As a Data analyst cleaning and wrangling is the the end process of data analysis. A data analyst is meant to tell the story of his/ her data to both technincal and non-technical audience and what better way to tell the story than through visualisation. But before i proceed with the visualisation some compelling questions need to ask

Questions

1. What dog has the highest favorite_count

2. Which dog has the highest and the lowest retweet_count

3. Were the algorithms able to classify the dogs or not

4. What breed of the dog has the highest retweet

5. Dog_stage that attracts the most retweet on an average

6. Dog_stage that attracts less retweets on an average

7. Dog_stage with highest average favorite_count

8. Dog stage with lowest average favorite_count

9. from which source did the tweets come from

10. Which dog dog_stage was rated well on an average

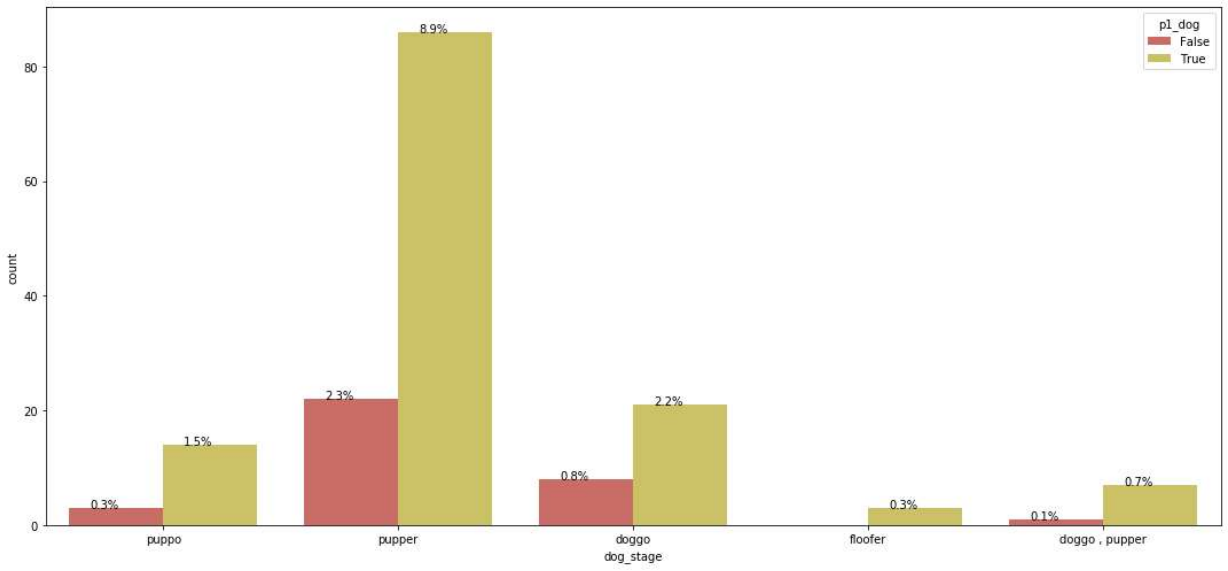
In [ ]:

Answers

1. Dog 744234799360020481 twitter_id has about 144033 favorite count, expected i should say since it is a Labrador retriever

