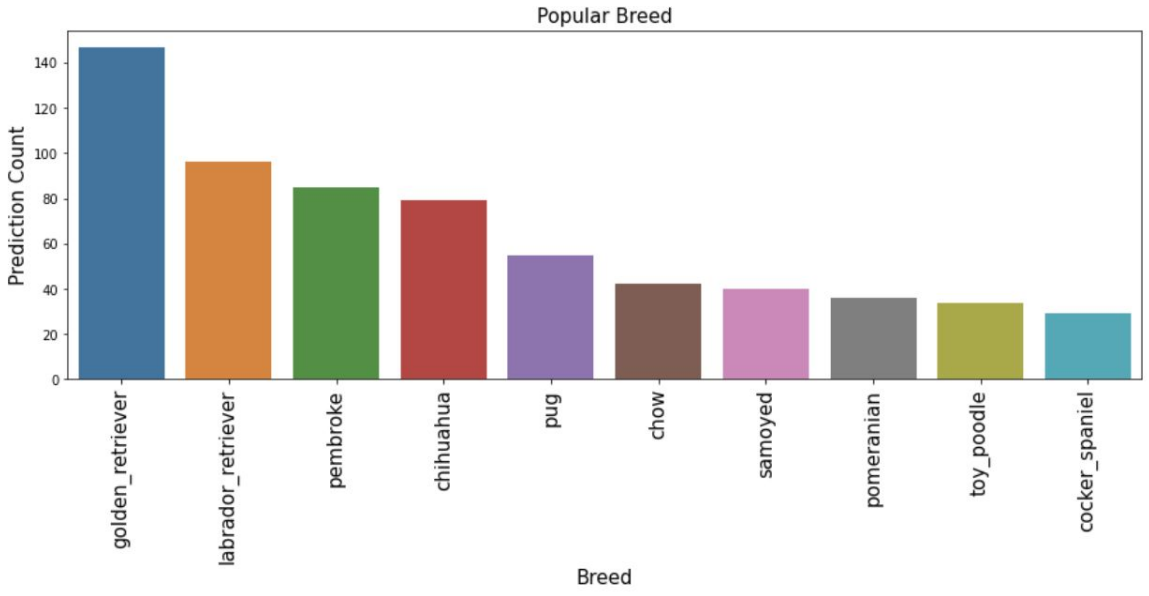
Insight

Finally, we're done with gathering, assessing, cleaning and analyzing using the cleaning WeRateDogs twitter data. People rate other people's dogs with comments on twitter. I discovered why the numerators through is almost greater than 10. It's because everyone thinks that dogs deserve at least 10 or more than that.

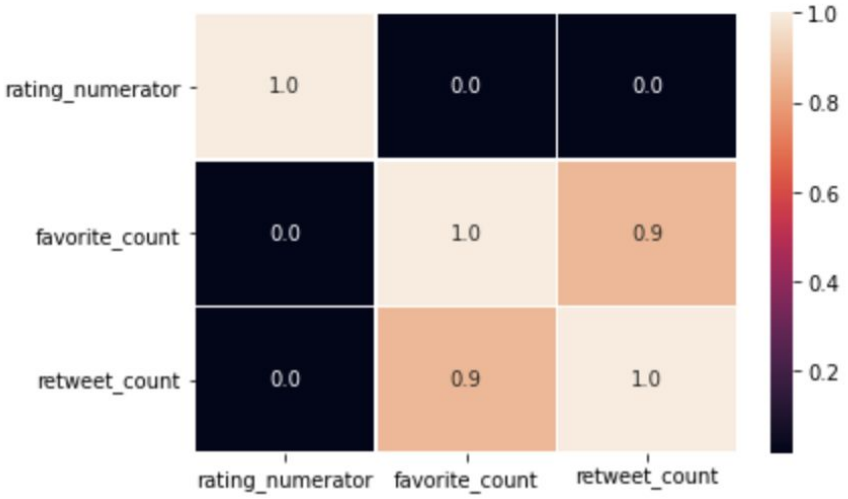
By using this cleaned dataset, how can we use this data? I was thinking we can get some information about popular breeds of dog, the relationship between retweets, favorites and ratings. We can offer the visualized data using python packages and libraries. They allow me to analyze the data more easily and effectively at the same time.

- What's the most popular dog breeds? : As you can see, the golden retriever are the most popular breed in this data.

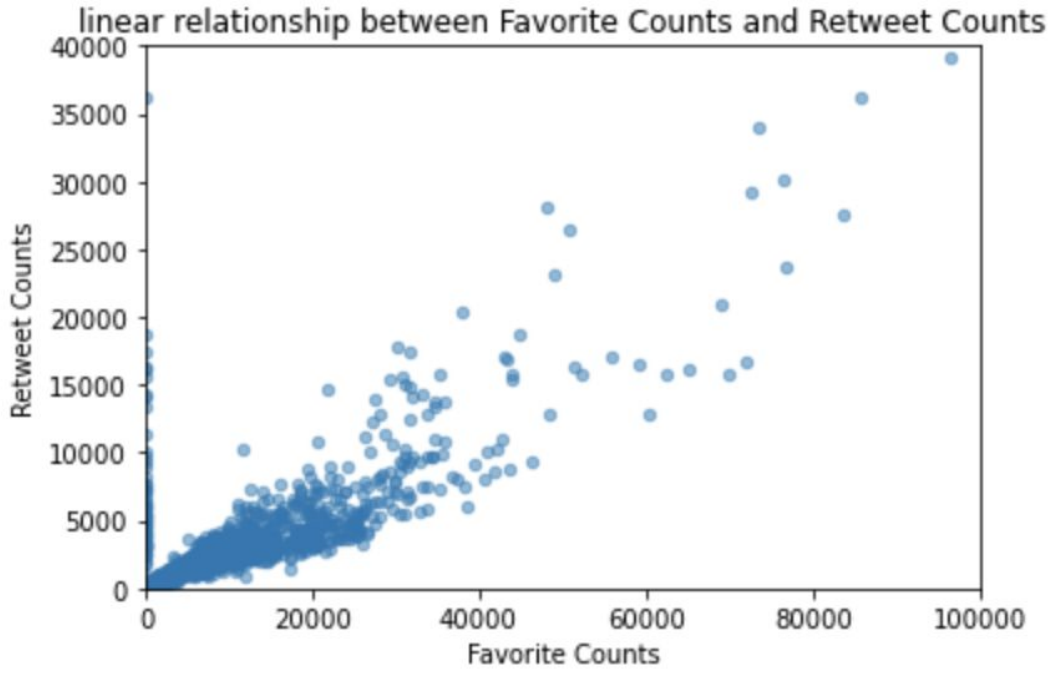


- What's the correlation between Retweet Count, Favorite Count, Ratings : You can see the strong positive correlation between Favorites and retweets.

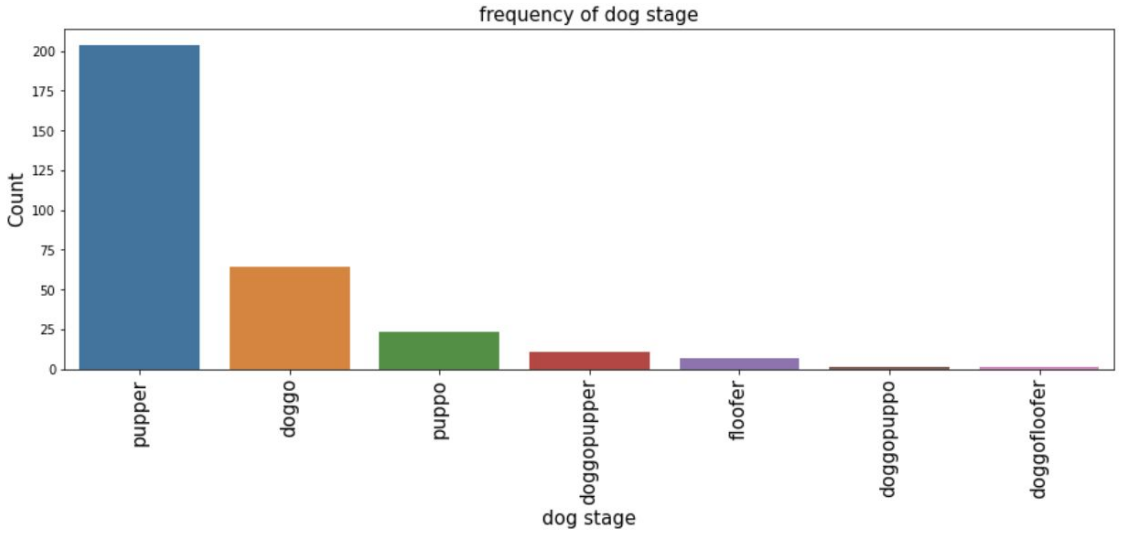
```
<matplotlib.axes._subplots.AxesSubplot at 0x7ff7d19f9160>
```



- What's the relationship between favorite counts and retweet counts : Yet we know about the clear relationship in the heat map. By doing that I discovered the linear relationship between favorite counts and retweet counts.



- Which dog stage appeared the most? The pupper are the most appeared in the data



- Where did you get this dog rate? I used the user definition function by tracking the path getting this data. As a result, ‘Twitter for iphone’ has appeared the most.