

Project: Wrangle and Analyze Twitter Data

By: Maithili Desai

Data Analysis

This section consists of insights on the dataset gathered, assessed and cleaned above. Below are the questions on which this analysis section is focused

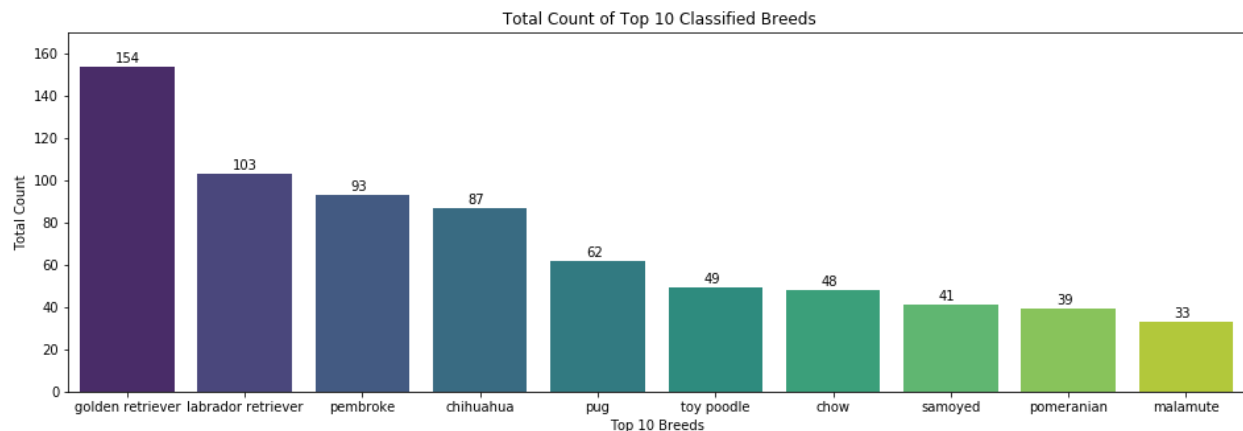
Questions:

1. Which breed is the most popular among WeRateDogs tweets and retweets?
2. How is the account activity throughout the time period available in the dataset?
3. What are the most active hours for the twitter account?
4. Which is the most common source used to post the tweets?

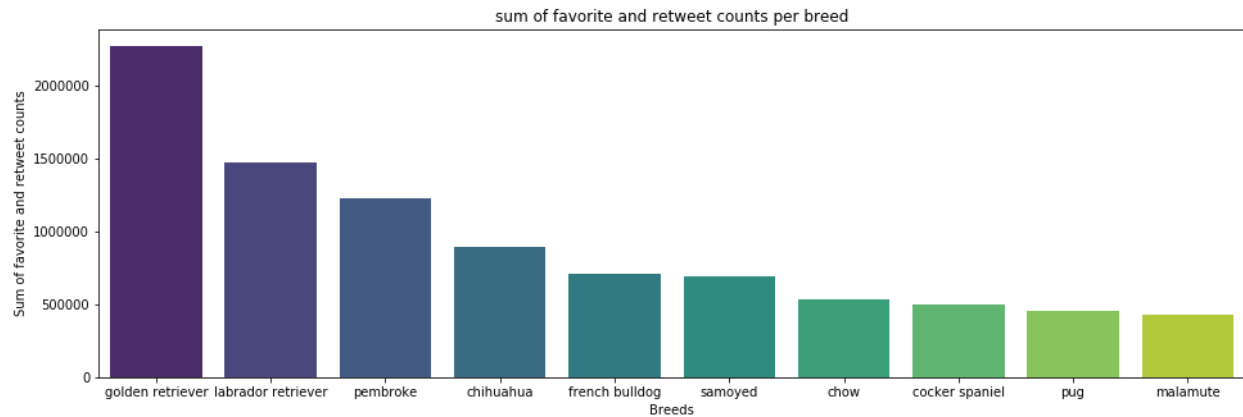
1. Which breed is the most popular among WeRateDogs tweets and retweets?

