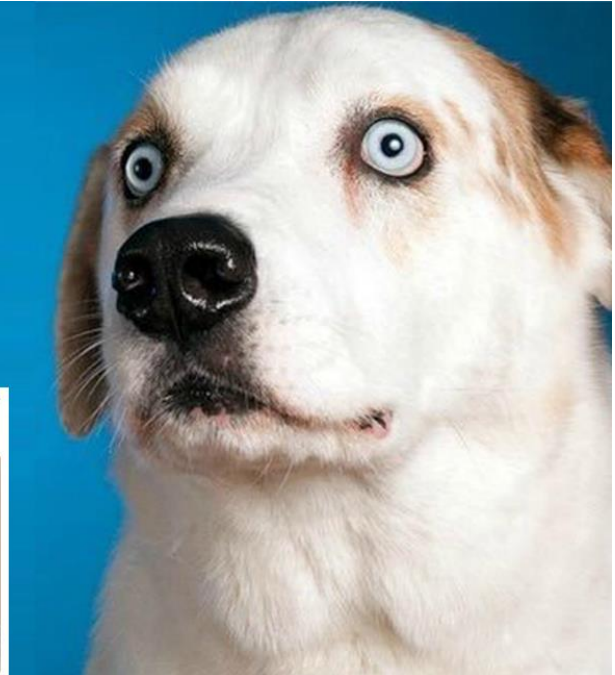


ACT REPORT

For the PROJECT : Wrangle & Analyze Data



Introduction

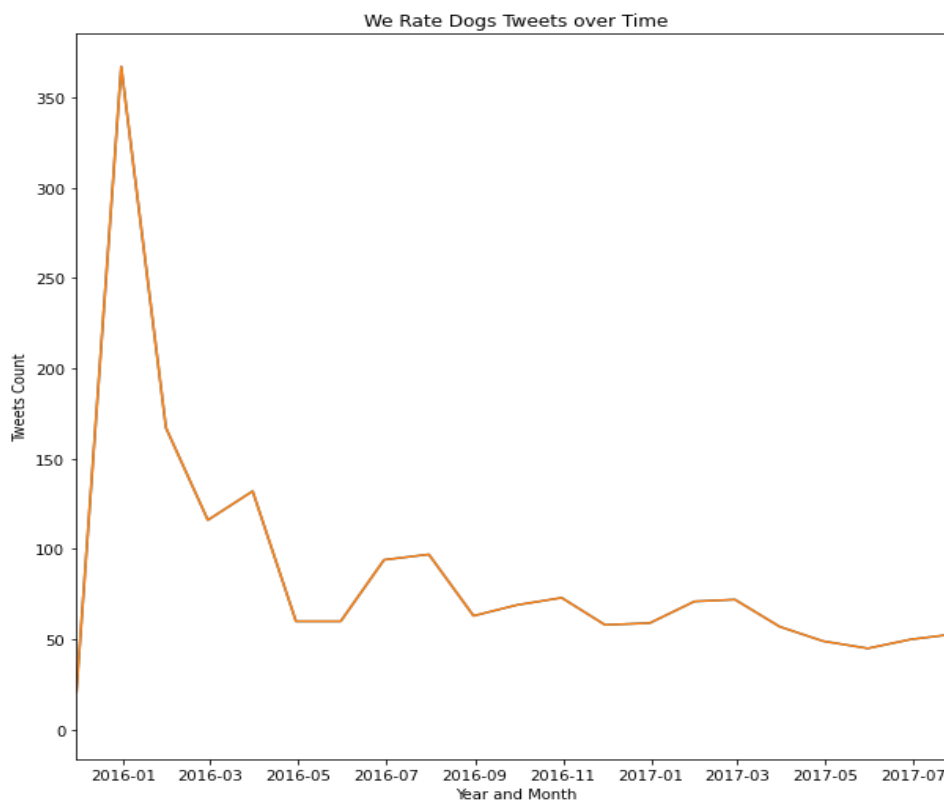
The main purpose of this project is to use real world data to wrangle (gather, assess, clean) and then apply analysis with visualizations.

