Pokemon Type and Stats

By Dani DiPinto

DS-SEA- 06

**1. Introduction and Background**

Nintendo released the games Pokemon Red and Pokemon Green for Gameboy Japan in 1996. Soon after, They were released in America as Pokemon Red and Pokemon Blue, with a Special Edition of Pokemon Yellow soon after. In this first generation of Pokemon Games, to fill your pokedex, you had to collect all 150 Pokemon (or 151 including Mew). Each of these Pokemon had a type and specific base stat that made it strong or weak against other Pokemon.

With each new game release, or new Generation of Pokemon, new pokemon and pokemon types are added to the list of existing ones. With the latest generation-- Generation VII, there are 802 Pokemon and 18 types. Currently, since Pokemon can have a primary and secondary type, there are 146 combinations.

Predicting the primary and secondary type is a bit out of scope for this project, so for the purpose of this project, we will be looking at Primary type only.

<http://bulbapedia.bulbagarden.net/wiki/Main_Page>

**2. Problem Statement and Hypothesis:**

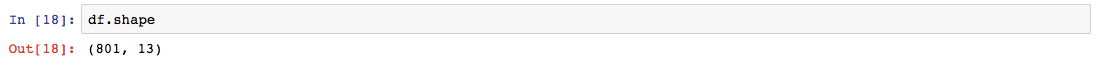
By comparing the pokemon base stats (including HP, attack, special attack, defense, special defense, and speed) we can predict the pokemon’s type.

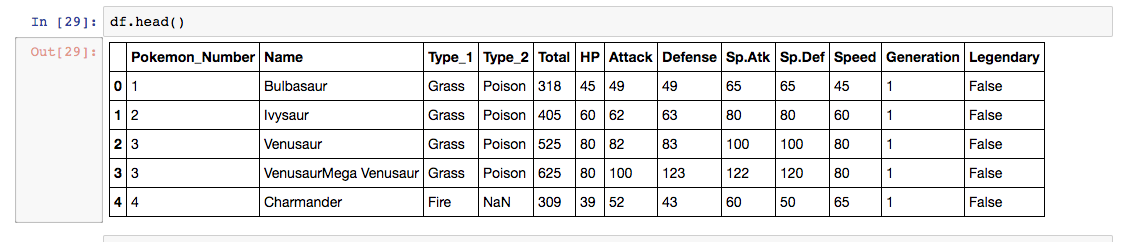
As I mentioned before, I will only be looking to find the Primary type or (Type\_1 in this dataset). It will be interesting to see how predicting (or not) the base stats of different Pokemon are.

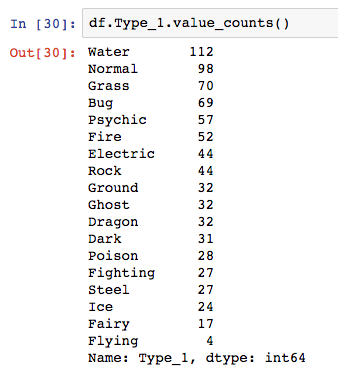
**3. The Data**

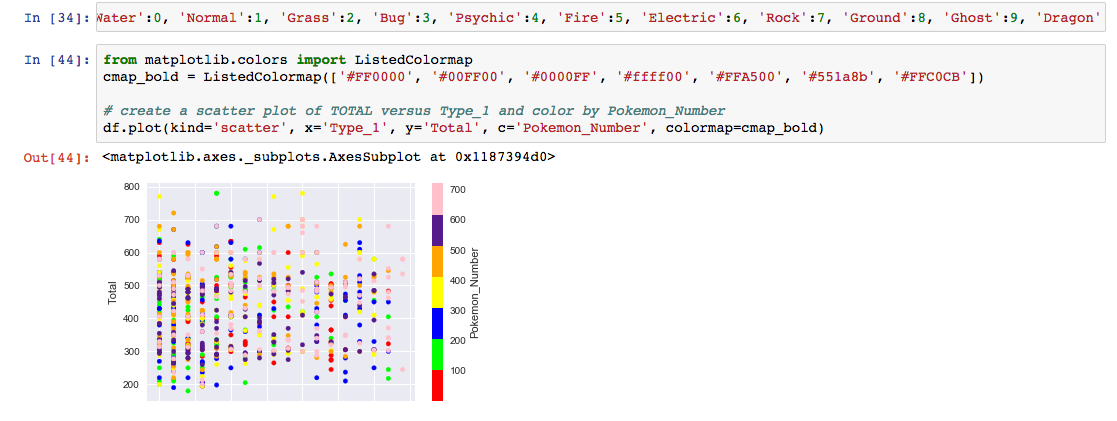
This data was found on Kaggle: <https://www.kaggle.com/abcsds/pokemon> and it is mentioned that the Kaggle user created the dataset from using a Pokemon website called Bulbapedia: <http://bulbapedia.bulbagarden.net/wiki/Main_Page>.

