Data Analysis Project — WeRateDogs®

WeRateDogs®
12.3K Tweets

WeRateDogs®
@dog_rates

Your Only Source For Professional Dog Ratings Instagram and Facebook ➪ WeRateDogs partnerships@weratedogs.com

「 DM YOUR DOGS 」 blacklivesmatters.carrd.co
Joined November 2015

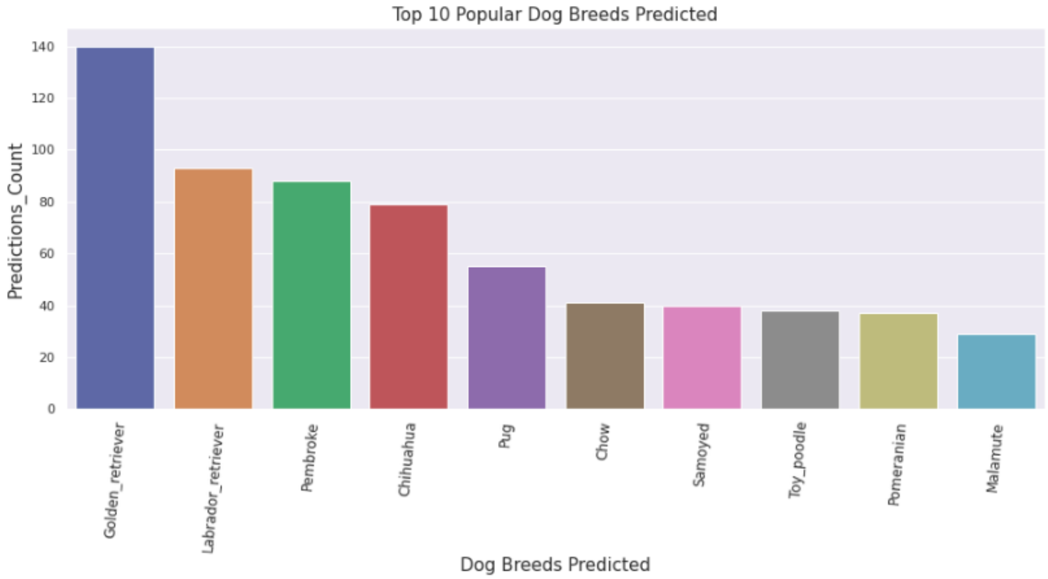
16 Following 8.7M Followers

A short Intro of @weratedogs:
WeRateDogs is a Twitter account that rates people's dogs with humorous comments about the dog. The account was started in 2015 by a college student Matt Nelson. This account has attracted 8.8 million+ followers ever since.

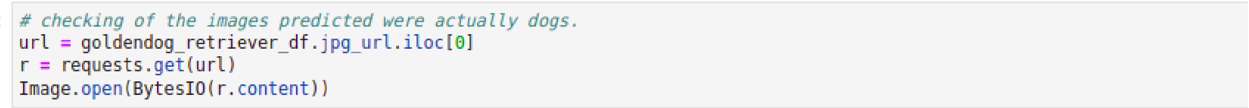
After gathering, assessing and cleaning my data I stored it into a master dataset which i analyzed then visualized to draw insights from the dataframe.

1.What are the top ten most popular dog breeds tweeted.

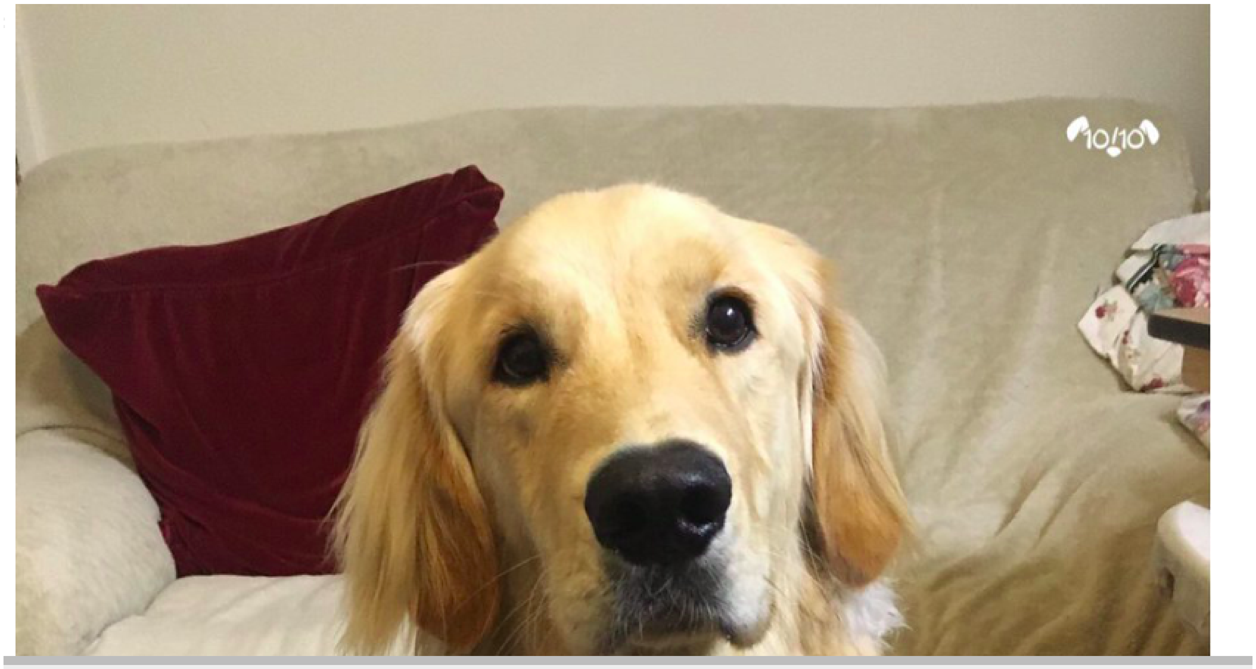
From the analysis, the Golden Retriever is the most popular dog. These dogs have the most tweets on them, which has led them on our most popular list. Used the



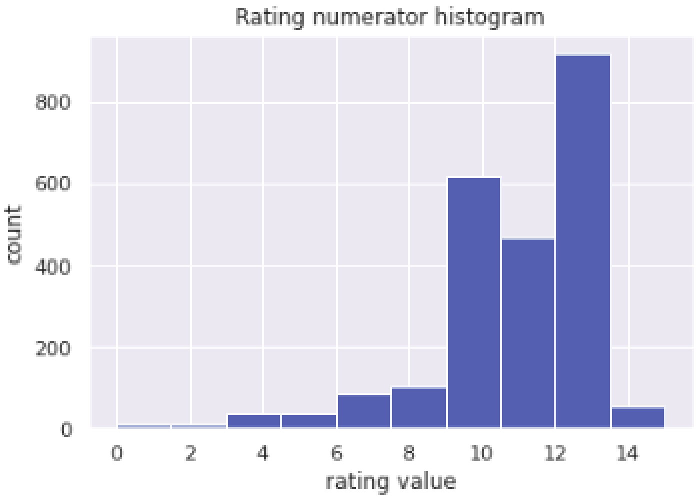
I tested to check if the predictions were really dogs and sure it was spot on. Though some were not really dogs.



```
# checking of the images predicted were actually dogs.
url = goldendog_retriever_df.jpg_url.iloc[0]
r = requests.get(url)
Image.open(BytesIO(r.content))
```

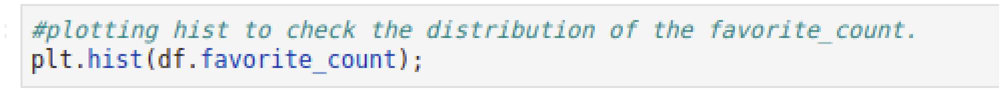


2. What is the distribution of the rating numerator?

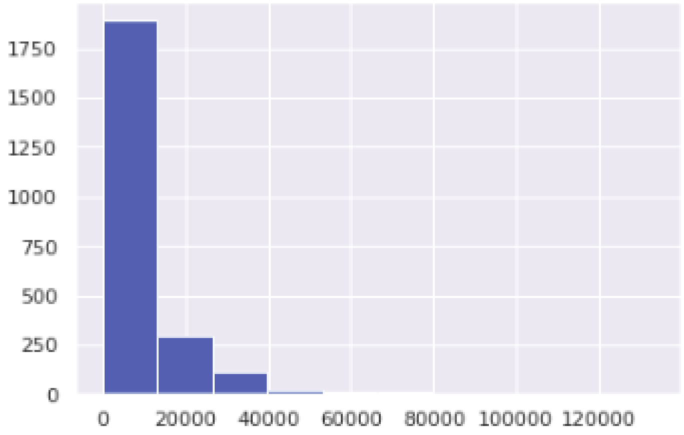


From the histogram plotted it revealed that the most dogs were rated between 10 and 13. The distribution is skewed to the right.

4.What is the distribution of the favorite_count?

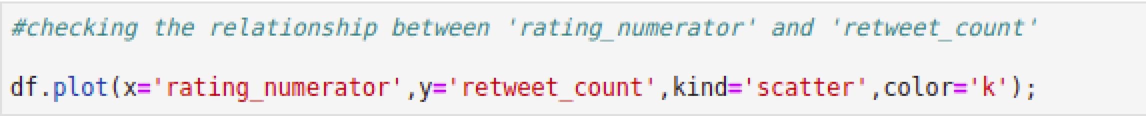


```
#plotting hist to check the distribution of the favorite_count.
plt.hist(df.favorite_count);
```



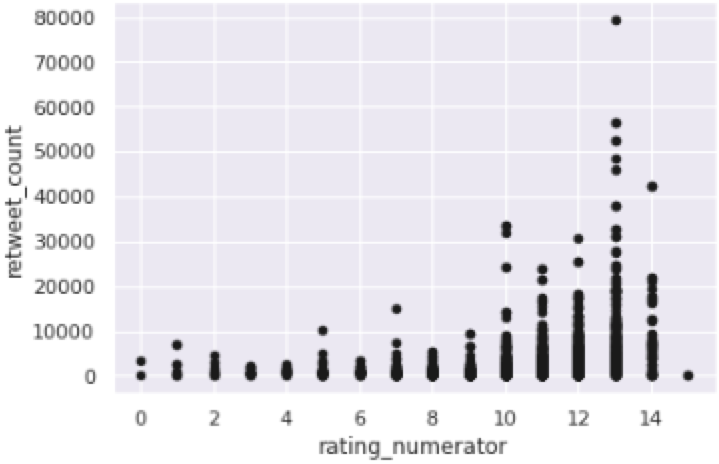
Most of the favourite_counts were less than 20000, The distribution is skewed to the right.

5.What is the relationship between 'rating_numerator' and 'retweet_count'?



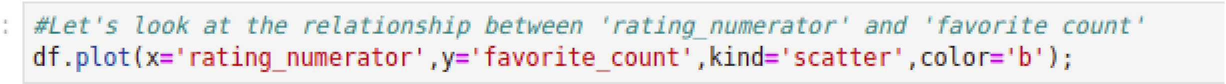
```
#checking the relationship between 'rating_numerator' and 'retweet_count'

df.plot(x='rating_numerator',y='retweet_count',kind='scatter',color='k');
```

