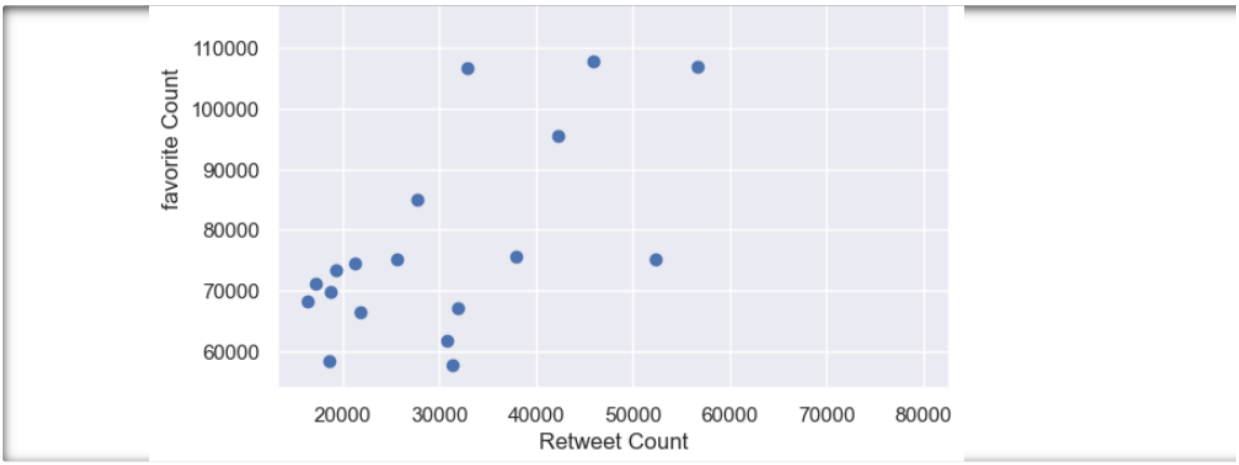
Visualisation and insights

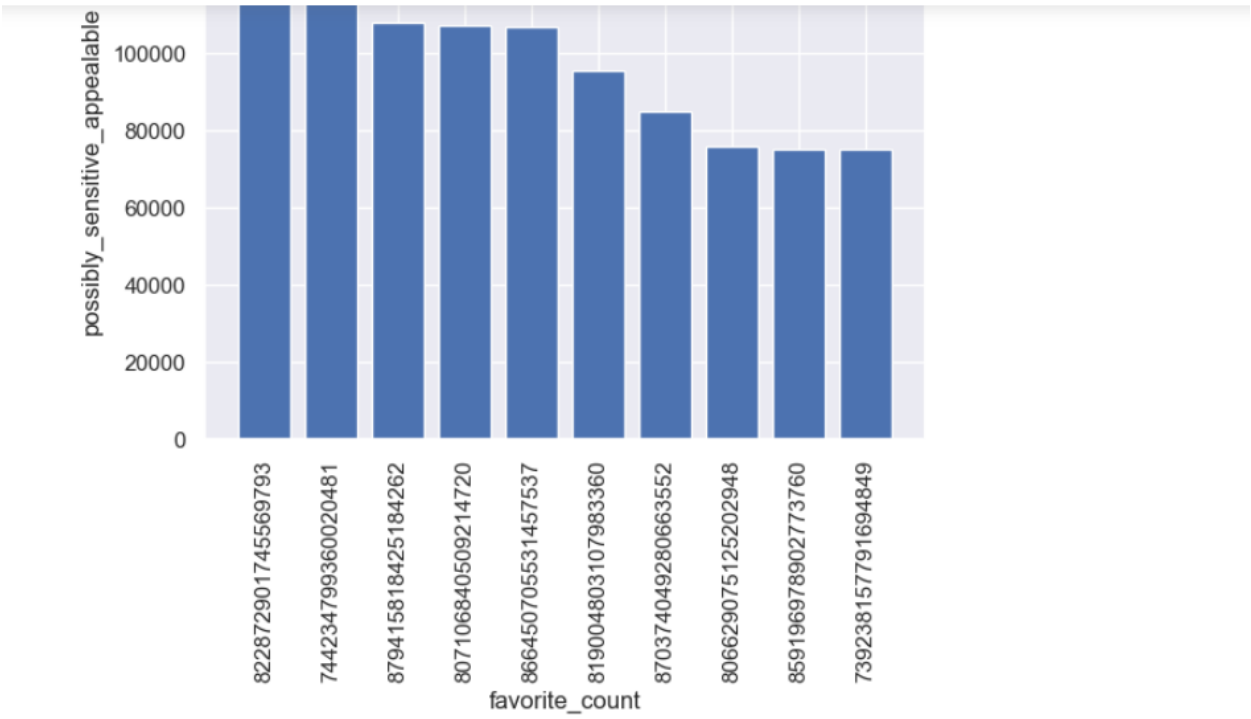
Correlation between retweet_count and favorite_count