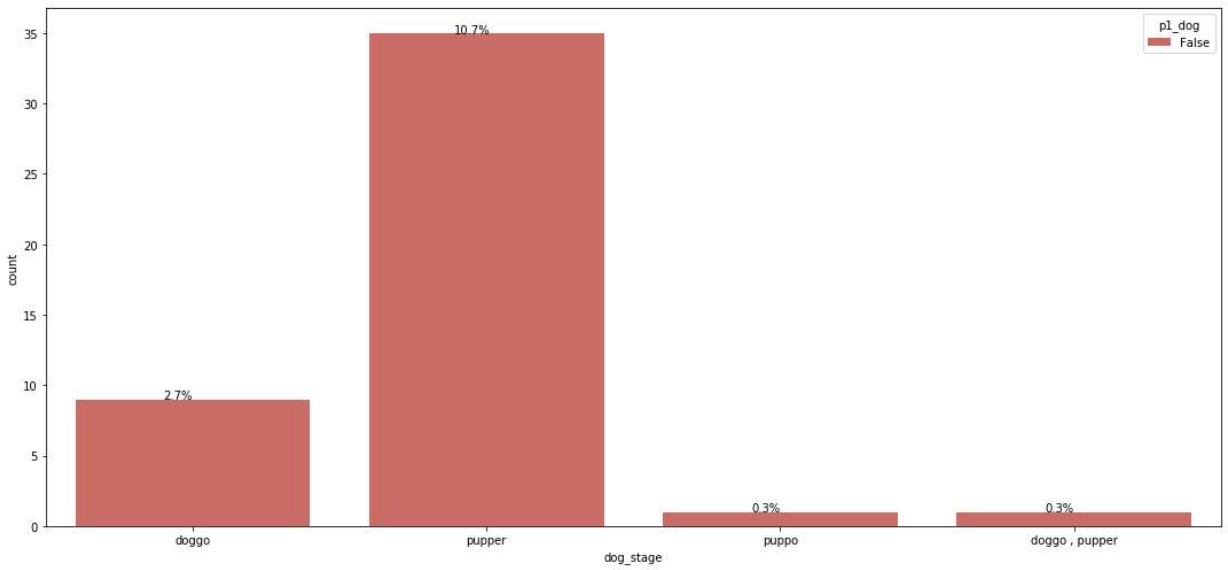
2. Dog 744234799360020481 twitter_id also has the highest retweet, but 666102155909144576 twitter_id has the lowest retweet of 11

In [11]:



pegging the algorithms at 0.6 benchmark for the algorithm to be accepted, we found about 4% of dog_stage were wrongly classified as false

In [12]:



3.

We can see how poorly the p1 model which has shown to have a higher confidence interval than the others has performed compared to human level performance , about 15 % were wrongly miss classified as not dogs

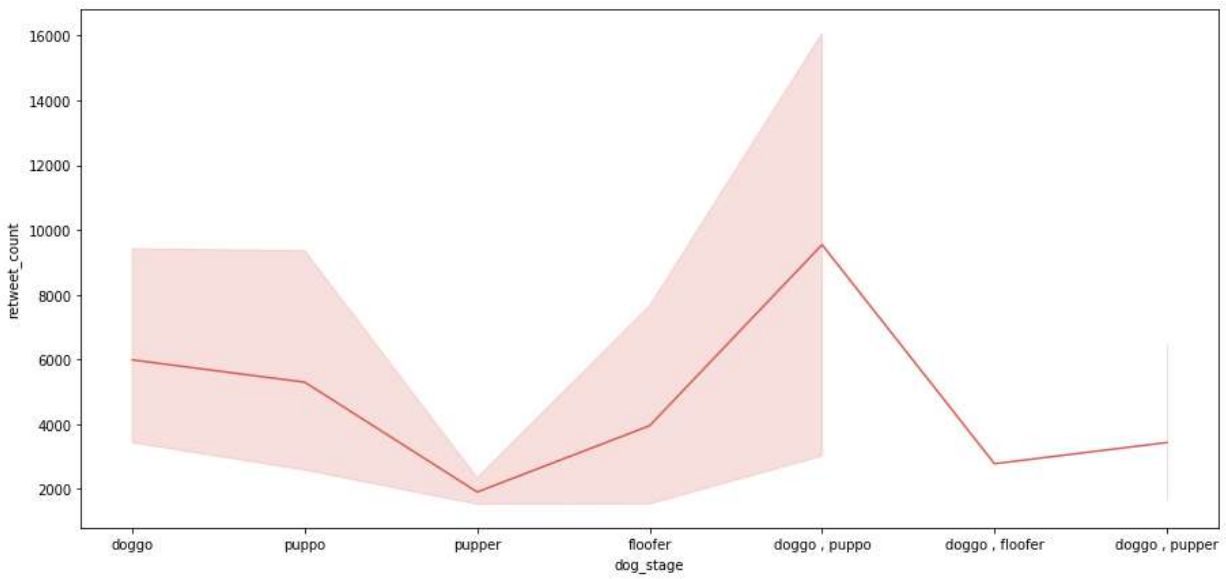
Therefore we can conclude that although the model has a higher performance but not high enough to surpass or equal the human-level performance which is one of the reasons of applying AI

4.

