

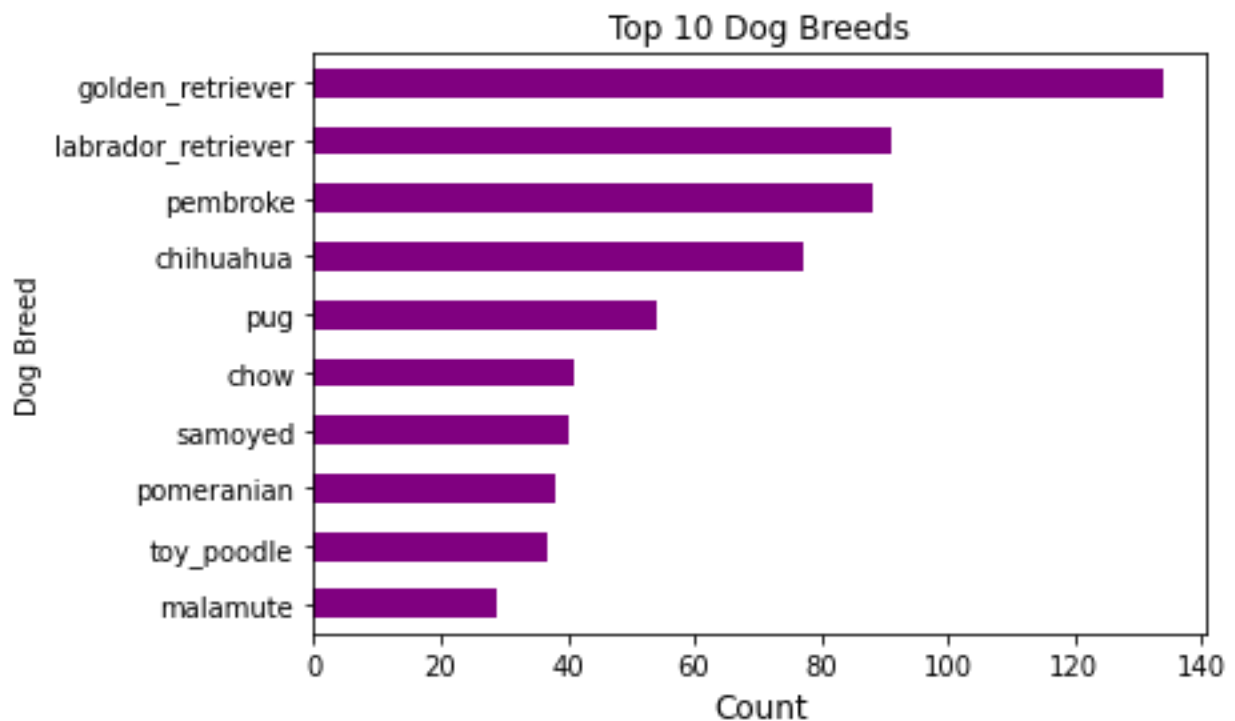
Analysis and Insights

Visualization

After wrangle the data by three main steps (gathering, assessing and cleaning), we can now perform some basic analysis to obtain insights.

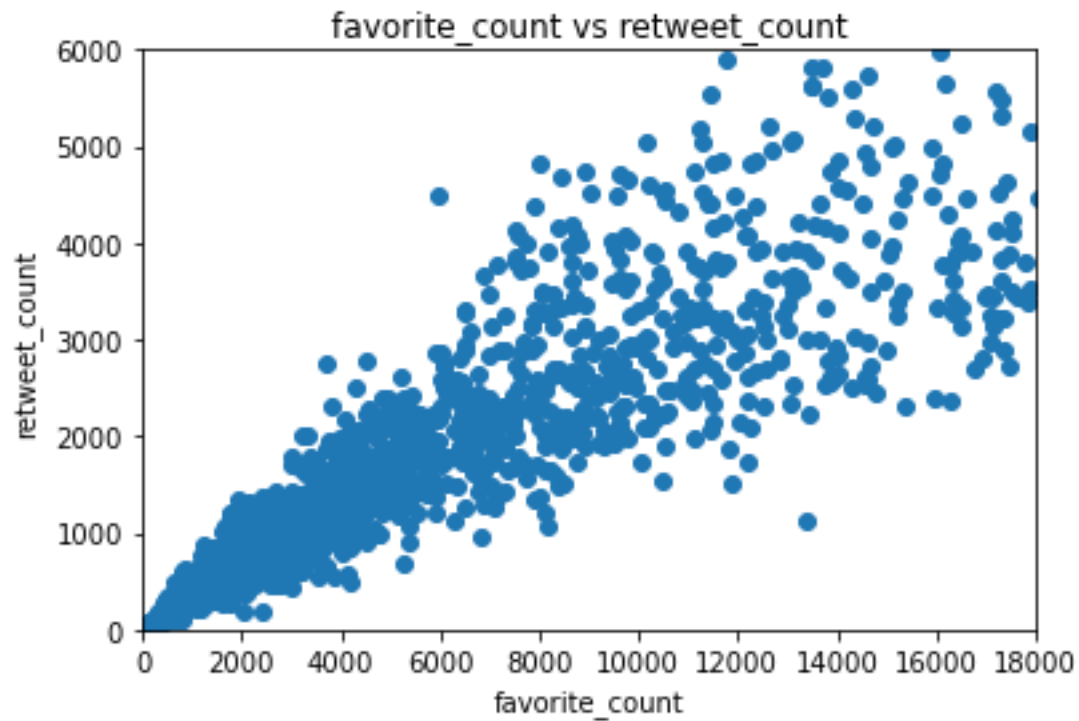
The following questions will be my analysis and to get insights

Which breed is most common/popular?



Golden_Retriever is the most common breed in this twitter dataset with a total of 134

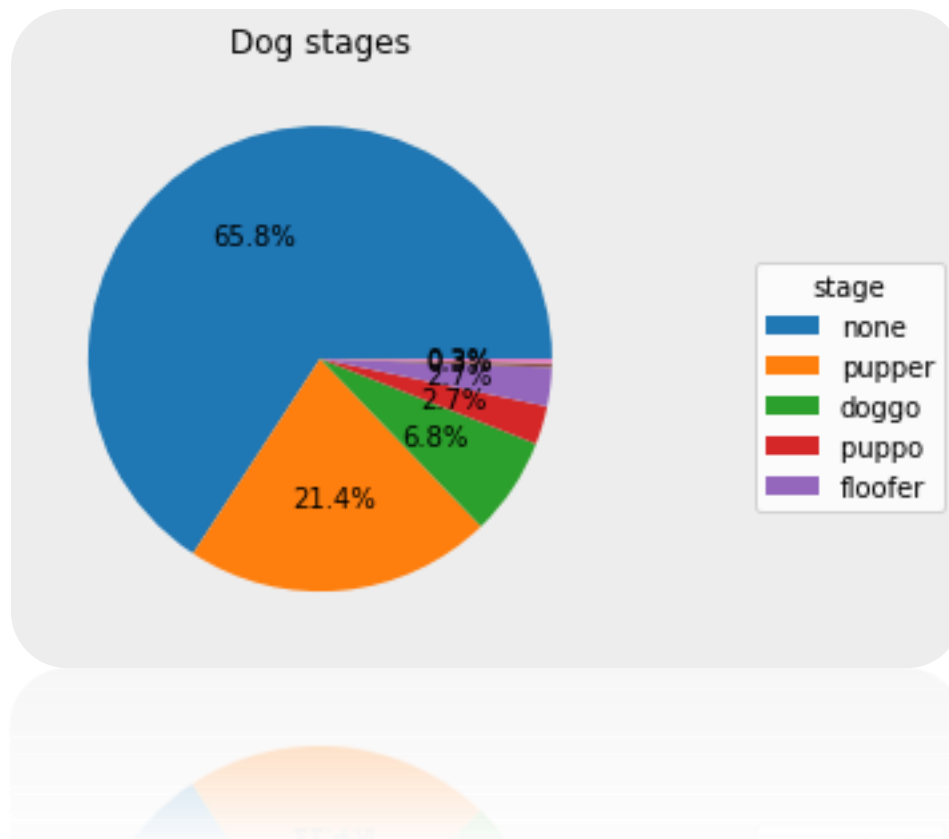
What is the relation between retweet and favorite?



There is a positive correlation between favorites and retweet counts with value of 0.911795

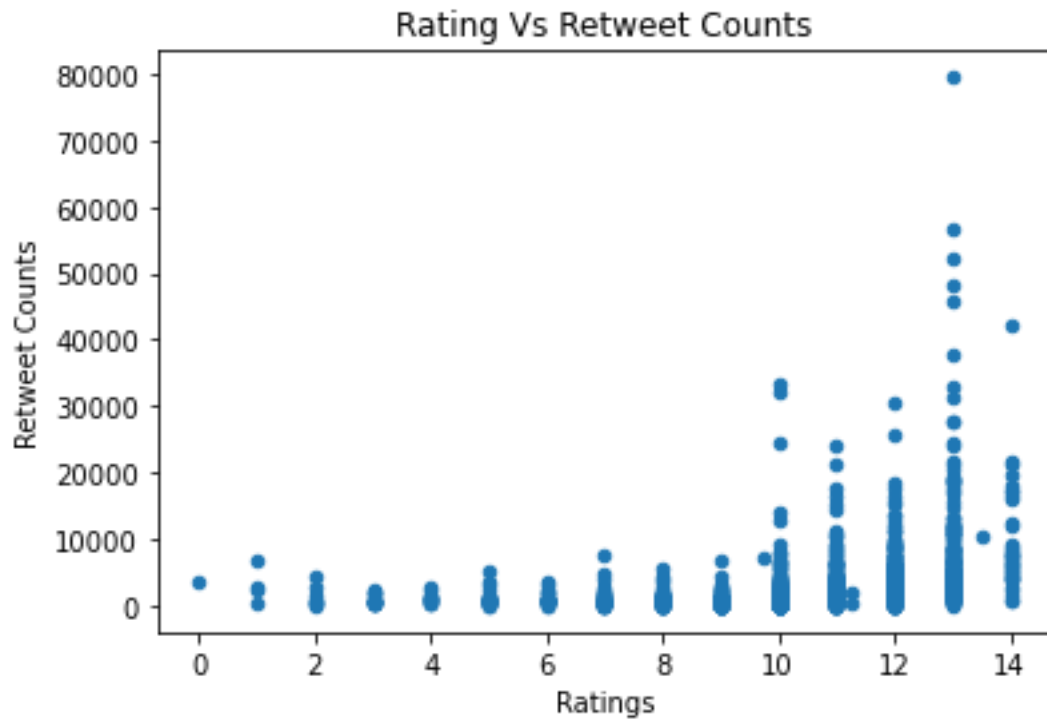
What are the most common dog stage?

```
pupper      221
doggo       72
puppo       23
floofer      9
doggo, pupper    9
doggo, puppo     1
doggo, floofer   1
Name: stage, dtype: int64
```

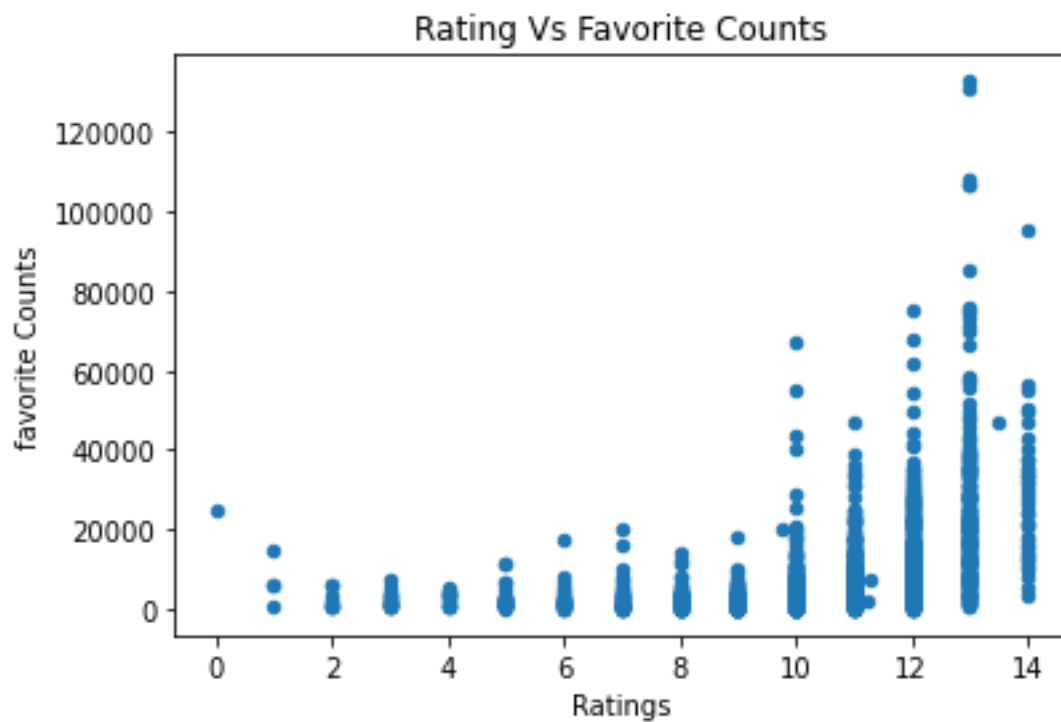


Pupper is the most common stages while floofer is the least

Are the ratings by WeRateDogs relative to the retweet count and favorite count?

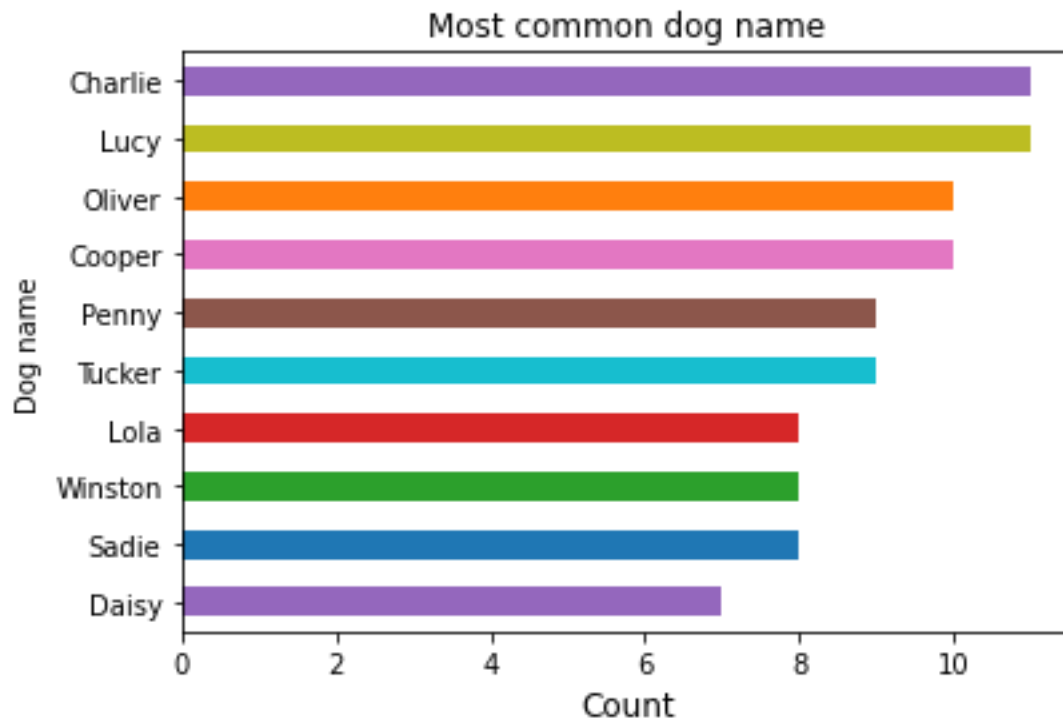


As a number of retweet increased, the rating increased too.



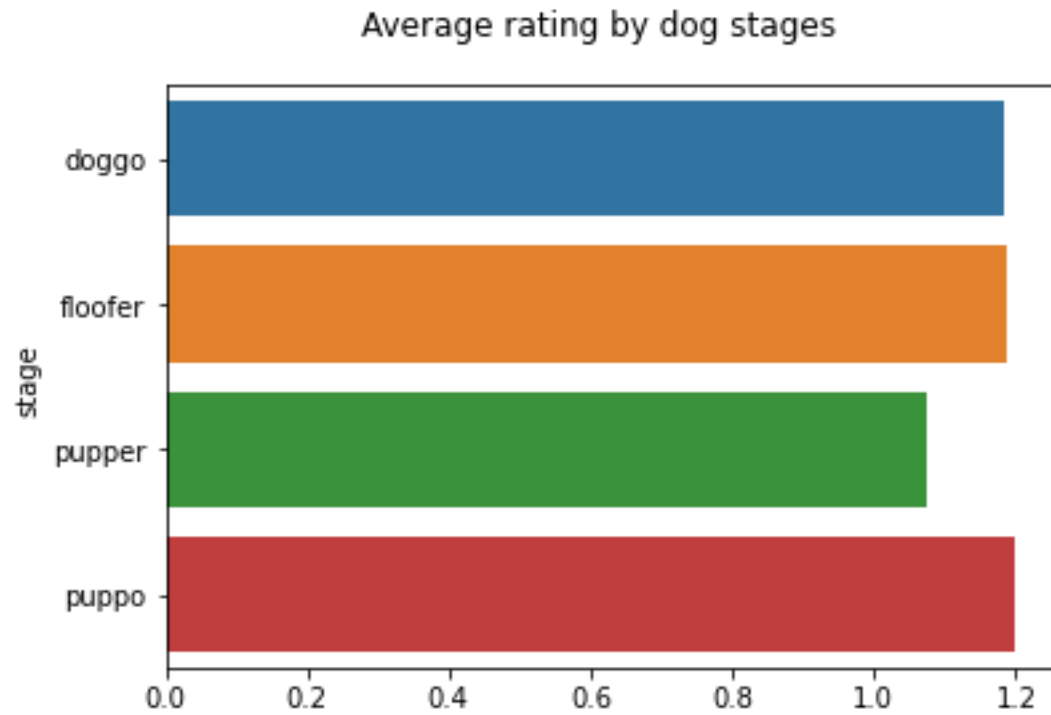
They are relative too. Also rating 13 got the highest favorite counts and retweet counts.

What are the most common dog names?



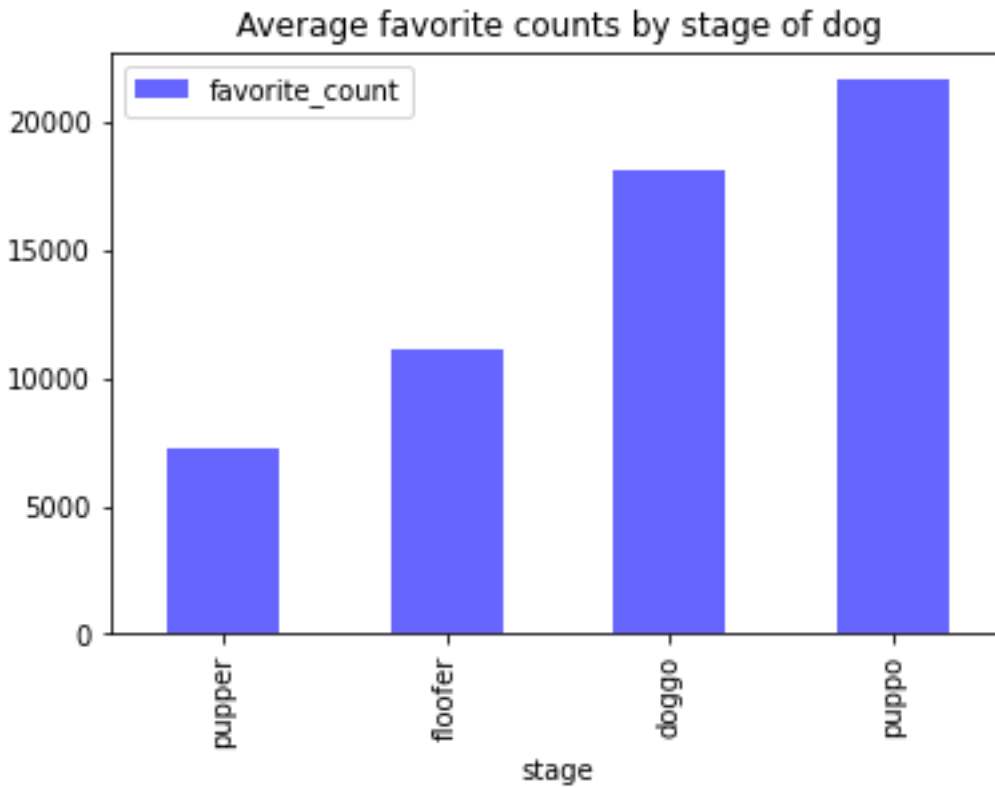
The most common dog names are Charlie, Lucy, and Oliver.

Which stage got the highest rating average?

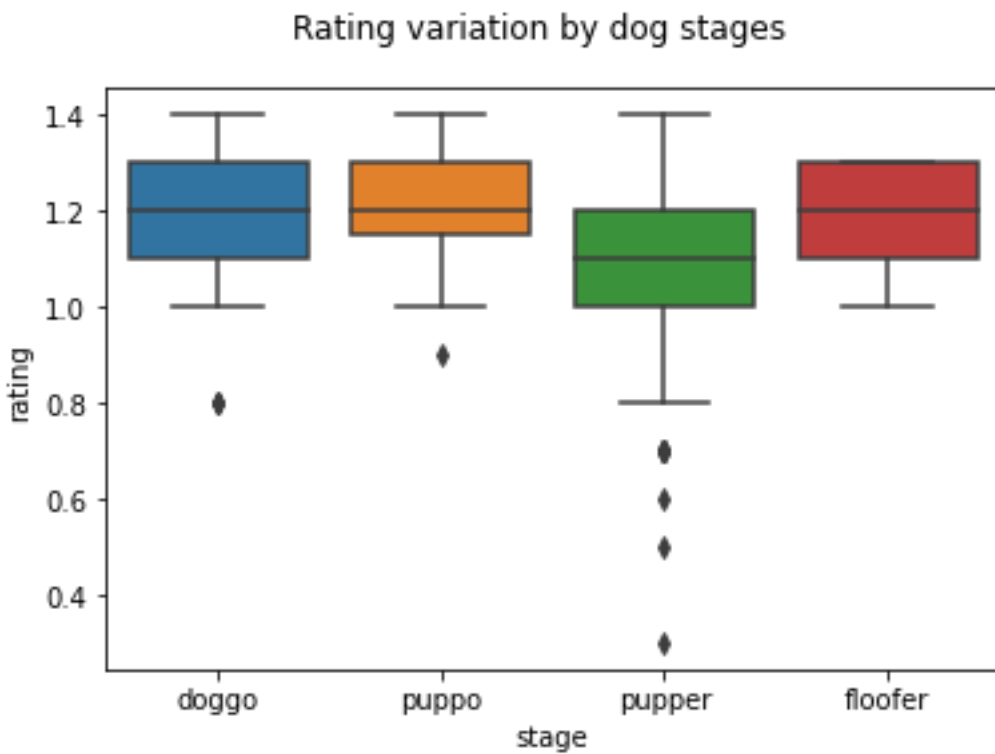


The highest rated dog is puppo

Which dogs stage got the highest favorite counts?



Puppo stage has the highest average favorite tweet with counts above of 20000



Insights:

	rating_numerator	rating_denominator	retweet_count	favorite_count	rating	img_num	p1_conf	p2_conf	p3_conf
count	2076.000000	2076.0	2076.000000	2076.000000	2076.000000	1649.000000	1649.000000	1649.000000	1.649000e+03
mean	10.608276	10.0	2838.117534	8977.551060	1.060828	1.214676	0.605535	0.137751	6.120135e-02
std	2.149631	0.0	4718.417254	12212.628718	0.214963	0.574543	0.266530	0.101764	5.178564e-02
min	0.000000	10.0	16.000000	81.000000	0.000000	1.000000	0.044333	0.000010	2.160900e-07
25%	10.000000	10.0	640.250000	2040.750000	1.000000	1.000000	0.379055	0.054790	1.553510e-02
50%	11.000000	10.0	1399.500000	4191.500000	1.100000	1.000000	0.606654	0.121607	4.990060e-02
75%	12.000000	10.0	3285.000000	11425.000000	1.200000	1.000000	0.852088	0.199544	9.469160e-02
max	14.000000	10.0	79515.000000	132810.000000	1.400000	4.000000	0.999984	0.467678	2.710420e-01

1. The minimum favorite tweet count is 81 and the maximum favorite tweet count is 132810.0
2. The minimum retweet count is 16 and the maximum retweet count is 79515.0
3. In terms of breeds, **Golden_Retriever** ,Labrador_Retriever and pembroke are the popular dogs.
4. About 65.8% of the dogs have no name.

Limitation

The breeds were predicted by the people who created the machine learning models was not accurate, I tried to delete the row if the three prediction was false ,then in my analysis I chose p1 since it's the high prediction. There are no 100% sure things in our conclusions.

