ACT_REPORT.PDF

INTRODUCTION:

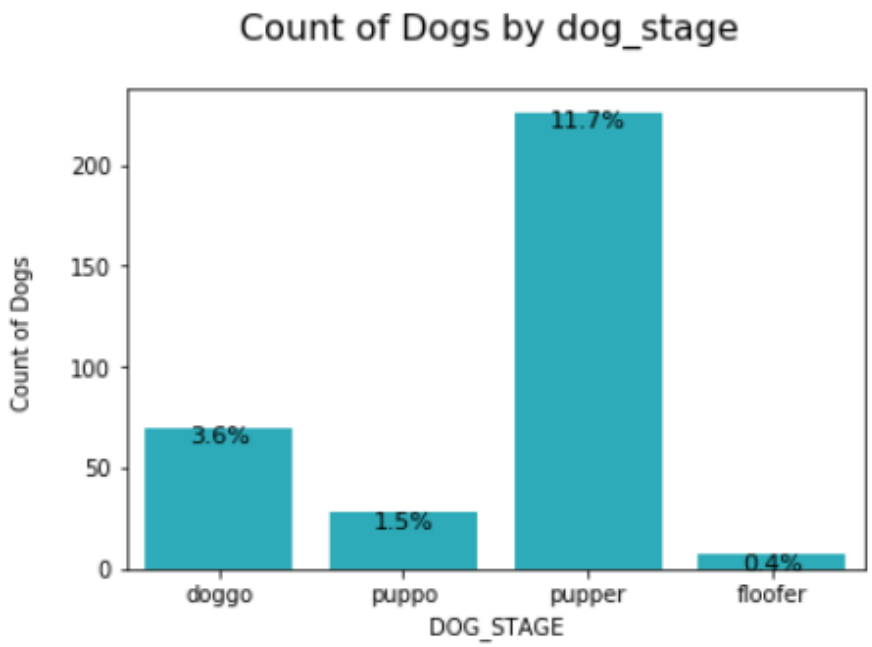
In this report, I will clean and analyze the WeRateDogs Twitter Archive dataset. To further aid my analysis, I will make use image_predicton dataset as well as the dataset from twitter api. In the end, I will come up with 3 insights and a visualization of my findings.

Insights:

1. Pupper is the most common dog stage
2. Charlie, Oliver, Cooper, Lucy and Tucker are the most popular dog names
3. Favorites and retweet counts are well correlated

Insights 1: Pupper is the most common dog stage

There 4 dog growth stages, Pupper is the most common dog stage



According to this bar count we can say that the dog more mentioned in the text tweet is in the Pupper stage having the highest proportion 11.7%.

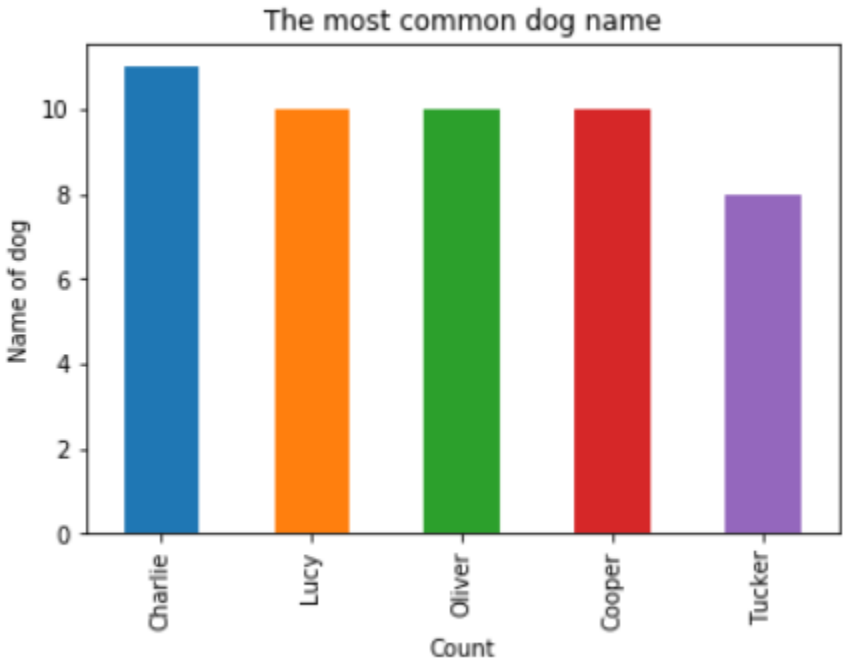
pupper 226
doggo 70
puppo 28
floofer 7
Name: dog_stage, dtype: int64

After grouping the rating ratio's mean for each dog stage, we have the result above which shows that the dog stage having the highest rate is Puppo.

Insights 2: Charlie, Oliver, Cooper, Lucy and Tucker are the most popular dog names

Charlie is the most popular dog name closely followed Lucy, Oliver, Cooper and Tucker

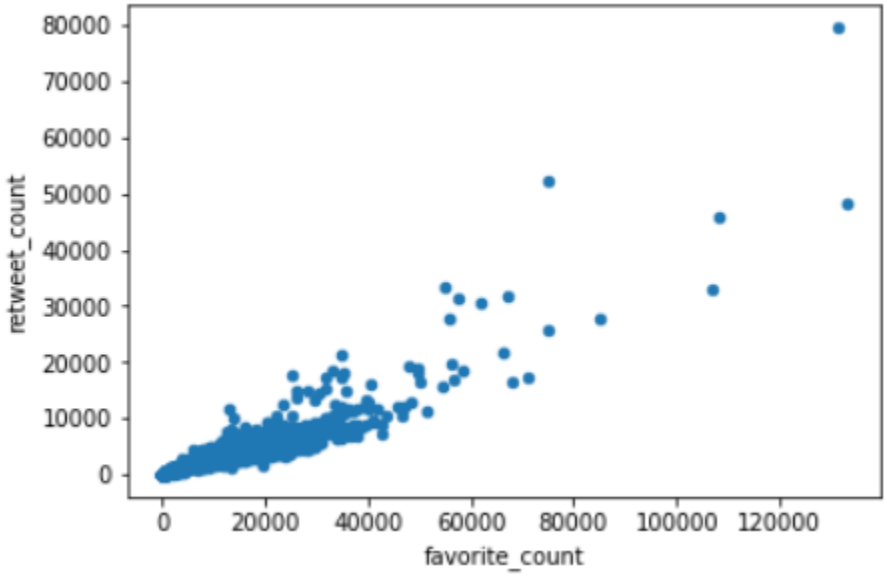
Charlie	11
Lucy	10
Oliver	10
Cooper	10
Tucker	8



Insights 3: Favorites and retweet counts are well correlated

Based on retweets and favorites count, there is a positive correlation between both values.

	retweet_count	favorite_count
count	1928.000000	1928.000000
mean	2514.632261	8296.298237
std	4187.214895	11371.277140
min	16.000000	81.000000
25%	611.750000	1873.750000
50%	1299.000000	3904.000000
75%	2911.250000	10280.000000
max	79515.000000	132810.000000



Looking at the scatter plot above we can see clearly that the retweet count and favorite count are positively correlated, when favorite count increases retweet count increases too.