Labrador retreiver has the highest retweet of 70252 and a twwet_id of 744234799360020481

In [ ]:

In [13]:



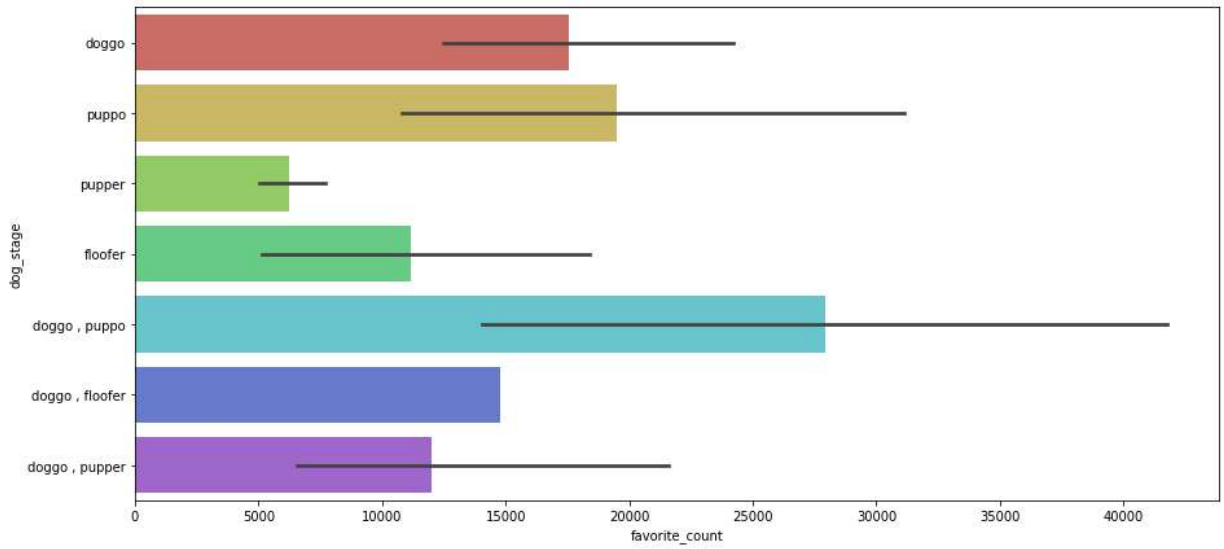
5.

On an average 'doggo, puppo' and 'floofer' have the highest retweet_count followed by puppo.

6.

While pupper has the lowest retweet_count.

