Data Analysis and Visualizations

Introduction

This document contains the insights that are derived from the analysis made using the cleaned data.
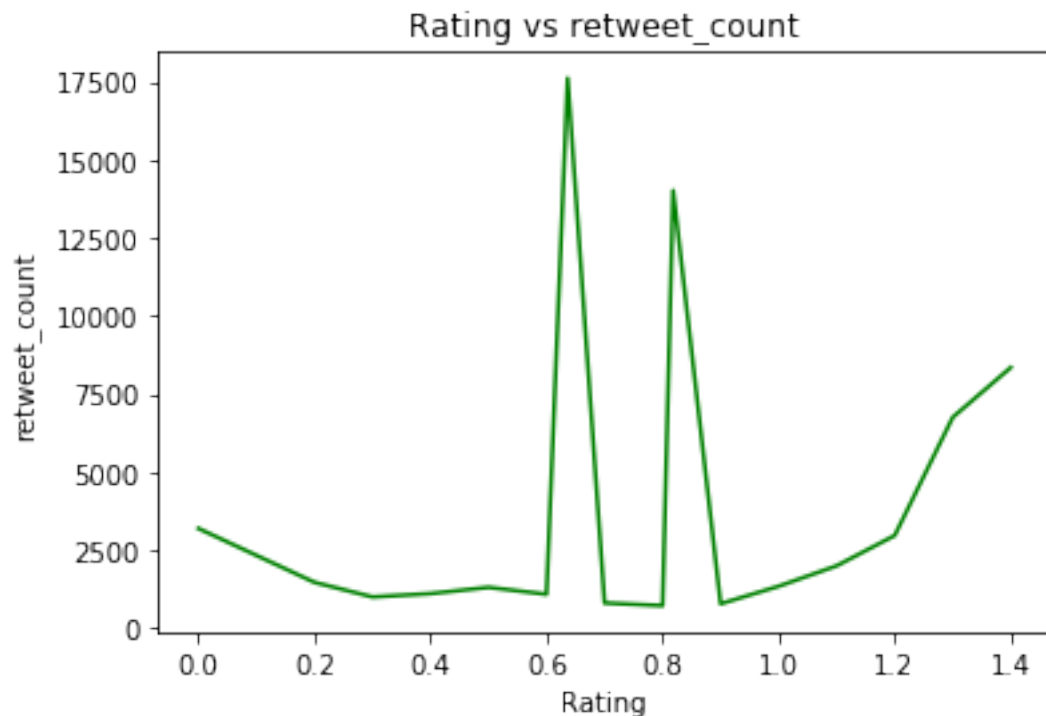
Analysis

1. **Most favorite dog breeds** – The most popular dog breed (by number of tweets) seems to be the golden retriever followed by Labrador retriever. The fact that the retriever are the most common and most popular breeds seems in the united states seems to be true.

| Breed	Count
Golden retriever	155
Labrador retriever	103
Pembroke	95
Chihuahua	91
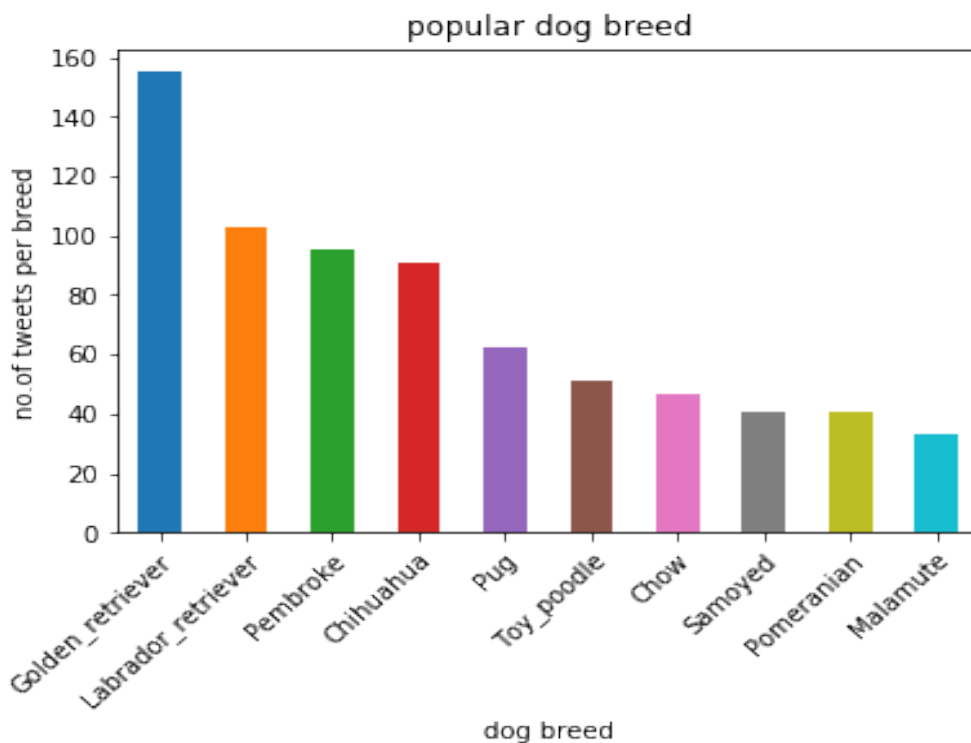
2. **Rating vs Retweet Count Relationship -** Each Rating (rating is num/den rating for each dog) is grouped and the mean of "retweets count" for each rating is calculated. From the analysis, it looks like there are some data issues (parsing issues) especially with ratings ranging from 0 – 1.0. This very reason might have caused to have spikes of higher "retweet count" values for ratings around 0.6 to 0.9. Apart from that in general, the trends seem to be, the higher the rating is, the more the retweet counts are starting from a rating of 1.0.

Rating vs retweet_count

3. **Dog stage vs Rating relationship -** Each of the dog stages are grouped and the mean of the total ratings is calculated. From the analysis it can be determined that the floofer dog stages has the highest mean rating. It can be derived from the analysis that the floofer (fluffier)- meaning the bigger, the chubbier or the furrier the dogs are, the more rating they get.

Dog_stage	Rating
doggo	1.182812
pupper	1.200000
floofer	1.086228
puppo	1.195238

Rating vs dog_stage
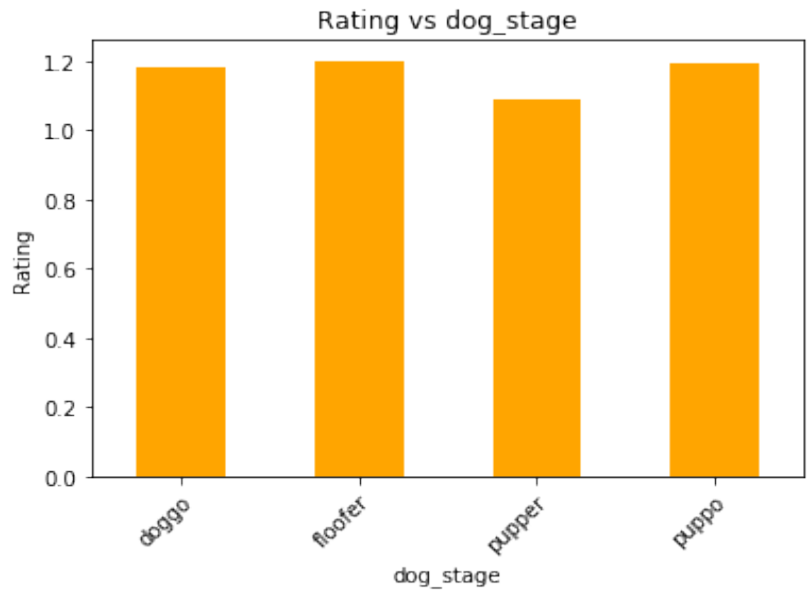
4. **What is the most common source for a tweet** - From the analysis, it can be clearly seen that the most common source of twitter activity is coming iphone. That totally makes sense that with the emergence of smart phones into today's world, most people tend to us their smart phone to do many things, twitter being one of it.

Source	Count
iphone	1636
Twitter Web	22
Twitter Deck	8

Source count