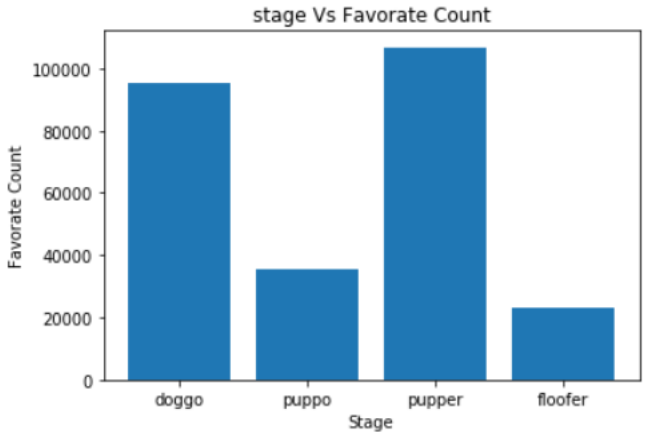
Act Report

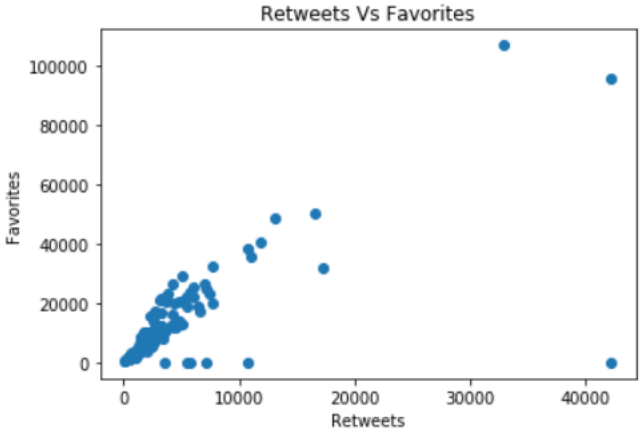
Finding and Visualization

1- Favorite count per stage.

The bar chart shows which dog stage gets the most likes. And it appears that the dog stage that getting the most likes is Pupper.



2- Correlation between retweet count and favorite count:
In this scatter we want to check if there are any correlation between the number of likes and we can see in most of the data points when retweets increase the likes increase also.



3- Average rating for each stage

After cleaning the data and analyze them, I have calculated the average rating by stage. The picture below shows the average rating by dog stage. It shows that on average Doggo gets more rating than any other dog stage. Although Pupper gets more likes, Doggo on average gets higher rating.

```
In [331]: master_df.groupby('stage').rating_numerator.mean()

Out[331]: stage
          doggo      12.350000
          floofer    11.600000
          pupper     11.229885
          puppo      12.000000
          Name: rating_numerator, dtype: float64
```

4- Count of dogs by stage

After analyzing the data and seeing the results of likes and average rating, I decided to find the counts of the dog stages. The count will make some sense specially in the like counts. Pupper and Doggo have the most likes which make sense because we have more observation of these two stages than any other stages.

```
In [343]: master_df.groupby('stage').count()['tweet_id']

Out[343]: stage
          doggo      40
          floofer     5
          pupper     87
          puppo      15
          Name: tweet_id, dtype: int64
```