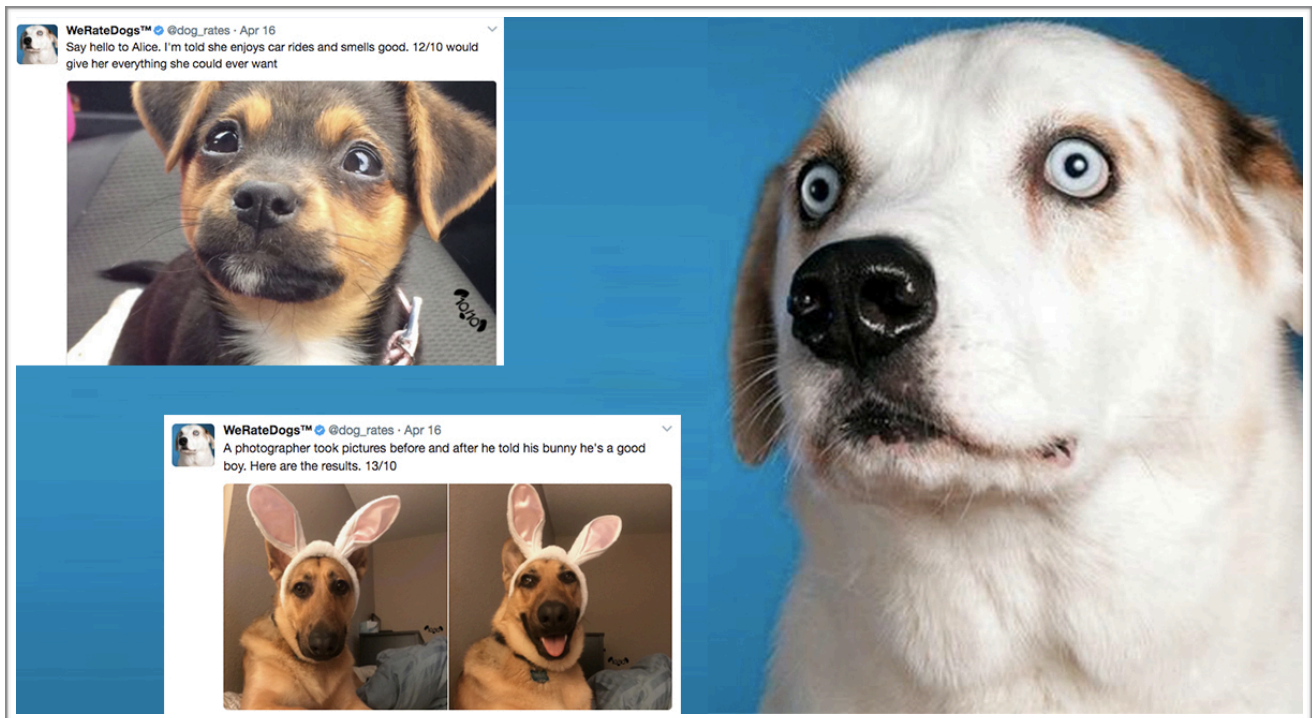
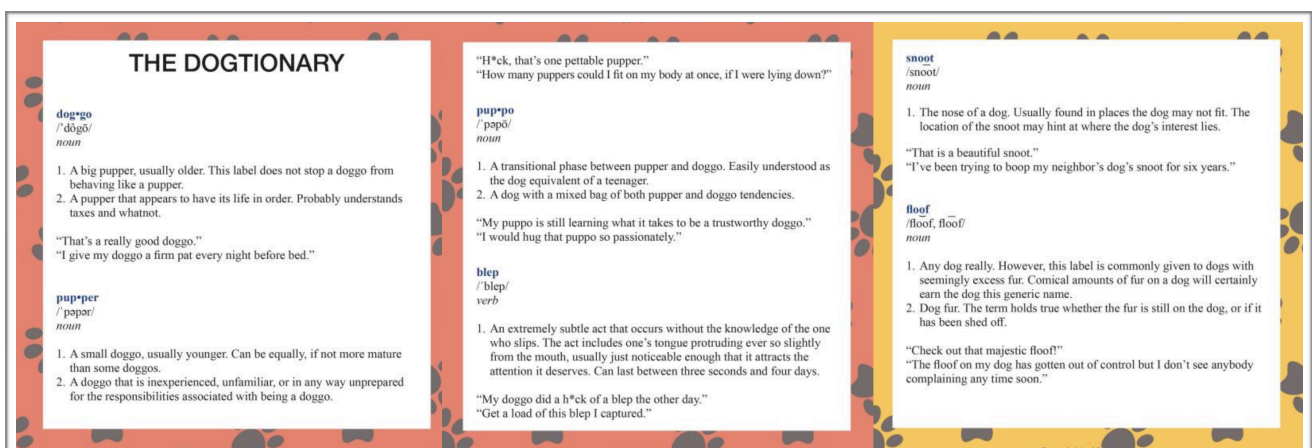