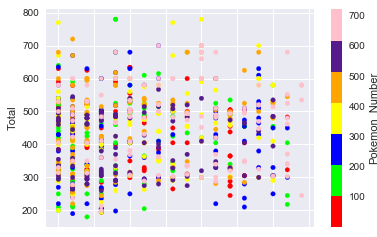
The dataset is very clean as is. It has 801 rows and 13 columns.











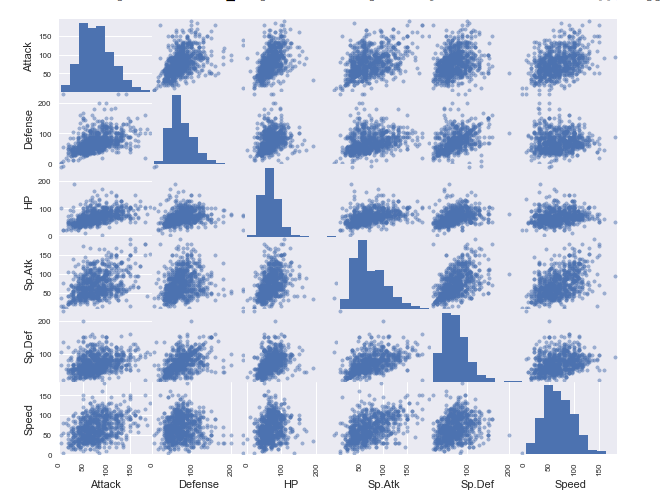
Exploring the Data some more-

Below I am going to graph the stats by pokemon number. The x axis will be pokemon number, and the y axis will be the stat (attack, defense, HP, speed, special defense, special attack).

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

Scatter Matrix to look at all features:





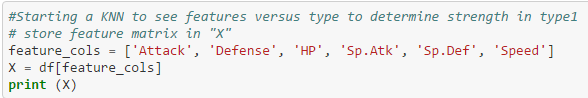
Basically, there’s a lot of variance among the different features. In the Kaggle forums, users said it wasn’t possible to determine the type accurately based on only one or two features. I’m going to try to use 6 in the first model. I’m going to start with KNN, and move on to more complex models afterwards.

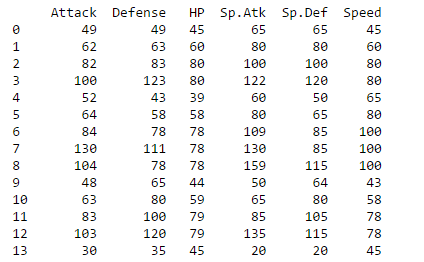
**4. Creating the Model**

Step 1.

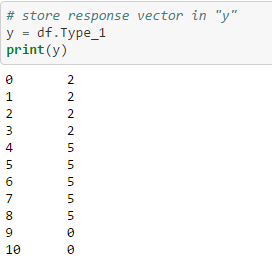
Define X and y.

First, we’re going to define the features and check X





Now that we have our features defined, we’re going to define the response vector, y:



Type\_1 is our y value

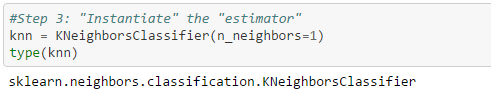
Step 2.

