

Analysis and Insight Report

September 17, 2020

0.1 Introduction

The process of analyzing and insight was done on the cleaned version of the twitter data set, whose raw data was extracted from the twitter user @dog_rates. This process required formation of questions and finding that answers through rigorous process of coding and plotting graphs to convey the insights.

0.2 Visualization

0.2.1 Insight 1 : The most frequent dog breed that was tweeted about

The first insight was to know which dog breed was the most frequently tweeted by the @dog_rates user.

Their absolute frequency were plotted and a bar graph was obtained as shown in the figure `dog_breed_abs_freq_graph`:

Conclusion: For this problem, we only visualized the top 15 most frequent dog breeds and found that Golden retriever was the most tweeted breed type followed by Labrador retriever, Pembroke, Chihuahua and pug, completing the top 5 frequently tweeted dog breeds.

Again, to see how they fare when plotted based on relative frequency, another barplot was plotted. Surprisingly, the most frequent dog breed is very less compared to total number of tweets, which refers that the tweets contained wide variety of dog breeds. The graph for relative frequency is plotted in the figure `dog_breed_relative_freq_graph`:

0.2.2 Insight 2 : Rating over the years

The ratings was calculated based on the ratio of its numerator and denominator ratings.

For task, the year was extracted from the timestamp column. Then ratings and years were combined to make a new dataframe and that dataframe was used to perform analysis and plot graph. A pointplot of seaborn library was used to observe the change of ratings from 2015 to 2017. The graph is shown in the figure `ratings_over_the_year`.

Conclusion: The rating increased from 2015 to 2017 with almost 0.6 difference on their average value.

0.2.3 Insight 3: relation between dogs stages , favourite_count, and retweet_count

Question 1: Which breed and stage name of dogs has the highest ever favorite and retweet count in a single tweet?

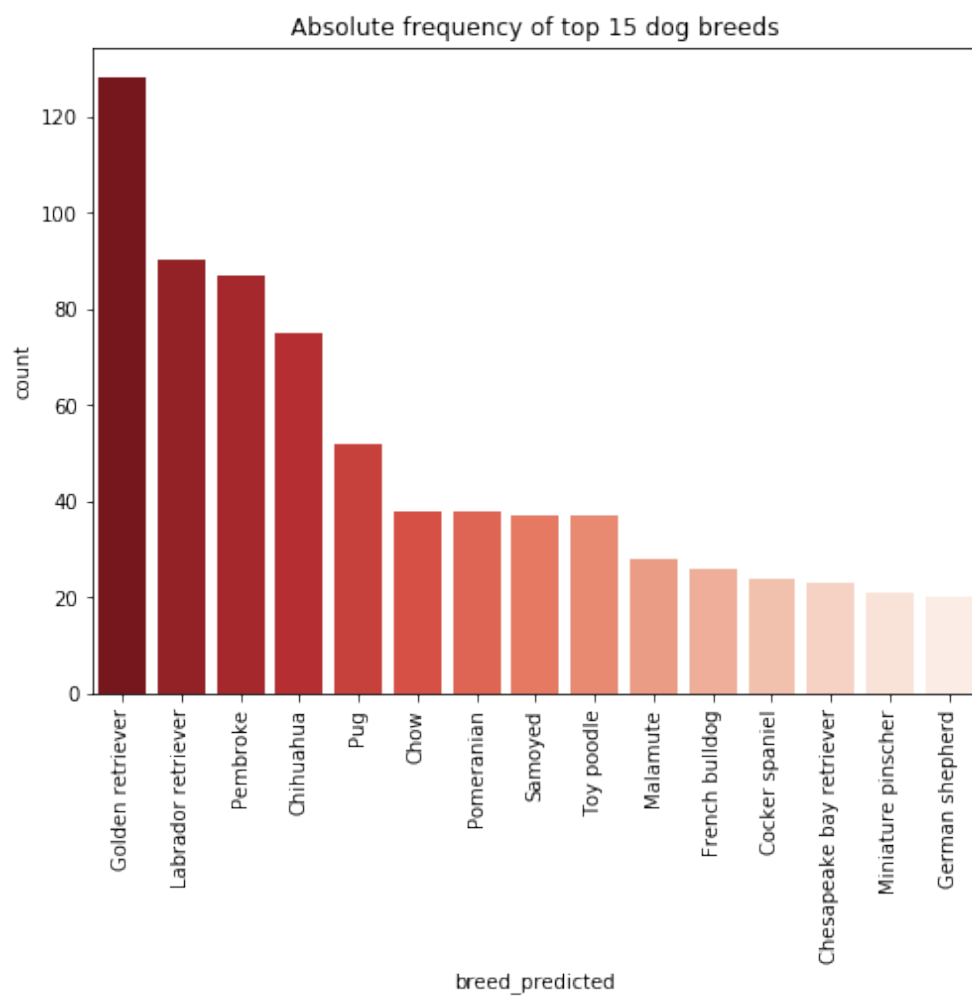


Fig: dog_breed_abs_freq_graph

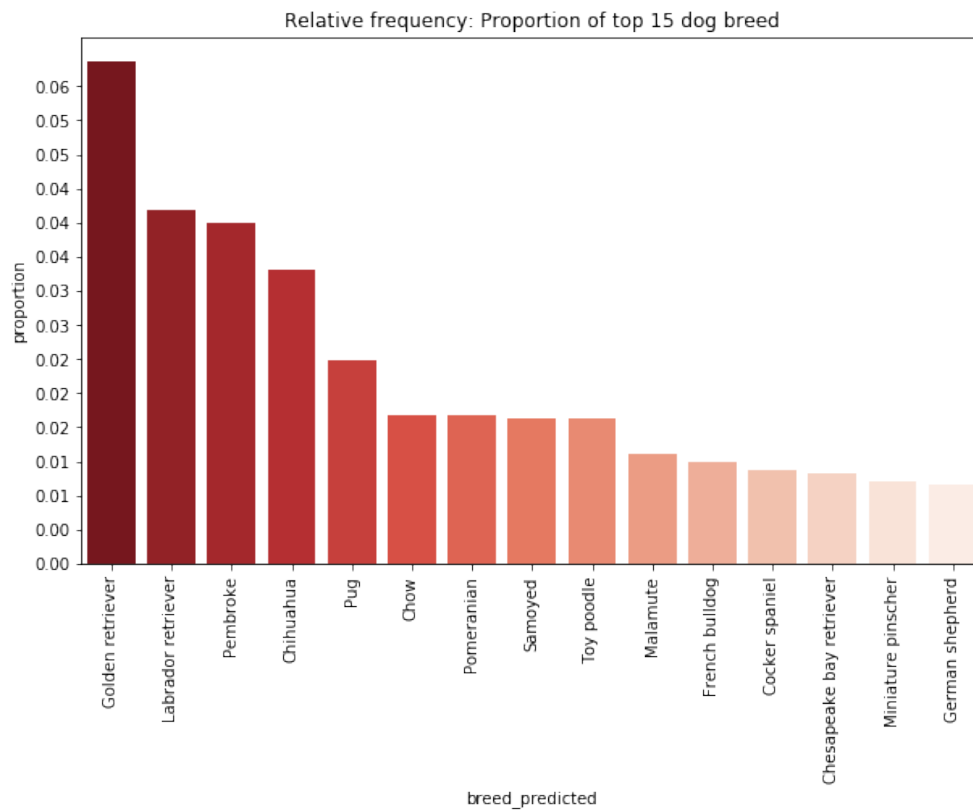


Fig: dog_breed_relative_freq_graph

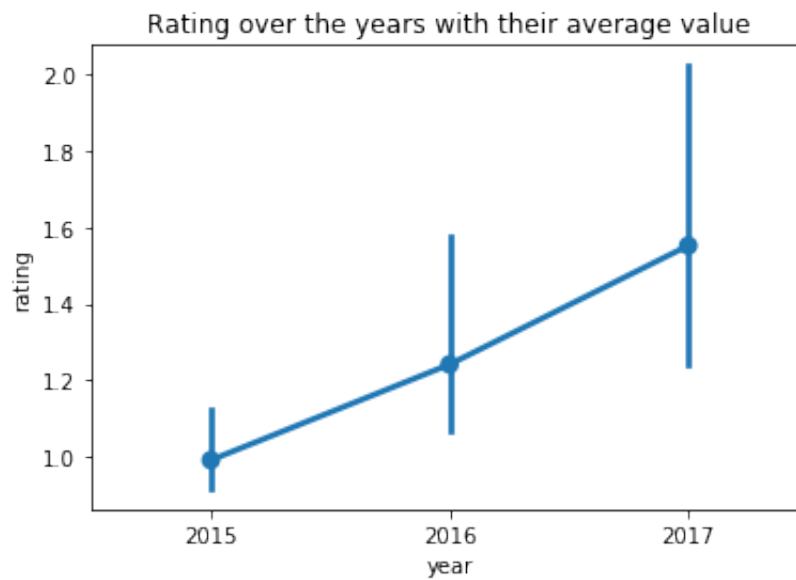


Fig: ratings_over_the_year

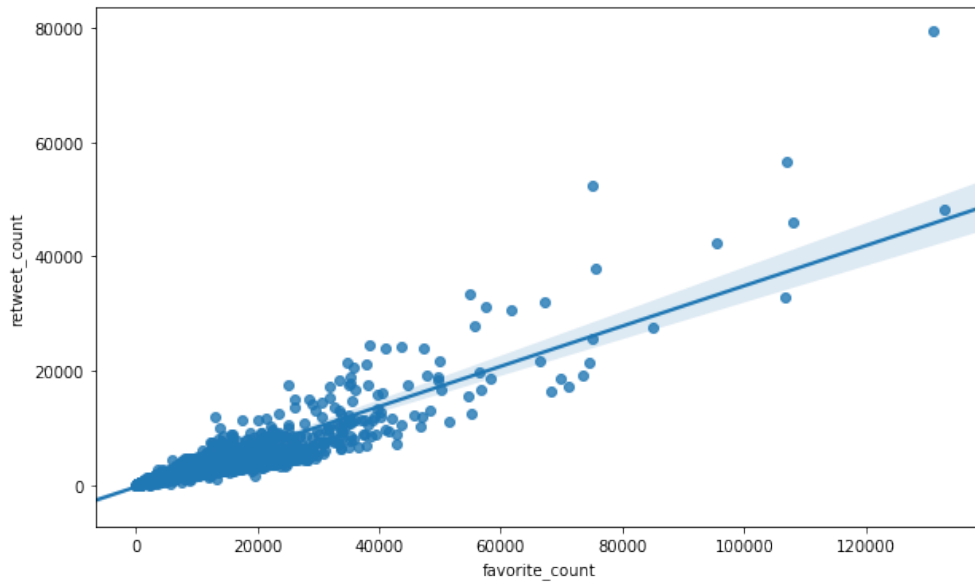


Fig: reg_plot_fav_retweet

Answer The analysis was performed and the following conclusion was made: - Breed and stage name of dogs with the highest ever favorite count is Lakeland terrier and puppo respectively. - Breed and stage name of dogs with the highest ever retweet count is Labrador retriever and doggo respectively.

Question 2: How are favorite count and retweet count related?

The regression plot was plotted to compare the relationships between favorite_count and retweet_count of each of the tweets and the following graph was obtained as shown in figure reg_plot_fav_retweet.

Conclusion - There is positive correlation (i.e. 0.912) between favorite_count and retweet_count showing that any person who like a tweet is very likely to retweet the very tweet.