

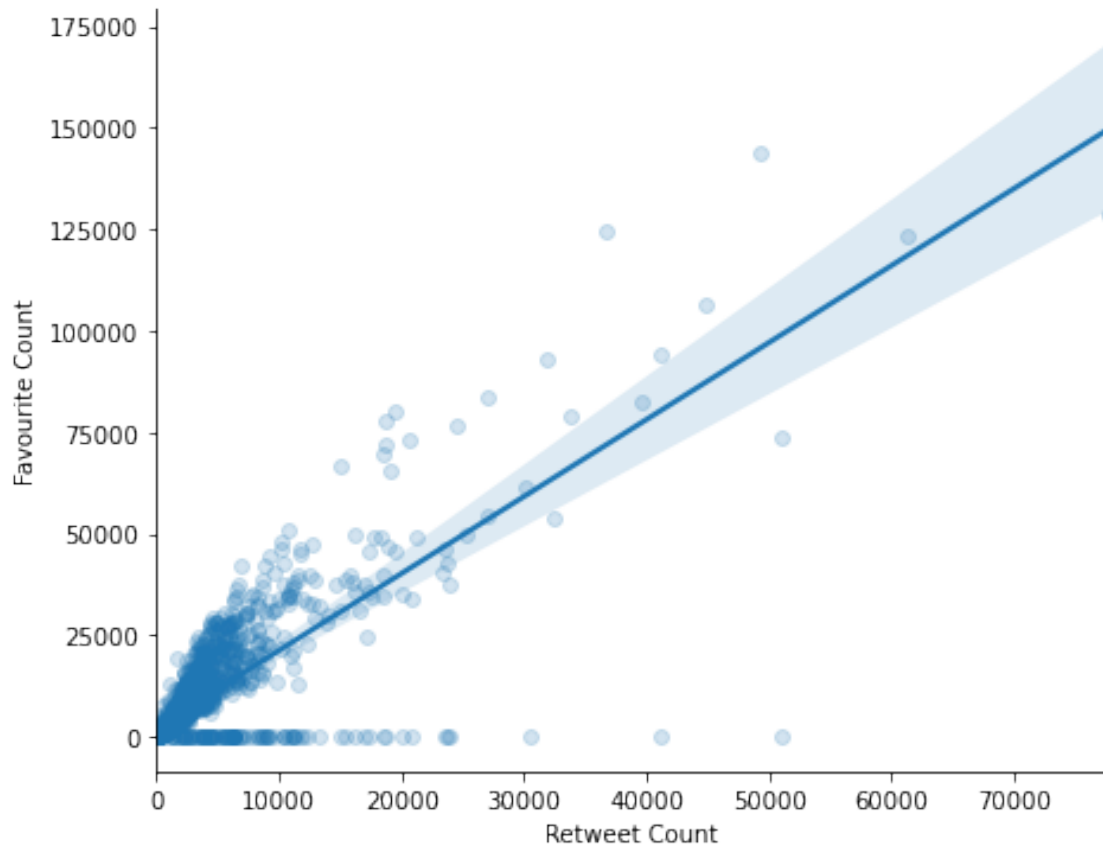
# wrangle\_act

December 18, 2020

## 0.1 Analyzing and Visualizing Data:

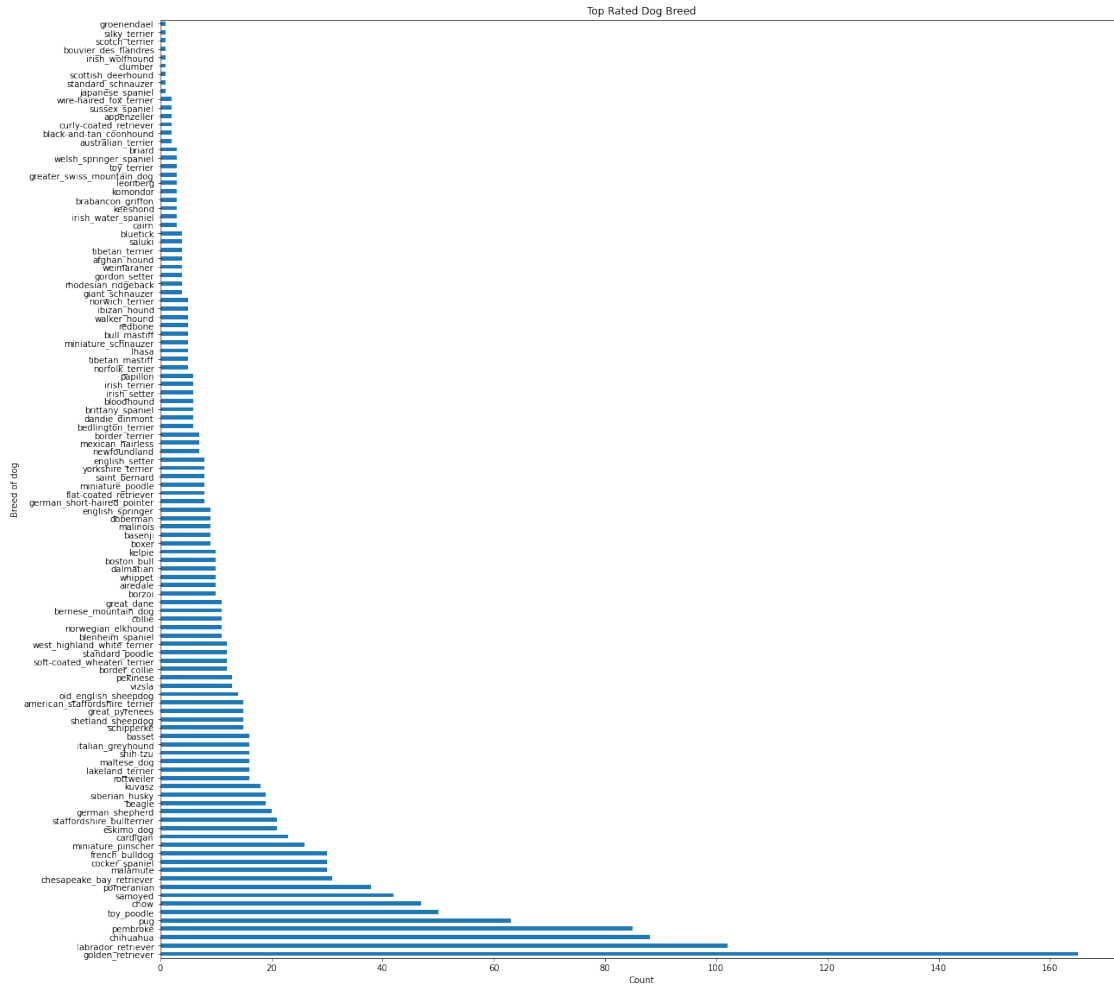
```
[374]: sns.lmplot(x="retweet_count",  
                y="favorite_count",  
                data=df,  
                size = 5,  
                aspect=1.3,  
                scatter_kws={'alpha':1/5})  
  
plt.xlabel('Retweet Count')  
plt.ylabel('Favourite Count');
```

```
C:\Users\Jabal\anaconda3\lib\site-packages\seaborn\regression.py:580:  
UserWarning: The `size` parameter has been renamed to `height`; please update  
your code.  
warnings.warn(msg, UserWarning)
```



- Favourite and retweet counts are positively correlated.
- The majority of the data falls below 40000 favorites and 10000 retweets.

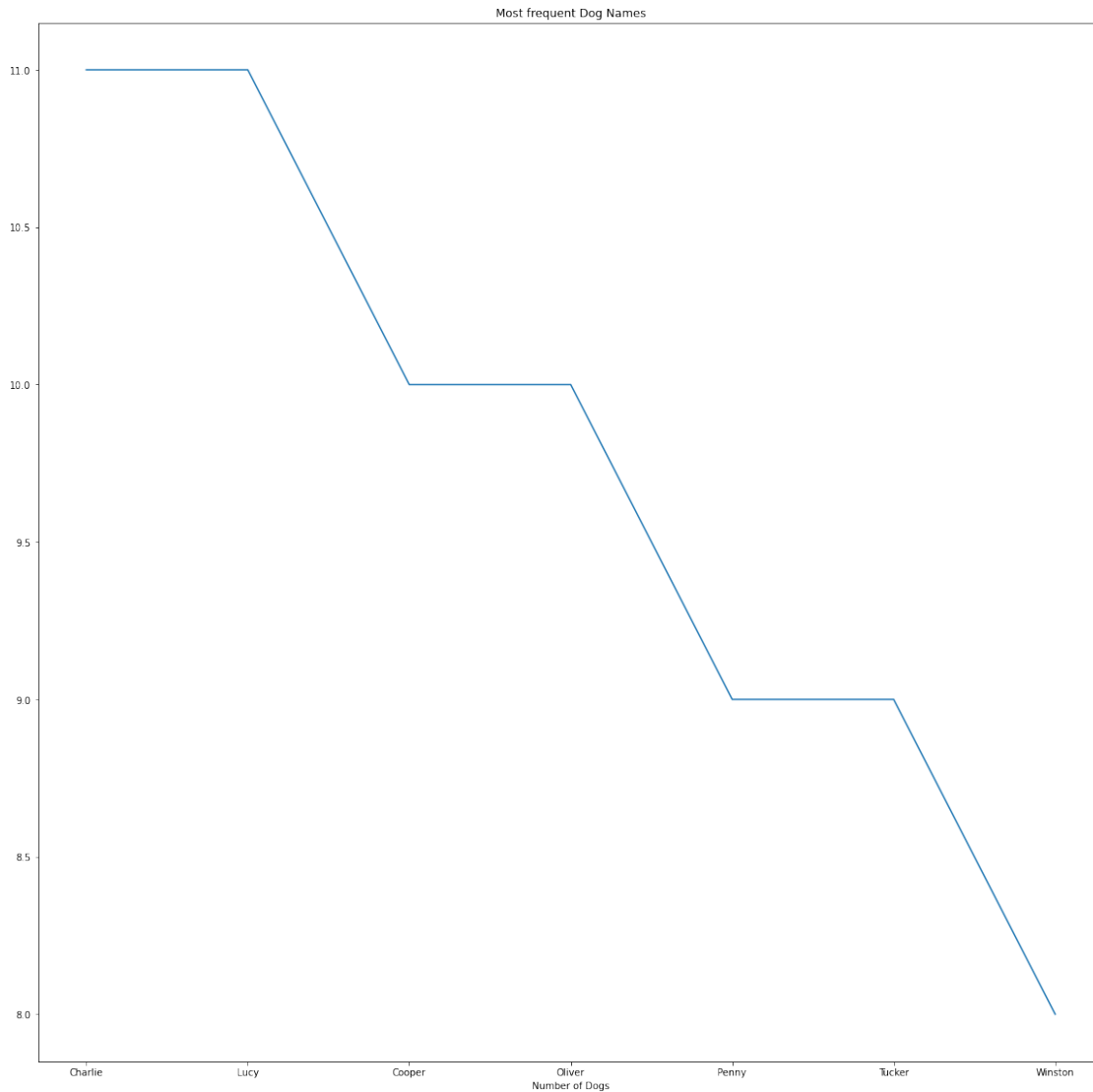
```
[375]: df['breed_pred'].value_counts().plot(kind = 'barh')
plt.title('Top Rated Dog Breed')
plt.xlabel('Count')
plt.ylabel('Breed of dog');
```



Visualisation above shows the most frequent dog breed in the dataset, which is golden retriever.

[ ]:

```
[376]: df.name.value_counts()[1:8].plot(title='Most frequent Dog Names').
        ↪set_xlabel("Number of Dogs");
```



Most frequent names are shown, Charlie, Lucy and Cooper.

### 0.1.1 Storing Wrangled Data

```
[377]: df.to_csv('twitter_archive_master.csv')
```

```
[ ]:
```