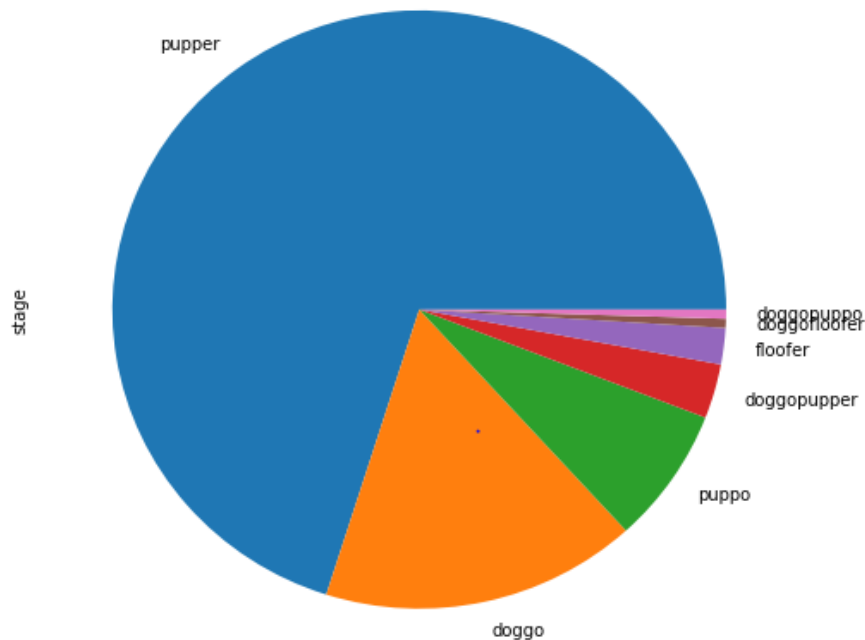


Act Report

After gathering, accessing, and cleaning data, I would analyze and provide my insights. Here are some questions I am interested:

1. How the dog stage was distributed?



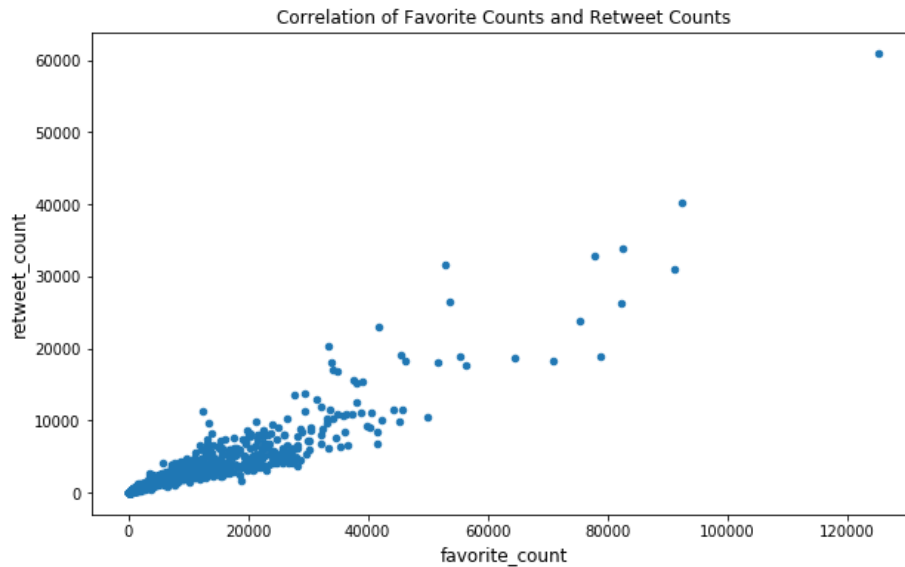
There are lots of tweet that showing dog in different stage. Here I used pie chart to represent the distribution. We can see that most of tweet are about pupper. Even though there are some 'doggo pupper', 'doggo puppo' and 'doggo floofer', we can ignore those since the amount is very small.

