

act_report

August 19, 2019

1 act_report

Introduction

This Wrangle and Analyze Data Project is part of Udacity's Data Analyst Nanodegree.

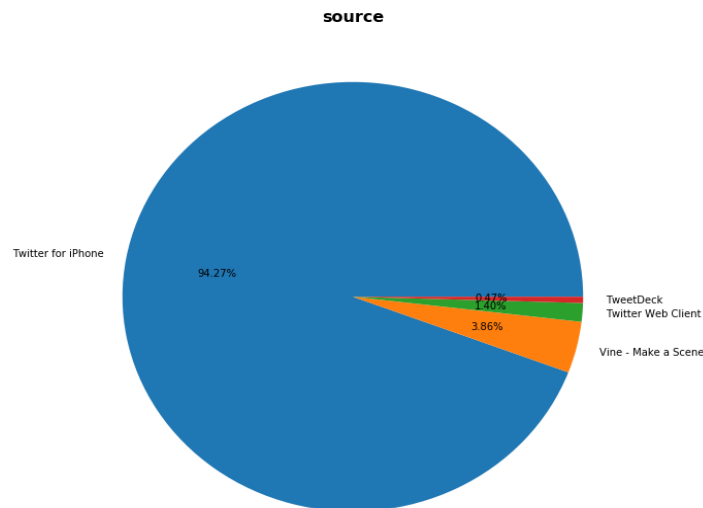
The project involves wrangling of data from various sources associated with tweets from the Twitter user @dog_rates, also known as WeRateDogs. After scraping together the data, quality and tidiness issues were assessed and then cleaned.

The idea behind the WeRateDogs account is that they ask people to send them photos of their dogs, and they will rate them on a scale of 1-10 with humorous comments.

Initially upon gathering we had 2356 observations but doing this project I really understand why data wrangling is so important as a lot of observations are repetitive, inappropriate, incomplete and messy but as project requirements only to clean up 8 quality issues and 2 tidiness issues so I can say I have completed somewhat.

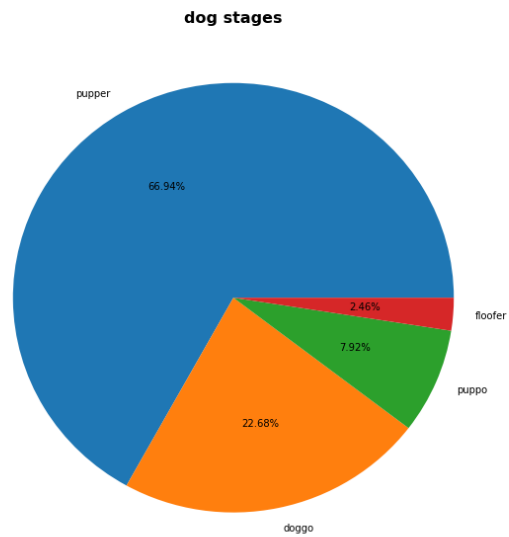
During analysis I found retweet and favourite have strong correlation as it's the nature of the Twitter platform, because if a tweet gets liked then it makes viral soon and then a lot of counts for them.

1-Most used Twitter source



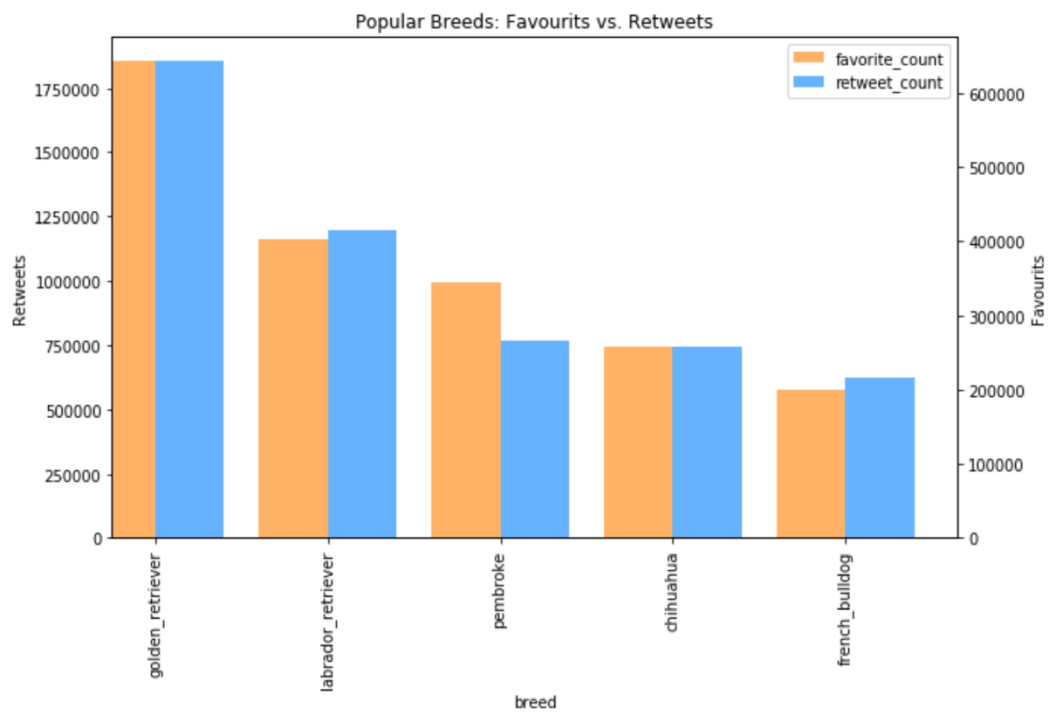
- Almost all tweets are sent via iPhone

2- Most Common Stage



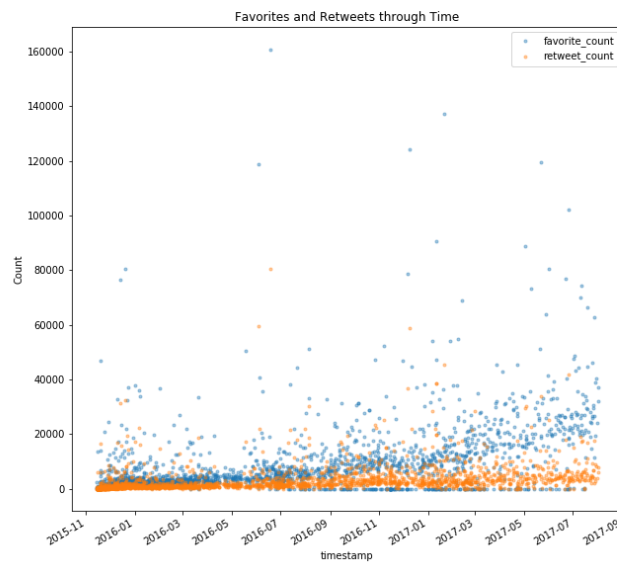
- Pupper is most common stage of dogs

3- The Most Popular Breed.



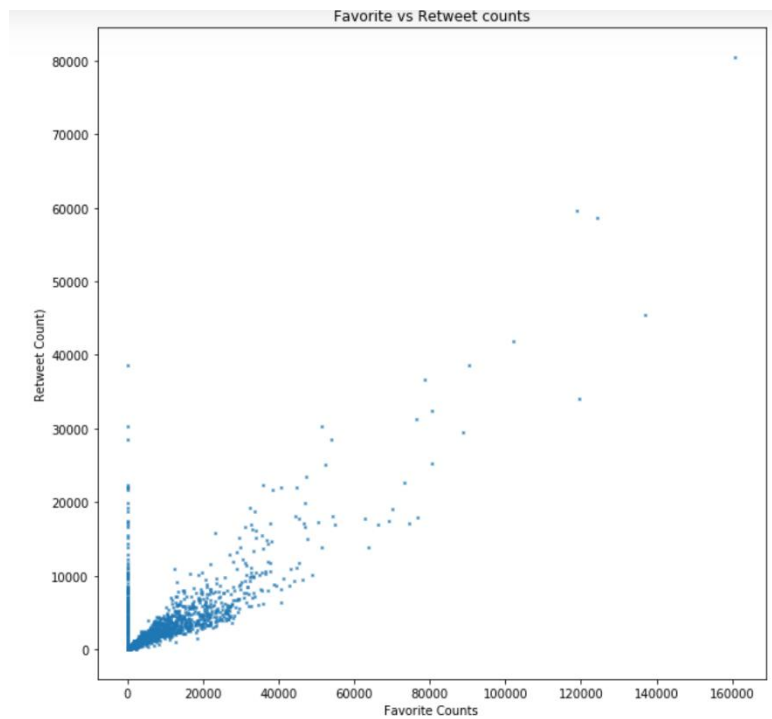
- Golden retriever is the most popular breed

4-Favorites and Retweets through Time



- The account got popularity over time More and more pictures got likes.
- Number of favorite_count is bigger than retweet_count

5- Corr. between retweets and favourites



Above seems like a strong relationship but obviously if tweet get liked then that's the one get retweeted thousands of times.

6- What is the most popular name (favorites, retweets)?

- Bo is the most popular name with 113720 retweets and 175340 favourites