I have answered this question based on individual and aggregated values for the following:

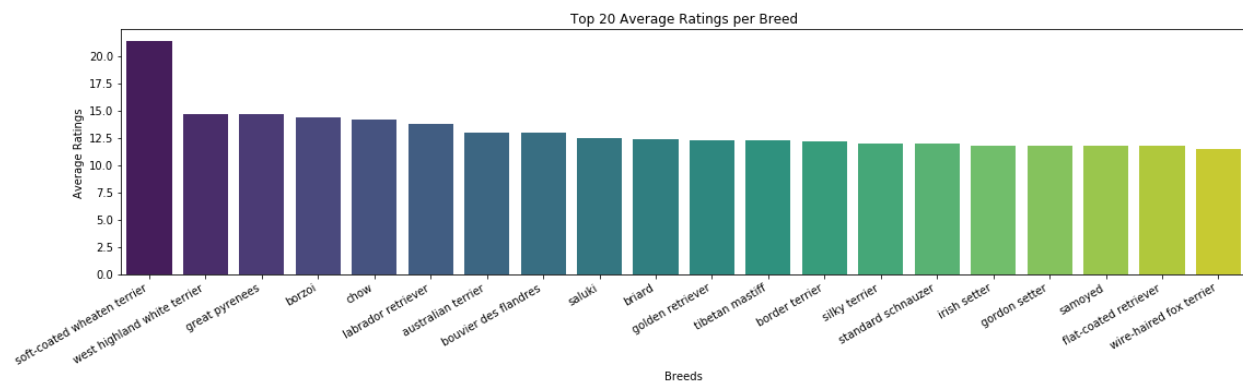
- Retweet counts and favourite counts
- Ratings



Insight: The distribution shows that the images in the dataset are mostly of Golden retrievers with a count of 154 and Labrador retrievers with a count of 103



Insight: Based on the sum of retweets and favorite counts, the top two are still: Golden retrievers and Labrador retrievers

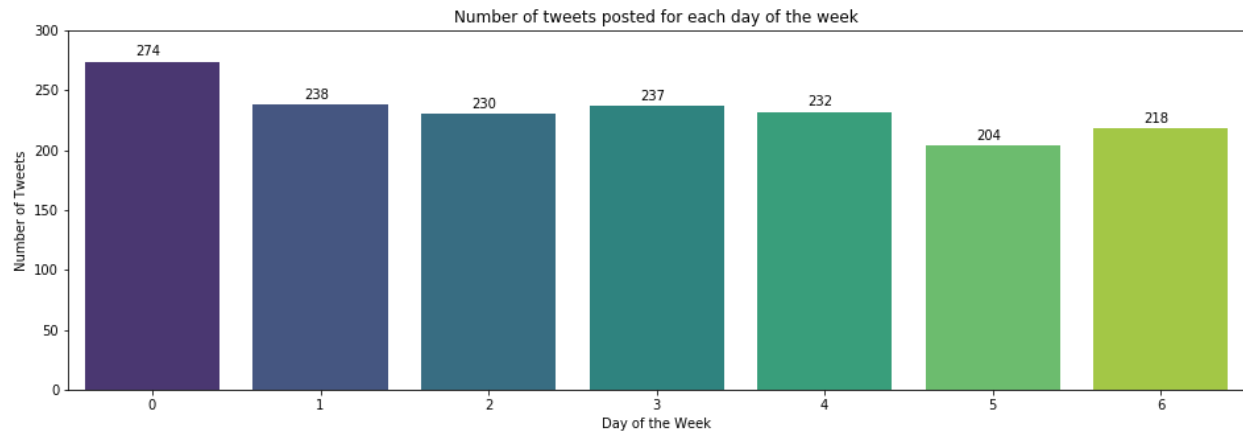


Insight: Most breeds follow a similar average range of ratings with an outlier of soft-coated wheaten terrier with a very high average rating of 21.35. To check this is not by chance we see the total number of soft-coated wheaten terrier dogs and the highest rating received by this breed.