Data Analysis report of Tweet Archive WeRateDogs

The tweet archive of Twitter user @dog_rates, also known as WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. WeRateDogs has over 4 million followers and has received international media coverage. Below image is an example of the tweets.



The interesting part of WeRateDogs is that there are humorous dog stages like pupper, doggo etc. The dogs are in a specific stage based on this ‘Dogtionary’ below. The dogs are rated as well but the ratings are always more than one, considering all of them are good dogs.



This Data Analysis report is the third part of the Data Wrangling process which involved gathering data from three different sources, cleaning and assessing to create interesting and trustworthy analyses and visualizations.

The three data sources are

1. Basic Twitter Archive available from WeRateDogs
2. Additional information from Twitter Archive obtained programmatically
3. Image Prediction neural network that predicts which breed a specific dog will be from.

After the gathering and cleaning process, we get a combined master datasource that has information that we can analyse effectively. The next sections will explore this data and present some insights and visualizations.

1. Understanding WeRateDogs Data

```
#   Column                               Non-Null Count  Dtype
---  -
0   tweet_id                             1469 non-null   int64
1   timestamp                             1469 non-null   object
2   retweeted_status_id                   53 non-null     float64
3   retweeted_status_user_id             53 non-null     float64
4   retweeted_status_timestamp            53 non-null     object
5   expanded_urls                         1469 non-null   object
6   rating_numerator                      1469 non-null   int64
7   rating_denominator                   1469 non-null   int64
8   name                                 1469 non-null   object
9   stage                                228 non-null    object
10  retweet_count                         1469 non-null   int64
11  favorite_count                        1469 non-null   int64
12  jpg_url                              1469 non-null   object
13  img_num                              1469 non-null   int64
14  p1                                    1469 non-null   object
15  p1_conf                              1469 non-null   float64
16  p1_dog                               1469 non-null   bool
17  p2                                    1469 non-null   object
18  p2_conf                              1469 non-null   float64
19  p2_dog                               1469 non-null   bool
20  p3                                    1469 non-null   object
21  p3_conf                              1469 non-null   float64
22  p3_dog                               1469 non-null   bool
dtypes: bool(3), float64(5), int64(6), object(9)
memory usage: 245.3+ KB
```

The dataset has per tweet id, the ratings numerator and denominators, favorite_count, retweet_count that help us understand the popularity of the dogs. p1, p2, p3 columns are the dog breeds predicted by the image prediction neural network, where as p1_conf and similar tell us how confident the prediction is (Above 90% means the prediction is very confident). p1_dog, p2_dog have true/false values, true meaning the prediction was actually a dog.

