Dog Craze ! - Visualizing >2k tweets by WeRateDogs

Hello world! I have spent crazy hours the past weeks gathering and cleaning data from the @WeRateDogs twitter acount... Hope you enjoy this article and some of the visualizations I managed to do.

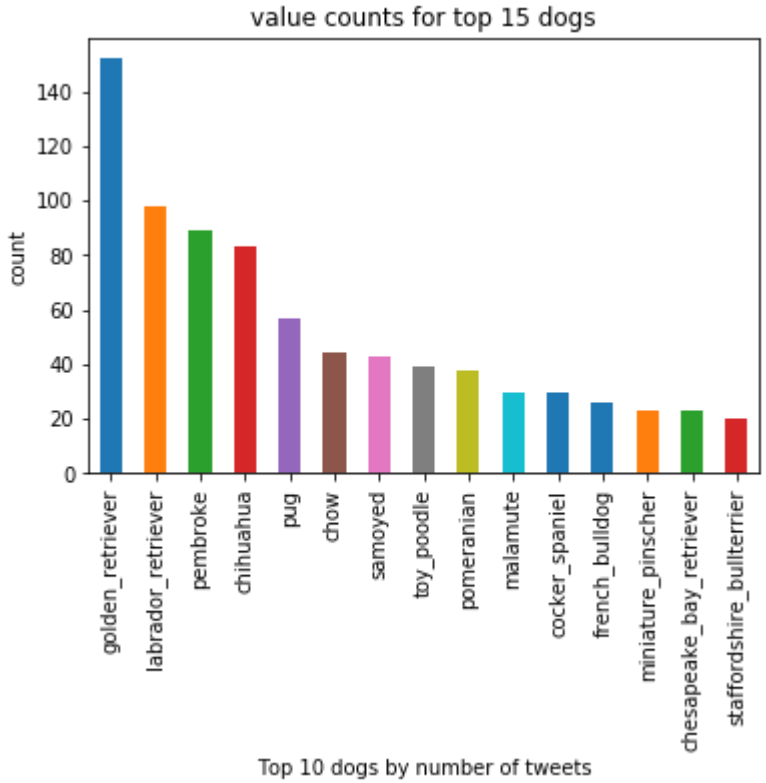
Before you start reading this... please go to WeRateDog's twitter page (https://twitter.com/dog_rates?lang=en) and look at the cute dogs!

So. now back to business. After gathering, assessing data using python, I started doing some visualizations. which you can see below

Top dogs on the block

In [66]:

```
df_combined.prediction_1.value_counts()[:15].plot(kind='bar',title='value counts for top 15 dogs')
plt.ylabel('count')
plt.xlabel('Top 10 dogs by number of tweets');
```



We rate dogs has thousands of tweets and dozens of dog breeds. But what are the top ones? No surprises here! it seems like the top 2 dogs were the golden retriever and the labrador.. a.k.a. America's favorite dogs. Quite surprisingly... the pug even made it to the top 15! That was... errm.... unexpected.. who likes pugs anyway? Oh wait... there's also the chihuahua. Let's not even talk about chihuahas... Although this data is not 100% accurate, the face recognitiion for the dogs was based on a neural network that analyzed thousands of images

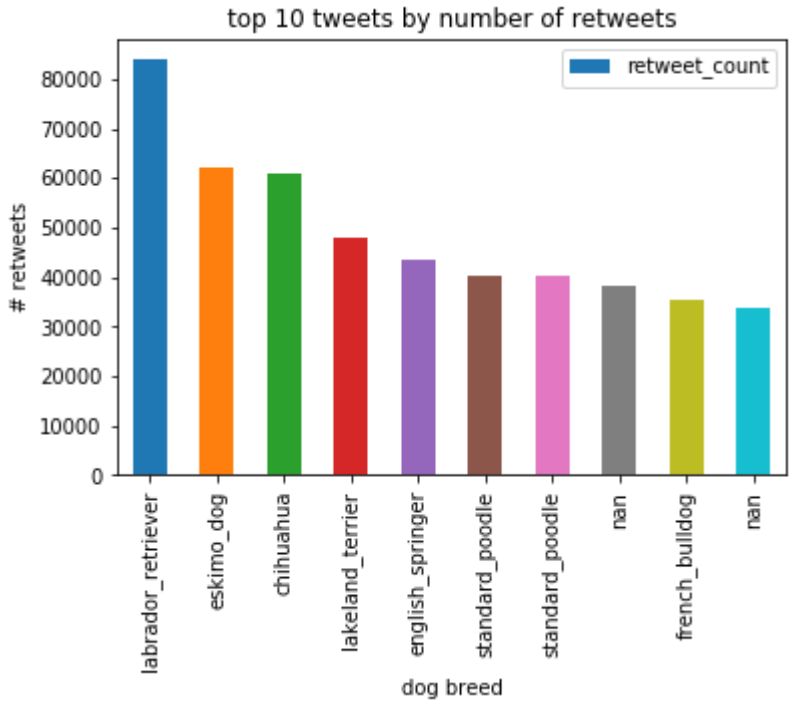
How popular are dogs?

A lot. Not only does WeRateDog boast 7.55 MILLION followers on instagram.. some of their tweets are loved, beloved, favorited, and retweeted by thousands of people.

Retweets

In [94]:

```
top_10_retweets = df_combined.sort_values(by=['retweet_count'],ascending=False)[:10][[
'prediction_1','retweet_count']]
top_10_retweets.plot(x='prediction_1',y='retweet_count',kind='bar')
plt.title('top 10 tweets by number of retweets')
plt.xlabel('dog breed')
plt.ylabel('# retweets');
```

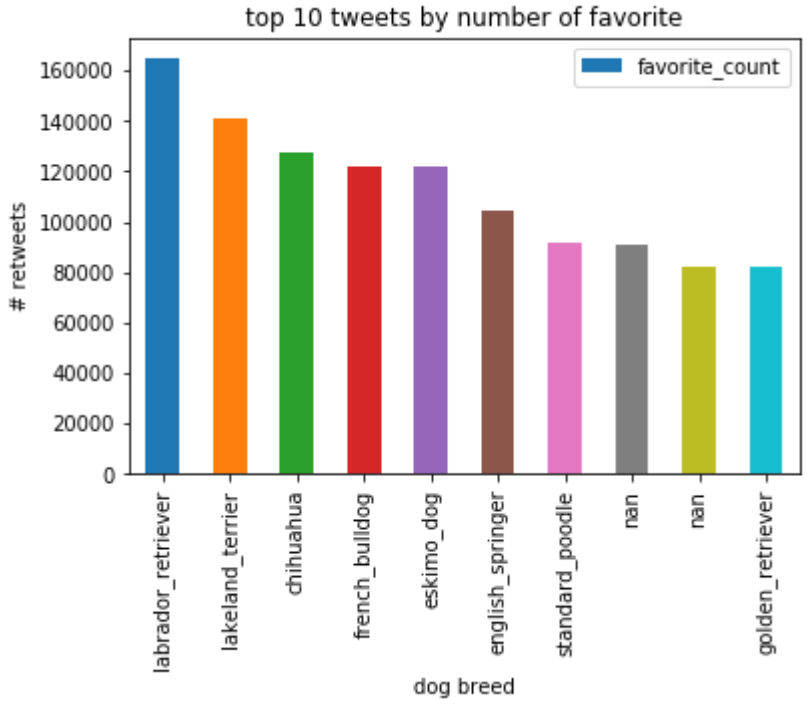


Look at that. I bet you're wondering what that top top tweet is, am I right? Yes. Obviously I'm going to show you. Here you go (@https://twitter.com/dog_rates/status/744234799360020481?lang=en). You're welcome. It's a labrador swimming in a pool. That adorable dog? He just fetched over 80k Retweets. That's crazy. But also off the charts. The median tweet for WeRateDog fetches 1385 retweets. Even this adorable french bulldog only had 40k retweets (half of our adorable labrador)

What about the favorites ?

In [101]:

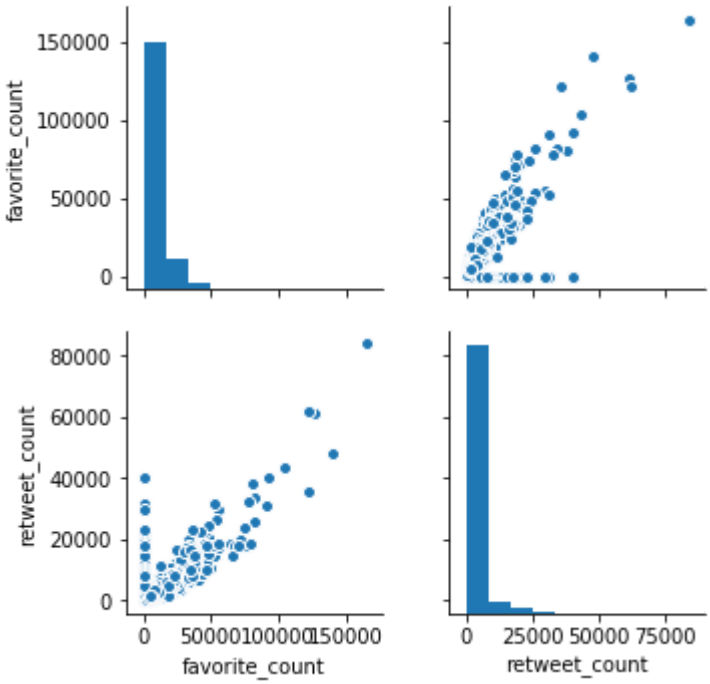
```
top_10_favorites = df_combined.sort_values(by=['favorite_count'],ascending=False)[:10]
[['prediction_1','favorite_count']]
top_10_favorites.plot(x='prediction_1',y='favorite_count',kind='bar')
plt.title('top 10 tweets by number of favorite')
plt.xlabel('dog breed')
plt.ylabel('# retweets');
```



Oddly similar, no? We could go down the rabbit hole to find out if those are exactly the same dogs from the favorite.. but deep deep inside we know that these variables are correlated... no? Let's put some more visualizations and science behind that assumption!

In [100]:

```
sns.pairplot(df_combined[['favorite_count','retweet_count']]);
```



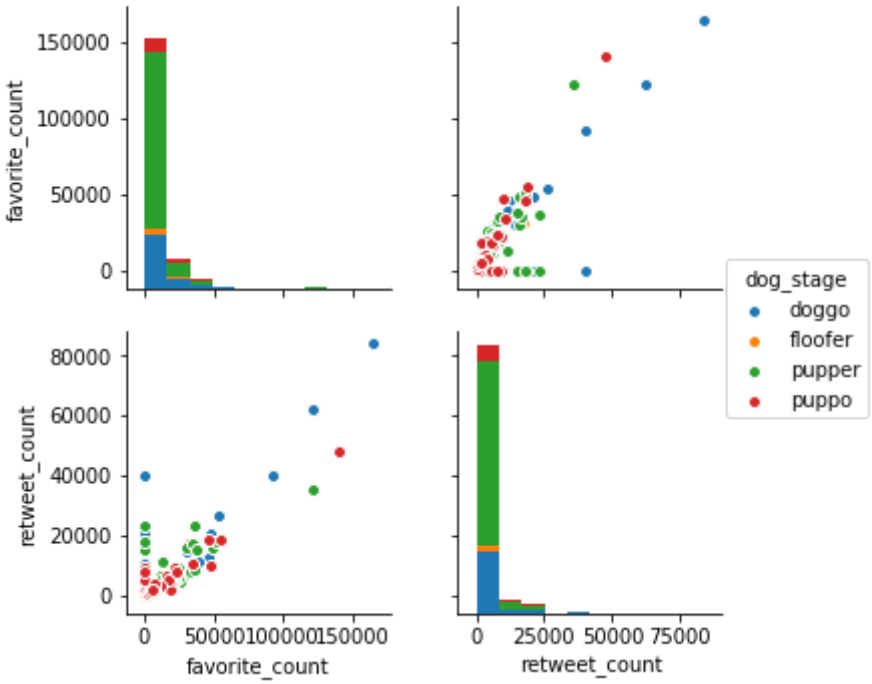
Beautiful chart, no? Here we can see that there seems to be a clear relationship between being favorited and being retweeted... sure, there are those outliers that seem to have a lot of retweets even if they were not favorited.. but let's forget about those and appreciate the strong correlation between favorites and retweets!