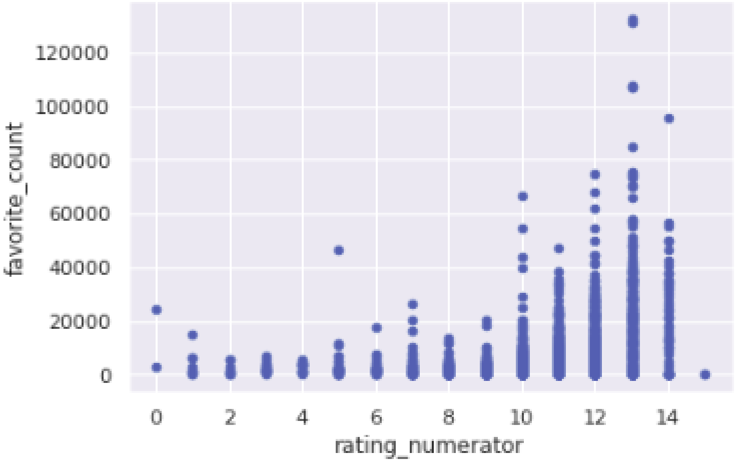


The ratings by WeRateDogs are pretty relative to the retweet count and favorite count.

6. What is the relationship between 'rating_numerator' and 'favorite count'?

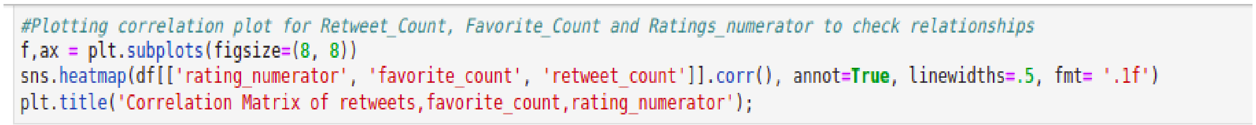


```
:  #Let's look at the relationship between 'rating_numerator' and 'favorite count'
   df.plot(x='rating_numerator',y='favorite_count',kind='scatter',color='b');
```

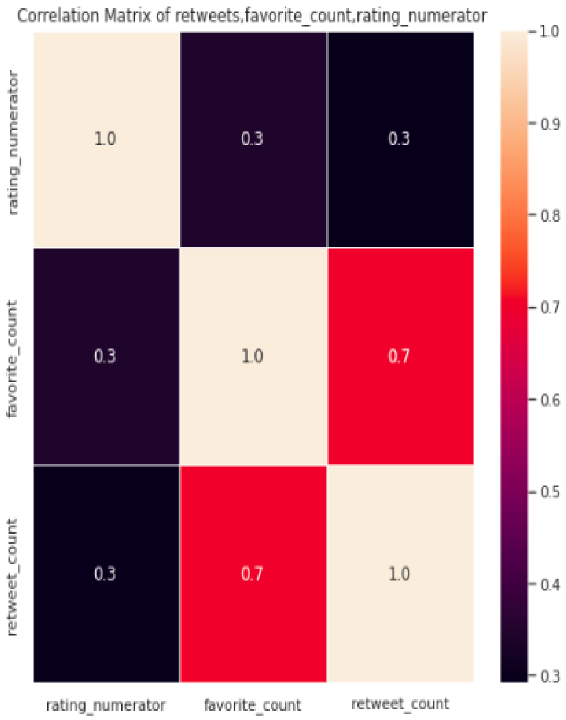


They are relative,too, and I also found that rating_numerator 13 got the highest favorite counts and retweet counts

7.Plotting the correlation of Retweet_Count, Favorite_Count and Ratings_numerator to check relationships.



```
#Plotting correlation plot for Retweet_Count, Favorite_Count and Ratings_numerator to check relationships
f,ax = plt.subplots(figsize=(8, 8))
sns.heatmap(df[['rating_numerator', 'favorite_count', 'retweet_count']].corr(), annot=True, linewidths=.5, fmt= '.1f')
plt.title('Correlation Matrix of retweets,favorite_count,rating_numerator');
```



There is a relatively strong correlation of 0.7 between retweet_count and favorite_count.