ACT REPORT

Ebenezer Maradesa
ALX-T Data Analytics
Project 2: Data Wrangling – Act Report

Table of Contents

INTRODUCTION

The Twitter page WeRateDogs ([@dogrates](#)) gets all the attention in this Analysis. WeRateDogs posts pictures of dogs with some caption and rates them, typically with rates like 12/10.

The tweet archive for WeRateDogs was gathered, and a lot of wrangling efforts have been spent on the data as detailed in 'wrangle_report.pdf.' The wrangled data was assessed and used to gain a lot of insights, and these insights are shared as below.

QUESTIONS

1. Which dog names are most common?

2. What is the mean rating of the dogs?

3. Is there any correlation between rating and number of likes?

4. Are tweets with more favorites likely to be retweeted more often?

5. Do hashtags increase engagement?

6. Which dog breeds are more common?

7. What are the most beloved dog breeds?

8. What is the trend of WeRateDogs tweets over time?

Overview

Before diving deep into questions, a general overview of the dataset is necessary. Here is a heatmap of the archive_clean dataframe.
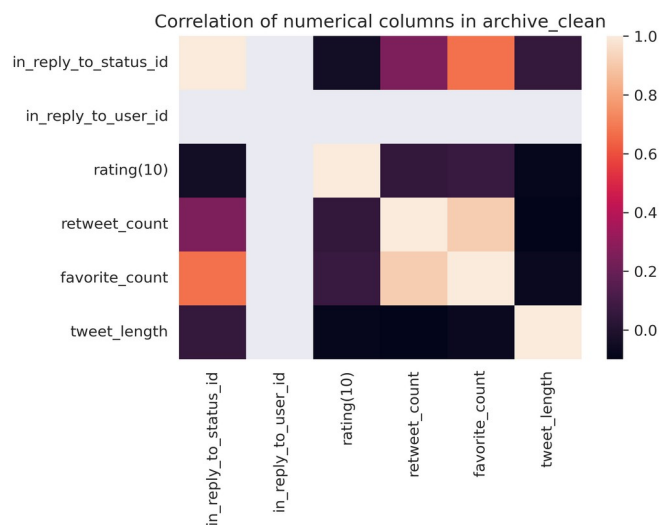

Figure 1

This heatmap will come in handy to answer some question later.

Q1: Which dog names are more common?

Though the names of dogs in a number of tweets could not be recovered, the ones collected still showed some insights as to the most common dog names. From this dataset, it appears that the **most common dog names are Cooper, Charlie, Lucy, Oliver, and Tucker.** Here is a bar plot to better prove that:
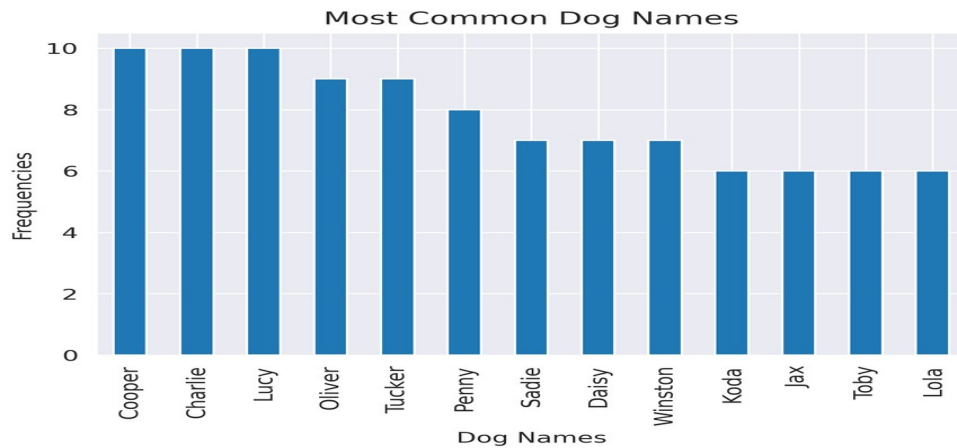
Figure 2

Q2: What is the mean rating of the dogs?

As expected, the mean rating is over 10, seeing as WeRateDogs often rate 10+/10. The **mean rating for the dogs is 11.47**. One dog was even rated 165/10.

