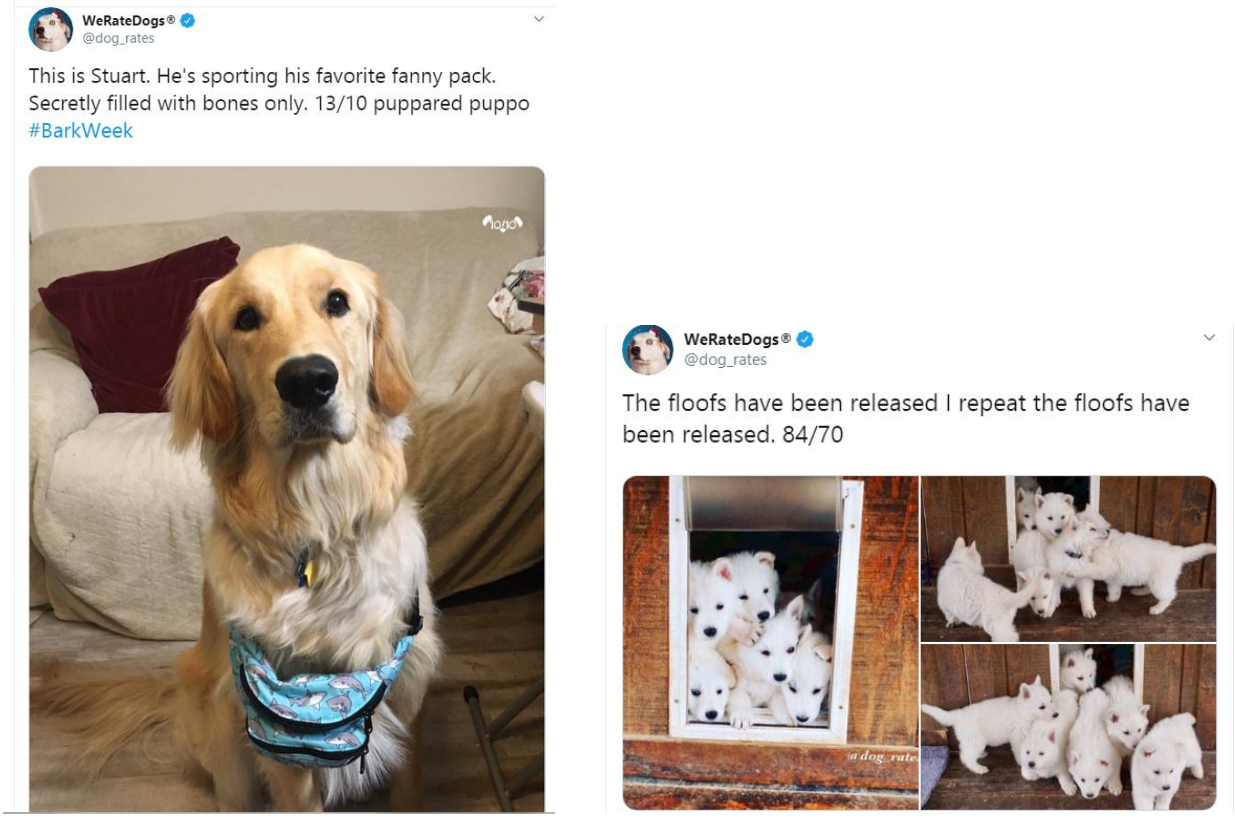


Examples of WeRateDogs Tweets

After gather and clean the WeRateDogs Twitter data, I have now obtained a well structured DataFrame containing basic tweet data for 1870 records of their original tweets with dog pictures. The information in this "twitter_archive_master" table includes:

"tweet_id": Tweet IDS

"timestamp": Time the Tweet has been published. From activation Day 1 of this Twitter account to August 1st, 2017

"text": Tweet content

"rating_numerator": Rating numerator for dog in the picture, normally larger than score 10

"rating_denominator": Rating denominator for dog in the picture. Normally on a scale of score 10, larger score scales are due to multiple dogs present in the picture

"name": The name of dog in pictures. Some not provided.

"stage": The stage of dog in pictures. Some not provided.

