Data Mining: Project 1

Akash Shanmugam

9/9/24

Dataset link:

https://www.kaggle.com/datasets/n2cholas/competitive-pokemon-dataset/data

Data Introduction:

The dataset for this project is called the “Complete Competitive Pokemon Dataset” which in total contains 2 .csv files for analysis of the competitive Pokemon information regarding both moveset data and pokemon data for use of analyzing the connection between moves and pokemon’s attributes and stats. All of the data was obtained from sources around 2018 which means that the latest data the dataset contains is from the generation 7 games (Pokemom Sun, Moon, Ultra Sun, and Ultra Moon) and all further additions and retroactive changes will not be reflected in the sets that were used. For the purposes of the project, only the move dataset will be used to reflect in the visualizations shown.

Pre-Processing Data:

The data in the move-data.csv file was not taken from something like a survey but instead was taken using the actual move data contained in the pokemon games. This means that the amount

of null values is essentially zero however I dropped possible ones anyway resulting in no change which meant I could move onto the next step which involved dropping the moves which did no damage and were listed as “None” under the power column. With this I was finished with the clean up.

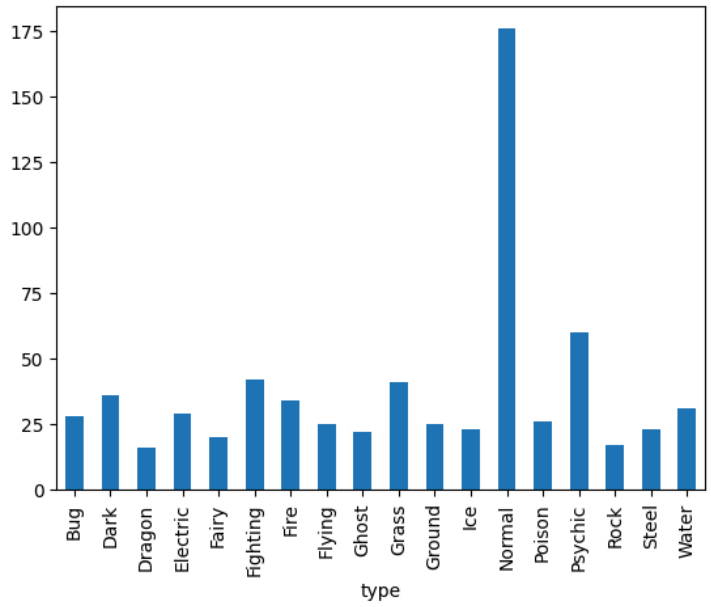
```
import matplotlib.pyplot as plt
df.dropna(how='any')
df.loc[df['Power'] == "None"]
print('Number of moves: ', len(df))
df.sample()
df.columns = ['index', 'name', 'type', 'category', 'contest', 'pp', 'power', 'accuracy', 'g
df.set_index('index')
df['power'].replace('None', 0, inplace=True)
df['accuracy'].replace('None', 100, inplace=True)
df['power'] = pd.to_numeric(df['power'])
df['accuracy'] = pd.to_numeric(df['accuracy'])

df = df[(df.pp != 1) | (df.name == 'Sketch')]

df.sample()
```

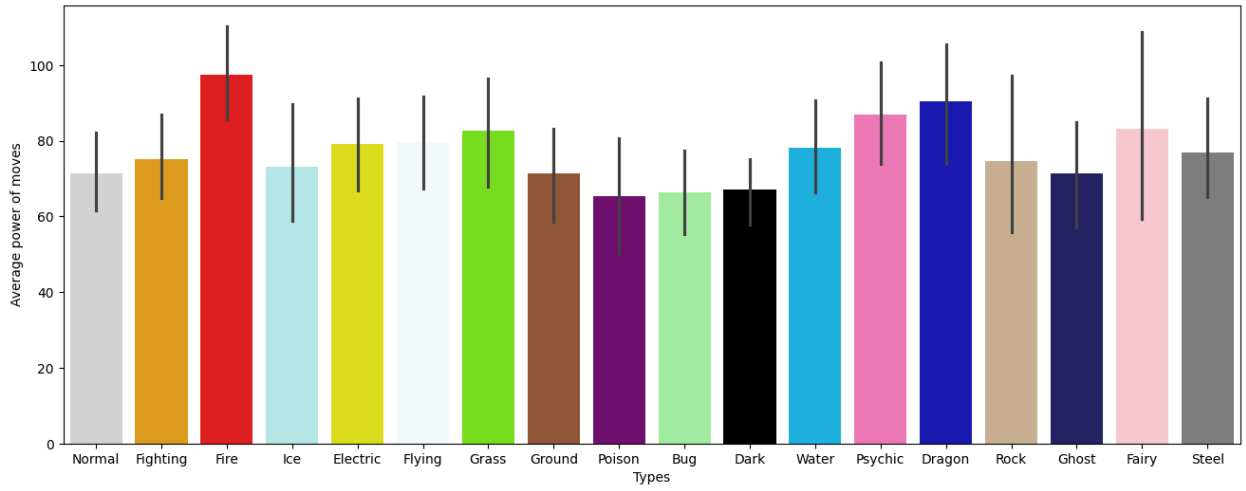
Vis 1:

This figure shows the distribution of the types of moves across the whole dataset and shows how the moves collected in the data have a massive tendency to be normal type across the board. The types behind normal tend to have a more consistent distribution with psychic being the highest that’s not an outlier. Dragon and rock being the ones that have the fewest amount of moves total. This fits with the fact that normal type moves are generally accessible to almost every pokemon and thus have more in general.



Vis 2:

My next task was to see what the average base power of moves in each type category accounted to. I used the data to plot a bar plot that showed what the power of each move tended towards and this lead to some interesting results. As you can see, the strongest type of move is fire type. Looking at the data on its own this makes sense as a lot of individual fire type moves have 150 base power which can contribute to how large the average is. The next thing I could find was that dragon type came in second place which makes sense as it is considered rarer and more powerful which is supported by the previous visualization. Fairy type was high as well but also had a large variance which indicates that the strongest move power individually is tied between fairy and fire.




```
plt.figure(figsize=(16,6))
plt.ylabel('Average power of moves')
plt.xlabel('Types')

datacolor =
['lightgray','orange','red','paleturquoise','yellow','azure','lawngreen','sienna','purple','palegreen','bla
ck','deepskyblue','hotpink','mediumblue','tan','midnightblue','pink','gray']

sns.barplot(x= 'type', y='power', data=df, palette=datacolor)
```