We can get similar insights about other things. Consider the table below:

	Rating(10)	Retweets	Favorites	Tweet Length
Mean	11.47	2817.11	9233.10	110.52
Minimum	0.00	16.00	81.00	32.00
25th Percentile	10.00	649.25	2163.25	93.00
Median	11.00	1437.50	4440.00	115.00
75th Percentile	12.00	3247.75	11656.75	136.00
Maximum	165.00	79515.00	132810.00	148.00

Table 1. Summary of Counts of Numerical Columns in the archive

Q3: Is there any correlation between rating and number of likes?

This is where the heatmap comes in handy. A quick glance at the heatmap (with a supporting look at a Pandas-generated table not shown here) tells us there is no correlation between the ratings and the retweet and favorite counts.

There is only a **correlation of 0.06** between ratings and number of likes and **0.05** between ratings and retweet counts. It appears that people's engagements with the post do not at all depend on the rating of the dogs.

Though this chart points in the otherwise direction, we cannot conclusively agree with it.
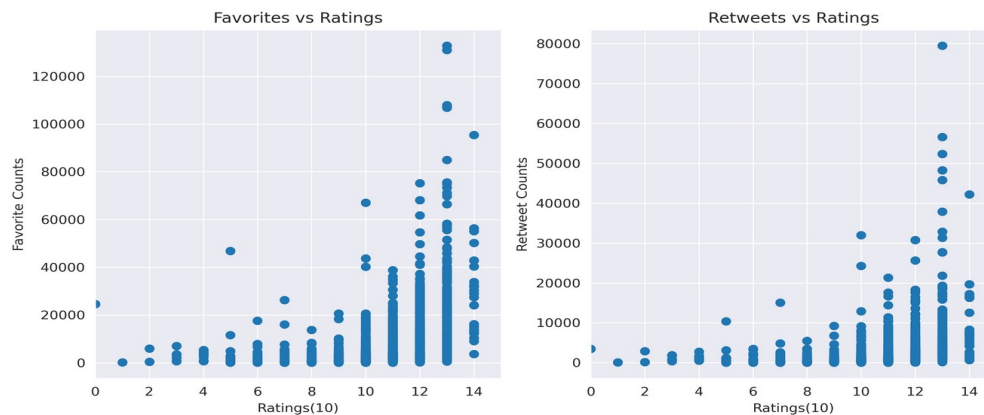

Figure 3

Q4: Are tweets with more favorite counts likely to be retweeted more often?

Again, the heatmap comes in handy. There is a **correlation of 0.91** between these two. Perhaps this is no surprise, as both favorite_counts and retweet_counts are measures of engagement. A scatter plot confirms that tweets with more favorites are retweeted more also.
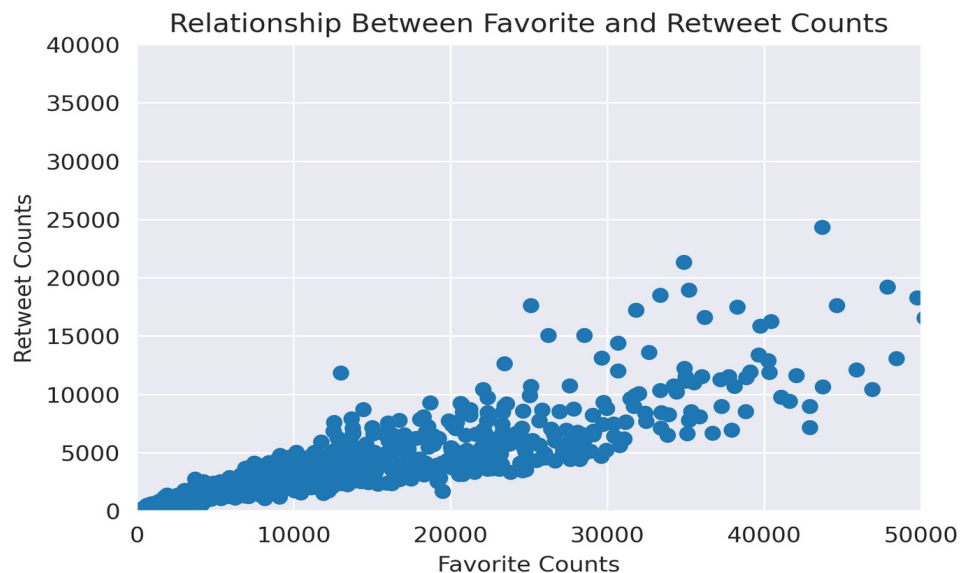

Figure 4

Q5: Do hashtags effectively increase engagements?

Social Media experts claim that hashtags can help increase engagement. Is that the case for WeRateDogs? It appears hashtags are not used frequently here (only 1.1% of tweets), but do those few have any difference? This bar plot tells the whole story.
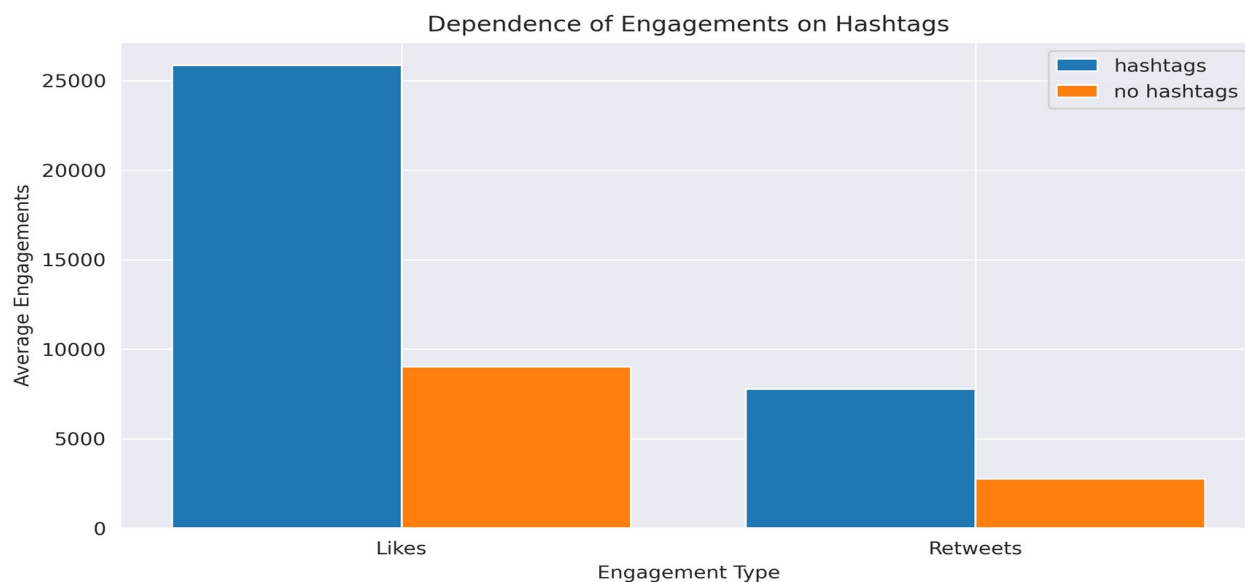

Figure 5

Clearly, **hashtags are effective**, and perhaps WeRateDogs should use them more.

Q6: Which dog breeds are more common?

There are a total of **111 unique dog breeds** that could be identified from the tweets. Which of them appear more frequently?
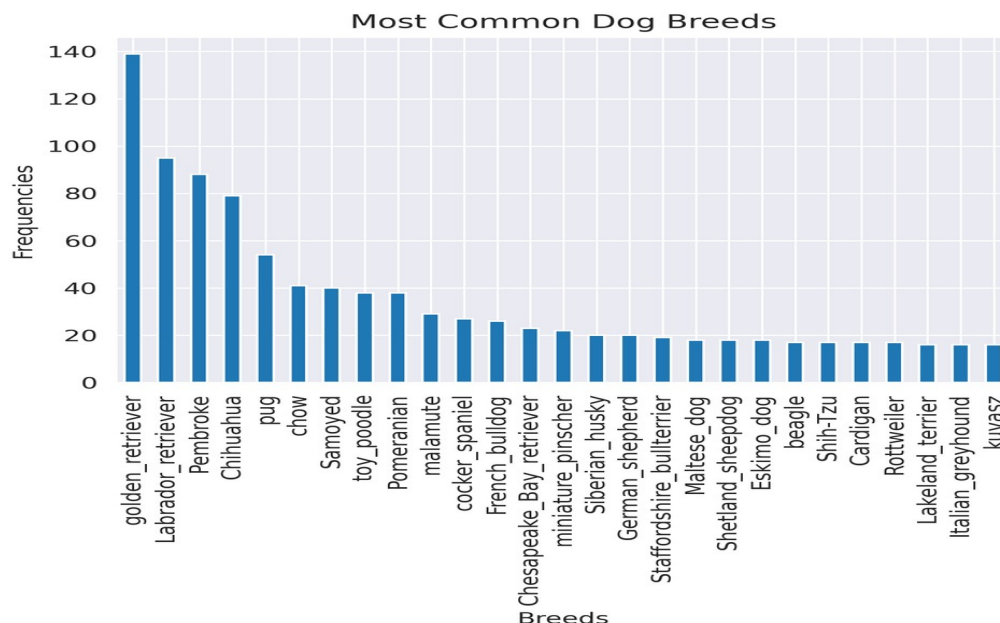
Figure 6

This is a plot of the more frequently appearing breeds. It appears **golden_retrievers are most common, appearing 139 times**, which is quite a gap to Labrador_retrievers, which appear 95 times, and so on. Other unpopular dog breeds seen include Scotch_terrier, EntleBucher, Japanese_spaniel, clumber, which, etc, which all appear just once.

Q7: Which are the most beloved dog breeds?

Then arises the question, are there any dog breeds that appear more frequently?

Chihuahuas seem to be quite beloved, as they have the **first and second highest favorites, 33819 and 25461 respectively, with corresponding retweets of 6514 and 4328.**

Q8: What is the trend of WeRateDogs tweets over time?

Has WeRateDogs consistently awarded distinctions to dogs for being dogs over time? Here, a plot of the monthly tweets over time helps to understand the trend.
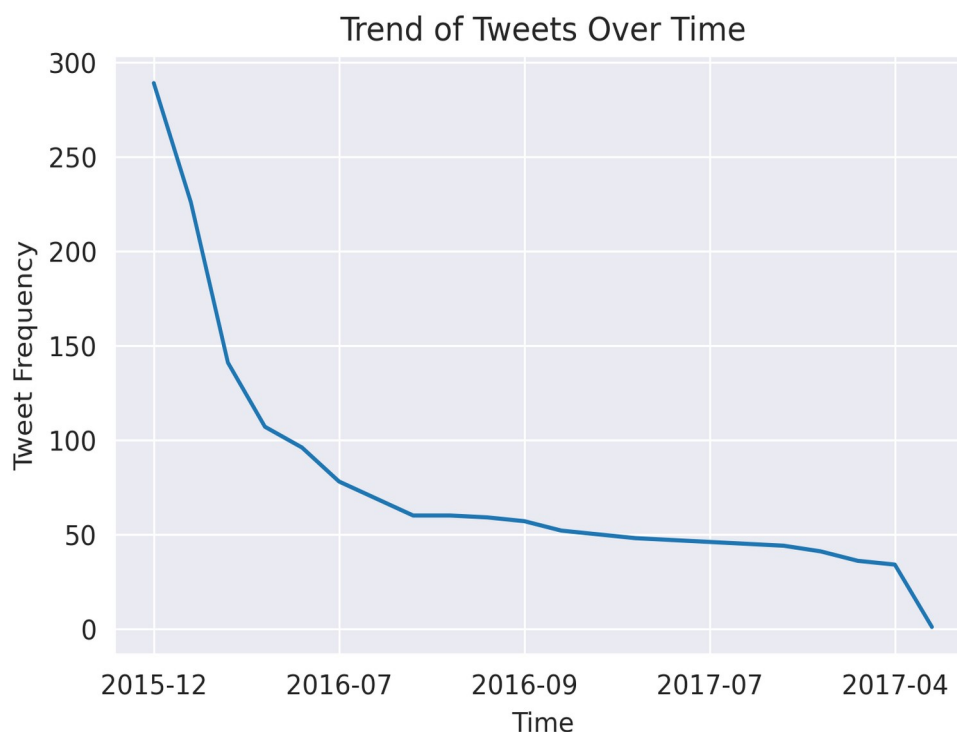
Figure 7

Sadly, it appears that WeRateDogs was most active in December of 2015 with almost 300 tweets in that month, and the **activity declined to 0 tweets** by April of 2017.

Perhaps WeRateDogs has been active again after then, but the dataset does not extend beyond this span of time, and it cannot be told from this dataset.

CONCLUSION

Dogs are indeed lovely, and WeRateDogs twitter page and the huge number of engagements that the page got are a testament to that.

Most common dog names, breeds, and the most beloved ones have been explored in this Analysis, alongside the effectiveness of hashtags in social media posts, among others. Full details of insights that have been drawn are included under each question.

Note that these insights have been drawn only from WeRateDogs twitter archive, and these are not necessarily absolute representations of all trends explored.