analysis**

Eli Sartin, [esartin@bellarmine.edu](mailto:esartin@bellarmine.edu)

Adham, [ahowaidy@bellarmine.edu](mailto:ahowaidy@bellarmine.edu)

1. **INTRODUCTION**

* This data set contains Pokédex information on the Pokémon from the first 7 generations of games. The data set can be found at the following address: (<https://www.kaggle.com/rounakbanik/pokemon>). The reason that we chose this data set is that it contained data about something that we both have an interest in. Our familiarity with the variables in the data set allowed us to be able to come to conclusions eagerly, as it piqued our curiosity.

1. **DATA SET DESCRIPTION**

* This data set contains 801 samples with 22 columns with various data types. The variables are shown in full in Table 1. Any missing data is simply data that doesn’t exist, either that Pokémon only has a single type, or gendered versions of it does not exist.

**Table 1: Data Types**

|  |  |
| --- | --- |
| Variable name | Data type |
| pokedex \_number | int64/Nominal |
| name | object/Nominal |
| japanese\_name | object/Nominal |
| classification | object/Nominal |
| type1 | object/Nominal |
| type2 | object/Nominal |
| abilities | object/Nominal |
| base\_total | int64/Ratio |
| hp | int64/Ratio |
| attack | int64/Ratio |
| defense | int64/Ratio |
| sp\_attack | int64/Ratio |
| sp\_defense | int64/Ratio |
| speed | int64/Ratio |
| height\_m | float64/Ordinal |
| weight\_kg | float64/Ordinal |
| base\_egg\_steps | int64/Interval |
| base\_happiness | int64/Ratio |
| capture\_rate | object/Ratio |
| percentage\_male | float64/Nominal |
| generation | int64/Nominal |
| is\_legendary | int64/Nominal |

1. **Data Set Summary Statistics**

* This section gives a brief statistical summary of the variables in the data set. A couple of variables are unique to each sample and provided no benefit in providing a percentage breakdown.

**Table 2: Summary Statistics**

**Table 3: Proportions**

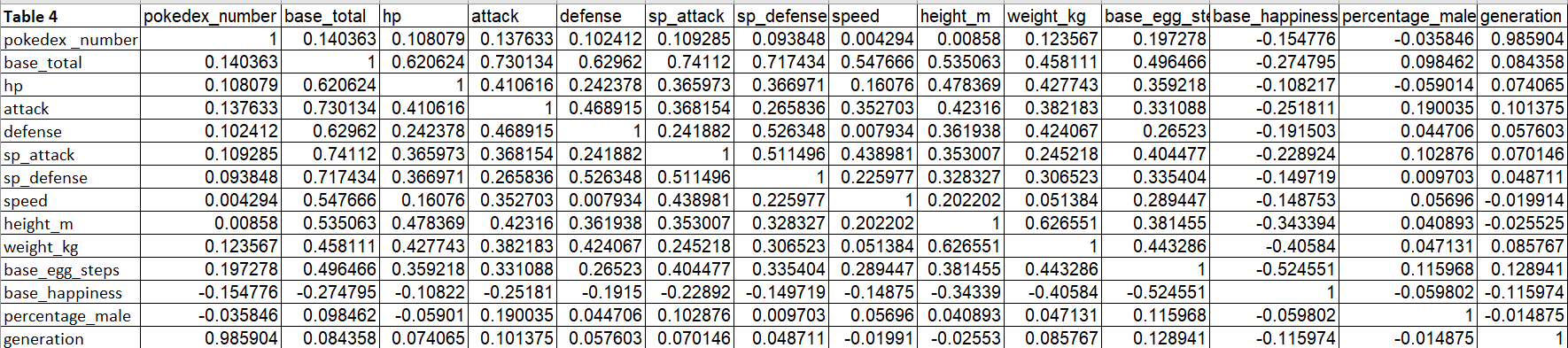
* **Type 1:**

|  |  |  |
| --- | --- | --- |
| Catergory | frequency | proportion |
| bug | 72 | 8.989% |
| dark | 29 | 3.620% |
| dragon | 27 | 3.371% |
| electric | 39 | 4.869% |
| fairy | 18 | 2.247% |
| fighting | 28 | 3.496% |
| fire | 52 | 6.492% |
| flying | 3 | 0.375% |
| ghost | 27 | 3.371% |
| grass | 78 | 9.738% |
| ground | 32 | 3.995% |
| ice | 23 | 2.871% |
| normal | 105 | 13.109% |
| poison | 32 | 3.995% |
| psychic | 53 | 6.617% |
| rock | 45 | 5.618% |
| steel | 24 | 2.996% |
| water | 114 | 14.232% |

**Type 2:**

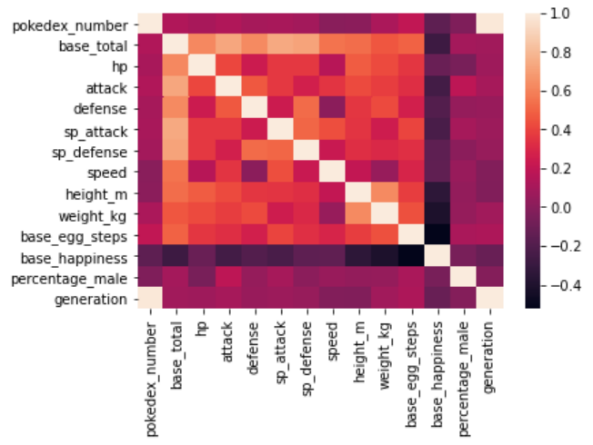
|  |  |  |
| --- | --- | --- |
| Catergory | frequency | proportion |
| bug | 5 | 0.624% |
| dark | 21 | 2.622% |
| dragon | 17 | 2.122% |
| electric | 9 | 1.124% |
| fairy | 29 | 3.620% |
| fighting | 25 | 3.121% |
| fire | 13 | 1.623% |
| flying | 95 | 11.860% |
| ghost | 14 | 1.748% |
| grass | 20 | 2.497% |
| ground | 34 | 4.245% |
| ice | 15 | 1.873% |
|  |  |  |
| normal | 4 | 0.499% |
| poison | 34 | 4.245% |
| psychic | 29 | 3.620% |
| rock | 14 | 1.748% |
| steel | 22 | 2.747% |
| water | 17 | 2.122% |

