

# act\_report\_final

September 12, 2022

## 1 Report: Wrangling and Analyze Data

### 1.0.1 from twitter WeReDogs dataset

### 1.0.2 Insights:

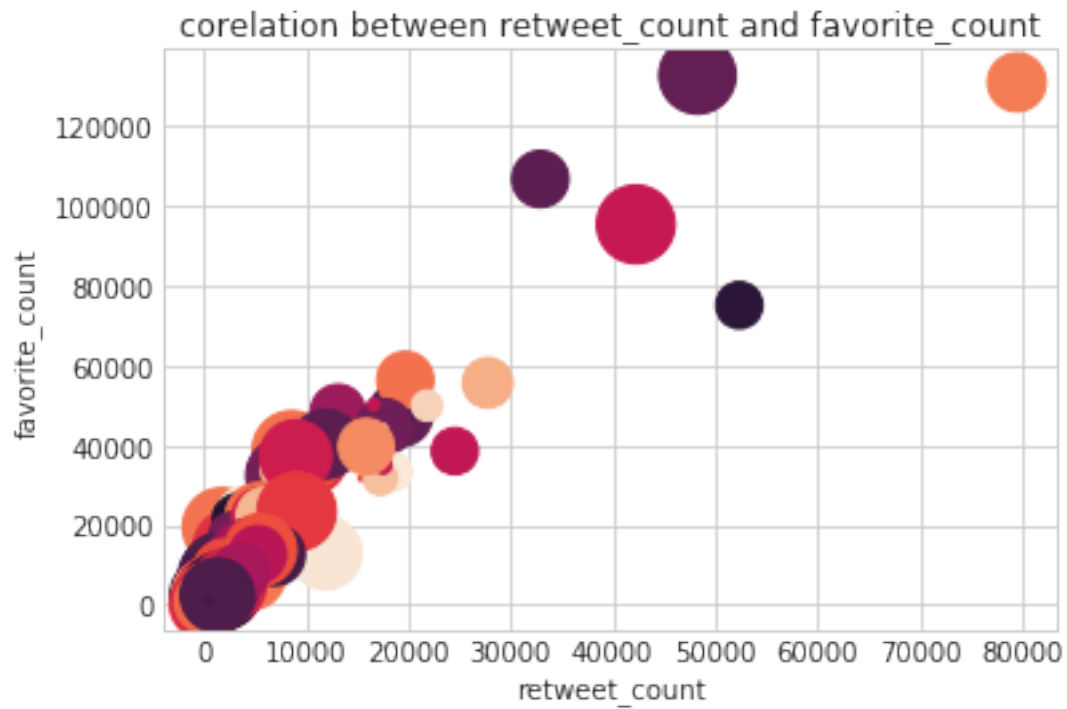
1. I was able to find that there a correlation between retweet\_counts and favorite\_counts
2. I noted that dog\_stage does not matter in giving higher or lower ratings

### 1.0.3 visualizations

```
In [47]: # for reproducibility
         np.random.seed(19680801)

         x = twitter_archive_master.retweet_count
         y = twitter_archive_master.favorite_count
         size = twitter_archive_master.shape[0]

         # adding colors
         colors = np.random.rand(size)
         area = (30 * np.random.rand(size))**2 # 0 to 15 point radii
         plt.scatter(x, y, s=area, c=colors, alpha=1.5)
         plt.title('corelation between retweet_count and favorite_count')
         plt.xlabel('retweet_count')
         plt.ylabel('favorite_count')
         plt.show()
```

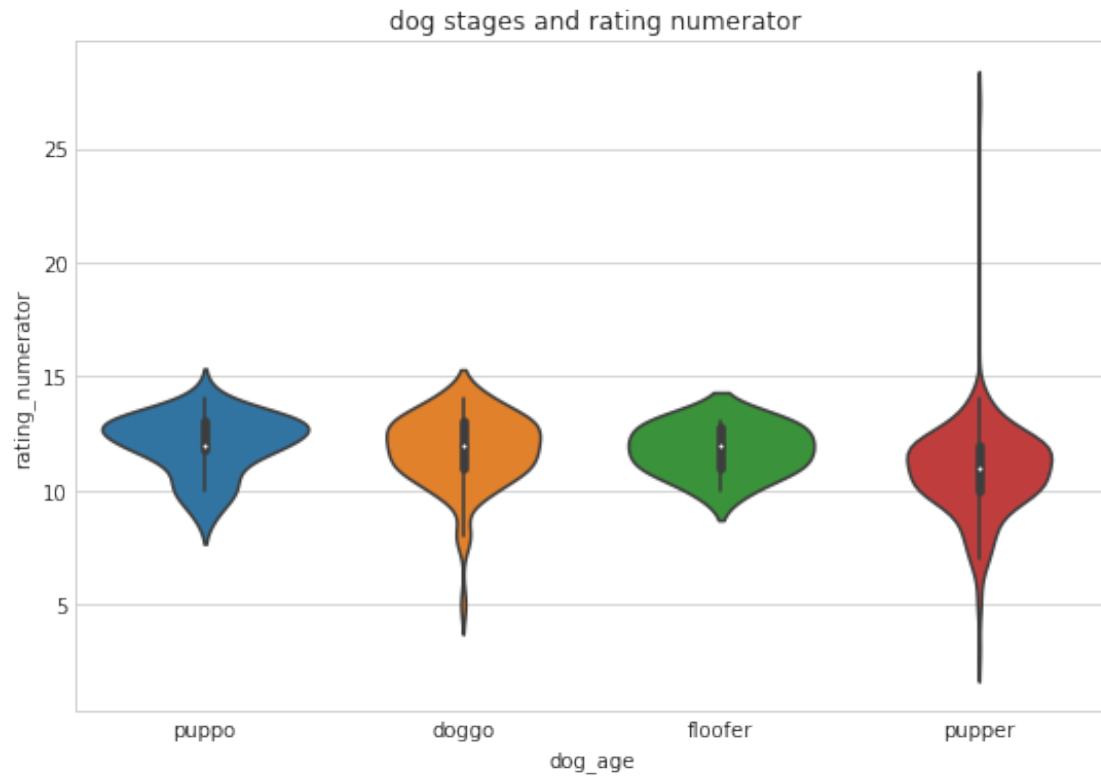


## 2. does dog\_stage matter in giving higher or lower ratings?

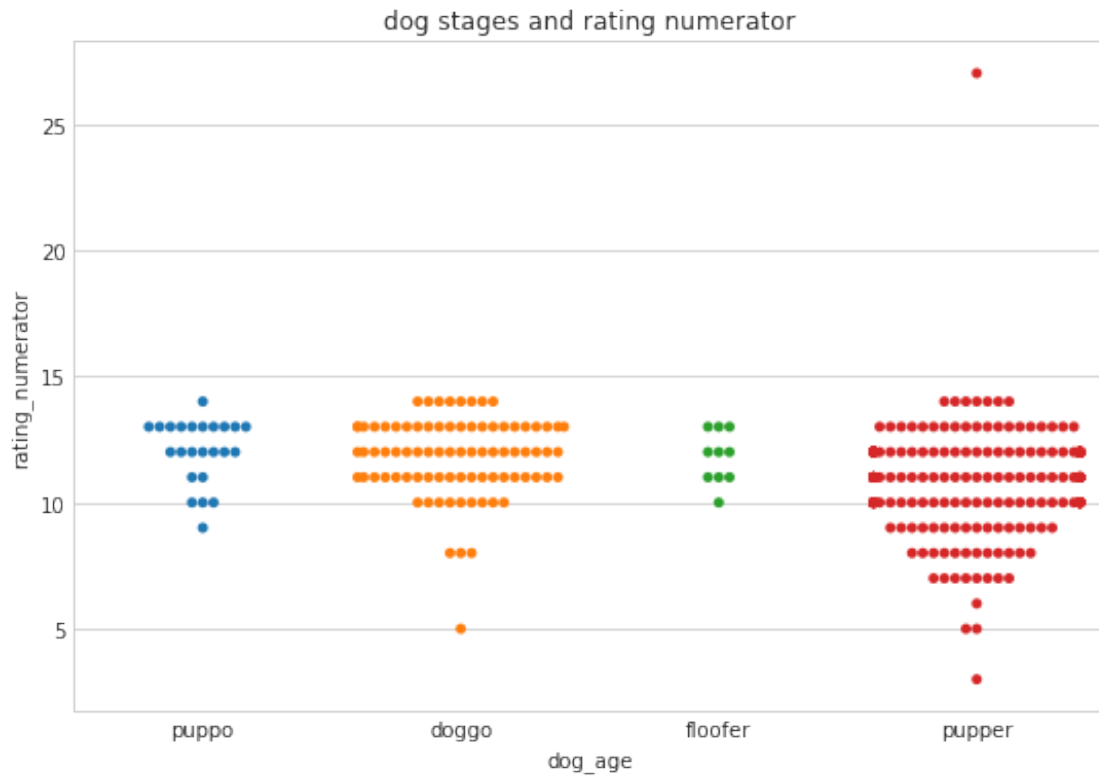
```
In [131]: ratings = twitter_archive_master[['rating_numerator', 'dog_age']]

# Set up the matplotlib figure
f, ax = plt.subplots(figsize=(9, 6))

t = ' dog stages and rating numerator'
# Draw a violinplot with a narrower bandwidth than the default
sns.violinplot(data=ratings, x = 'dog_age', y = 'rating_numerator').set_title(t);
```

