

act_report

July 23, 2020

0.1 Analysis

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('twitter_archive_master.csv')
```

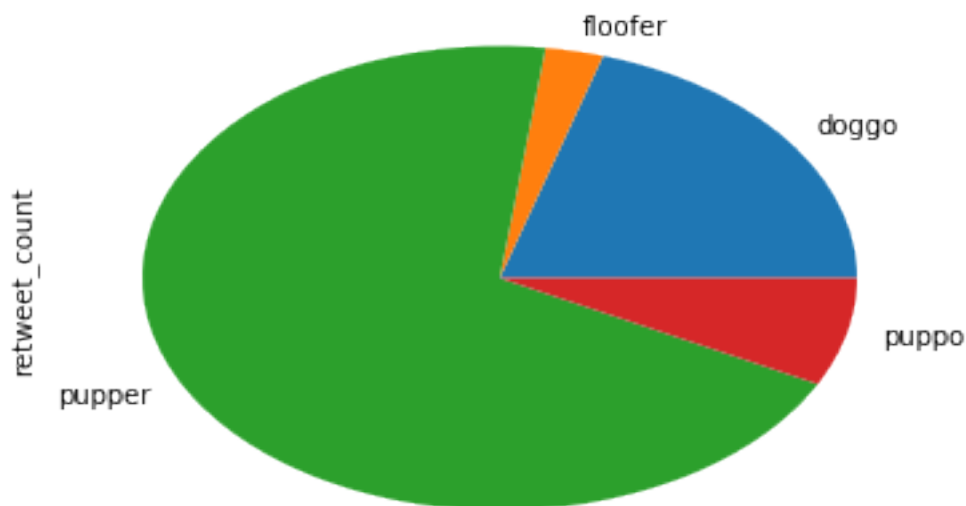
What is the most retweeted dog phase? Many people love dogs but it is really interesting to see what kind of dogs do people love the most. People like to share dogs photos that they like the most. After carrying out an analysis we found out that pupper are the ones that are most retweeted dogs. Let us look at the below analysis and chart to see it.

```
In [2]: df.groupby('dog_phase').retweet_count.count()
```

```
Out[2]: dog_phase
doggo      62
floofer     8
pupper    212
puppo      23
Name: retweet_count, dtype: int64
```

```
In [3]: df.groupby('dog_phase').retweet_count.count().plot(kind = 'pie')
plt.plot(figsize = [10,3])
plt.title('Dog phases and their retweet count', fontsize = 18);
```

Dog phases and their retweet count

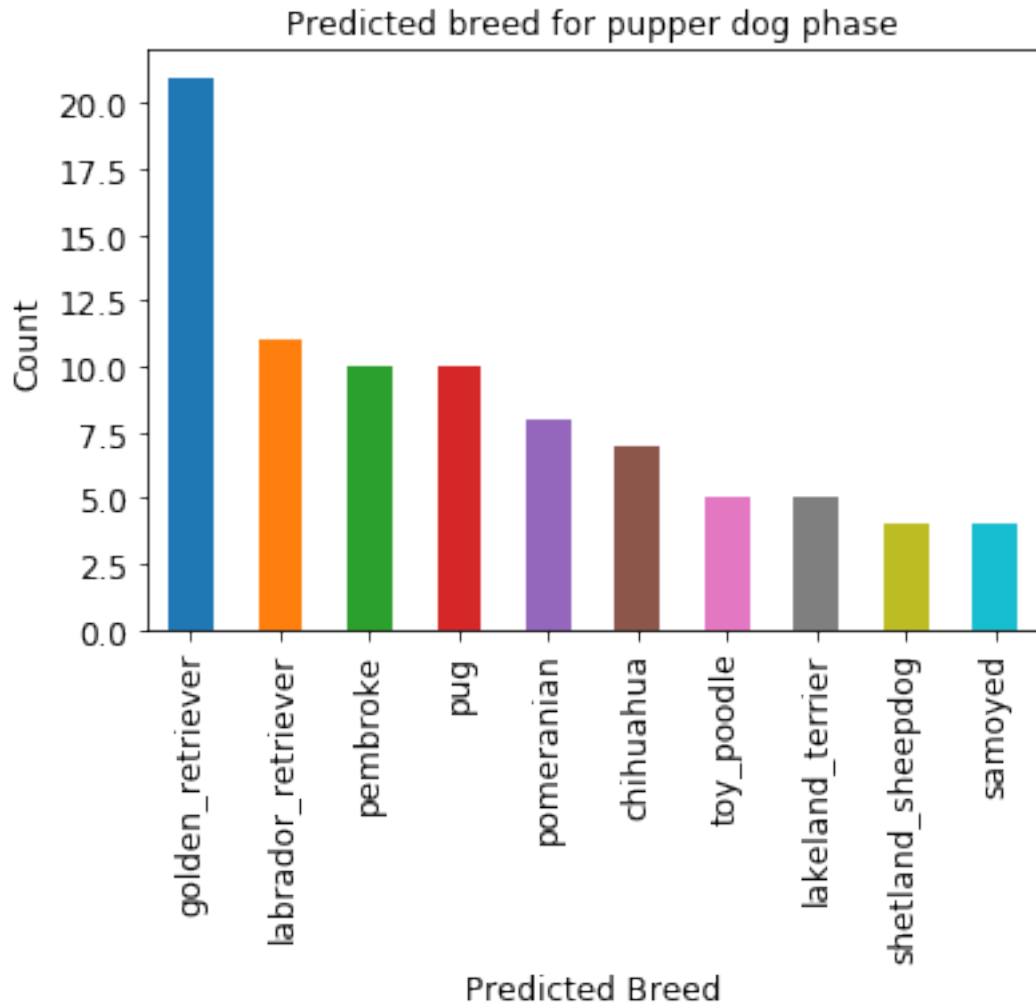


What are the 10 most predicted dog breeds in dog phase - pupper We already found out that pupper was the most retweeted dog phase that people loved to share. There are many dog breeds. It is interesting to see the predicted dog breed that was in pupper dog phase. Through the analysis and visualization carried out below we found that golden retriever was the most predicted dog breed in the pupper dog phase.

