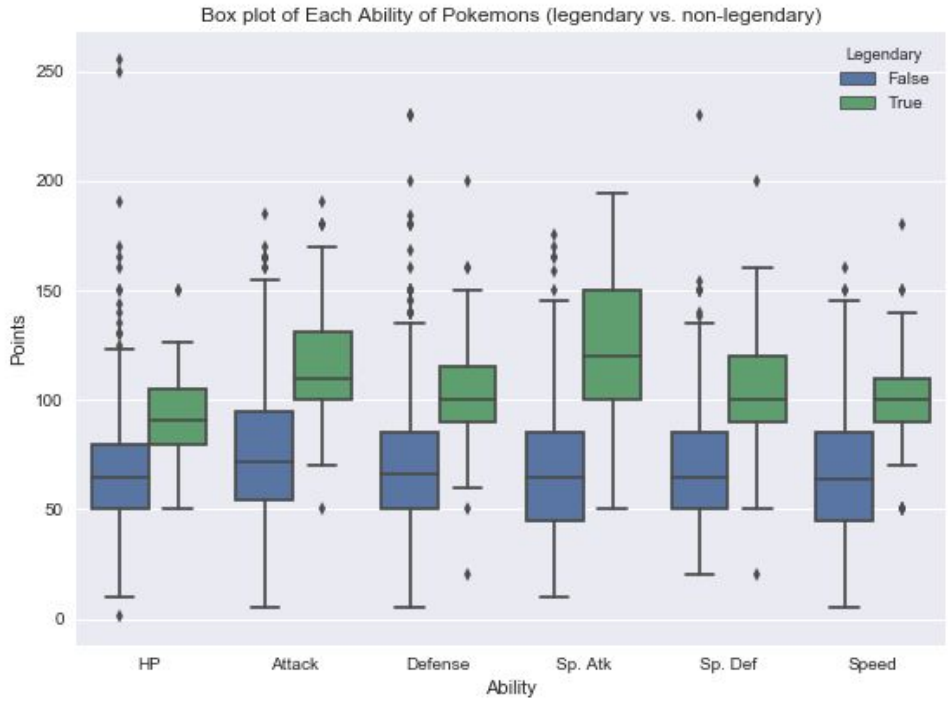
Classify Pokemon Legendary Status

CITINA LIANG, YAN KANG, ZHENG WANG

Our Dataset

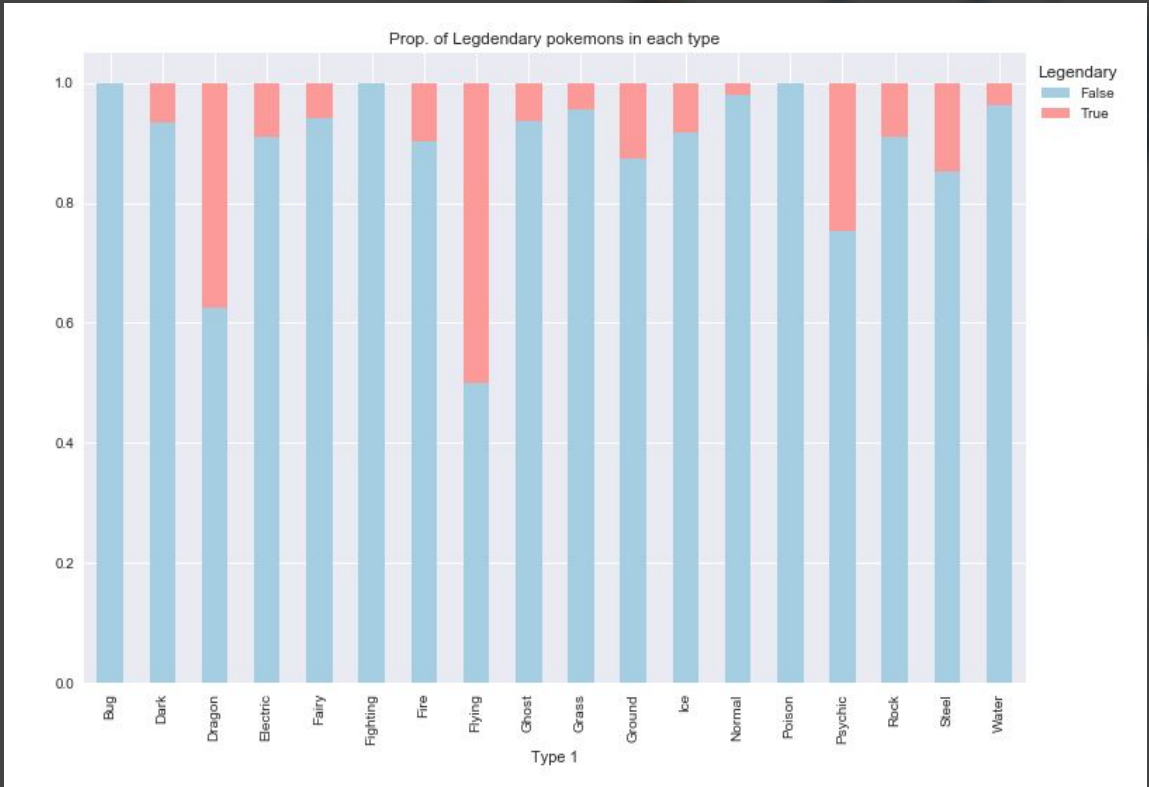
- The dataset is from the kaggle competition which consists All strength stats for each pokemon that appears from generation 1 to generation 6.
- The dataset includes 800 observations and 12 variables.