2. How is the account activity throughout the time period available in the dataset?

I am going to analyse the account activity based on the following parameters:

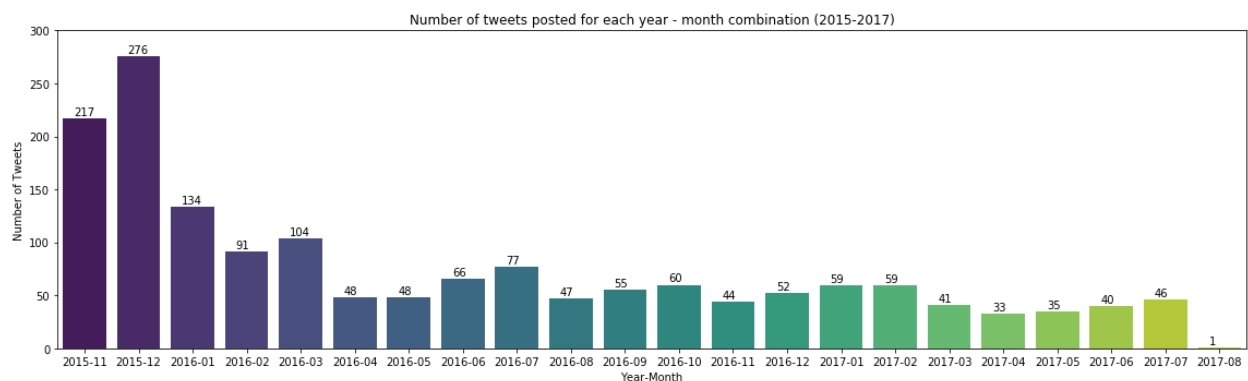
- Popular day of the week
- Popular month
- Popular month-year
- Retweet count
- Favorite count
- Number of uploaded images
- Tweet text length



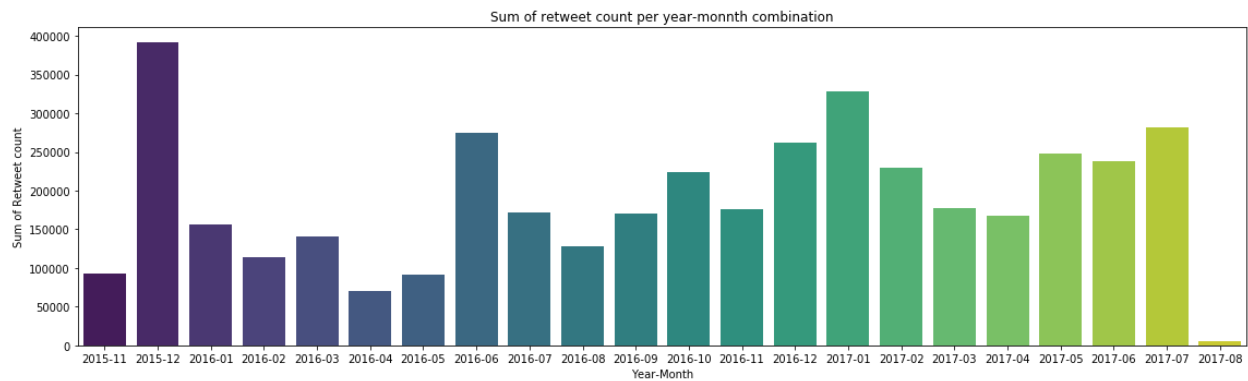
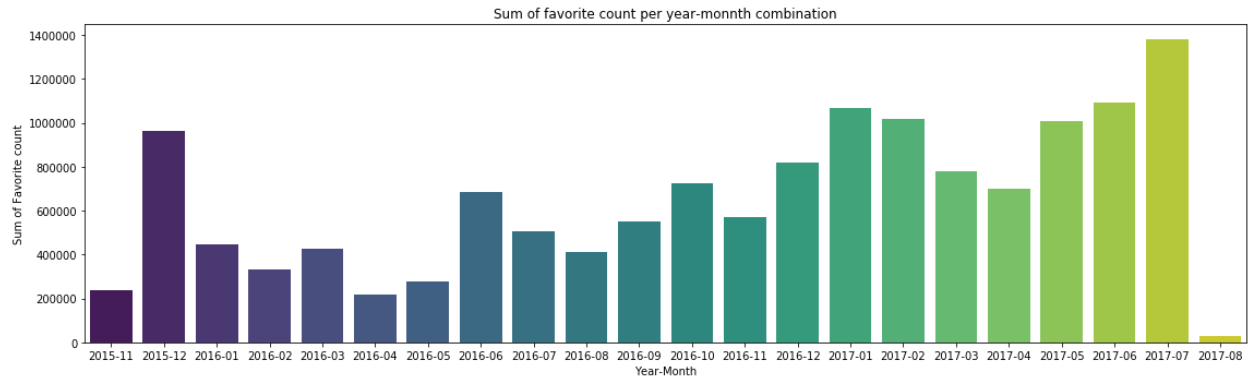
This plot shows that

