From the visualizations and analyses shown above, there are a few observations and insights that stand out. First, the top three dogs that appear the most in WeRateDogs twitter posts are: Golden Retrievers, Pembrokes (Corgis), and Labrador Retrievers.

```
#finding top 10 dog appears most
dognames = df['p1'].value_counts()
dognames.head(10)
```

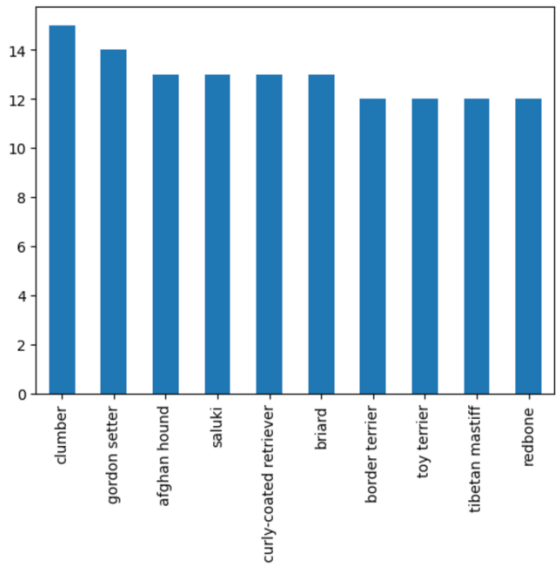
```
golden retriever      99
pembroke              61
labrador retriever    53
pug                   31
chihuahua             29
pomeranian            25
french bulldog        18
samoyed               18
malamute              17
toy poodle            17
Name: p1, dtype: int64
```

When I looked at the highest average ratings, at first, I was surprised to see the dogs who appear the most were not rated the highest, the top three dogs with the highest average ratings were Clumber (spaniel), Gordon Setter, and Afghan Hound. But realized the average rating probably was not the best statistic to explain the dataset.

```
#seeing which dog gets the highest average rating
groupby_rating = df.groupby('p1')['numerator'].mean().sort_values(ascending=False).head(10)
groupby_rating
```

```
p1
clumber                   15.0
gordon setter             14.0
afghan hound              13.0
saluki                    13.0
curly-coated retriever    13.0
briard                    13.0
border terrier            12.0
toy terrier               12.0
tibetan mastiff           12.0
redbone                   12.0
Name: numerator, dtype: float64
```

```
#top ten dogs with the highest average rating
df.groupby('p1')['numerator'].mean().sort_values(ascending=False).head(10).plot(kind='bar', x='p1', y='numerator');
```

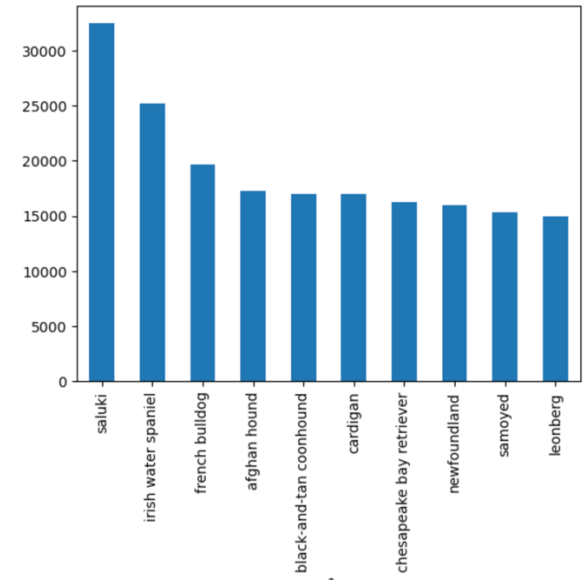


I also looked at highest average retweets and favorites. This would take how many times a dog appears in a tweet and how many favorites or retweets it gets on average. The highest for average favorites was a Saluki, then an Irish Water Spaniel, and finally a French Bulldog.

```
#top ten dogs with the highest average favorites
groupby_favorites = df.groupby('p1')['favorite_count'].mean().sort_values(ascending=False).head(10)
groupby_favorites
```

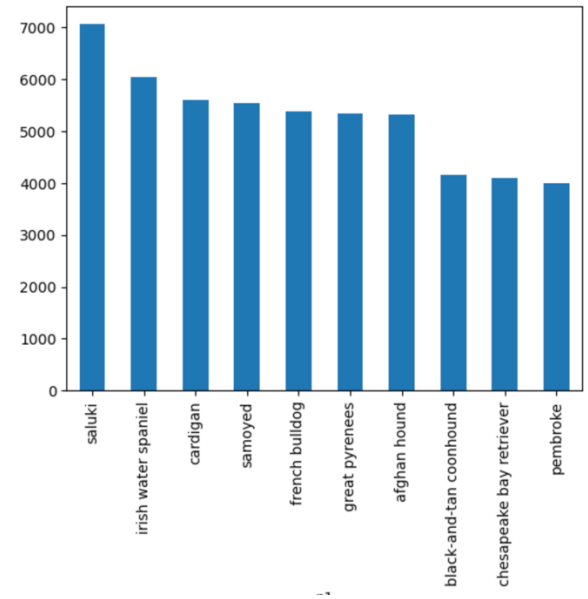
```
p1
saluki                        32444.500000
irish water spaniel           25225.000000
french bulldog                19619.833333
afghan hound                  17305.000000
black-and-tan coonhound       17012.000000
cardigan                      16993.300000
chesapeake bay retriever      16286.888889
newfoundland                  16014.500000
samoyed                       15335.222222
leonberg                      14934.333333
Name: favorite_count, dtype: float64
```

```
#visual for top ten dogs with the highest average favorite count
df.groupby('p1')['favorite_count'].mean().sort_values(ascending=False).head(10).plot(kind='bar', x='p1', y='favorite_cc
```



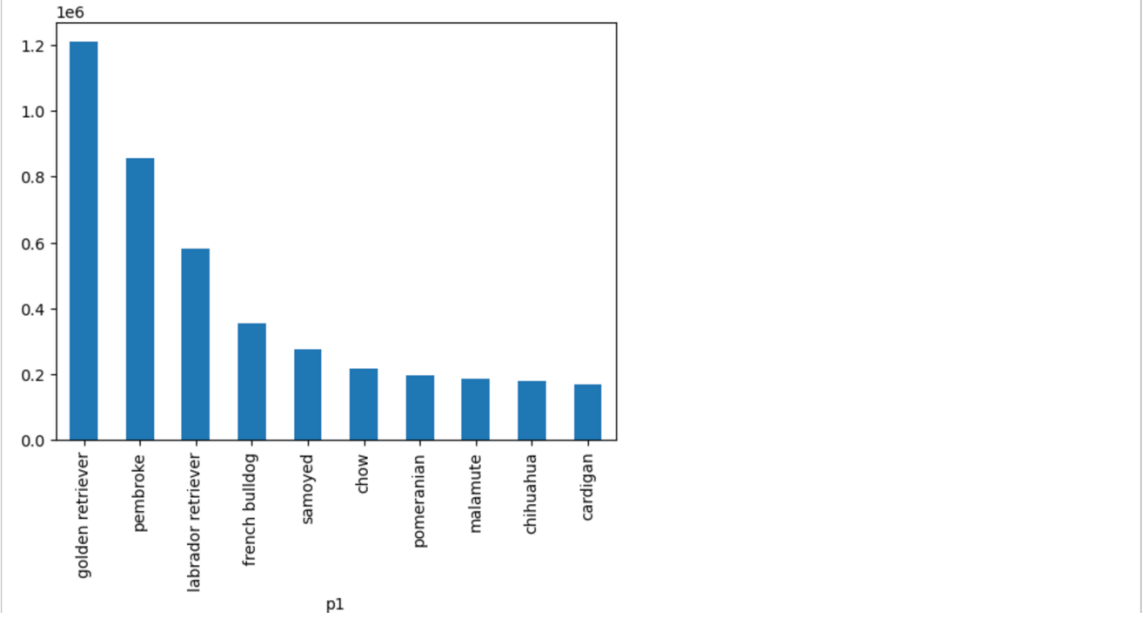
And the highest for retweets was just a little different, with Saluki still at the top, then Cardigan (Corgi), then the Irish Water Spaniel.

```
#visual for top ten dogs with highest average retweet count
df.groupby('p1')['retweet_count'].mean().sort_values(ascending=False).head(10).plot(kind='bar', x='p1', y='retweet_cour
```