Exploratory Data Analysis



- All abilities of legendary Pokemons are significantly higher than non-legendary Pokemons
- The difference of special attack of legendary and non-legendary Pokemon is the largest

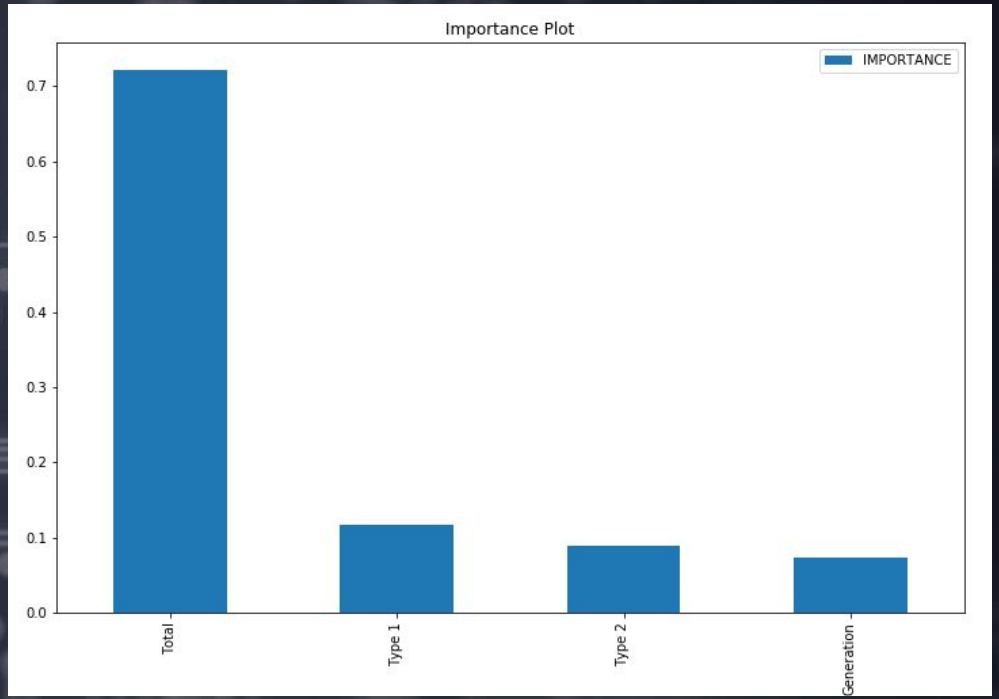
Exploratory Data Analysis



- Flying pokemons have largest proportion of Legendary Pokemon
- There is no Legendary pokemons of type Bug, type Fighting, and type Poison
- There is large variation in the proportions of Legendary pokemons for each type

Modeling & Methodology

- Split 70% of the data as training data. Save the rest of 30% data as testing data.
- Perform 10-fold cross validation to absolve overfitting.
- Use f1 score instead of raw accuracy.
- Perform the TPOT for optimize our existing model.



* The Sum of all abilities scores is the most important variable.

Result & Summary

Result:

- Accuracy of classifying Legendary status: 0.95
- 10-fold Cross Validation Score: 0.95165
- F1 score before TPOT: 0.86604
- F1 score after TPOT: 0.95169

Limitation:

- Unbalanced split between legendary and non-legendary classes
- Small dataset
- More hidden variables might need to be included

Thank You!