- Most tweets have been posted on Monday
- Constant number of tweet from Tuesday through Friday
- Weekends show comparatively lower tweets posted

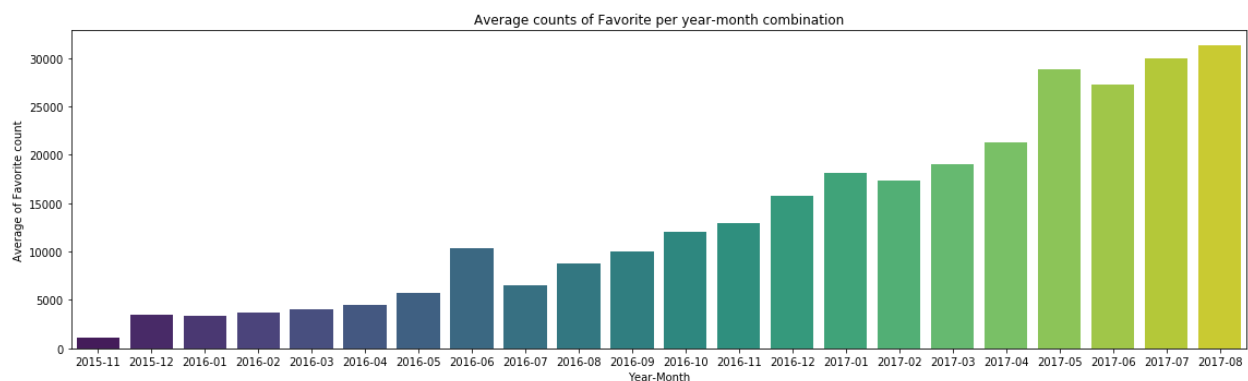
Insight: In the year 2016, most tweets were posted in the month of January (134). The overall tweets posted in 2016 is showing a declining trend. This might suggest that maybe twitter users stopped sending in dog images to rate or the account reduced the amounts of ratings given towards the end of the year.

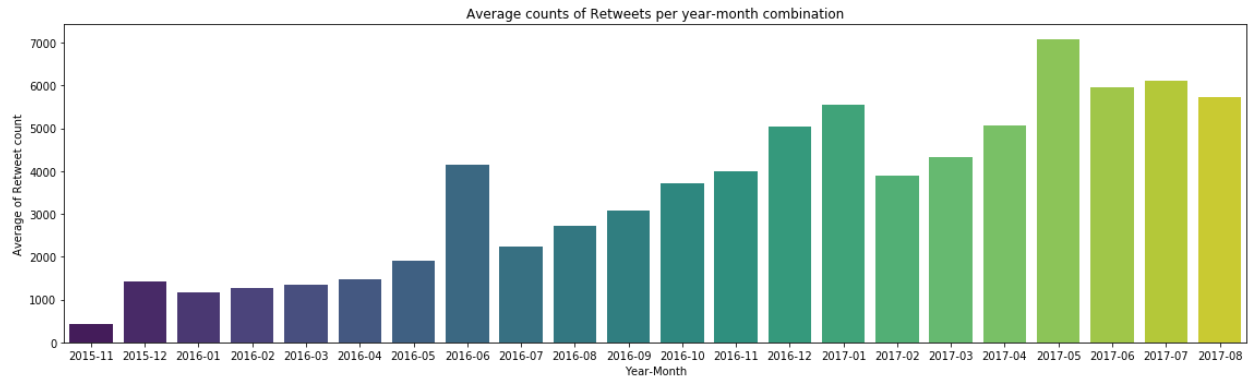


Insight: Looking at the entire time period from 2015 to 2017, the declining trend in tweets is clearer. In April 2016, the number of tweets posted dropped and since then it has been relatively constant. To see if the performance of the Account has decreased I will take a look further on the favorites and retweets that the posts get.



