WeRateDogs Twitter Archive - Act Report

-By Piyush Kumar

WeRateDog

1. WerateDog twitter archived data contains 2365 rows and 17 columns from the @dog_rates twitter account. Which is downloaded manually from the link provided in the section "Project Details".
2. Then we created a request to download the "image-predictions.tsv file".
3. Then we extracted the data using twitter api and some specific data is extracted such as "tweet_id, retweets, favourites"

Wrangling of Data

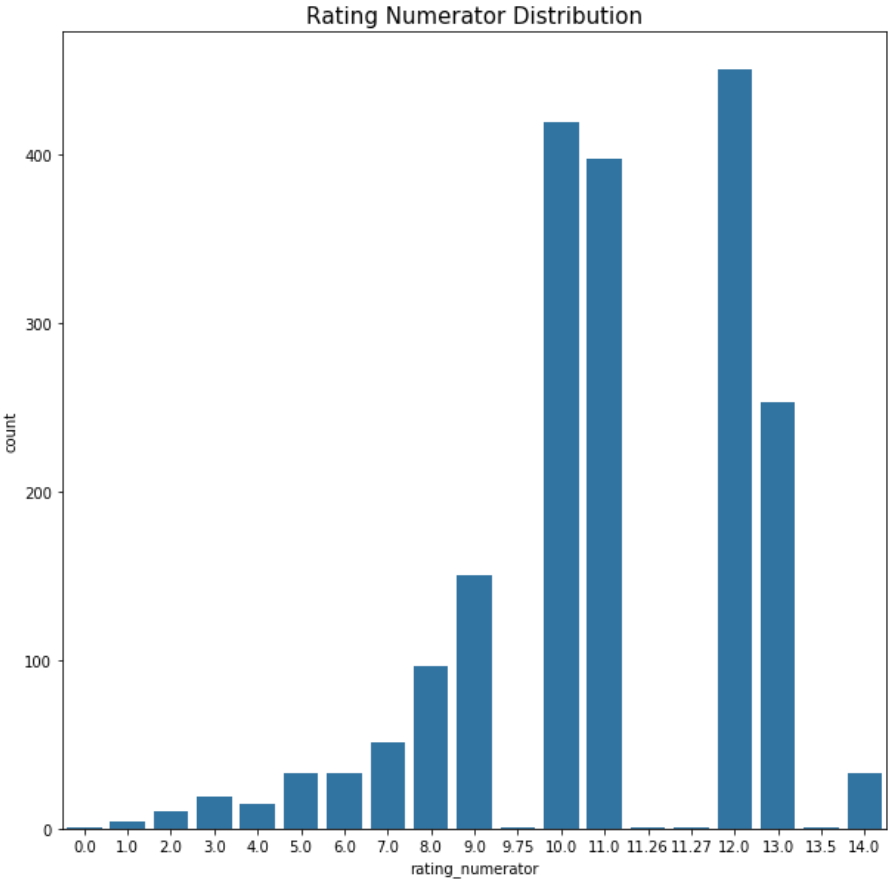
1. Before cleaning the Dataset some useful insights are drawn to distinguish whether data is having:
    a. Quality Issues
    b. Tidiness Issues
2. All three Datasets are combined With respect to their Tweet_Id's.
3. After removing all the errors and storing them in a separate csv file. And observing this csv we concluded that only 1968 rows and 14 columns were useful for Analysis and drawing useful insights from it.

Visualizations

1. Which rating was given to most of the dogs?

Code:

```
try:
    plt.figure(figsize=(10,10))
    plt.title ('Rating Numerator Distribution', size=15)
    plt.xlabel('Rating Numerator')z
    plt.ylabel('Number of Ratings')
    base_color = sns.color_palette()[0]
    sns.countplot(data = df, x = 'rating_numerator', color = base_color)
except:
    print("error")
```



Observations:

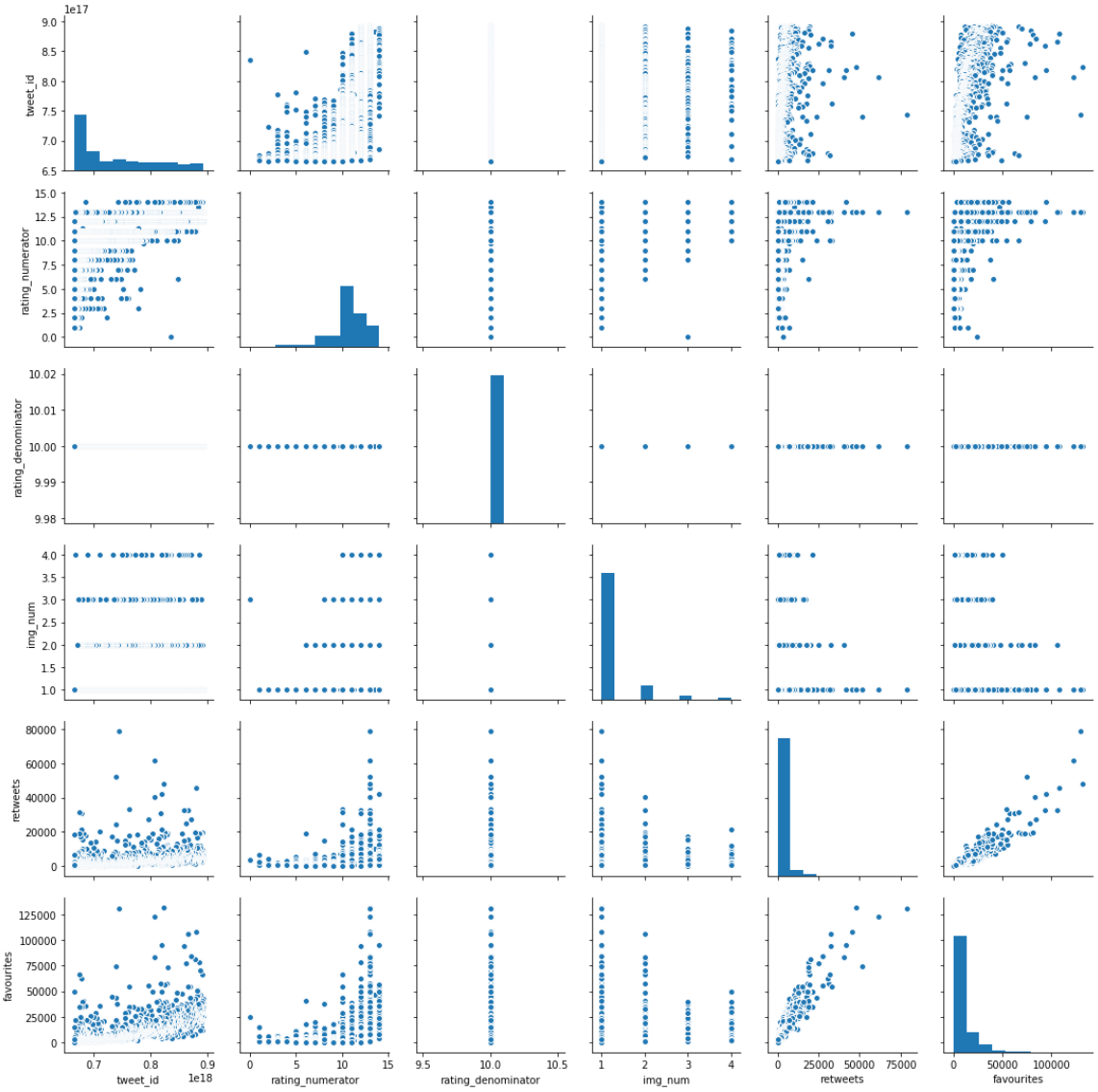
This visualization indicates which rating was given to most of the dogs and we can conclude that:

1. Most of the dogs get rating_numerator in between 10-13
2. There is very few dogs even 1 or 2 with rating 0 and numerator rating as float

2. Favourites and Retweets

Code:

```
sns.pairplot(data= df)
```



Observations:

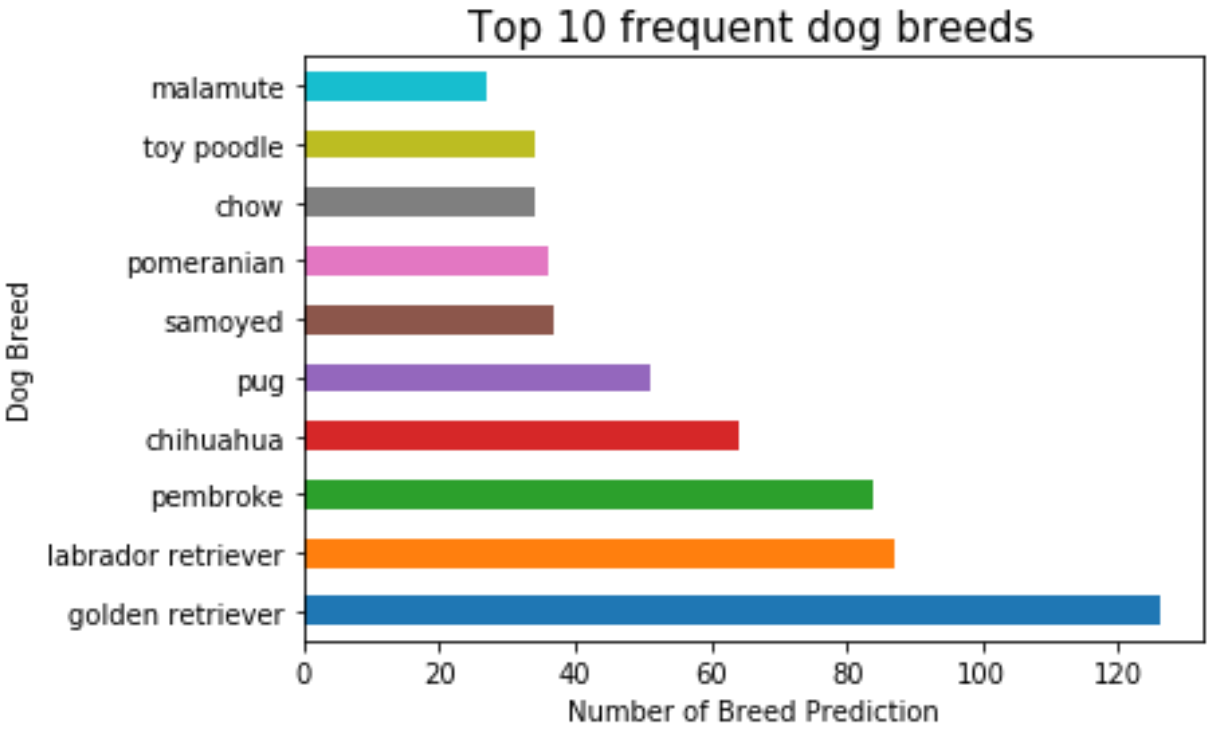
This visualization indicates which the Favourites and Retweets with respect to other data attributes:

1. Rating_numerator vs favourites/retweets are right skewed.
2. Retweets vs favourites is linear positive.

3. What are the 10 most frequent predicted dog breeds?

Code:

```
try:
    df['pred_breed'].value_counts()[0:10].sort_values(ascending=False).plot(kind = 'barh')
    plt.xlabel('Number of Breed Prediction')
    plt.title('Top 10 frequent dog breeds', size=15)
    plt.ylabel('Dog Breed')
    plt.plot();
except:
    print("error")
```



Observations:

This visualization indicates What are the 10 most frequent predicted dog breeds

1. Most of the dogs have golden retriever, labrador retriever as breed which all are rated.

Conclusion

1. Most of the dogs get rating_numerator in between 10-13
2. There are very few dogs even 1 or 2 with rating 0 and numerator rating as float.
3. Most of the dogs have golden retriever, labrador retriever as breed which all are rated.
4. The max no of tweets are: 78839
5. The max no of tweets are: 14
6. Rating_numerator vs favourites/retweets are right skewed.
7. Retweets vs favourites is linear positive.