Act Report

Edim Ernest
26th June, 2022
Data Wrangling, Analysis and Visualization of Twitter Data from WeRateDogs

Introduction
The @dog_rates Twitter handle has been around since November 15th, 2015. It is a twitter handle that rates dogs based on their body and facial characteristics. The handle allows users to rate dogs from over 10 with a denominator of 10 on a scale from cute to ugly. The site's goal is to have an open forum for people to discuss their love for dogs in a casual manner. Since we all have our favorite kinds of dogs, no specific breed or type is favored over another one. Some dog pictures have witty and funny comments on them.

The @dog_rates twitter handle has gained over 9 million followers on twitter. WeRateDogs shared their twitter archive with Udacity containing basic tweet data like tweet_id, source, retweet_count, favorite_count and more. More data was gathered using twitter API to gain better insights based on analysis and visualizations.

I got a cleaned dataframe after identifying both quality and tidiness issues which I defined, cleaned and tested during the process of data wrangling.
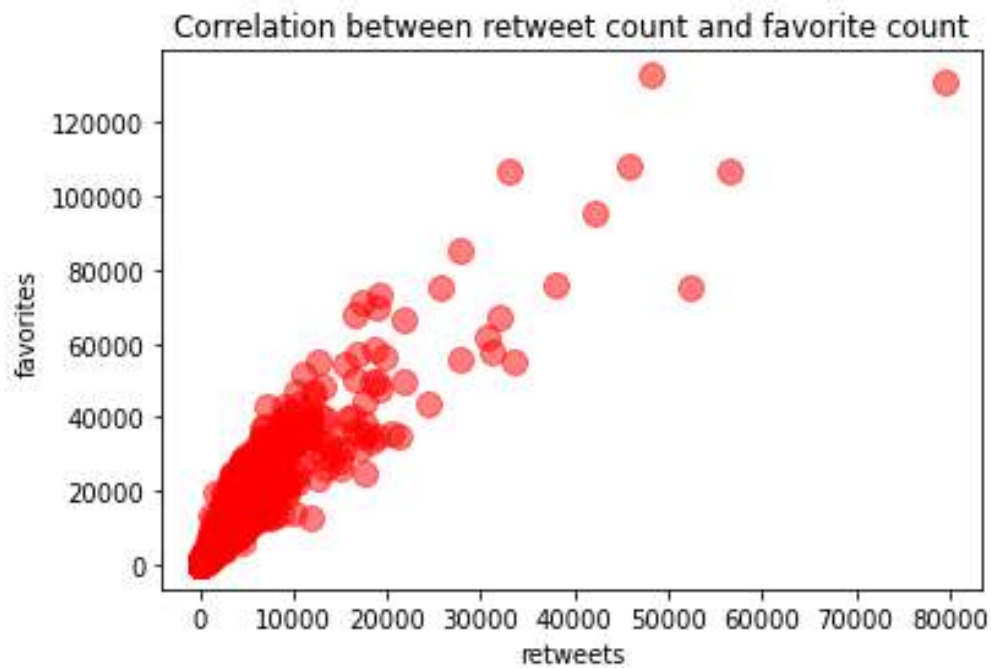
Insight 1

Following through with my project analysis, Stephan, a Chihuahua breed with over 56k retweet counts turned out to be the dog with the highest number of retweets. The tweet about Stephan states "This is Stephan. He just wants to help. 13/10 such a good boy". Turns out Stephan was trying to be of help as seen in the photo that follows. Who knows, his eagerness to help out may have earned him the high


Insight 2

If you have been following the trends of social media, you know that likes and retweets are good indicators of popularity. So it's no surprise that I was interested to see if there was a relationship between retweets and favorites when testing the WeRateDogs data. I analyzed the relationship between retweet count (the number of retweets) and favorite count and found that there is a stronger correlation between the two metrics than might be expected by simple causality. The correlation coefficient between them is 0.9.

0.91294662741837

Correlation between retweet count and favorite count

Insight 3
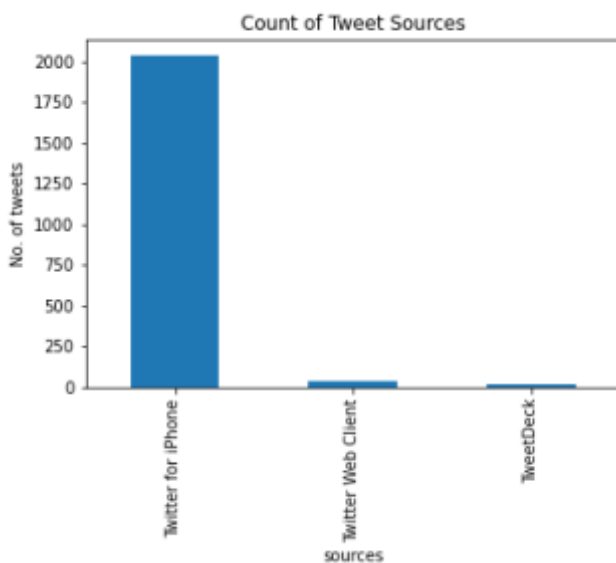
From the data collected, it is important to understand the different sources of tweets. The tweet sources came from three different Twitter clients, namely: Twitter for iPhone, Twitter Web Client and TweetDeck. From my visualization I found out that twitter for iPhone is the source of over 98% of all tweets leaving less than 2% which is shared among Twitter Web Client and TweeDeck. This is visually represented on the bar chart by the left wing.

Insight 4

According to data from the @dog_rates Twitter account, dog tweets increase significantly during these times 12 am, 1 am and 2 am. These are the hours of the day people tweet about dogs. At 12 am, there are far more tweets about dogs at this time — despite the fact that not many other people are awake. Perhaps there is something special about this time that attracts dog lovers?
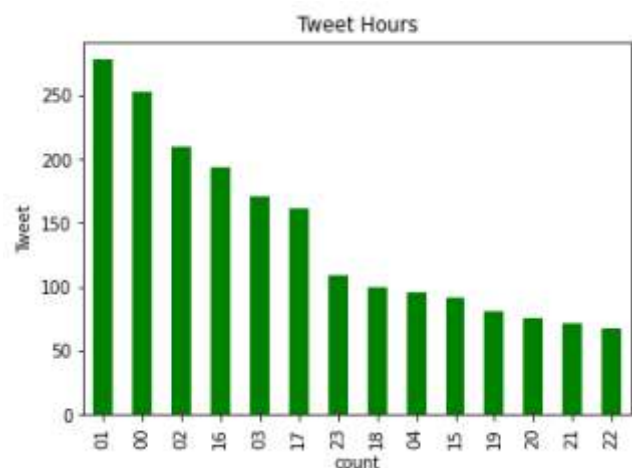
Summary of Findings

- Stephan, a Chihuahua, had the most retweets
- There is a strong correlation between retweets and likes (favorites)
- Twitter for iPhone has the highest number of tweet source
- There are more tweets during the hours of 12AM, 1AM and 2AM