2. Exploration of Data

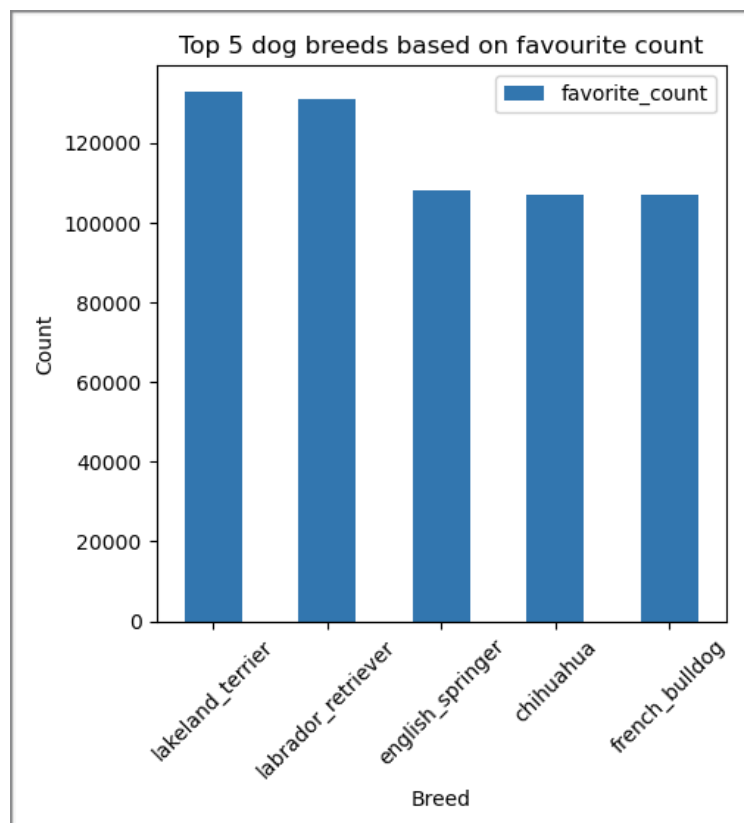
Following insights or questions are quite interesting to explore -

1. What are the most favourite dog breeds?
2. Which dog stages are most common or popular?

Insight1 : What are the most favourite dog breeds?

The top 5 dog breeds which are most favorites are lakeland_terrier, labrador_retriever, english_springer, chihuahua and french_bulldog .

	p1	stage	favorite_count	retweet_count
256	lakeland_terrier	puppo	132810	48265
660	labrador_retriever	doggo	131075	79515
51	english_springer	NaN	107956	45849
343	chihuahua	NaN	107015	56625
97	french_bulldog	pupper	106827	32883



Insight1 : What is the most favourite dog breed based on correct prediction and favourite count?

