

ANALYSIS AND INSIGHTS

Udacity DAND: Wrangle and Analyze Data Project

By: Vatsal Prakash

INTRODUCTION

This project called Wrangle and Analyze Data is part of the Udacity's Data Analyst Nanodegree program. The project involves gathering, assessing and cleaning data from various sources. All the data is associated with the tweets from WeRateDogs. The ratings for WeRateDogs are often higher than 10/10 which is a part of their popularity. Once the data was scraped from various sources, it was visually and programmatically assessed for quality and tidiness issues. These issues were then cleaned. Finally, some visualizations were created that can be found below.

RETWEETS vs FAVORITES

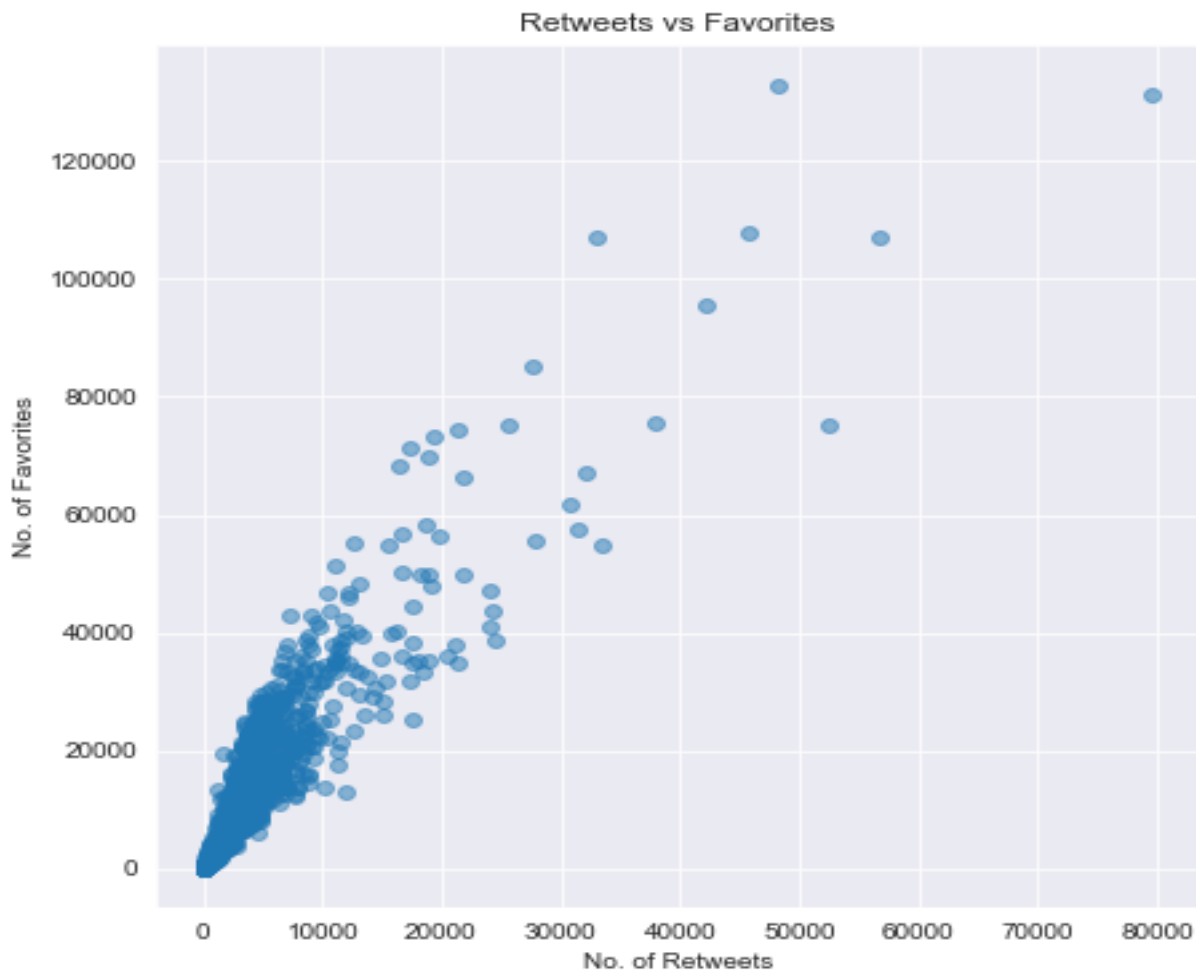


FIGURE 1. Retweet count vs Favorite count of 2098 tweets from WeRateDogs

From this figure we can infer that there is a positive correlation between the number of retweets and the number of times the tweet is favorited. Majority of the data lies below 30,000 retweets and 60,000 favorites.

DOG RATINGS

For our analysis a new column was created that contained the dog ratings which was obtained by dividing the rating_numerator column by rating_denominator column. WeRateDogs is pretty popular for having greater numerator rating than the denominator and most of the dogs receive a rating of 11 or 12 as they're 'Good Dogs'

Source:- <https://knowyourmeme.com/memes/theyre-good-dogs-brent>

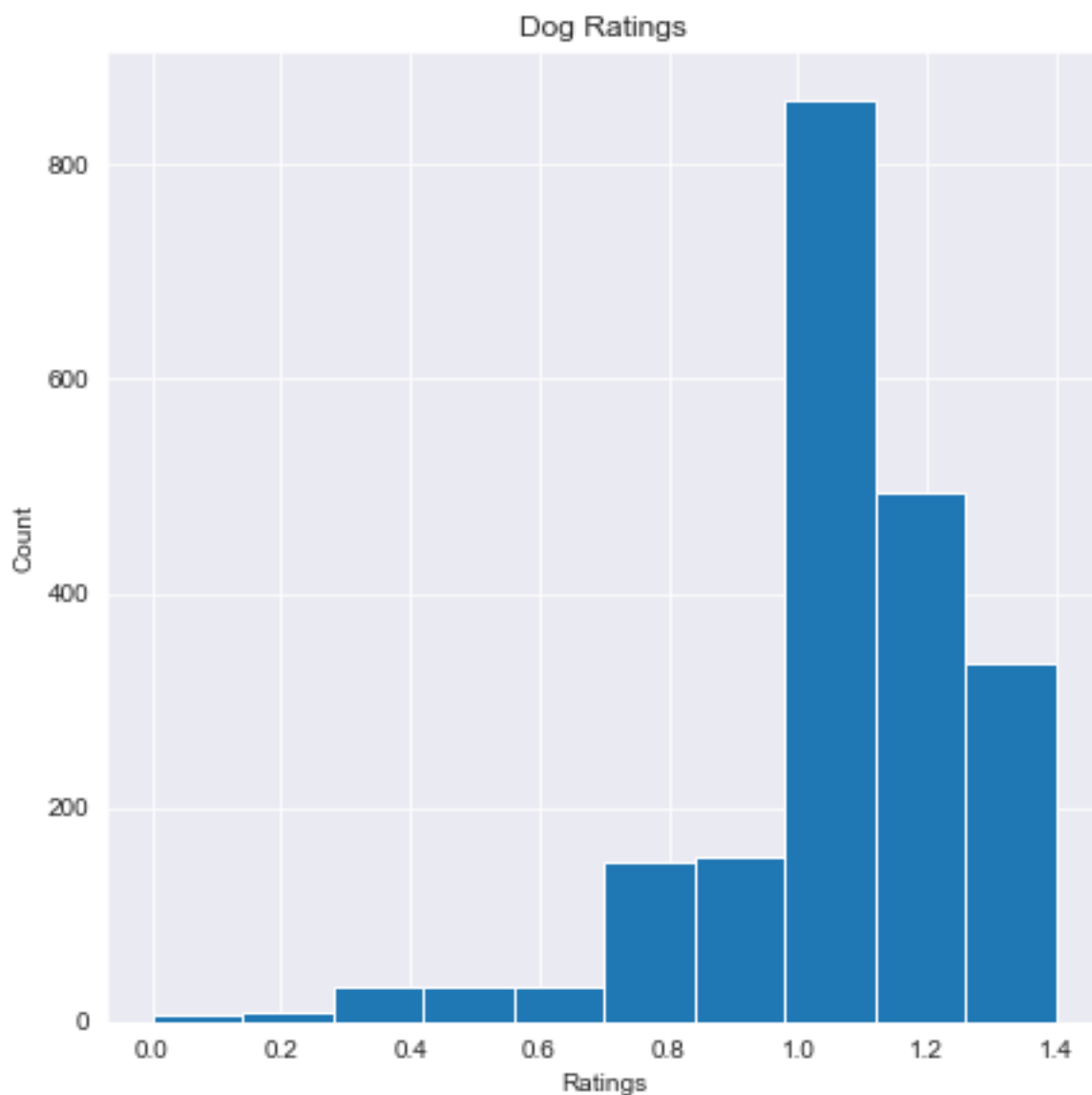


FIGURE 2. Histogram of 2098 dog ratings

We see that majority of the rating falls between 1 and 1.2 with the histogram being skewed to the left. As most of the dogs get 11s and 12s out of 10 this confirms that they're 'Good Dogs'

Top 10 dog breeds from top 1000 favorite and retweet posts

The algorithm's #1 prediction(p1) from image_predictions dataset was used to find out the top 10 breeds of most retweeted and favorite posts.

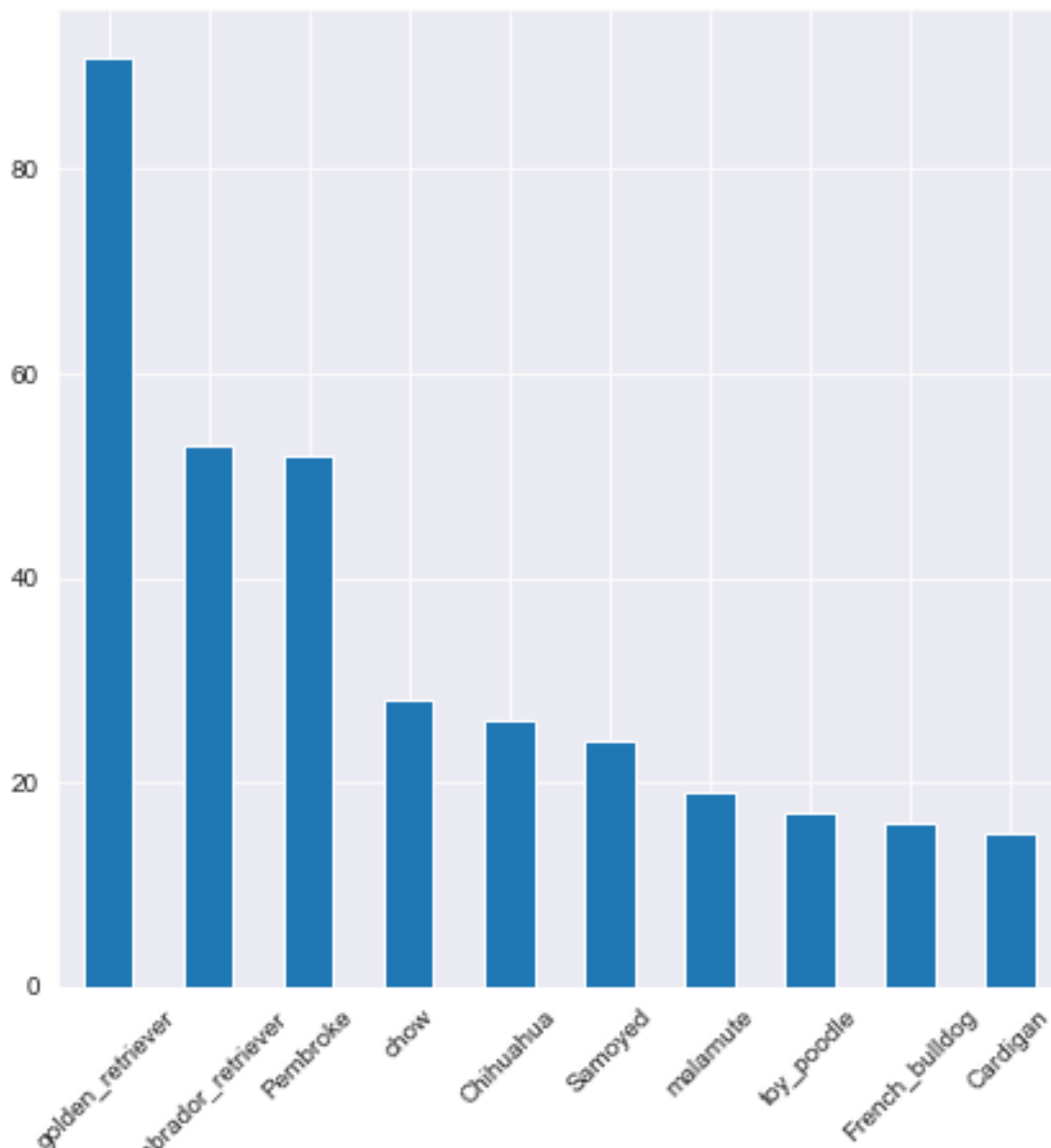


FIGURE 3. Bar chart of top 10 dog breeds from 1000 favorite posts

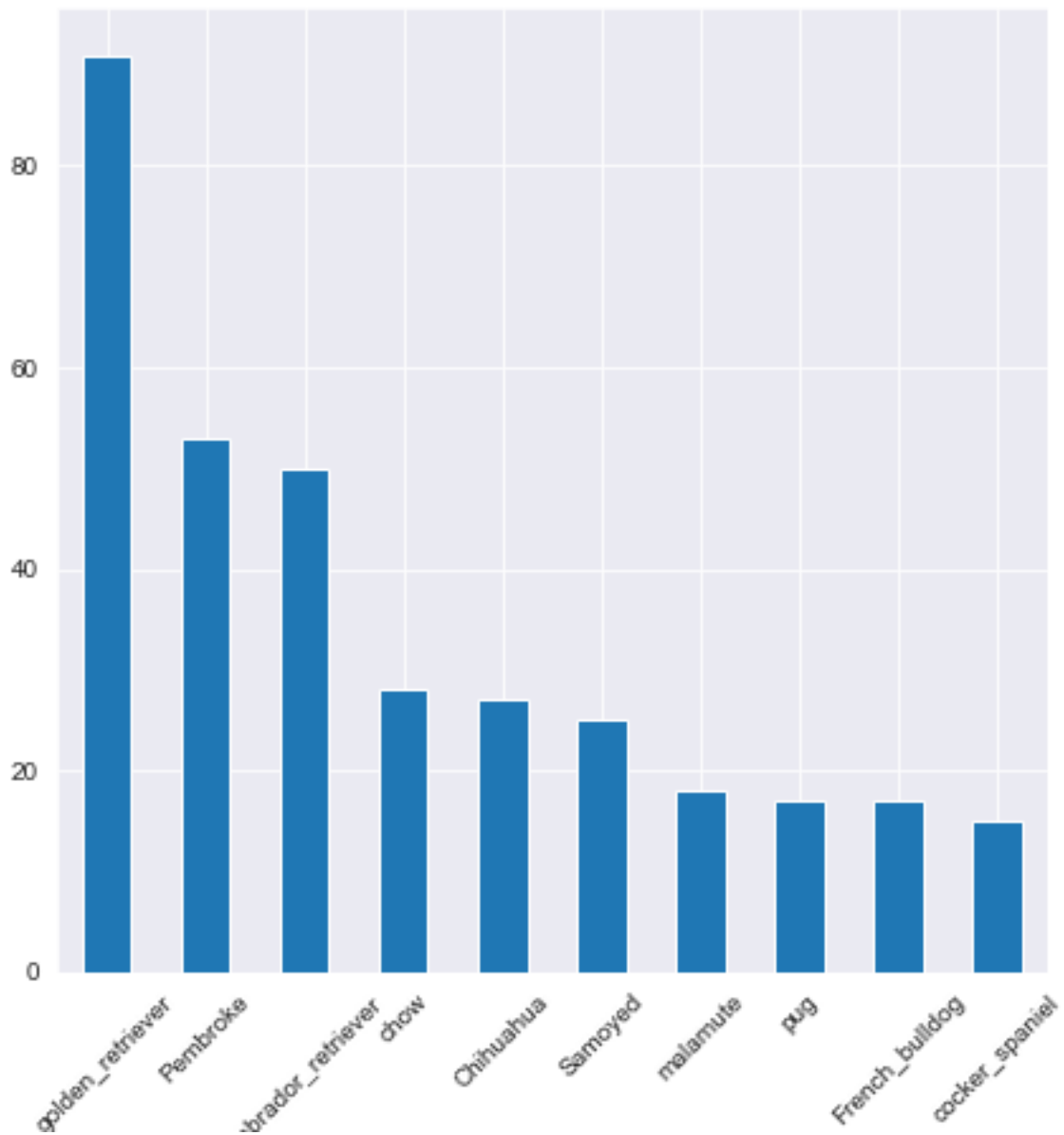


FIGURE 4. Bar chart of top 10 dog breeds from 1000 retweeted posts

It can be inferred from the data that the most favored breed is golden retriever, followed by Labrador and Pembroke while the most retweeted breed is also golden retriever followed by Pembroke and Labrador. Chow, Chihuahua also seems to be one of the top 10 most favored and retweeted dogs.