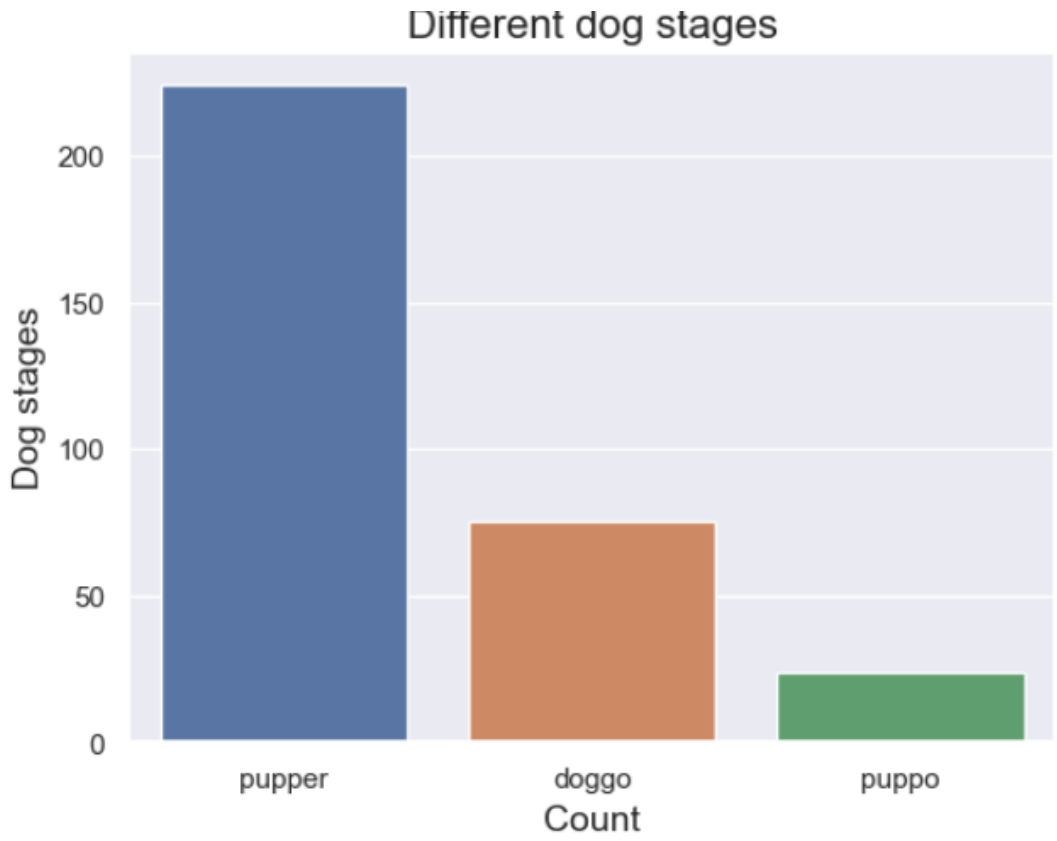
The scatter plot show a positive correlation between favourite_count and retweet_count which could mean that the more a tweet is liked the more chances it has of it being retweeted.¶

2. Counted favorite and a possible sensitive appealable

The above data was done to analyses if one of the reasons why an appeal or a possible appeal might have happened was based on the favorite tweet count. It seems like the more the tweet has been counted as favorite the better chance it has for an appeal of the votes done.



3. Different dog stages



The division of dog stages.

The stage similarly, I check the division of dog stages. It shows that ‘pupper’ (a small doggo, usually younger) is the most popular dog stage, followed by ‘doggo’ and ‘puppo’. It could be due to the young and unmatured dog is usually cuter than the adult dog. It should also be noticed that there’s huge amount missing data in dog stages, thus the distribution may not reflect the truth.

3. Image prediction

```
In [313]: image_df_copy['second_prediction'].value_counts().head(10)

Out[313]: Labrador_retriever          104
          golden_retriever             92
          Cardigan                     73
          Chihuahua                    44
          Pomeranian                   42
          Chesapeake_Bay_retriever     41
          French_bulldog               41
          toy_poodle                   37
          cocker_spaniel               34
          miniature_poodle             33
          Name: second_prediction, dtype: int64
```

The dog breeds above is the top 10 breed this model predicted. The Labrador_retriever is the second one. It could be because those are the common breeds in U.S. There are more image data on those brees and hence better results.¶