**Abilities and Classifications are in the excel file attached since they are too large to fit in this document.**

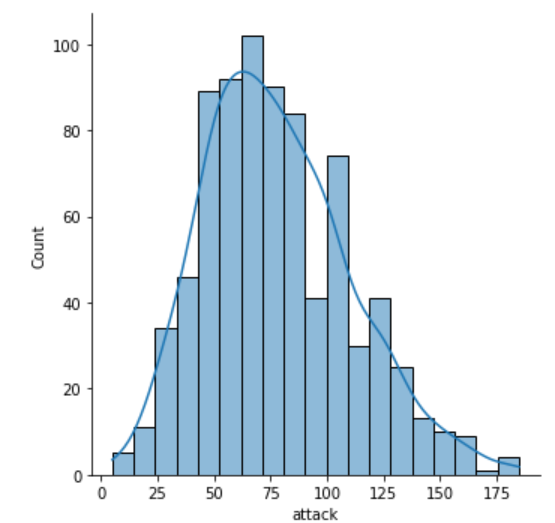
**Table 4: Correlation/Table**

1. **Data Set Graphical Exploration**

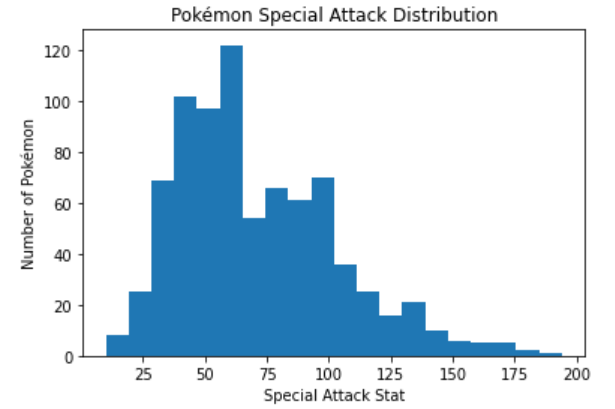
* Our data set is a population and in turn, gives us some interesting insights into the way that all of the Pokémon compare to each other. We used various types of graphs to gain insight into how the Pokémon stats are distributed. Figure 2 showed us that the attack was unimodal with a slight right skew. The concentration right above 175 vs. below it was very interesting to me. More Pokémon exist at the far end of the attack spectrum than right before it. Figure 3 was interesting to observe because the special attack stat has a more extreme max, but is distributed lower. Figure 4 is very interesting because height and weight have a relatively positive correlation, but the heaviest Pokémon are not very tall. Figure 5 is interesting because, despite a slight positive correlation between the two variables, the weight seems to not have a huge impact on the speed of a Pokémon. Figure 6 shows a strong correlation between HP and Weight with a couple of outliers with both high weights and low HP, as well as low weight and high HP. Figure 7 shows a distinct difference in the power between Legendary and non-Legendary Pokémon, as expected. It is interesting that Legendries do not excel above standard Pokémon, particularly in one stat. Figure 8 shows us that primary typing does not have a big effect on any particular stat. Figure 9 shows us that the secondary typing of Pokémon does not have much sway on any particular offensive or defensive stat. Figures 10 and 11 show the staggering difference between primary and secondary typing. The most common secondary types vary greatly from the most common primary types with the exception of psychic types appearing relatively high in both.
* Figure 1: Heatmap