```
In [4]: pupper = df.query("dog_phase == 'pupper'")['predicted_breed'].sort_values(ascending = False)
pupper
```

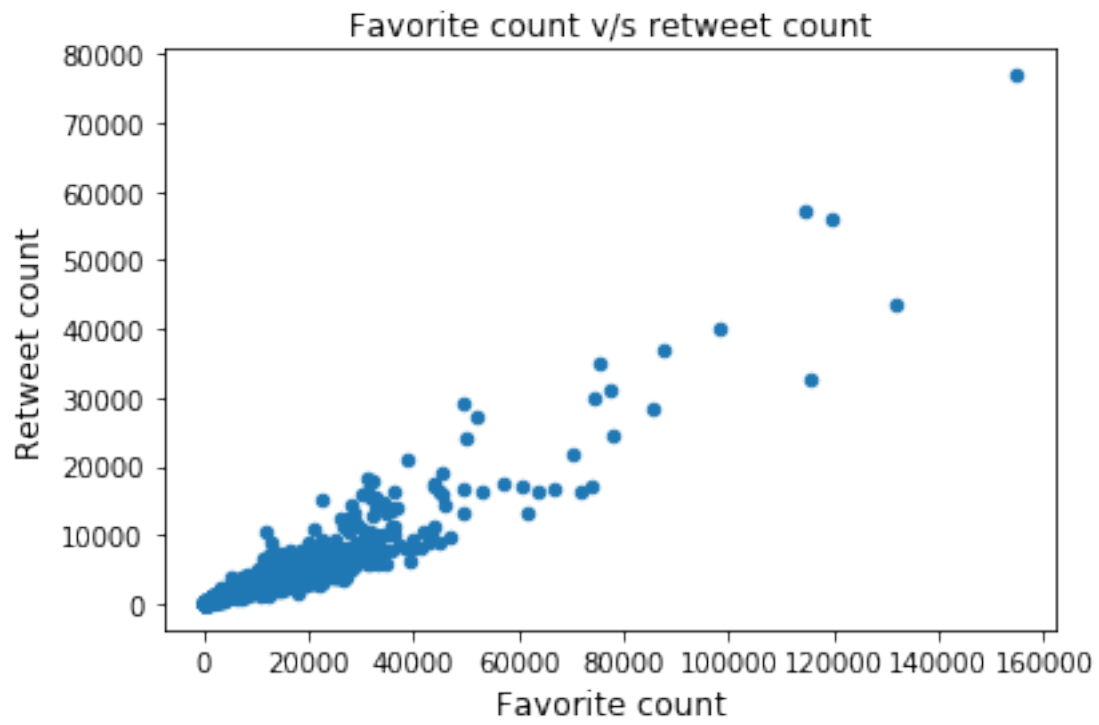
```
Out[4]: golden_retriever      21
labrador_retriever          11
pembroke                    10
pug                         10
pomeranian                   8
chihuahua                    7
toy_poodle                   5
lakeland_terrier             5
shetland_sheepdog            4
samoyed                      4
Name: predicted_breed, dtype: int64
```

```
In [5]: pupper.plot(x= 'predicted_breed', y = 'count', kind = 'bar', title = 'Predicted breed for
plt.xlabel('Predicted Breed', fontsize = 12)
plt.ylabel('Count', fontsize = 12);
```



What is the relationship between favorite count and retweet count? People who love dogs mark dog pictures as favorites and also retweet them since they probably like to share it with others. That is exactly what even the visualization shows. There is a positive correlation between favorite count and retweet count.

```
In [6]: df.plot(x = 'favorite_count', y = 'retweet_count', kind = 'scatter', title = 'Favorite c
plt.xlabel('Favorite count', fontsize = 12)
plt.ylabel('Retweet count', fontsize = 12);
```



In []: