We rate dogs data - results

I assumed that if a post had more retweets, that it would also have been favourited more times. When I look at the correlation there was a 0.968 correlation coefficient which shows a strong positive relationship. The scattergraph (figure 2) also showed a positive correlation and therefore it can be said that this is the case.

	retweet_count	favorite_count
retweet_count	1.000000	0.967678
favorite_count	0.967678	1.000000

Figure 1: Shows the correlation between the retweets and favorites using spearmans correlation.

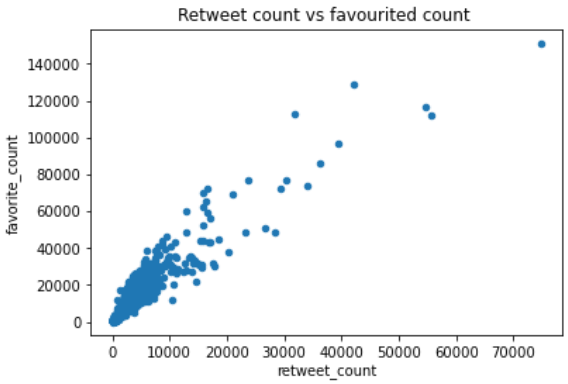

Figure 2: Scatter graph of retweet vs favourited count

I wanted to look at the source of the data too, as I found earlier there were only 3 types of source.

Twitter for iphone 0.981779
Twitter Web Client 0.013398
TweetDeck 0.004823

This shows that 98% of the data comes from twitter for the iphone. Therefore if they wanted to focus on one source then this would be the best one. The scatter plot (figure 3) also shows the timings of when people use the sources. Web client and tweetdeck seem to have been used very sparsely, with twitter for iphone always being popular.

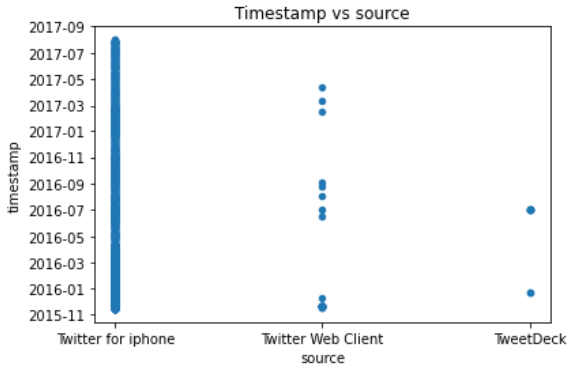

Figure 3: Scatter graph of timestamp vs source of tweet

I wanted to see at what dog stage are people more likely to post about their dog.

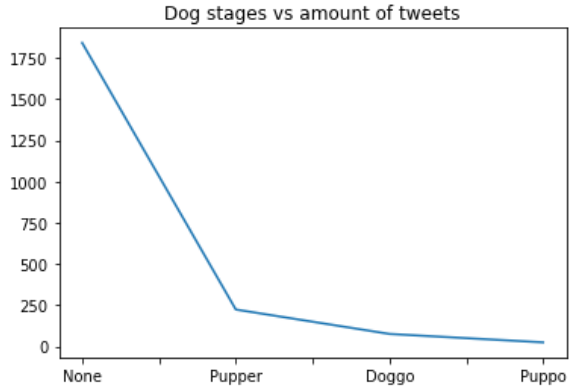

Figure 4: Scatter graph of dog stages vs amounts of tweets

I decided to remove the None group to see the percentages.

Pupper 0.702422
Doggo 0.217993
Puppo 0.079585

Figure 4 shows that most people do not mention their dog's stage. The most common is Pupper which is the youngest, 70% of those that mentioned the dog stage were this. Therefore, it can be said that people are more likely to post about their dog when it is younger. It is hard to actually confirm a correlation seeing as the majority do not mention the stage of their dog, therefore it could be that

when you have a Pupper you are more likely to include the dog stage than when they get older.

The dog names were pulled from the text and so were the ratings. Some of them were not pulled over and therefore this program may have to be looked at to make it more accurate as the ones that were not correct had to be removed, making less data to look at. The most common wrong name was 'a'. With the ratings, it did not help that the rating system was not standardised and some rated multiple dogs together.

Figure 5: Image from https://pbs.twimg.com/media/CU6xVkbWsAAeHeU.jpg

The image prediction model needs work. For example the image above was predicted to be food and not a dog - I understand why as the donut takes precedence but maybe the model could be improved?
However all of the images left in the dataframe had at least one correct prediction so therefore it does work a lot of the time.

One would assume that over time a twitter page may receive more likes and retweets once it gets more known.

The scatter graphs show this is true for this twitter page, with favourites increasing the most over time.

This shows that the page is getting more popular over time.

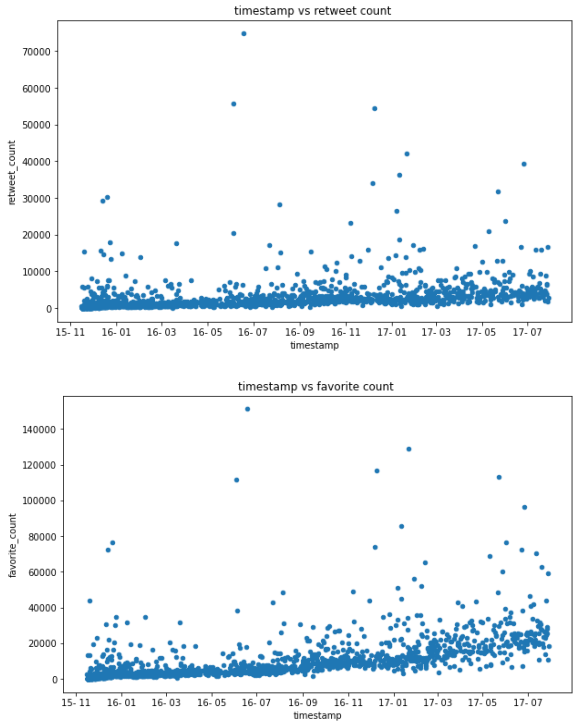

Figure 6 and 7: Scatter graphs of timestamp vs retweet count and favourite count

---

Conclusion - overtime the twitter page has become more popular. There is a correlation between the retweets and favourites. Not many people mention the stage of their dog but if they do its more likely to be a Pupper. The models for the images, names and ratings may need to be improved, although for the majority of posts they worked. People are more likely to use the twitter iphone app and therefore this is where they would get more interaction.