"retweet_count": Numbers of retweets

"favorite_count": Numbers of being liked by clicks

"predicted _dog _breed": The most confident prediction of dog breed for the dog in picture using AI classifier

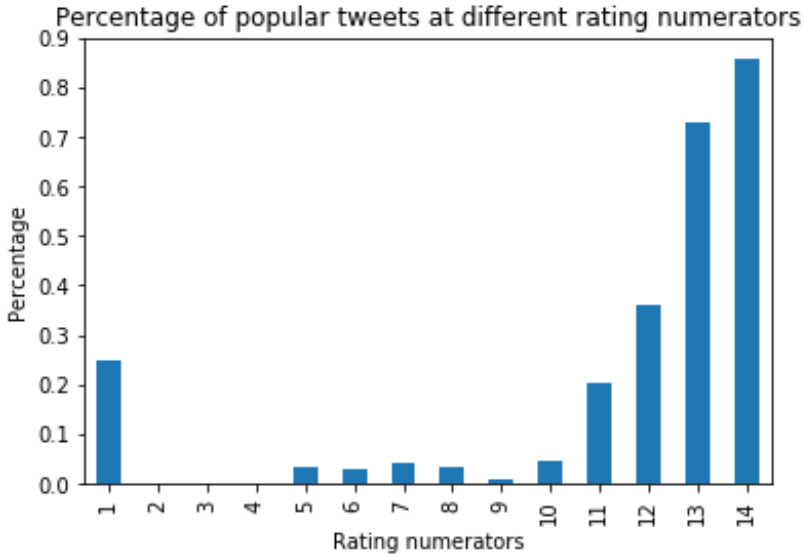
I have successfully produced 4 insights by analyzing the above Tweets data:

1) The most common dog names identified in this Twitter archive are Oliver and Cooper.

```
twitter_archive_master.name.value_counts().head()
```

```
Not given    569
Cooper        10
Oliver        10
Lucy           9
Tucker         9
Name: name, dtype: int64
```

2) I have regarded Tweets with retweet and favorite counts more than 75th percentile to be popular Tweets. By analyzing their dog ratings, I have found that original tweets with higher rating numerators are more likely to be popular tweets, which makes sense. Visualization is provided for this analysis.



3) There are more than 300 predicted dog breeds in this Twitter archive and the most common type is golden retriever! This insight confirms this dog breed is widely beloved.

```
twitter_archive_master.predicted_dog_breed.value_counts().head(10)
```

```
golden_retriever      134
Labrador_retriever     86
Pembroke               83
Chihuahua              77
pug                    51
toy_poodle             37
Pomeranian             37
chow                   37
Samoyed                36
malamute               28
Name: predicted_dog_breed, dtype: int64
```

```
twitter_archive_master.predicted_dog_breed.nunique()
```

```
363
```

4) WeRateDogs Twitter account is getting more and more popular, as it's shown in calculation of my "wangle_act.ipynb" file that the percentage of tweets receiving high engagement of retweets and favorite counts is increasing each year!

```
#Find out the number of records in each year in the twitter archive

Archive_Years=list()
Archive_Counts=dict()
for i in twitter_archive_master.index:
    year=twitter_archive_master.timestamp[i].year
    Archive_Years.append(year)
for year in Archive_Years:
    Archive_Counts[year]=Archive_Counts.get(year,0)+1
Archive_Counts
```

```
{2017: 327, 2016: 918, 2015: 646}
```

```
#Find out the number of records in each year in the pupolar Tweets

Popular_Years=list()
Popular_Counts=dict()
for i in twitter_popular.index:
    year=twitter_popular.timestamp[i].year
    Popular_Years.append(year)
for year in Popular_Years:
    Popular_Counts[year]=Popular_Counts.get(year,0)+1
Popular_Counts
```

```
{2017: 281, 2016: 167, 2015: 25}
```

```
for year in [2015,2016,2017]:
    print ("The percentage of popular tweets at year %d is: %f" % (year, Popular_Counts[year]/Archive_Counts[year]))
```

```
The percentage of popular tweets at year 2015 is: 0.038700
The percentage of popular tweets at year 2016 is: 0.181917
The percentage of popular tweets at year 2017 is: 0.859327
```