* Figure 2: attack\_kde



* Figure 3: sp\_attack\_hist



* Figure 7: pairplot\_legendary

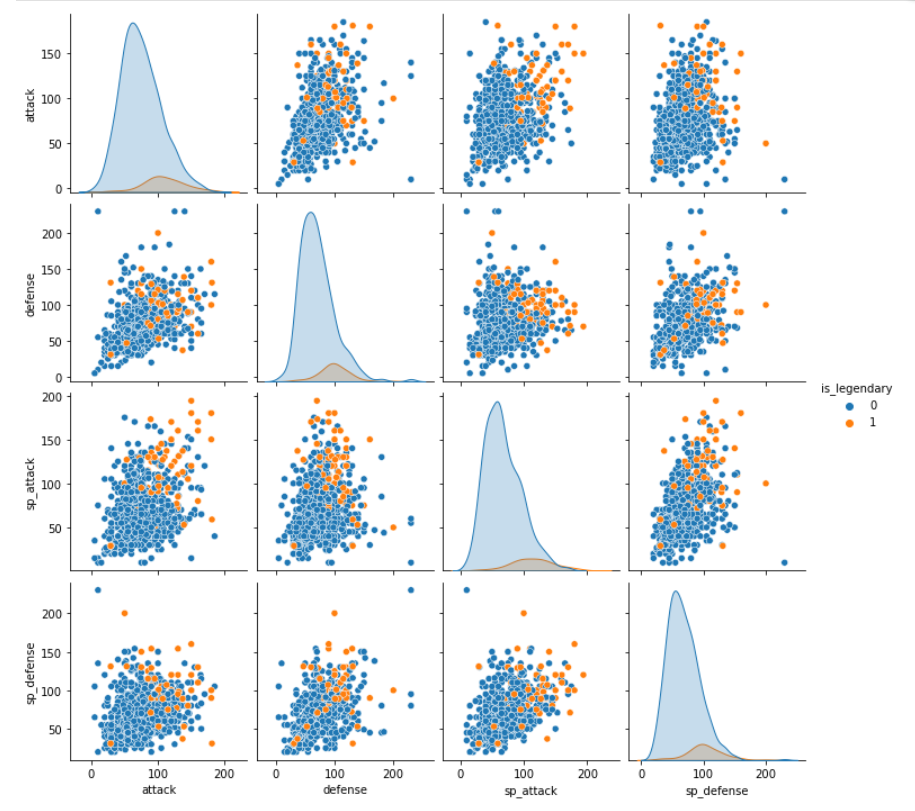
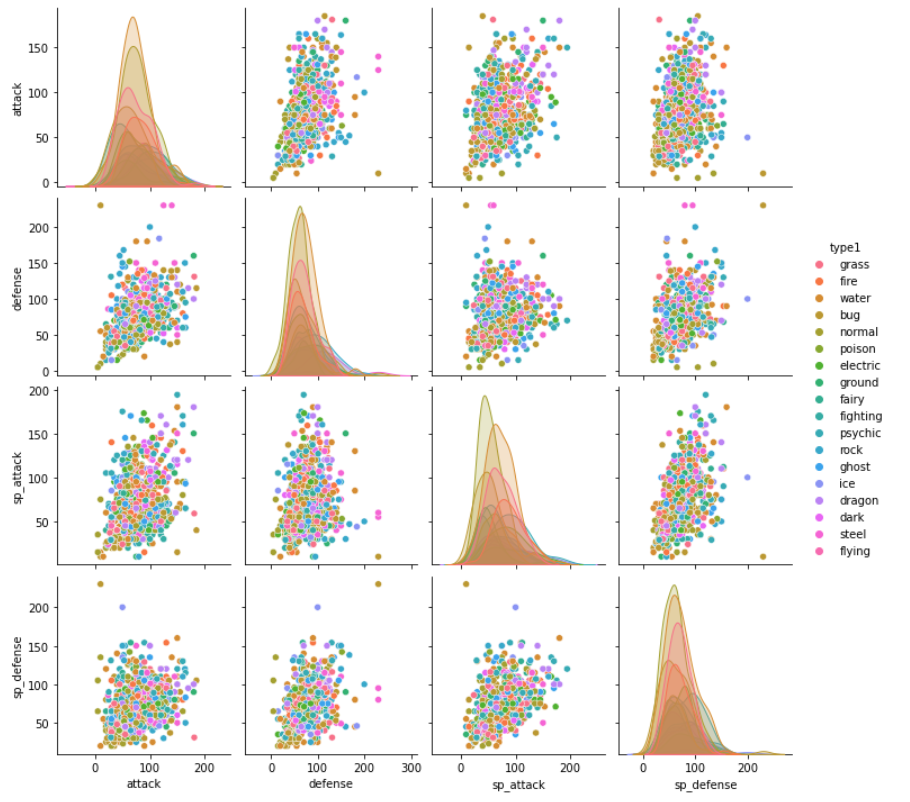
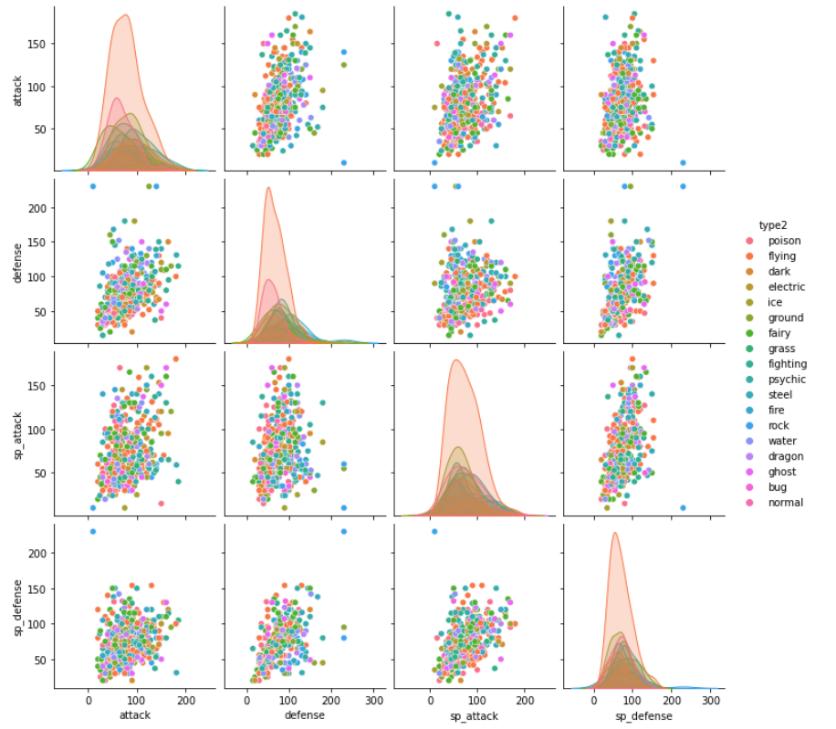