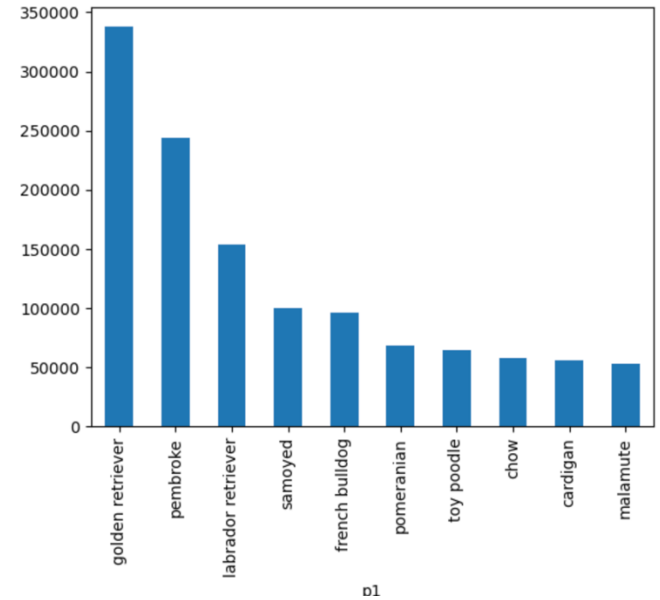


When I looked at the sum of retweets and favorites, it was no surprise that the dogs who were in the most number of tweets were the dogs with the most amount of favorites. We can see clearly that the three dogs mentioned earlier, Golden Retriever, Pembroke, and Labrador Retrievers had the highest number of favorites and retweets.

```
#now looking at summed highest favorite count
df.groupby('p1')['favorite_count'].sum().sort_values(ascending=False).head(10).plot(kind='bar', x='p1', y='favorite_cou
```

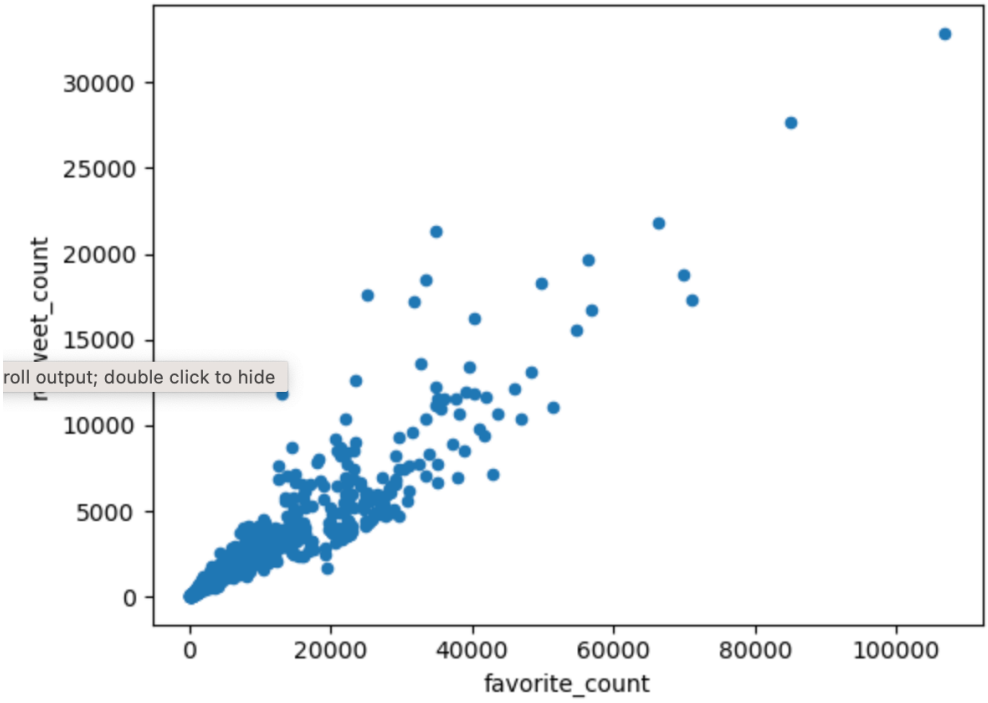


```
#summed highest retweet count
df.groupby('p1')['retweet_count'].sum().sort_values(ascending=False).head(10).plot(kind='bar', x='p1', y='retweet_count
```



Next, I looked at the relationship between favorite and retweet count. I found a positive relationship. As retweet count goes up, favorite count goes up.

```
df.plot(kind='scatter', x='favorite_count', y='retweet_count');
```

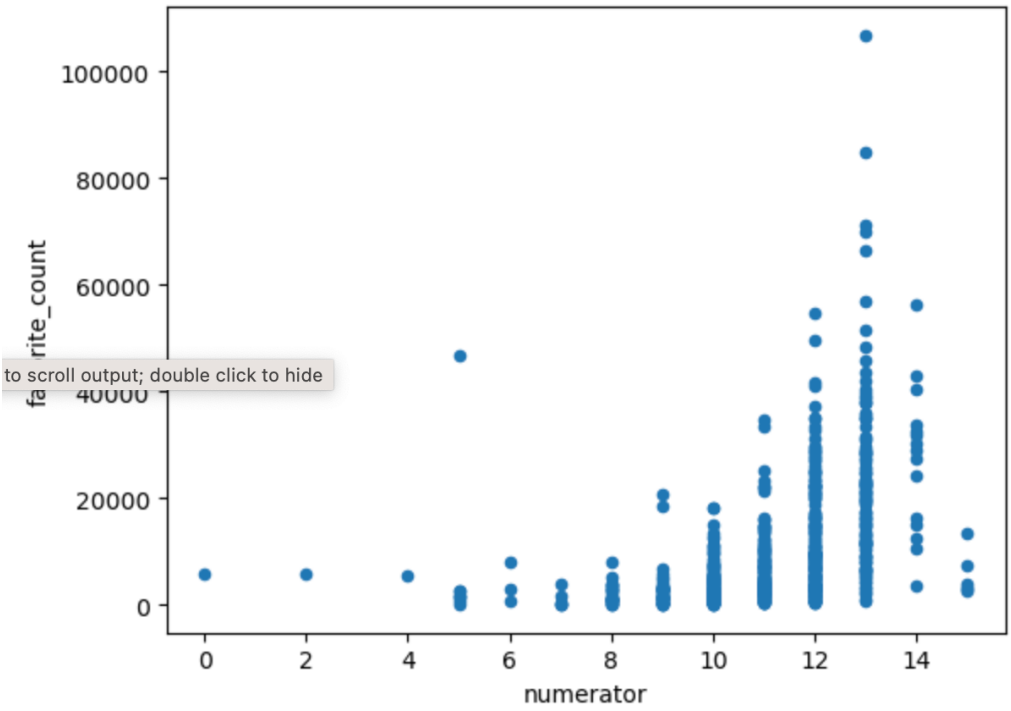


roll output; double click to hide

Finally, I looked at the relationship between rating (numerator only), and favorite count. I was curious to see whether the favorite count increased when the numerator increased. I found that it does.

```
df.plot(kind='scatter', y='favorite_count', x='numerator')
```

```
<Axes: xlabel='numerator', ylabel='favorite_count'>
```



Without cleaning the dataset in its entirety, these insights are tentative. It is important to note that there are many dog breeds which are very similar such as a Pembroke and Cardigan Corgi. In the analysis, these two were separated, but together may have resulted in more favorites, retweets, etc.