2. what are some popular dog names?

```
None      419
Oliver     8
Tucker     7
Winston    7
Penny      6
Name: name, dtype: int64
```

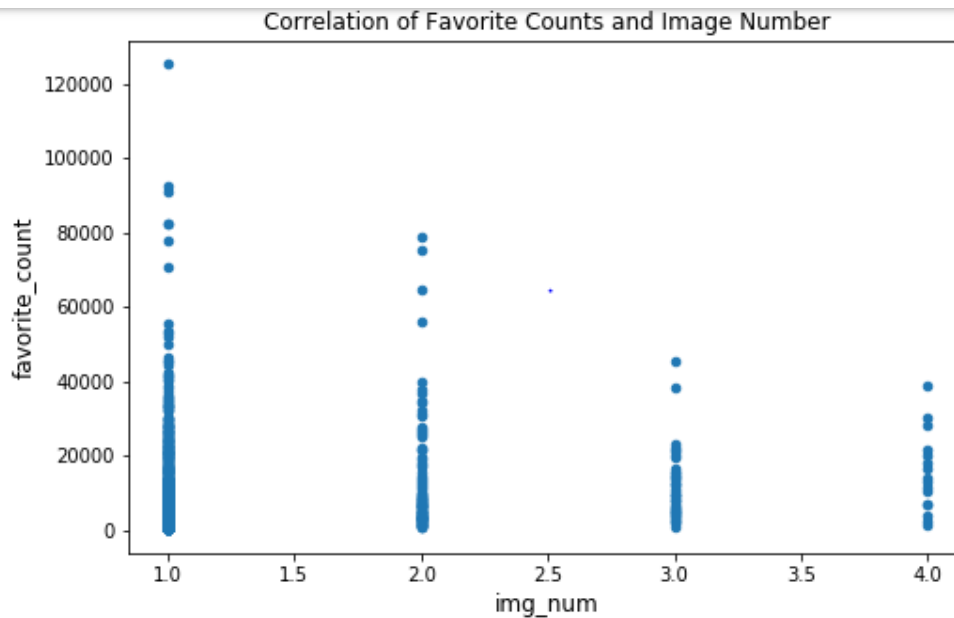
Even though I cleaned the data, there are still many missing values for dog name. From the chart we can see that there are many None name. However, it shows that Oliver, Tucker, Winston, and Penny are popular name for dog besides None.

3. What is the relationship between favorite_count and retweet_count?



Before I did this project, I know people would usually like the tweet but retweet it. According to the graph, we can see that the more favorite_count are, the more retweet_count. If people like the tweet, they would like to share to other people. I used scatter plot to show the correlation between two variables. It is increasing.

4. What is the relationship between favorite_count and numbers of image?



From this graph, we can see that the less images people post on twitter the more likely people would like the tweet. This is a very interesting fact that I never had thought about. I used to think that more images can catch people's attentions. However, it is the opposite if we just look at this database.

5. which dog has most favorite_count?

```
Final_master[Final_master['favorite_count'] == 127729]
```

tweet_id	timestamp	source	rating_numerator	rating_denominator	name	stage
232						
807106840509214720	2016-12-09 06:17:20	<a href="http://twitter.com/download/iphone" r...	13	10	Stephan	NaN https://pbs.twimg

Since we are talking retweet counts and favorite counts for tweet, I wanted to know which dog has most favorite_count. First, I used describe() to see the maximum of favorite count. It turns out to be 127729. Then, I selected the row with 127729 favorite counts. It is a dog called Stephan and it has 61375 retweet counts. The rating is 13/10.

Insights:

- Among all the tweets in database, most tweets show dogs in pupper stage.
- Oliver, Tucker, Winston, and Penny are popular name for dog besides None.
- The more people like the tweet, it would more likely to be retweeted.
- The less images people post on the tweet, the more favorite counts they would get.
- A dog named Stephan has most favorite count among other dogs in the database