ANALYSIS AND VISUALIZATION OF WeRateDogs Dataset

Introduction

This project involves wrangling of various datasets associated with tweets from WeRateDogs.

About Data set:

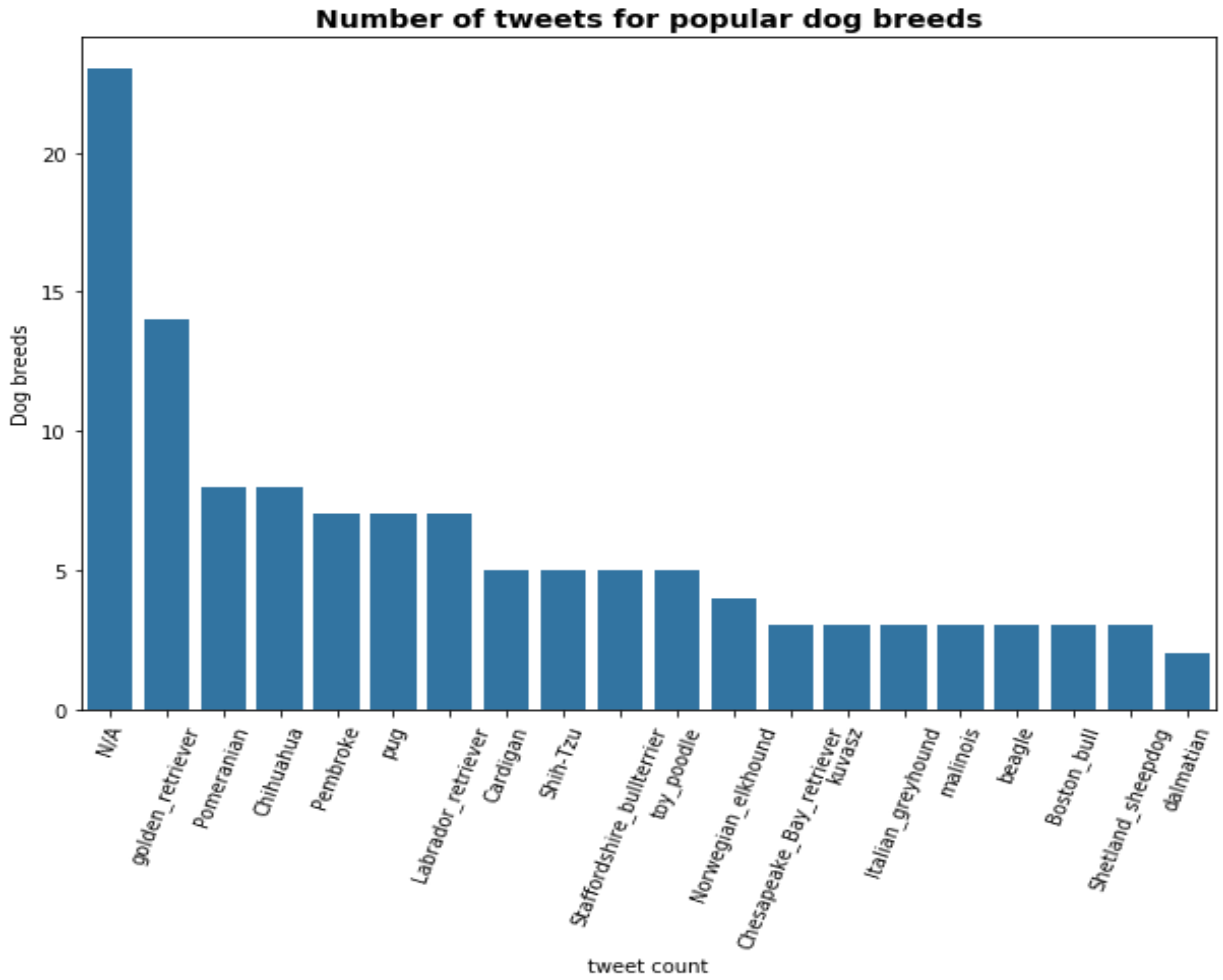
- WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. The account was started in 2015 by college student Matt Nelson, and has received international media attention for its popularity.
- WeRateDogs asks people to send photos of their dogs, then tweets selected photos rating and a humorous comment. Dogs are rated on a scale of one to ten, but are invariably given ratings in excess of the maximum, such as "13/10".

The dataset after cleaning had 2100+ rows. Each row is associated with a unique tweet and doesn't contain retweets.

