Act Report

Documentation of Analysis and Insights into Final Data

Introduction

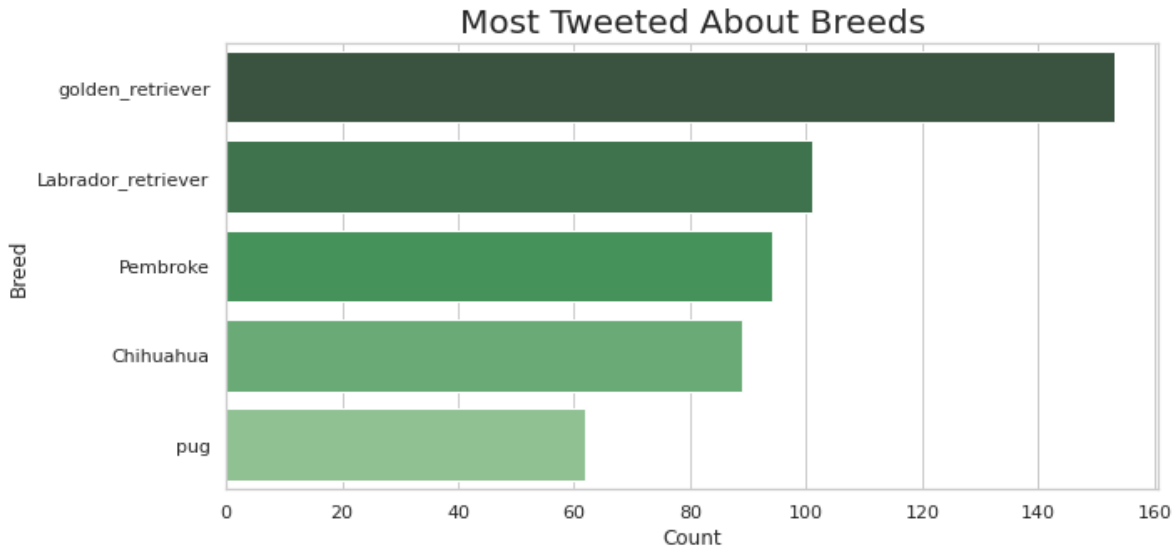
WeRateDogs (@dog_rates) is a Twitter Account that rates dogs' pictures. Its rating system is a large contributor to the account's popularity since dogs are rated out of ten, but can have numerator higher than 10. Why? Because they're good dogs.

As of 2021 this Twitter account has 8.9M Followers.

Note on Limitations of this Analysis: One of the three datasets wrangled together, "Enhanced Twitter Archive" only has dog breeds and nicknames up till August 2017. This might skew some of the results we have.

Analysis

Which dogs breeds were the most tweeted about?



What is the average rating for all dogs?

```
In [34]: print("Averge rating for all dogs is {}".format(df_clean.rating.mean()))
         Averge rating for all dogs is 10.531925849639547
```

The average rating for all dogs is 10.53. They're good dogs.

On average, Which dog breeds were the "Goodest? (highest rated)

```
goodest_dogs = df_clean.groupby(['breed']).rating.mean()
goodest_dogs.sort_values(ascending=False)
```

```
breed
Bouvier_des_Flandres           13.000000
Saluki                         12.500000
briard                         12.333333
Tibetan_mastiff                12.250000
Border_terrier                 12.142857
                                  ...
Tibetan_terrier                 9.250000
Scotch_terrier                  9.000000
Walker_hound                    9.000000
soft-coated_wheaten_terrier     8.166667
Japanese_spaniel                5.000000
Name: rating, Length: 112, dtype: float64
```

Which dogs were the most retweeted?