In [14]:
```
bar_plot(df_clean, 'favorite_count', 'dog_stage')
```



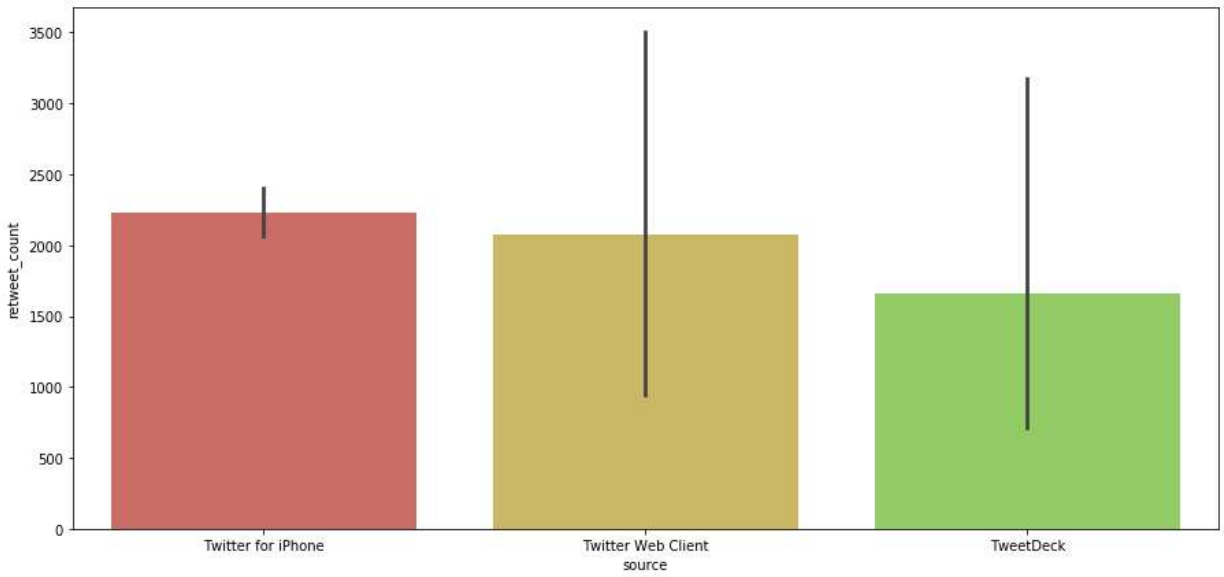
7.

'doggo, puppo' has the highest favorite_count followed by puppo

8.

while Pupper has the lowest favorite_count, it seems users feel a little bit disinterseted with dog_stage pupper

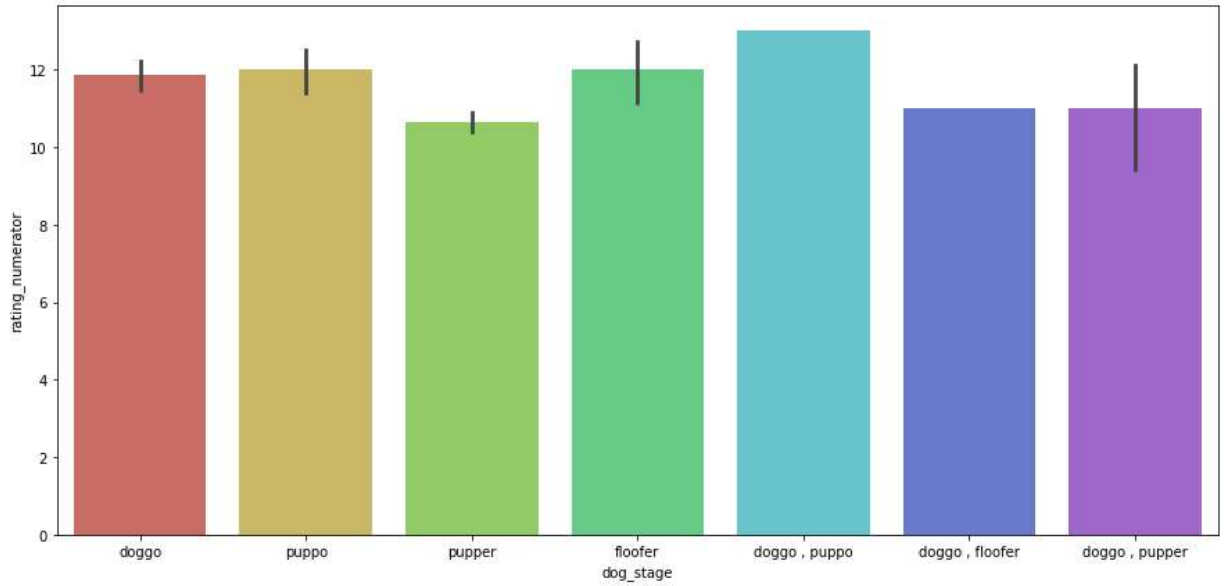
In [15]:



9.

Most users twitted from Iphone and Web Client

In [17]:



10.

On an average 'doggo, puppo' was rated 13/10 which is the highest on an average.

In [ ]: