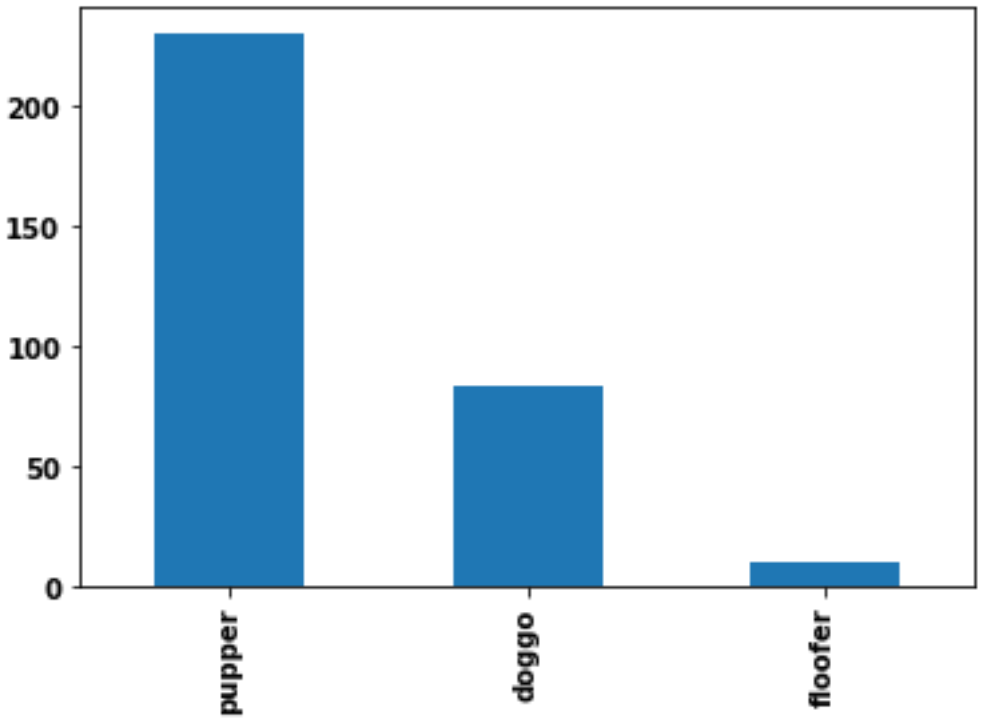
Analyzing and visualization

This report aimed to the insights that we get it from the datasets we cleaned about WeRateDogs account. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. WeRateDogs has over 4 million followers and has received international media coverage.

1- which one of dog life stage was the most common in our dataset?



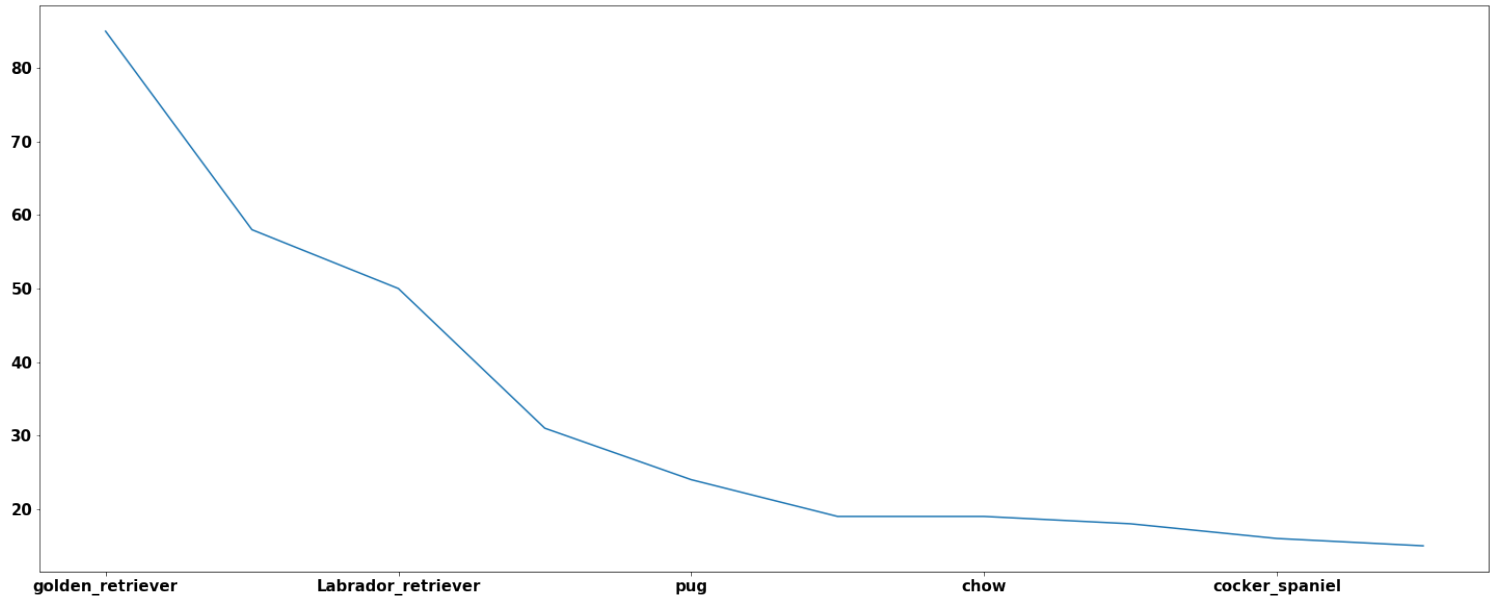
Popper was the most common dog life stage.

2- Since we have three algorithms use neural network to predict the dog type in the **tweet image predictions** dataset, so my question was which one of these algorithms was more confident rate?

- algorithm 1: 0.628
- algorithm 2: 0.143
- algorithm 3: 0.0612

according to the above results algorithm 1 was the highest conf rate (after I filter the results to show only the True prediction)

3- based in the previous point I decide to take the types of dogs for algorithm 1 to see the top 10 dogs type we have.



The figure shows that the top 10 most common dogs' types, golden-retriever was the most common dog type coming after that the Pembroke in second place.

4 - what is the top 5 dog type that got highest rate (single dog)?

In the dataset there are some tweets that contain more than one dog, so in this result I focus only to the tweets that have single dog.

12	448
10	385
11	381
13	242
9	128
8	87
7	39
14	29
5	22
6	19
3	10
4	8
2	3

First of all, I display the frequency of rating for single dogs, the maximum one was 14/10

(the left column refers to the rate and the right one is the frequency of that rate)

After that I filter the data to see which type of dogs that get 14/10?

Pembroke	3
French_bulldog	2
golden_retriever	2
Pomeranian	1
Bedlington_terrier	1

The results show that these are the top 5 dogs' type that get maximum rate