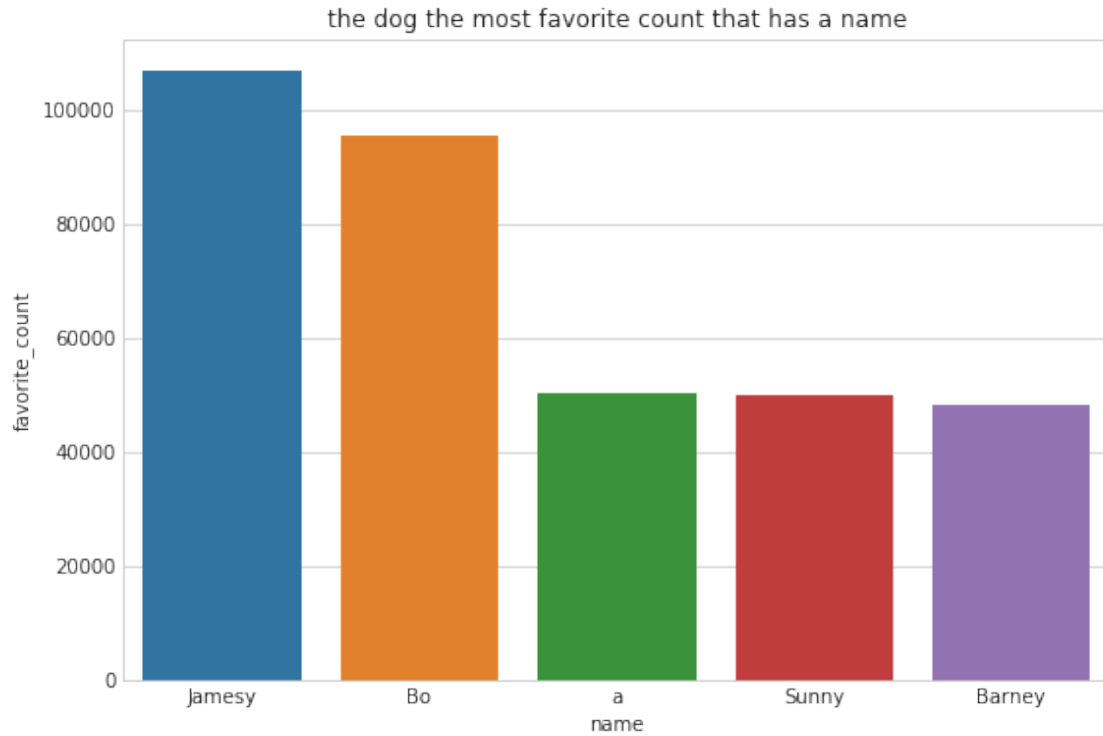


```
In [130]: # using swarmplot
f, ax = plt.subplots(figsize=(9, 6))
title_ = 'dog stages and rating numerator'
sns.swarmplot(data=ratings, x='dog_age', y='rating_numerator').set_title(title_);
```



3. which dog name has the most popular favorite\_counts?

```
In [128]: #plotting a
f, ax = plt.subplots(figsize=(9, 6))
sns.barplot(x=popular.name, y=popular.favorite_count ).set_title('the dog the most fav
```



#### 1.0.4 Report conclusions

- we can note that there is a positive correlation between favorite\_count and retweet\_count. With all other factors that are likely to affect retweet count set aside, we can safely say that a picture of a dog liked as favorite is most likely to be retweeted.
- Apart from other dogs that had the most favorite counts but didn't have names we can see that Jamesy was the dog with a name with the most favorite counts.
- although we can see much outlier rating in pupper dog age, we can safely say there is not much of a relationship between the dog's age and its rating