Act Report

In this project I had to gather, asses, clean and analyze three data frames from a twitter account called WeRateDogs, where they rate a dogs.


I also used a json file from an algorithm that predicted the breed of each dogs .
At the end the 3 databases were all merged into one.

Insights

1. Based on the describe method of the merged dataframe, most predictions were made based on 1 image of the dog. 75% of the predictions were with just one picture and the max images they had was 4.

	id_str	retweet_count	favorite_count	possibly_sensitive	possibly_sensitive_appealable	rating_denominator	img_num	p1_conf	p2_co
count	1.896000e+03	1896.000000	1896.000000	1896.0	1896.0	1896.000000	1896.000000	1896.000000	1.896000e+
mean	7.375736e+17	2800.941983	9042.924051	0.0	0.0	10.458333	1.207278	0.594885	1.337713e-
std	6.779611e+16	4711.308079	12248.136520	0.0	0.0	6.485735	0.564350	0.273147	1.005117e-
min	6.660209e+17	16.000000	81.000000	0.0	0.0	7.000000	1.000000	0.044333	1.011300e-
25%	6.766148e+17	648.750000	2083.000000	0.0	0.0	10.000000	1.000000	0.360137	5.378615e-
50%	7.101290e+17	1386.500000	4206.500000	0.0	0.0	10.000000	1.000000	0.588030	1.167160e-
75%	7.909564e+17	3261.000000	11520.500000	0.0	0.0	10.000000	1.000000	0.850559	1.940848e-
max	8.924206e+17	79515.000000	132810.000000	0.0	0.0	150.000000	4.000000	1.000000	4.880140e-

2.Based on the next graph, the In mayority, people defined dogs as pupper with 221 puppers and floofler was the least used with 8 floofers.
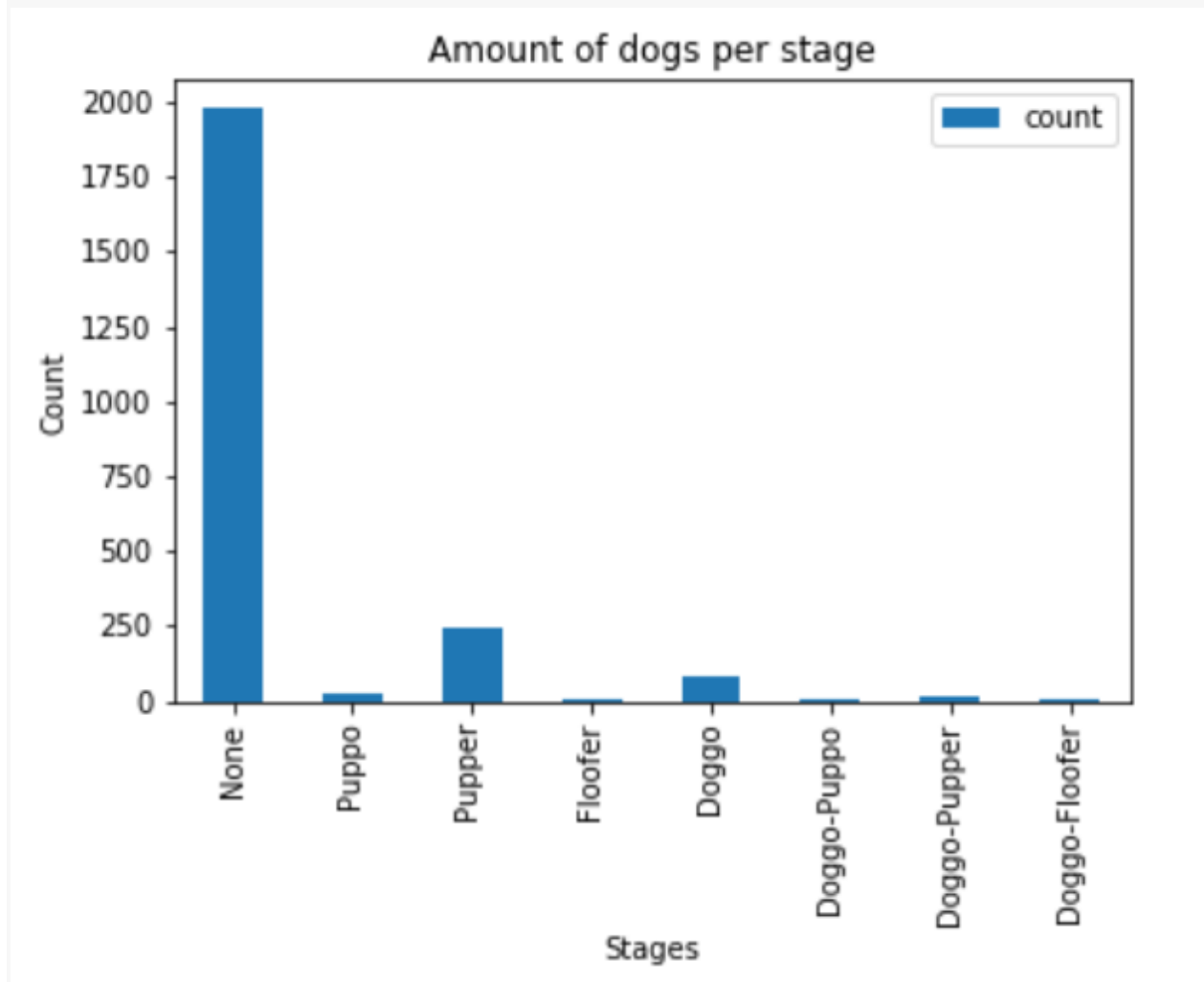
This graph shows the amount of dogs per stage of dog and the amount of tweets that were not described as any stage. The total amount od stages of dogs are 380 out of 2356 values. The most common stage is Pupper with 245, and the least is doggo-floofer and doggo-puppo, both with 1.

3. The algorithm to predict the breed of dog, the variable p1_dog, p2_dog and p3_dog would define whether or not the prediction is a breed of dog based on confidence. This algorithm would not always come up with a breed of dog, since all predictions also had False results. On each prediction the algorith would not come up with the dogs breed on around a 3th of the dogs..

```
In [92]: merged.p1_dog.value_counts()

Out[92]: True     1407
         False     489
         Name: p1_dog, dtype: int64

In [93]: merged.p2_dog.value_counts()

Out[93]: True     1425
         False     471
         Name: p2_dog, dtype: int64

In [94]: merged.p3_dog.value_counts()

Out[94]: True     1383
         False     513
         Name: p3_dog, dtype: int64
```