Gotta Plot 'Em All | Pokemon

Brandon Reid | Team Lead | Architect - Lead in the development, manage dev-op

Shreya Reddy Buchireddy | Analyst | Engineer - Provide data analysis, statistics, plotting of data

Sravani Vangala | Analyst | Engineer - Provide data analysis, statistics, plotting of data

Datasets

https://www.kaggle.com/rounakbanik/pokemon

https://www.kaggle.com/vishalsubbiah/pokemon-images-and-types

https://www.kaggle.com/terminus7/pokemon-challenge

Workflow & Collaboration

The collaboration of this project will be done via slack communication and the shared codebase will be managed through a Github repository. The **Github** repository will contain all project contents, including visualizations and the project reports. Most of the project is anticipated to be done locally, but if greater processing power is needed (if this project extends into more machine learning/image recognition) we can use the **Google Cloud Platform** which provides many options for machine learning like **Tensorflow**, which is our choice for training models. For general data visualization, we will most likely be using **Python**.

The Github repository will consist of the following:

- README file
- Project datasets
- Project codes
  - Jupyter Notebook
  - Python Scripts
- Output files
  - Visualizations
  - Tables
- Project Report

Abstract

Pokemon is a global icon for children and adults everywhere. It is a TV series that has expanded into video games, card games, movies, merchandise, and everything in-between. The motivation behind this analysis is to further understand the dynamics of the pokemon universe through data, while also having fun in the process. There's a plethora of sample projects and tutorials on the web using Pokemon data, we hope to combine some of the favorites that we have found along the way.

The desired outcome of this project will be to incorporate different data science concepts using an accumulation of pokemon data provided through Kaggle. The main goal is to provide **statistical analysis** and **data visualization**. If time permits, we hope to incorporate basic machine learning concepts with the given dataset and train a model to provide some type of prediction based output. We also have the option to use image recognition, a great example is providing a model with an image of a pokemon and it will output it's evolved state. The options here are endless.

The following are possible questions to answer with the given dataset:

- Is it possible to build a classifier to identify legendary Pokemon?
- How does the height and weight of a Pokemon correlate with its various base stats?
- What factors influence the Experience Growth and Egg Steps? Are these quantities correlated?
- Which type is the strongest overall? Which is the weakest?
- Which type is the most likely to be a legendary Pokemon?
- Can you build a Pokemon dream team? A team of 6 Pokemon that inflicts the most damage while remaining relatively impervious to any other team of 6 Pokemon

Project Design and Milestones

As stated above, we anticipate this project will use **Python** for Data Analysis, and Data Visualization. We will manage **Jupyter workbooks** and any other project related code through a Github repository. If this project extends into more machine learning/image recognition concepts, we can use the **Google Cloud Platform** which provides many options for machine learning like **Tensorflow**, which will most likely be our tool of choice for machine learning purposes.

The following are anticipated milestones:

1. **Parse and merge any desired pokemon datasets** - this includes cleaning of the data, removing missing values, etc.
2. **Visualize the data**
    1. Distribution of pokemon types
    2. Single type vs dual types

3. heatmap to show a correlation between base stats
   4. Distribution of different attack types - there are two types of attacks in the game: Attack and Special Attack.
   5. Plot legendary vs non-legendary pokemon
3. **Exploratory Data Analysis Conclusions**
   1. Be able to return the strongest and weakest 10 pokemon by a given base stat.
   2. Be able to return pokemon with the highest and lowest total stats.

Bonuses if time permits:

1. **Bonus: Find that pokemon**
   1. Merge photosets to existing data and attempt to interpret images or find evolved pokemon from an unevolved image of pokemon.
2. **bonus: Machine Learning**
   1. Build a model that will predict the likelihood of a pokemon winning a battle.

Related projects:

https://www.kaggle.com/ishmaeln/statistical-analysis-pokemon

https://towardsdatascience.com/whos-that-pok%C3%A9mon-39d1150aedfe

https://www.kaggle.com/mmetter/pokemon-data-analysis-tutorial