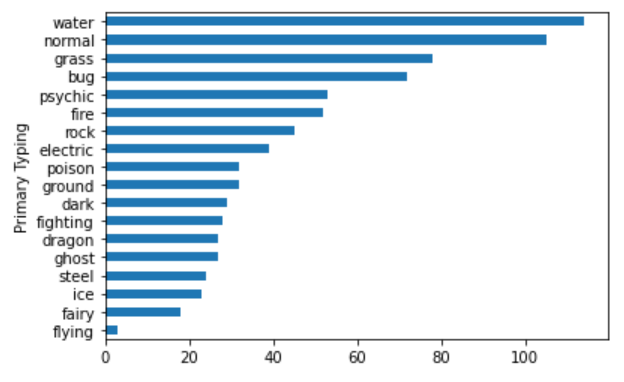
Figure 8: pairplot\_type1



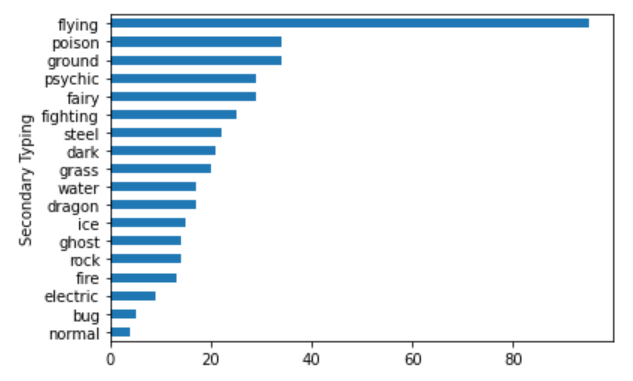
* Figure 9: pairplot\_type2



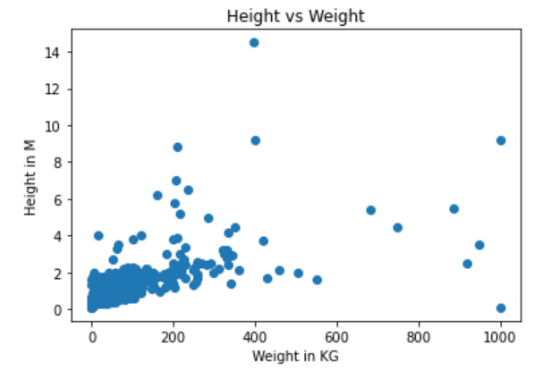
* Figure 10: type1\_bar



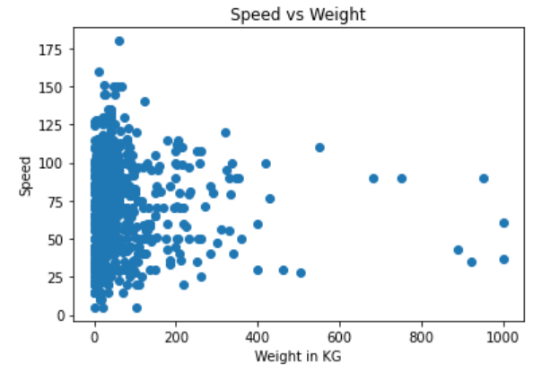
* Figure 11: type2\_bar



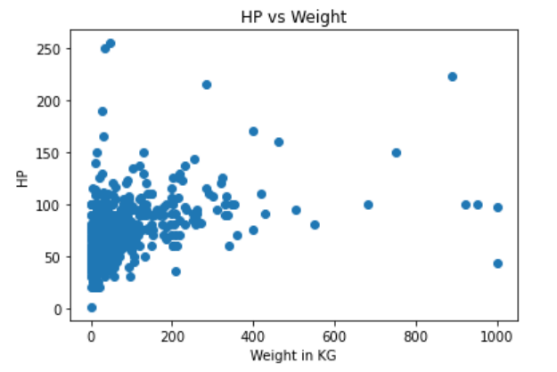
* Figure 4: height\_vs\_weight



* Figure 5: speed\_vs\_weight



* Figure 6: hp\_vs\_weight



1. **SUMMARY OF FINDINGS**

Throughout our Exploratory Data Analysis of this data set, we observed some interesting trends in Pokémon stats and typing. In general, Pokémon have a rather even distribution of stats, with no stat being heavily swayed by typing of the Pokémon. Legendries are stronger than the average Pokémon across the board without particularly excelling in one stat across the board. The weight of Pokémon plays a decently large role in the agility and size of Pokémon with a few outliers. Primary and secondary typing of Pokémon varies greatly with not much overlap in the distribution of the two. The overall distribution of stats between Pokémon appears to be normal, resulting in a balanced competition field.