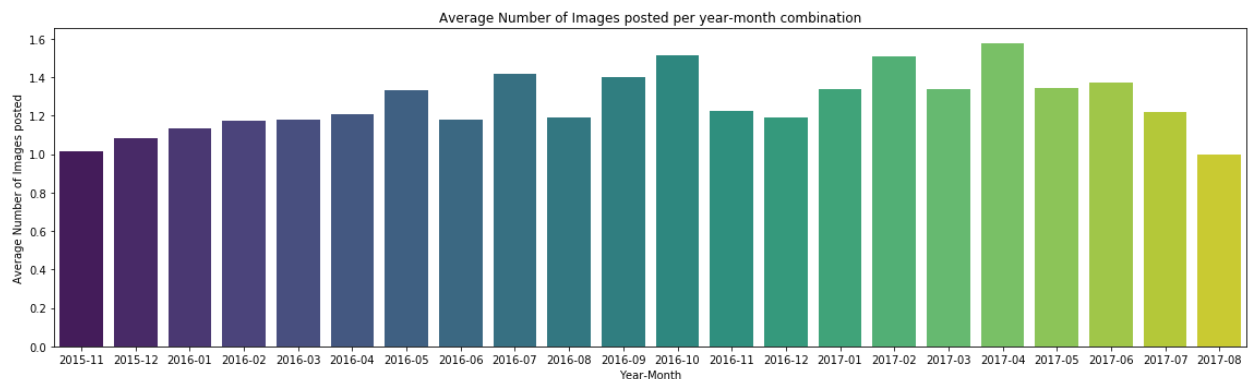
Insight: Though the number of tweets posted per month are decreasing, the count for retweets and favorites is increasing.



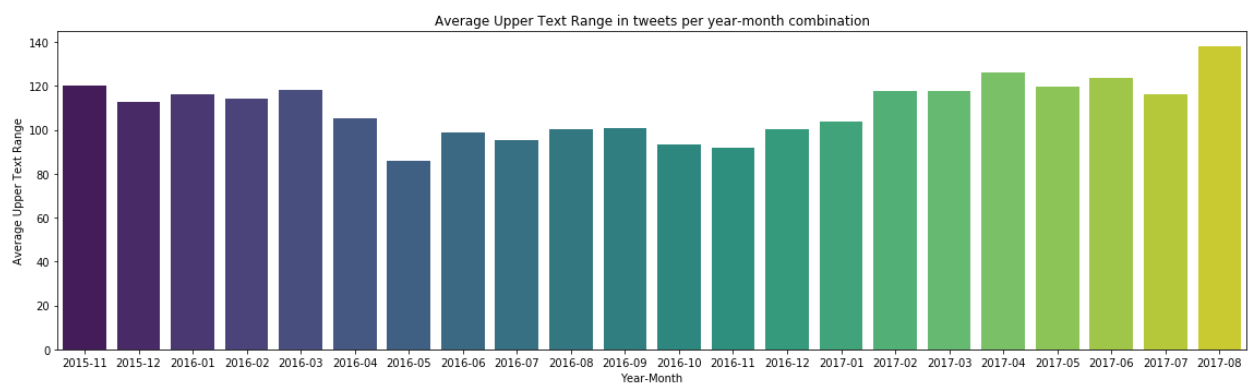


Insight:

- By taking the average of the counts of favorite and retweets the uptrend becomes even clearer than before.
- Now since retweets and favorite counts are not the reasons for low profile activity of tweets posted we will check the images posted per month and the average of the upper text range.