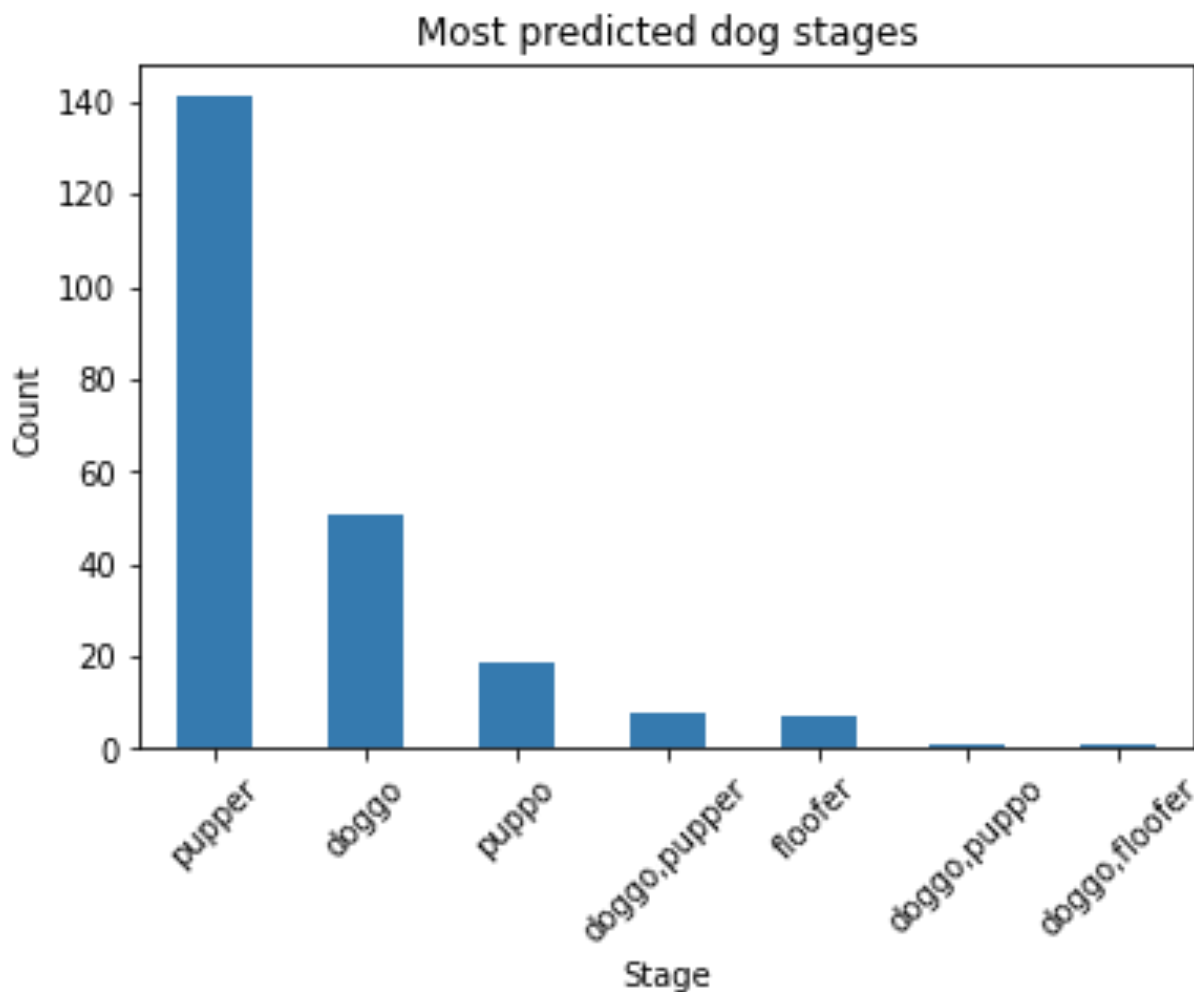
The p1_dog columns tells us if the prediction actually a dog or not. Let us see which is the popular dog breed based on averaged favourite count and given that the prediction was actually a dog.

```
p1
black-and-tan_coonhound    33911.000000
saluki                     24060.000000
french_bulldog             17913.000000
flat-coated_retriever      16791.625000
irish_water_spaniel        16400.000000
...
groenendael                1896.000000
tibetan_terrier            1841.000000
japanese_spaniel           1362.000000
ibizan_hound               1114.666667
brabancon_griffon          885.000000
Name: favorite_count, Length: 111, dtype: float64
```

From the above table, black-and-tan_coonhound and saluki seem to be the two most favorite dog breeds¶

Insight2: Which dog stages are most common or popular?

Pupper is the most popular dog stage followed by doggo



Insight2: What are the ratings of Pupper and Doggo?

Below table shows Pupper's average mean rating as 11.

	tweet_id	retweeted_status_id	retweeted_status_user_id	rating_numerator	rating_denominator
count	1.410000e+02	4.000000e+00	4.000000e+00	141.000000	141.0
mean	7.260760e+17	7.865482e+17	3.275939e+09	11.106383	10.0
std	6.201641e+16	8.282127e+16	1.842090e+09	1.929849	0.0
min	6.721600e+17	6.798284e+17	5.128045e+08	7.000000	10.0
25%	6.801913e+17	7.501178e+17	3.275939e+09	10.000000	10.0
50%	6.950743e+17	7.947568e+17	4.196984e+09	11.000000	10.0
75%	7.514560e+17	8.311872e+17	4.196984e+09	12.000000	10.0
max	8.863661e+17	8.768508e+17	4.196984e+09	27.000000	10.0

Below table shows Doggo's average mean rating as 10.

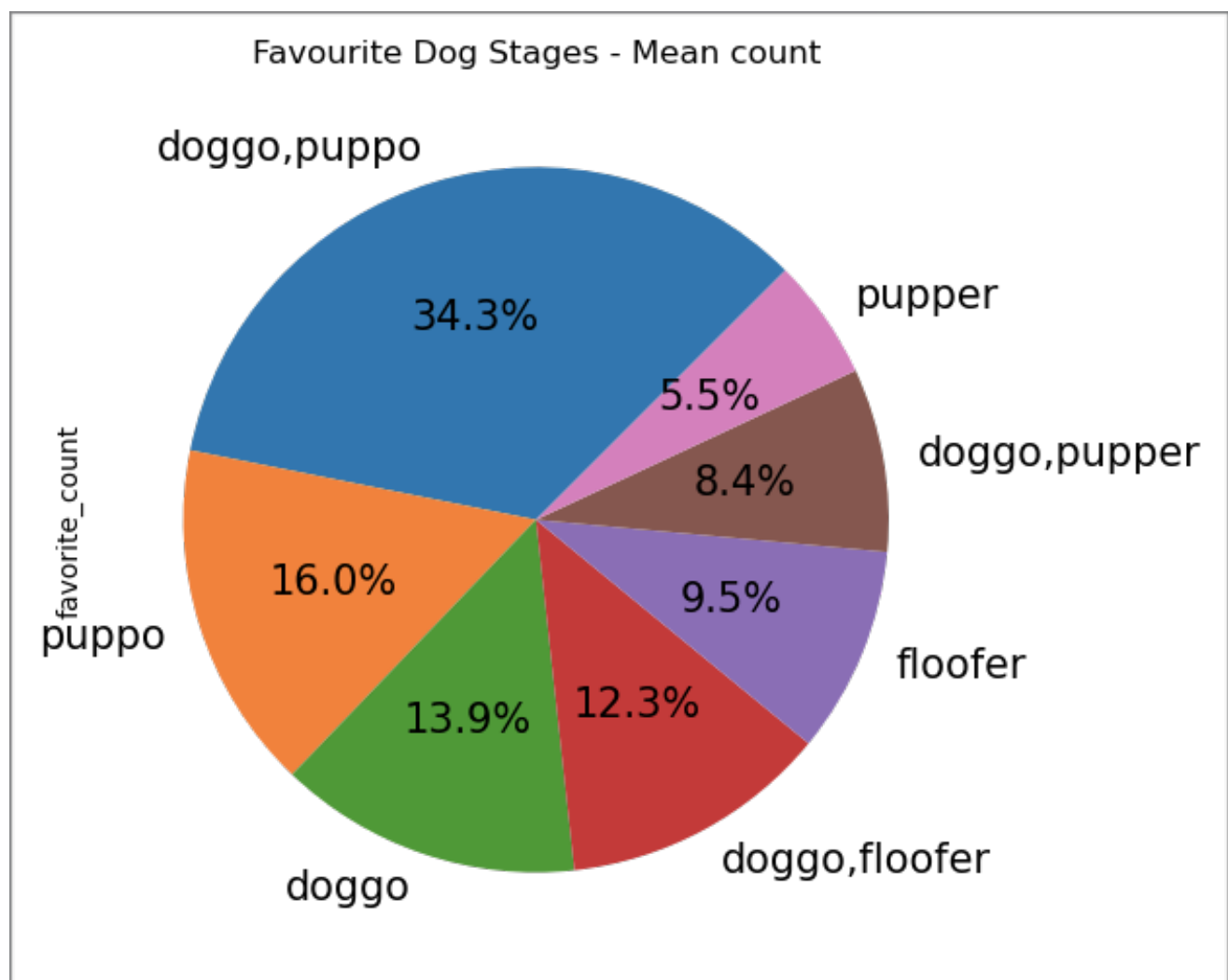
	tweet_id	retweeted_status_id	retweeted_status_user_id	rating_numerator	rating_denominator
count	5.100000e+01	3.000000e+00	3.000000e+00	51.000000	51.0
mean	8.056420e+17	8.251126e+17	4.196984e+09	12.078431	10.0
std	4.526203e+16	5.425658e+15	0.000000e+00	1.383375	0.0
min	7.160809e+17	8.190048e+17	4.196984e+09	8.000000	10.0
25%	7.681294e+17	8.229817e+17	4.196984e+09	11.000000	10.0
50%	8.190153e+17	8.269587e+17	4.196984e+09	12.000000	10.0
75%	8.316313e+17	8.281665e+17	4.196984e+09	13.000000	10.0
max	8.902403e+17	8.293743e+17	4.196984e+09	14.000000	10.0

From the above two tables, we can see that pupper has average mean average rating as 11 and doggo has 10. There is only a slight difference between the two.

Insight3 - What dog stage is most favorite?

An assessment of mean count based on favorite_count of the tweets reveals doggo,puppo as the favoured stage, followed by puppo and doggo as individual stages.