On Average

```
df_clean.groupby(['breed']).retweet_count.mean().sort_values(ascending=False)
```

```
breed
Bedlington_terrier    7002.166667
Afghan_hound          4988.000000
standard_poodle       4680.545455
French_bulldog        4604.900000
English_springer      4590.200000
                            ...
EntleBucher            586.000000
Tibetan_terrier        465.250000
Japanese_spaniel       378.000000
Brabancon_griffon      242.333333
groenendael            234.000000
Name: retweet_count, Length: 112, dtype: float64
```

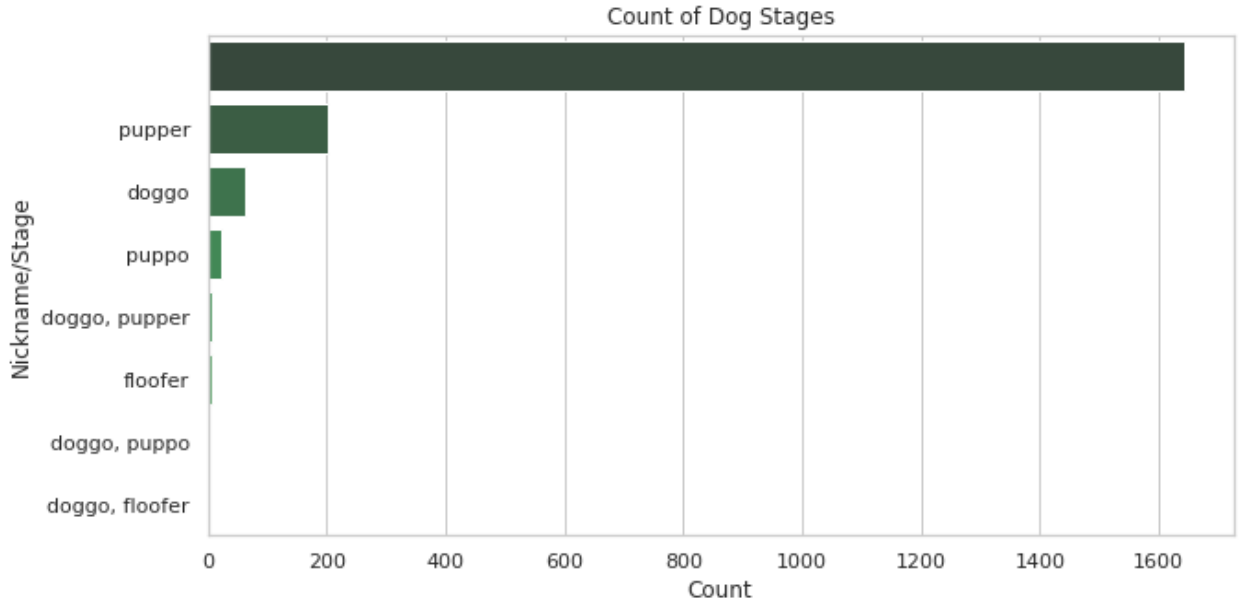
In Total

```
df_clean.groupby(['breed']).retweet_count.sum().sort_values(ascending=False)

# Keep in mind that Golden Retrievers were also the most tweeted about.
# The data is skewed due to the number of tweets for each breed being different.
```

```
breed
golden_retriever      483509
Labrador_retriever    326956
Pembroke              247358
Chihuahua             204271
Samoyed               162855
                         ...
Brabancon_griffon        727
Irish_wolfhound          692
EntleBucher              586
groenendael              468
Japanese_spaniel         378
Name: retweet_count, Length: 112, dtype: int64
```

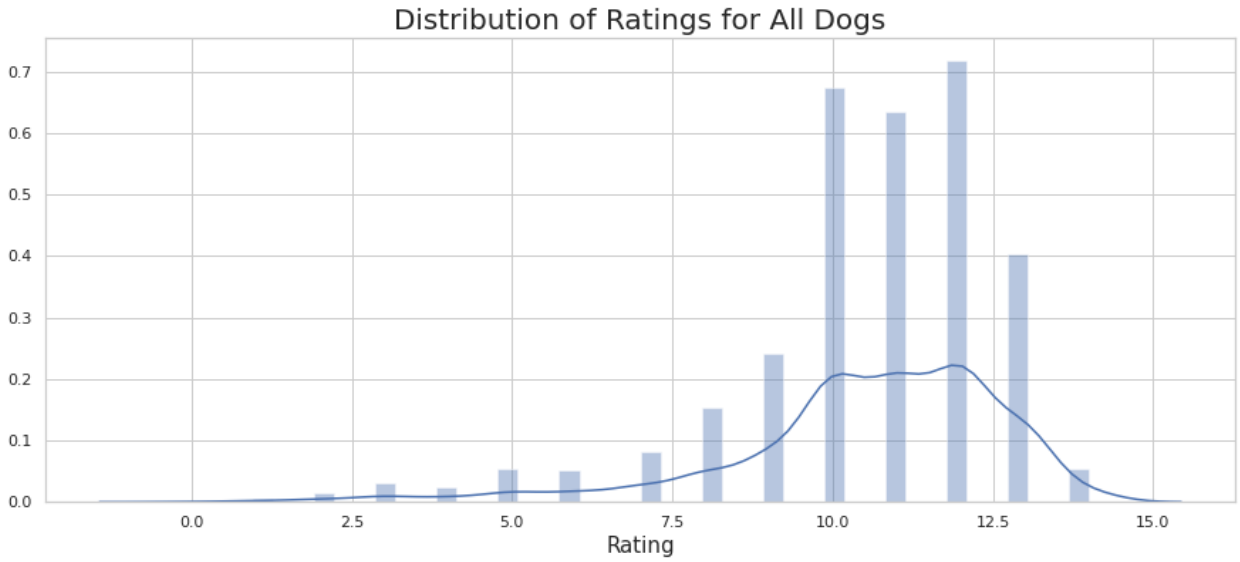
How many Floofers, Doggos, Puppos and Puppers?