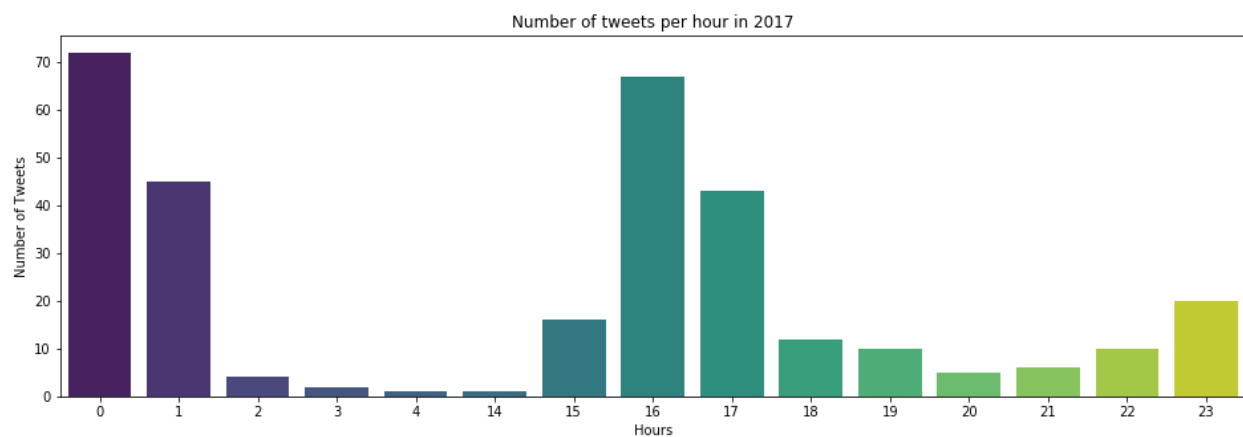
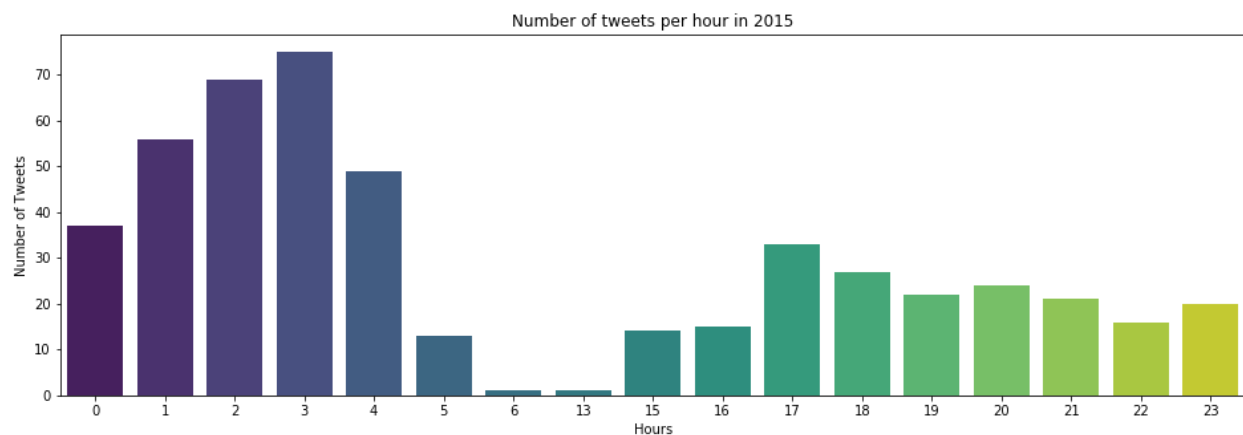
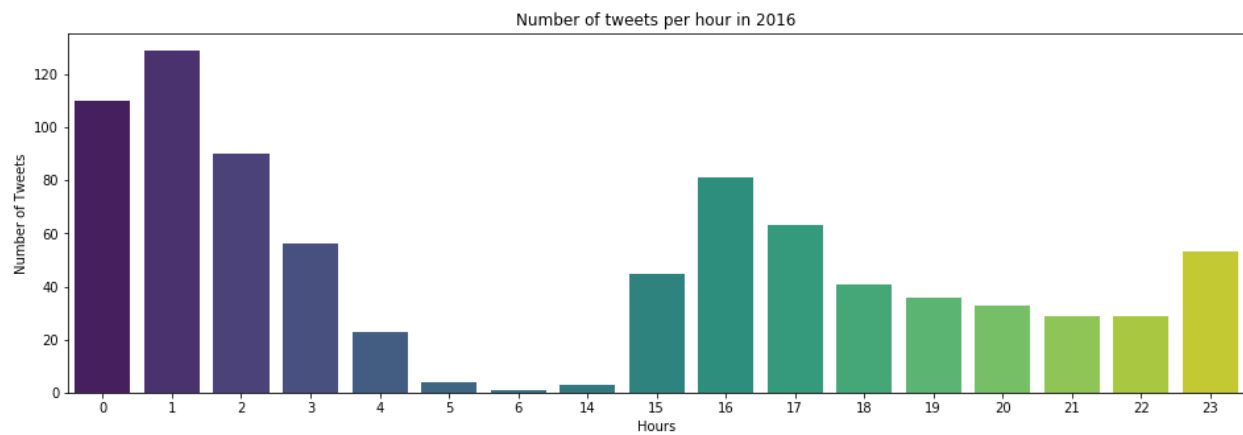