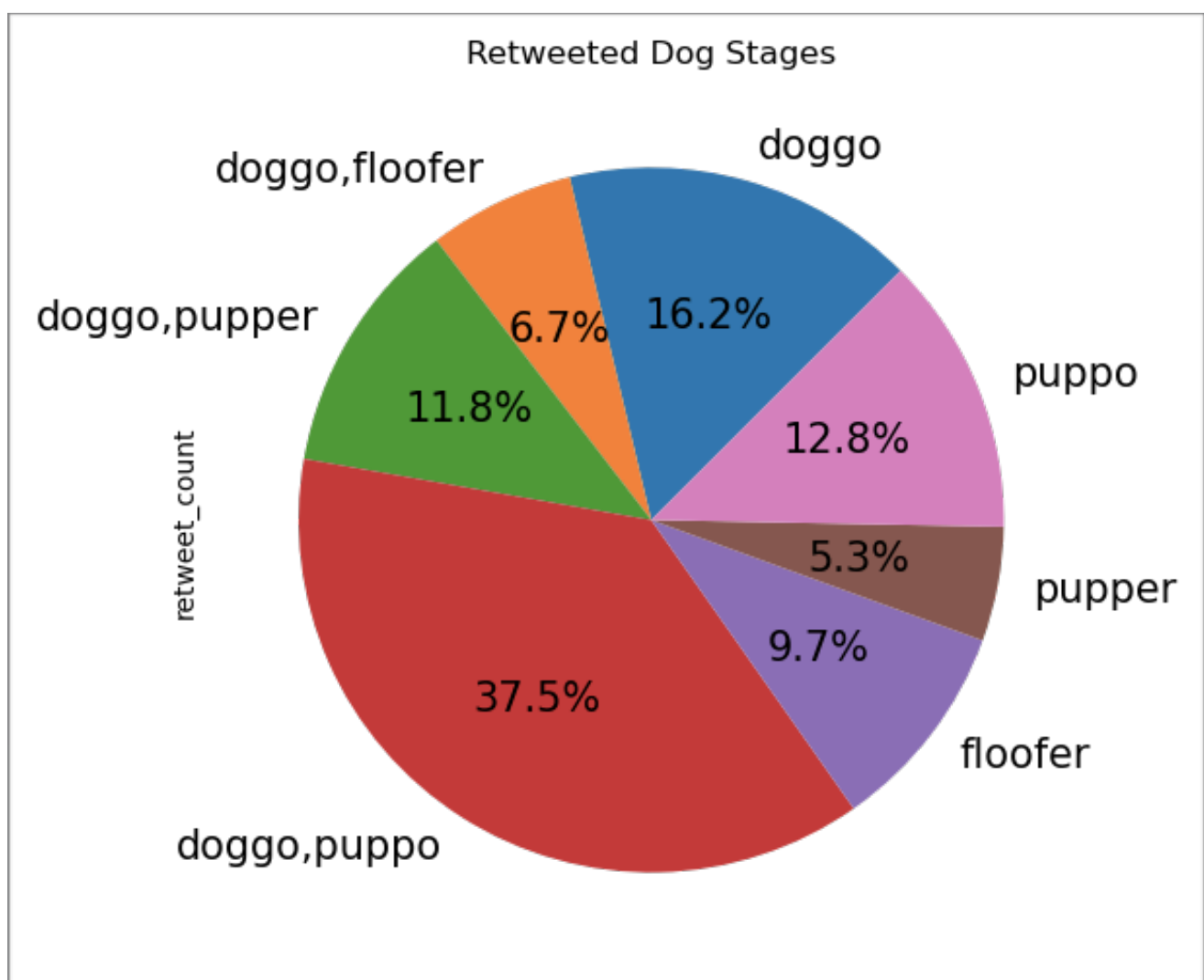
```
: stage
doggo,puppo      47844.000000
puppo            22308.210526
doggo            19347.156863
doggo,floofer    17169.000000
floofer          13206.000000
doggo,pupper     11715.125000
pupper           7729.361702
Name: favorite_count, dtype: float64
```



Insight4 - What dog stage is most retweeted?

From below table and graph, its clear that doggo, pupper have highest retweet count.

```
stage
doggo      8298.901961
doggo,floofer  3433.000000
doggo,pupper  6036.375000
doggo,puppo  19196.000000
floofer    4968.714286
pupper     2687.702128
puppo      6551.473684
Name: retweet_count, dtype: float64
```



Conclusion/Summary

From our exploration of the WeRateDogs, some interesting insights were explored. To summarize -

1. The top 5 dog breeds which are most favorites are lakeland_terrier, labrador_retriever, english_springer, chihuahua and french_bulldog .
2. Pupper is the most popular dog stage followed by doggo, with mean ratings of 11 and 10 respectively
3. Doggo, puppo combined was the most favoured stage, followed by puppo and doggo as individual stages based on both favorite_count and retweet_count.