act_report

February 18, 2023

0.1 Report: act_report

- Create a **250-word-minimum written report** called "act_report.pdf" or "act_report.html" that communicates the insights and displays the visualization(s) produced from your wrangled data. This is to be framed as an external document, like a blog post or magazine article, for example.

0.1.1 INTRODUCTION

The dataset that was analyzed and visualized was the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. Three different datasets were wrangled, multiple quality and tidiness issues identified and cleaned. A master dataset is now created from the combination of cleaned datasets, presented for further analysis and visualizations.

Codes below attempt to load the cleaned master dataset so visualizations and analysis results are possible in this report.

```
In [4]: import pandas as pd
        import matplotlib.pyplot as plt
        from IPython import display
        % matplotlib inline

        df_twitter_archive_master = pd.read_csv('twitter_archive_master.csv')
```

Analysing and Visualizing1: Top 5 most correctly predicted outcome based on neural network analysis

1) golden_retriever
2) Labrador_retriever
3) Pembroke
4) Chihuahua
5) pug

Using the first prediction attempt of the neural network, code above generate top 5 most correctly predicted dog breed. This result might speak to the data used to train the neural network model or simply highlighting the fact that most people world owns one or more of the listed dog breeds and therefore there is abundants of data available to properly train the neural network.

Analyzing and visualizing2: Top 10 most rated dogs

```
In [3]: df_Most_rated = df_twitter_archive_master[df_twitter_archive_master['prediction1']!= 'no
        df_Most_rated.groupby('prediction1')['rating_numerator'].value_counts().nlargest(10)
```

```
Out[3]: prediction1         rating_numerator
        golden_retriever    12                 54
                            13                 28
        Pembroke            12                 27
                            11                 26
        Labrador_retriever  12                 25
        golden_retriever    11                 24
        Chihuahua           12                 22
        Labrador_retriever  10                 22
                            11                 22
        golden_retriever    10                 20
        Name: rating_numerator, dtype: int64
```

The retriever breed dominates the sppace of most rated dogs which further confirms the initial theory of their popularity and abundance in the world.

Analyzing and visualizing3: Top 5 comments/dogs with most retweet/favorite counts People appears more emotionally drawn to English Springer, Chihuahua, and French bulldog above the most popular Retriever based on the number of favorite tweets.

```
In [13]: display.Image('https://pbs.twimg.com/ext_tw_video_thumb/879415784908390401/pu/img/cX7XI
```

Out[13]:

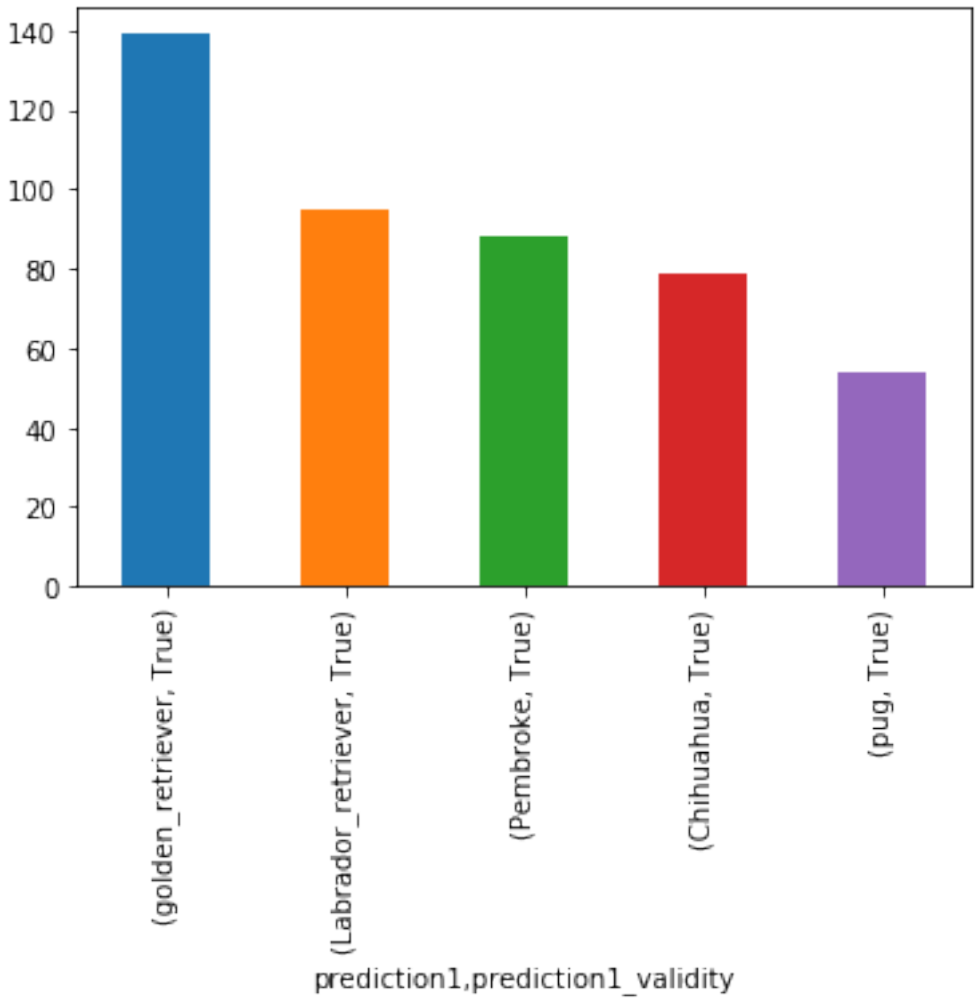
English_springer

0.1.2 Insights:

1. The retriever (Golden and Labrador) appears to the most valued dog bread based on the high number of positive ratings from the general public.
2. The neural network analysis has the most success with dog breeds like Pembroke, Chihuahua, Labrador_retriever, Golden_retriever, and Pug
3. Retriever breeds might be highly rated and popular but people appears more emotionally drawn to English Springer, Chihuahua, and French bulldog based on the number of favorite tweets.

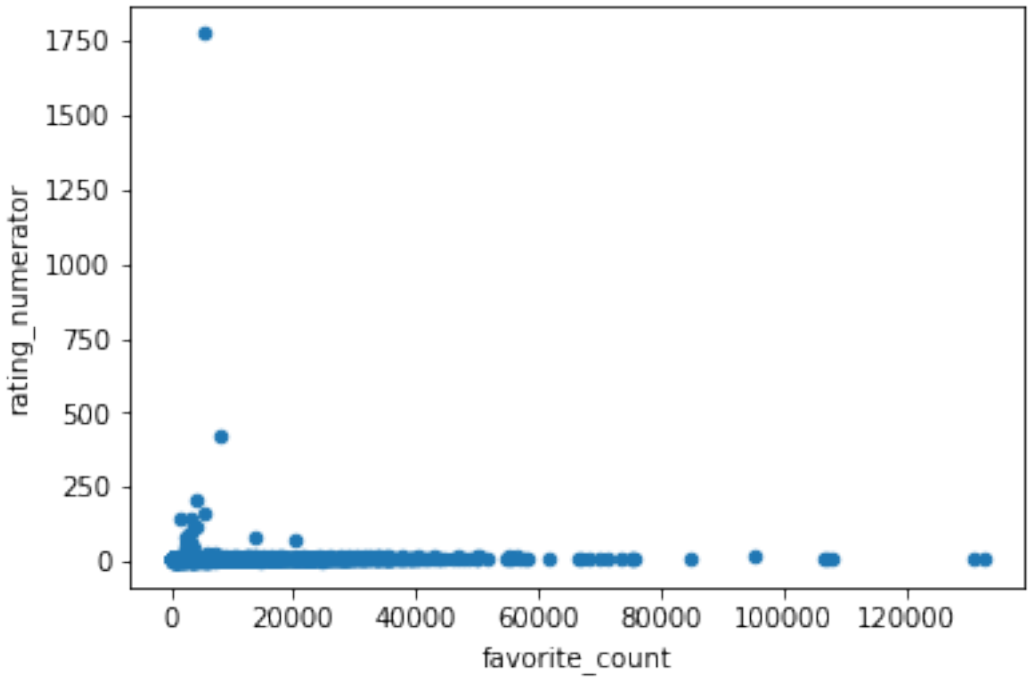
0.1.3 Visualization

```
In [4]: df_top_predicted = df_twitter_archive_master[df_twitter_archive_master['prediction1_vali
        df_top_predicted.groupby('prediction1')['prediction1_validity'].value_counts().nlargest(
```



Visualization of the top 5 most correctly predicted outcome based on neural network analysis

```
In [10]: df_twitter_archive_master.plot(x='favorite_count', y='rating_numerator', kind='scatter'
```



Examining relationship between the dog ratings and how adorable people think they are. It appears theere is a strong relationship between thse two variables as depicted above.