Insight: The number of images posted seem pretty stable. There are months where there are more and months where there are less posted images, but overall there is no clear uptrend or downtrend visible.



Insight: The tweet length seems to have increased over the second half of the dataset from an average of 106 to 113.

3. What are the most active hours for the twitter account?

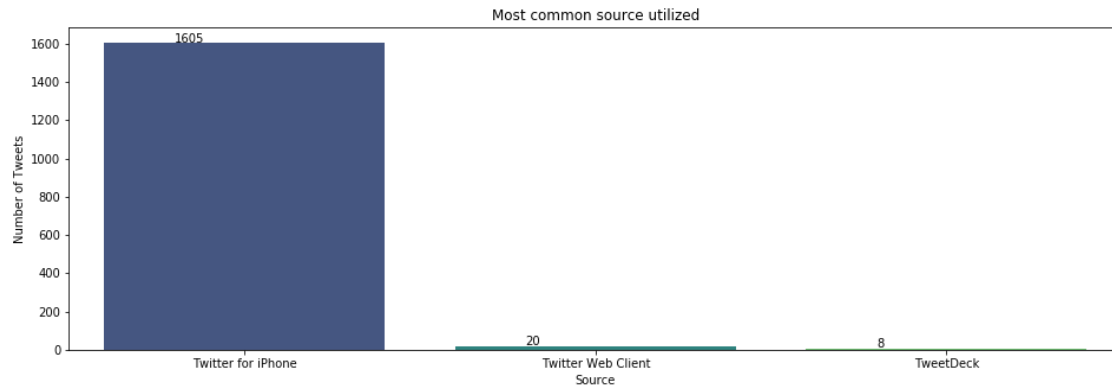


Insight:

- As seen in the above graphs for 2015 and 2016, the number of tweets posted are concentrated between 0:00 and 4 o'clock. This is the most active time for the account (global maximum).
- After 4 o'clock, there are very few tweets between 6:00 and 14 o'clock. In fact, there are no tweets for hours 7:00-12:00.

- The tweets do rise after 14 o'clock but it forms only the local maximum which further turns into a plateau between 18:00-22:00.
- In the year 2017, there are two peak hours between 0:00-1:00 and between 16:00-17:00. Overall the number of tweets posted are less compared to 2015 and 2016.

4. Which is the most common source used to post the tweets?



Insight: As we can see, the most common source utilized for posting tweets is iPhone (1605 times). The Twitter Web Client is the next with 20 counts and TwitterDeck is rarely used (only 8 times)

Summary and Conclusions

- The cleaned master data is analysed to answer the below 4 questions

Questions:

1. Which breed is the most popular among WeRateDogs tweets and retweets?
Insight: Labrador retriever.
2. How is the account activity throughout the time period available in the dataset?
Insight: The number of tweets per month decreased, while the retweets and favorites show an uptrend. For the image numbers there is no clear trend visible, the length of the tweets got a little bit closer to the maximum of 130 in the second half of the dataset.
3. What are the most active hours for the twitter account?
Insight: Here we found that between 5 and 15 'o clock there are nearly no tweets at all. The most tweets are during the time from 0 - 4 and then from 15 - 23, but from 15 - 23 are less tweets than between 0 - 4.
4. Which is the most common source used to post the tweets?
Insight: Twitter iPhone is the most common source utilized by WeRateDogs account.