After the data gathering, assessment, cleaning processes, I did some analysis on datasets obtained and here are the findings:

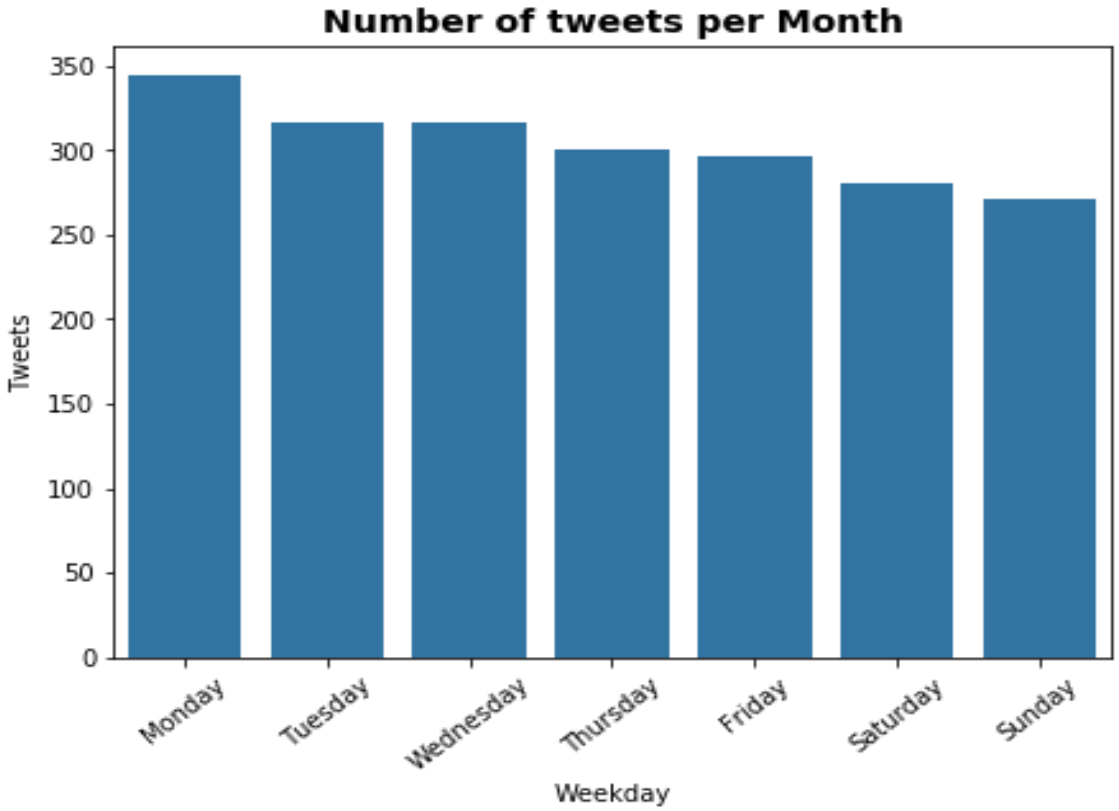
1. NUMBER OF TWEETS ASSOCIATED WITH DIFFERENT DOG TYPES

In all these tweets around 60+ dog breeds were mentioned. I have analysed counts of top 20 breeds by count. Ignoring tweets for which breed information is not present, we can see that Golden Retriever is the most tweeted dog breed.



2. NUMBER OF TWEETS ON DIFFERENT DAYS OF WEEK

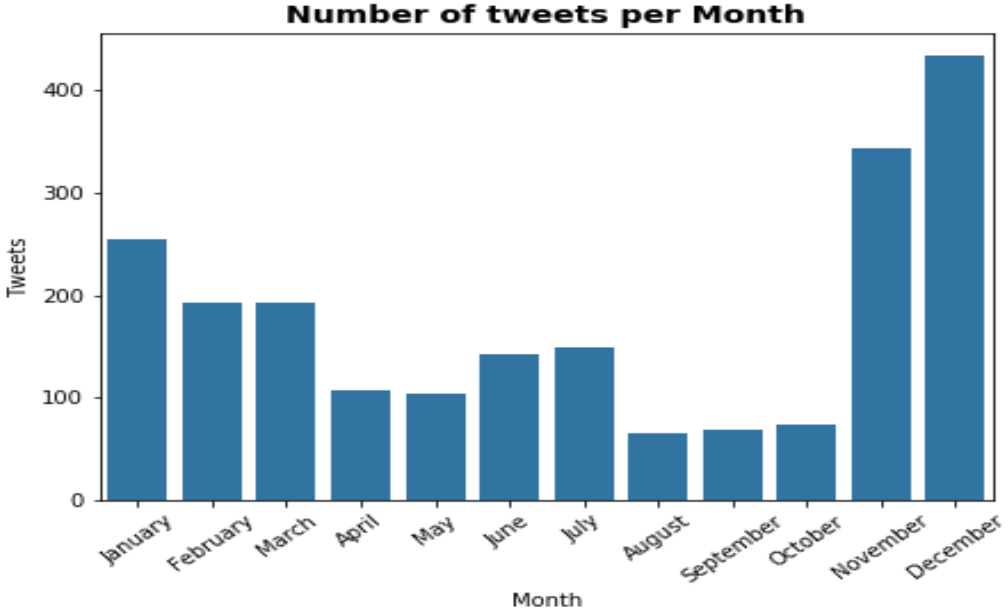
Let's take a look at the number of tweets on each weekday.We would like to see if there are more dog ratings on one day than others.



The number of tweets on Monday are the highest. But the distribution is almost equal. We can't say number of tweets on a particular day are going to be more than others.

3. NUMBER OF TWEETS ON DIFFERENT MONTHS

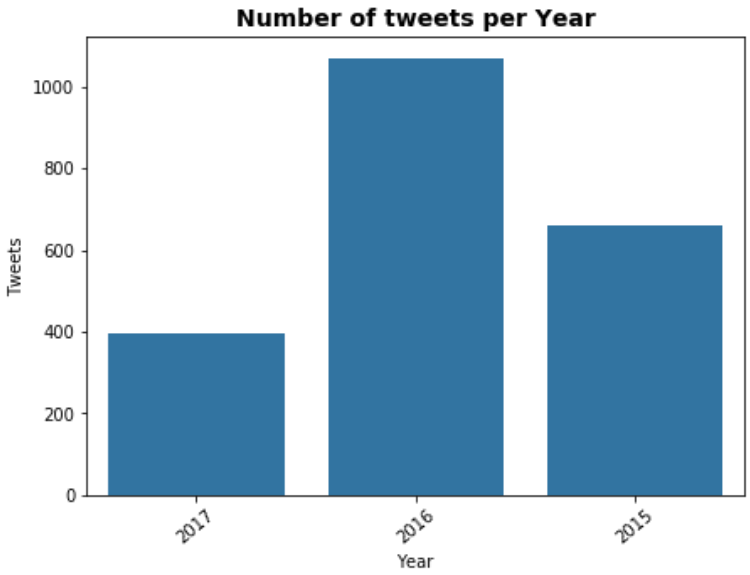
Let's take a look at the number of tweets on different months of the year. We would like to see if there are more dog ratings in one month than others.



The number of tweets in the month of December are the highest followed by November and January. It may be because people spend more time with family and friends on holidays and which results in a lot of Doggo Tweets. As we haven't run statistical analysis hence these are inferences just by visual analysis.

4. NUMBER OF TWEETS ON DIFFERENT YEARS

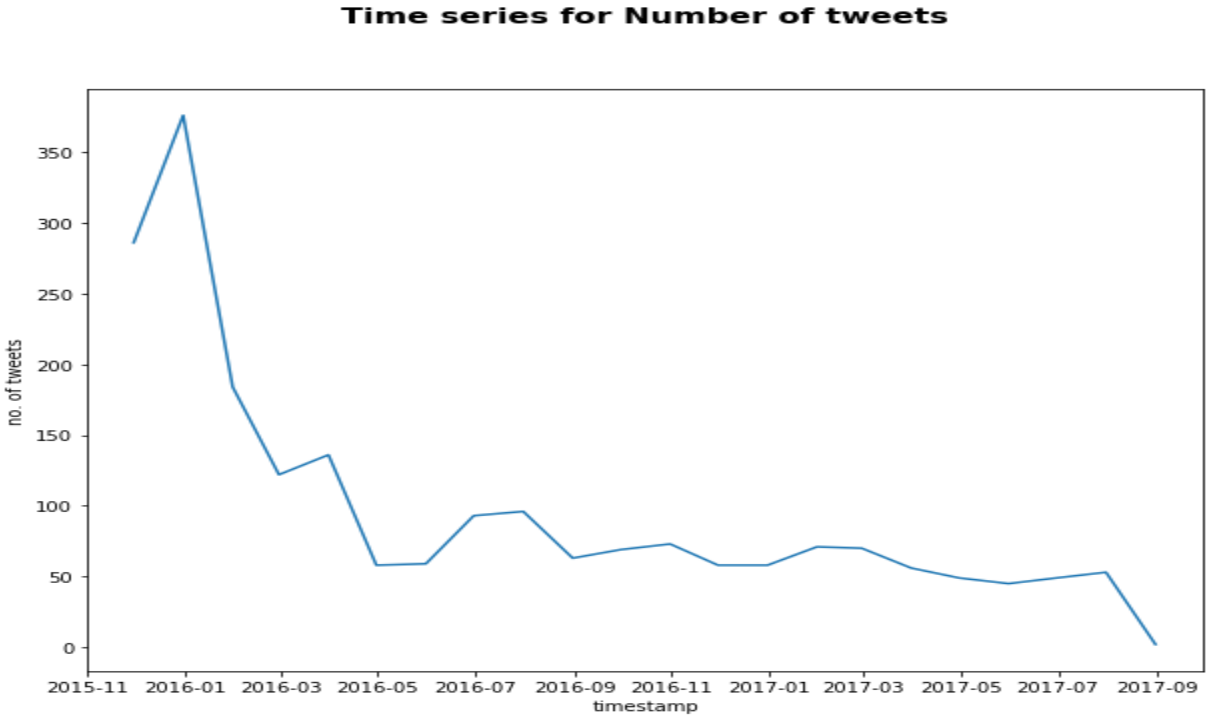
Let's take a look at the number of tweets in years. We would like to see if there are more dog ratings in one one year than the other.



The Number of Tweets were highest in 2016.

4. TWEETS OVER TIME

Let's see the trend in the number of tweets by WeRateDogs over the span of 3 years.



We can see that the number of tweets were highest in January 2016. After this the tweet count started decreasing gradually with small spikes in April 2016 and August 2016.