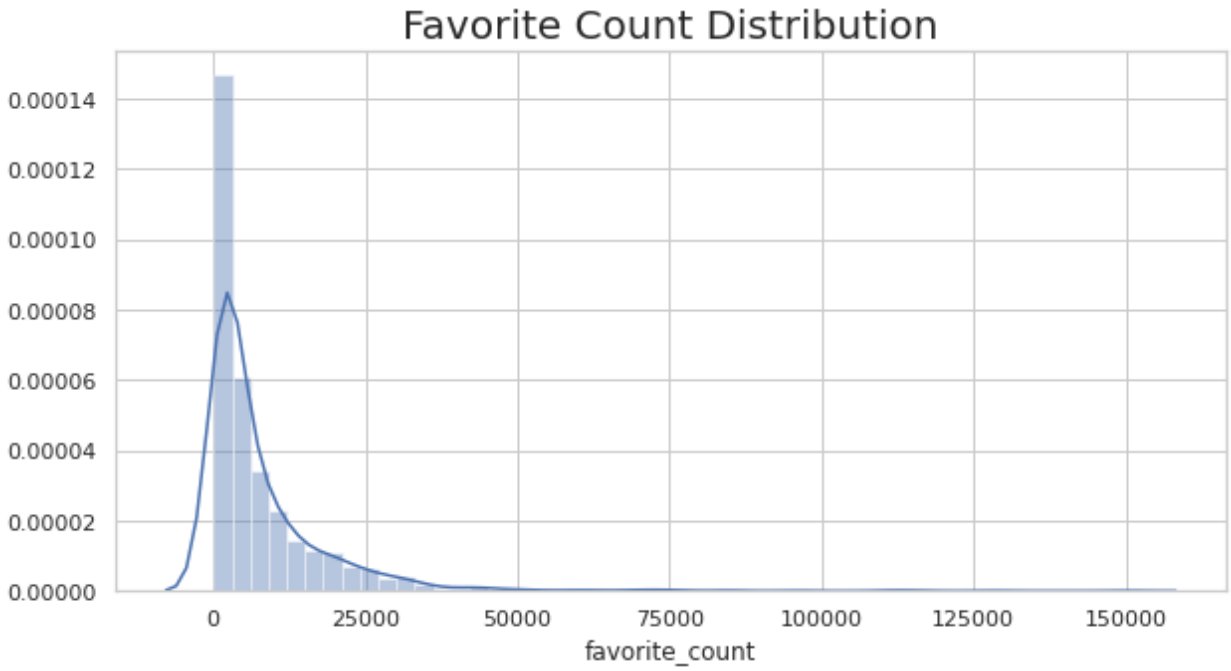
Looking at the plot above, we note the following:

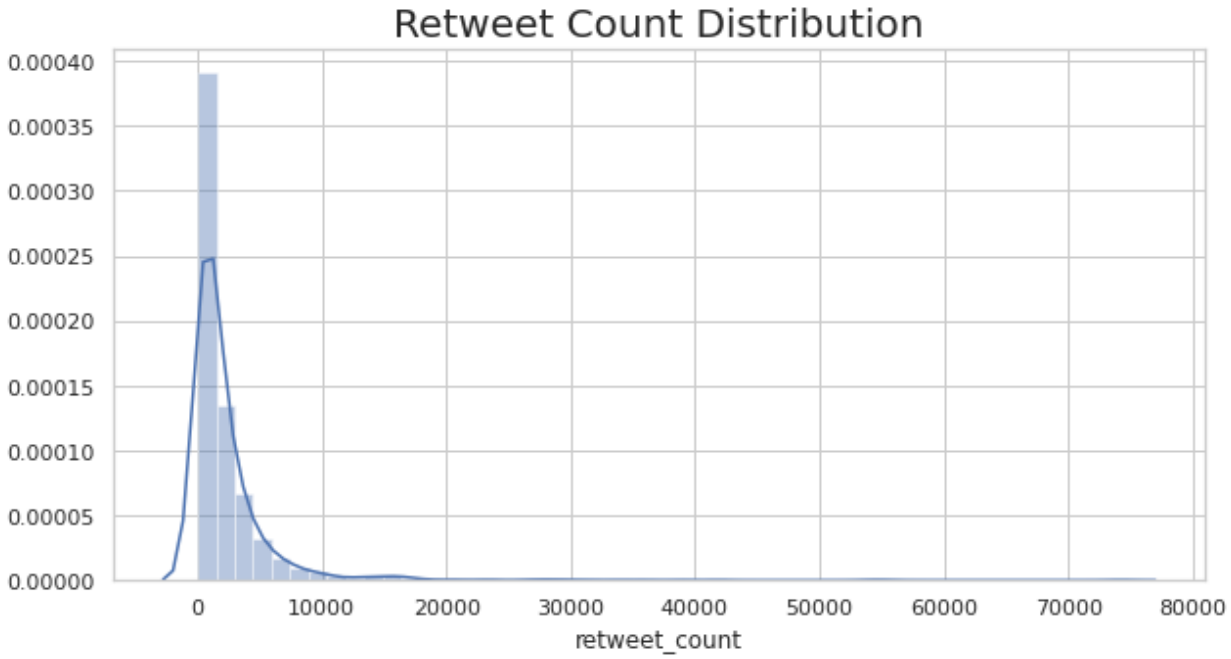
- The Pupper stage has the highest count. Doggo|Pupper and Floofer have the lowest.
- The most retweeted dog breed is the Golden Retriever, while the lowest is Japanese Spaniel. On average however, the most retweeted dog breed is Bedlington Terrier.

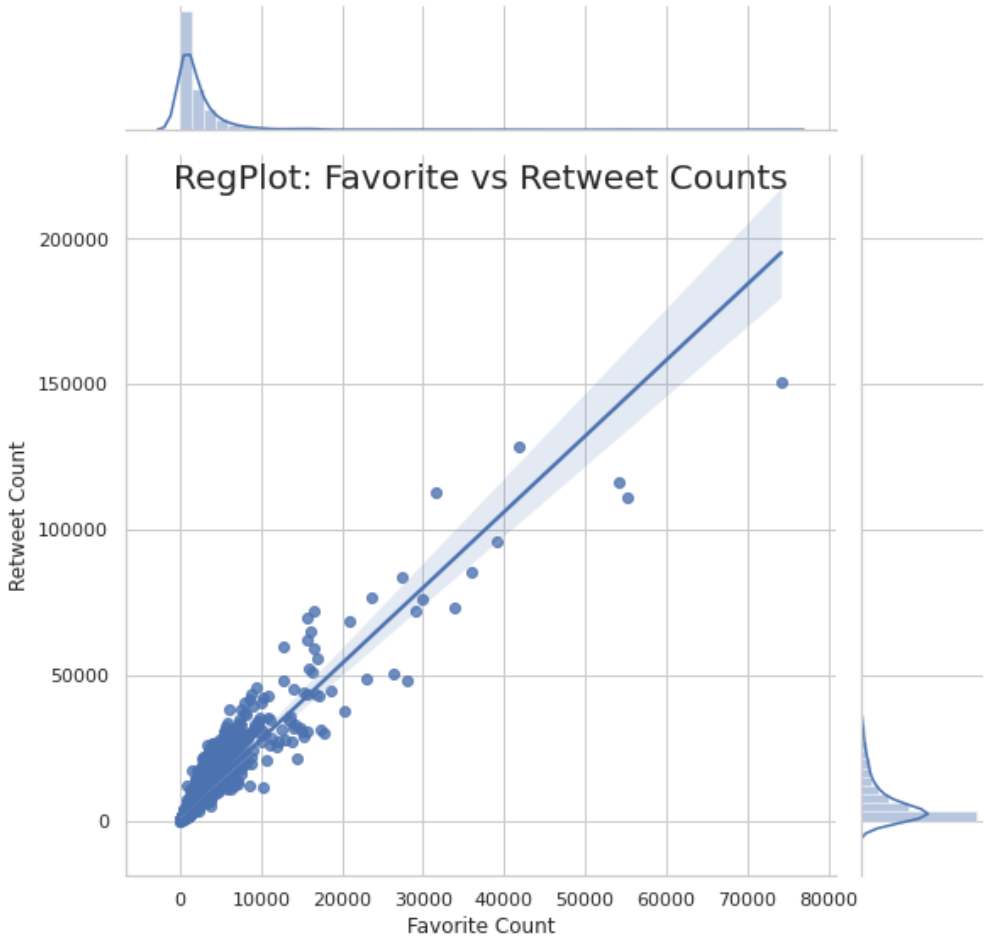
What is the distribution of ratings across the entire dataset?



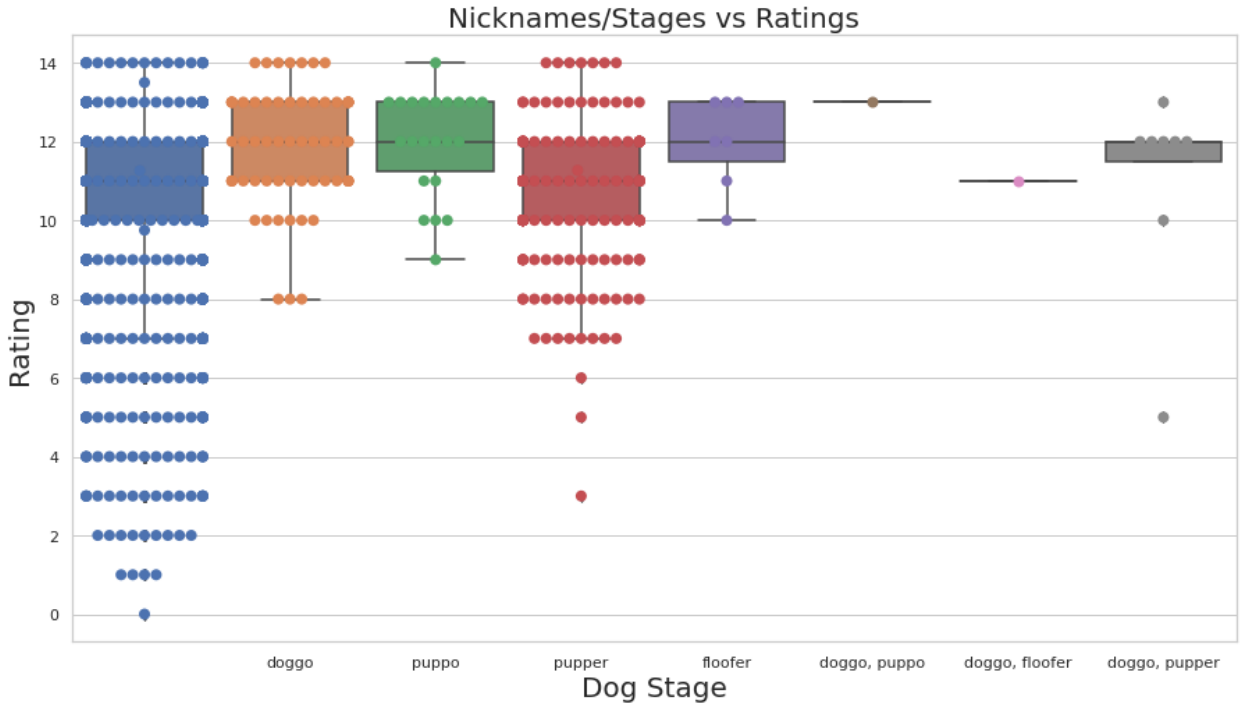
What is the distribution of Favorites and Retweets?