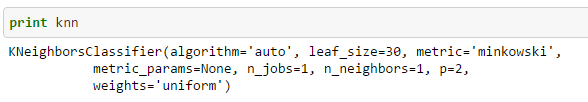
My next step is to import KNN Classifier from Sklearn. This is the “Importing the Estimator” portion.



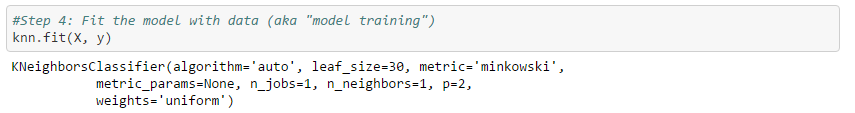
Step 3.

The next step is to instantiate the estimator

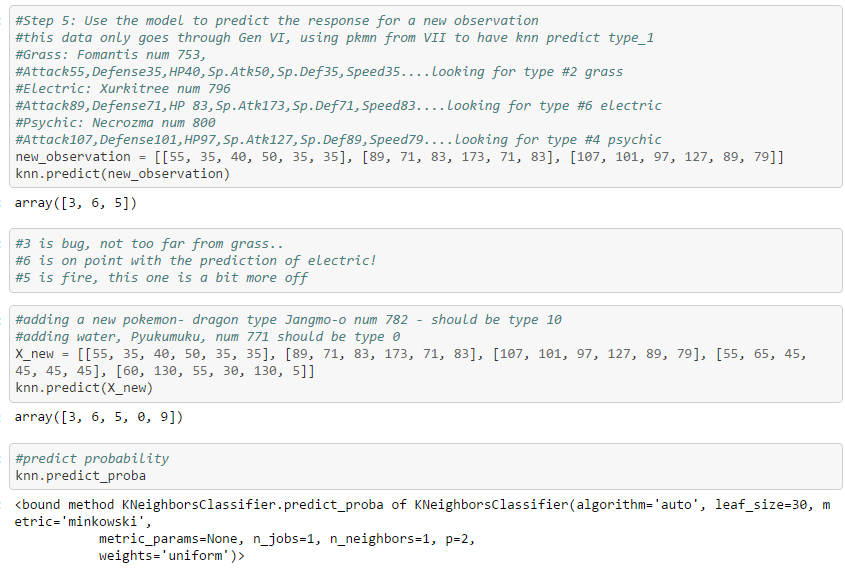




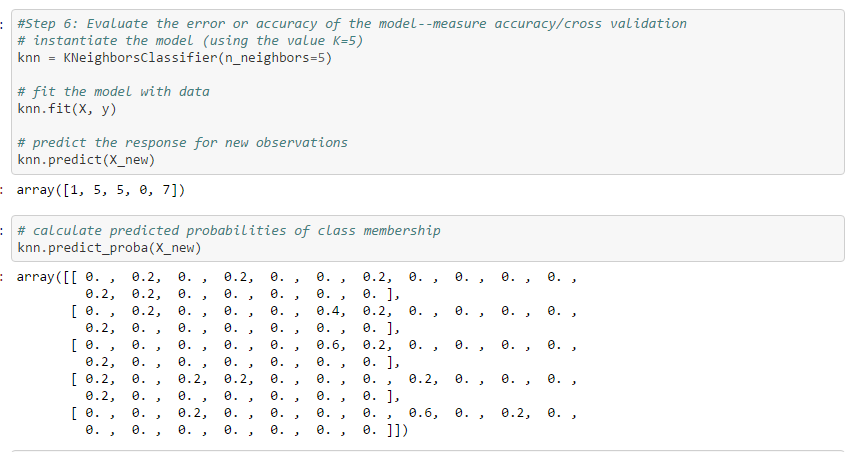
Step 4. Fit the model



Step 5. Predict the response for a new observation



Step 6. Evaluate



**5. Conclusion**

The KNN was just a start. I can predict *some* types based on the stats, but not very many. I’m planning on using a few other models to work on increasing the accuracy. I plan on using decision trees to check on most important features, clustering, and random forests.

* Conclusions and key learnings
* Your challenges and successes
* Possible extensions or business applications of your project