Doggos, floofers, pupper, puppo..

What the... ? What are those?? Don't worry, I have your back. I was also confused at first.. here's a video (@http://www.tastefullyoffensive.com/2017/07/the-difference-between-doggo-pupper-and.html) that explains this overly complicated categorization. Anyway, in our data set we had the four categories.

In [102]:

```
sns.pairplot(df_combined[['favorite_count','retweet_count','dog_stage']],hue='dog_stag
e');
```



Ok, so we see that there are a lot lot of puppers and doggos but not many floofers (:( ) nor puppos. Go back to the video to see the difference. Interestingly, we see that the outliers, that is the ones that have the most most retweets ad favorites are mostly doggos.

What other insights do you get from this? I leave it up to you :) !

Code used

How did I do all this? Here's the code!

In [23]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

%matplotlib inline
```

In [18]:

```
df_combined = pd.read_csv('weratedogs_combined.csv')
```

In [19]:

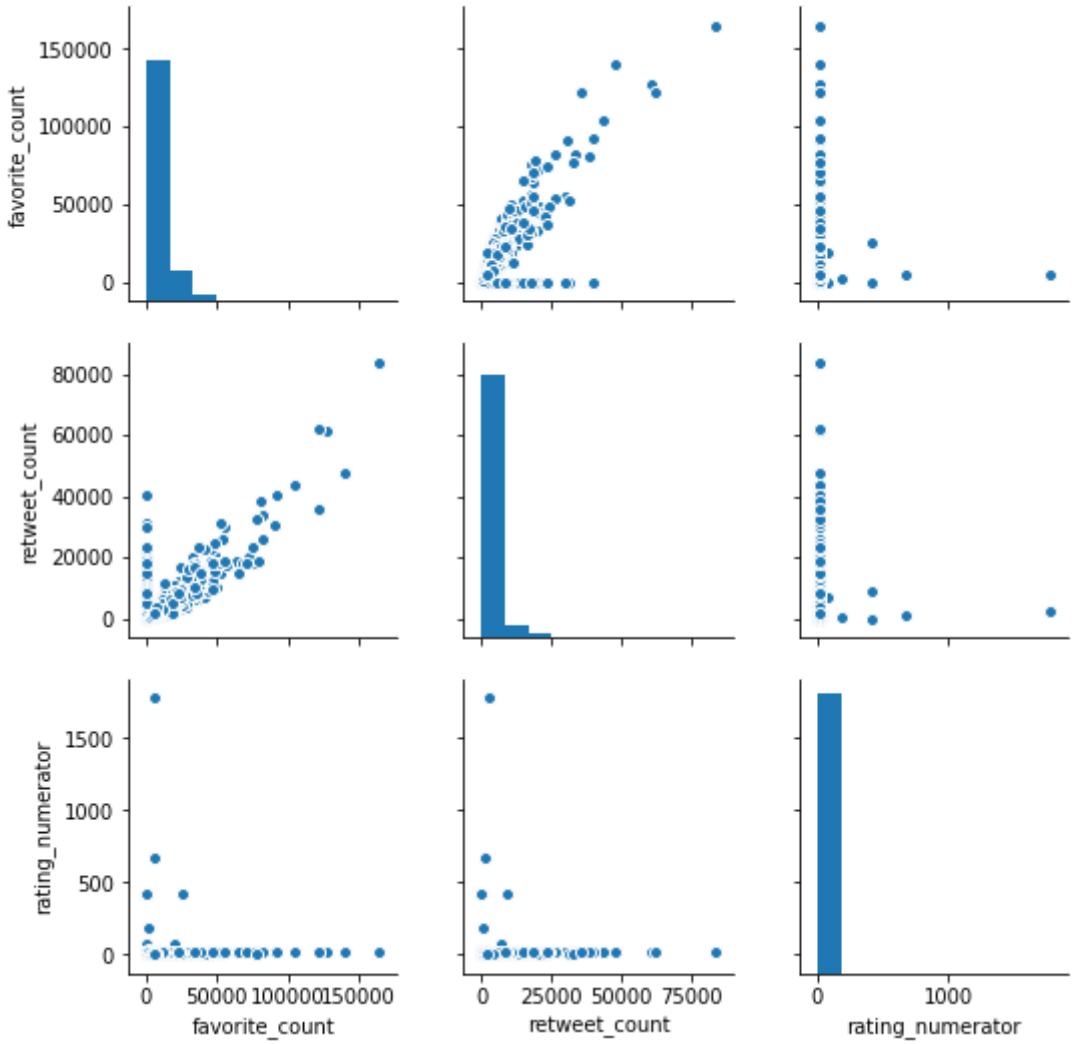
```
df_combined.dog_stage.value_counts()
```

Out[19]:

```
pupper     257
doggo       97
puppo       30
floofer     10
Name: dog_stage, dtype: int64
```

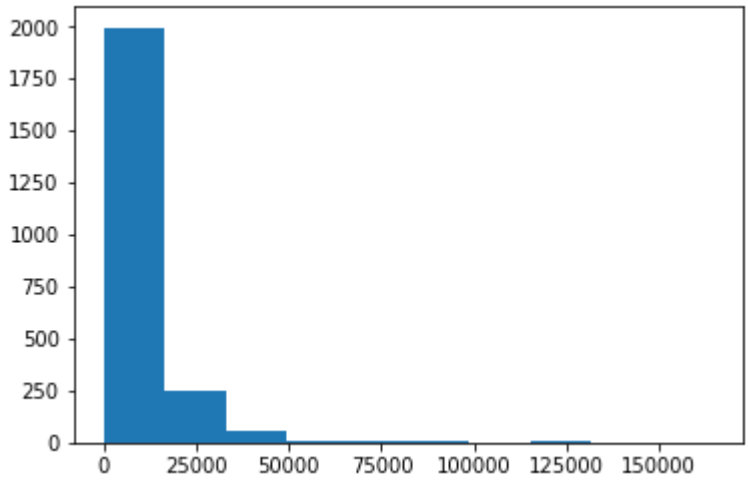
In [21]:

```
sns.pairplot(df_combined[['favorite_count','retweet_count','rating_numerator']]);
```



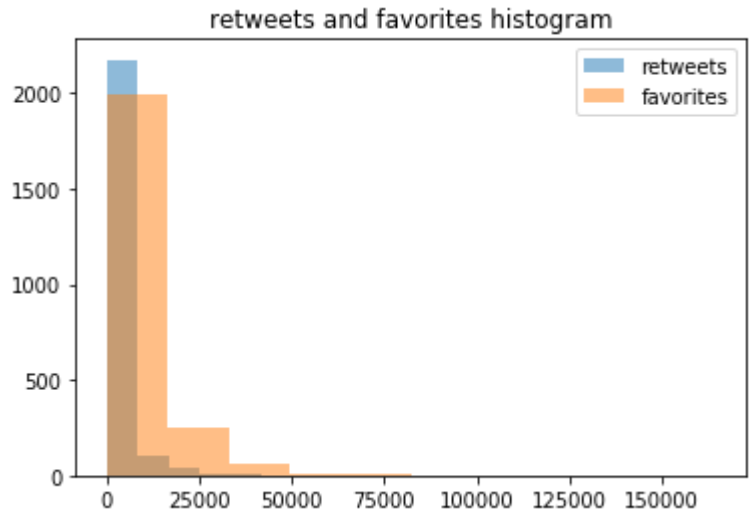
In [7]:

```
plt.hist(df_combined.favorite_count);
```



In [6]:

```
plt.hist(df_combined.retweet_count,alpha = 0.5, label='retweets')
plt.hist(df_combined.favorite_count,alpha = 0.5, label='favorites')
plt.title('retweets and favorites histogram')
plt.legend();
```



In [55]:

```
top_dogs = df_combined.prediction_1.value_counts()[:10]
```

In [56]:

```
df_combined.prediction_1.value_counts()
```

Out[56]:

```
golden_retriever                  152
labrador_retriever                 98
pembroke                           89
chihuahua                          83
pug                                57
chow                               44
samoyed                            43
toy_poodle                         39
pomeranian                         38
malamute                           30
cocker_spaniel                     30
french_bulldog                     26
miniature_pinscher                 23
chesapeake_bay_retriever           23
staffordshire_bullterrier          20
german_shepherd                    20
cardigan                           19
siberian_husky                     19
beagle                             18
shetland_sheepdog                  18
eskimo_dog                         18
rottweiler                         17
maltese_dog                        17
shih-tzu                           17
lakeland_terrier                   17
italian_greyhound                  16
kuvasz                             16
dalmatian                          15
pekinese                           13
american_staffordshire_terrier     13
                                  ...
curly-coated_retriever              4
mexican_hairless                    4
gordon_setter                       4
afghan_hound                        4
keeshond                            4
norwich_terrier                     4
brabancon_griffon                   3
greater_swiss_mountain_dog          3
giant_schnauzer                     3
briard                              3
scottish_deerhound                  3
ibizan_hound                        3
irish_water_spaniel                 3
leonberg                            3
komondor                            3
cairn                               3
welsh_springer_spaniel              3
australian_terrier                  2
black-and-tan_coonhound             2
sussex_spaniel                      2
wire-haired_fox_terrier             2
appenzeller                         2
toy_terrier                         2
standard_schnauzer                  1
silky_terrier                       1
clumber                             1
scotch_terrier                      1
groenendael                         1
```

```
entlebucher                         1
japanese_spaniel                    1
Name: prediction_1, Length: 111, dtype: int64
```

In [97]:

```
df_combined.sort_values(by=['retweet_count'],ascending=False)[:10][['text','prediction_
1']]
```

Out[97]:

	text	prediction_1
928	Here's a doggo realizing you can stand in a po...	labrador_retriever
959	Here's a doggo blowing bubbles. It's downright...	eskimo_dog
482	This is Stephan. He just wants to help. 13/10 ...	chihuahua
2327	Here's a super supportive puppo participating ...	lakeland_terrier
60	This is Duddles. He did an attempt. 13/10 some...	english_springer
407	This is Bo. He was a very good First Doggo. 14...	standard_poodle
404	RT @dog_rates: This is Bo. He was a very good ...	standard_poodle
485	"Good afternoon class today we're going to lea...	NaN
2070	This is Jamesy. He gives a kiss to every other...	french_bulldog
1511	This made my day. 12/10 please enjoy https://t...	NaN

In [89]:

```
df_combined['retweet_count'].median()
```

Out[89]:

```
1385.0
```