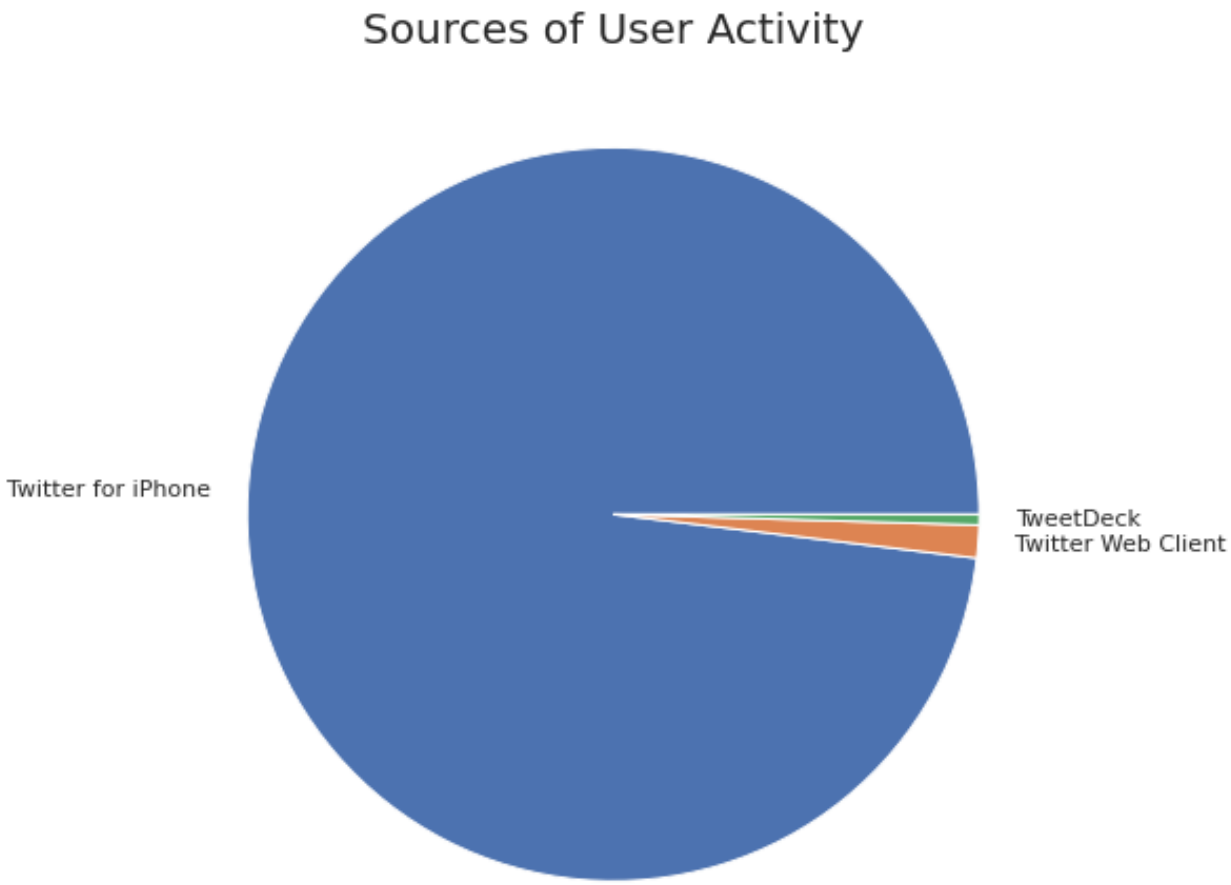






We notice a strong positive correlation above between Tweets that were Favorited and Retweeted. This makes sense since the more a tweet is retweeted/favorited the more likely it shows up for others, leading to more engagement.


Nicknames/Stages vs Ratings
Rating
14
12
10
8
6
4
2
0
doggo
puppo
pupper
floofer
doggo, puppo
doggo, floofer
doggo, pupper
Dog Stage



We notice most of the user activity came from Twitter for iPhone.