The dataset used is the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog with a denominator of 10 and the numerator is usually higher than 10 to show how lovely the dog is.

After get the data cleaned in the wrangling part, I did data analysis and visualizations to derive some insights.

This section provides an analysis of the data set, and corresponding visualizations to draw valuable insights. I will document these insights below.

Insight 1: Does the number of tweets increase or decrease over time ?



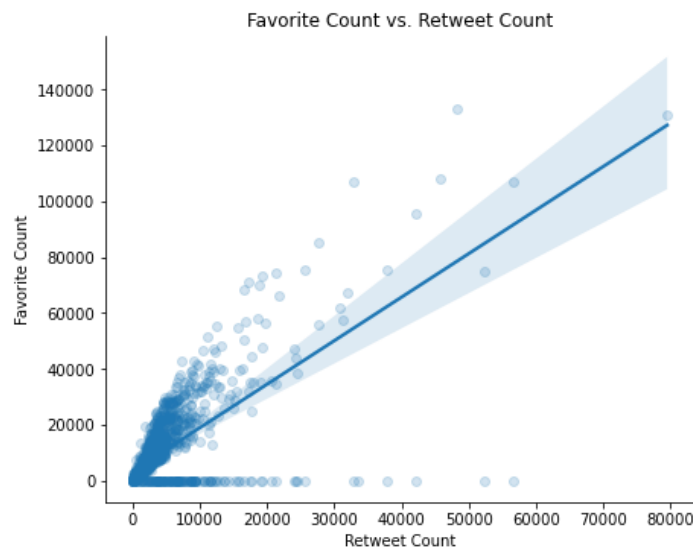
Over the time tweets decreased significantly.

Spikes in activity were registered during the January of 2016, and started decreasing a lot from there except a second spike during March 2016.

Insight 2: Is there a relationship between retweet counts and favorite counts ?

From the Data Assessment Part we know that a lot of the tweets have been retweeted.

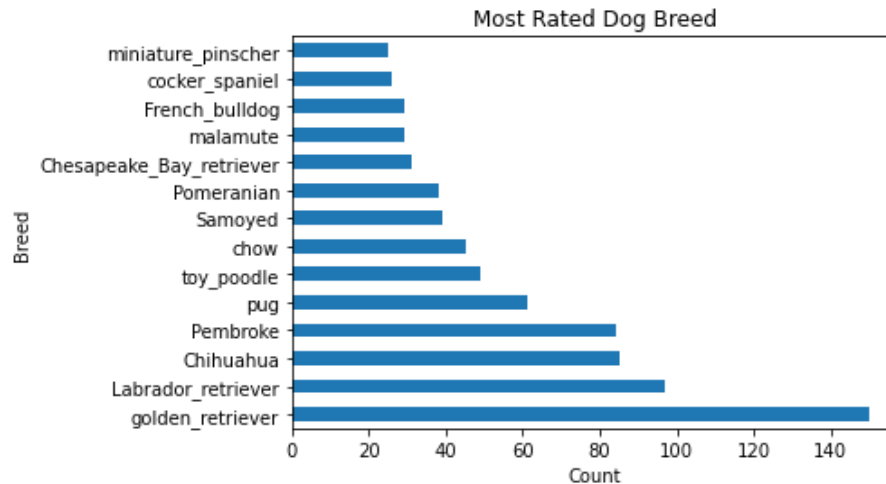
I tried to find out if there is any relationship between the Retweet counts and favorite counts through a scatterplot.



Favorite counts are correlated with retweet counts - this is a positive correlation. That means every increase in retweets leads to an increase in likes.

Insight 3: Which dog breed people love the most ?

To solve this question, I manipulated the data by groupby function to comparing popularity among different dog breeds. Then I used the histogram from the matplotlib library.

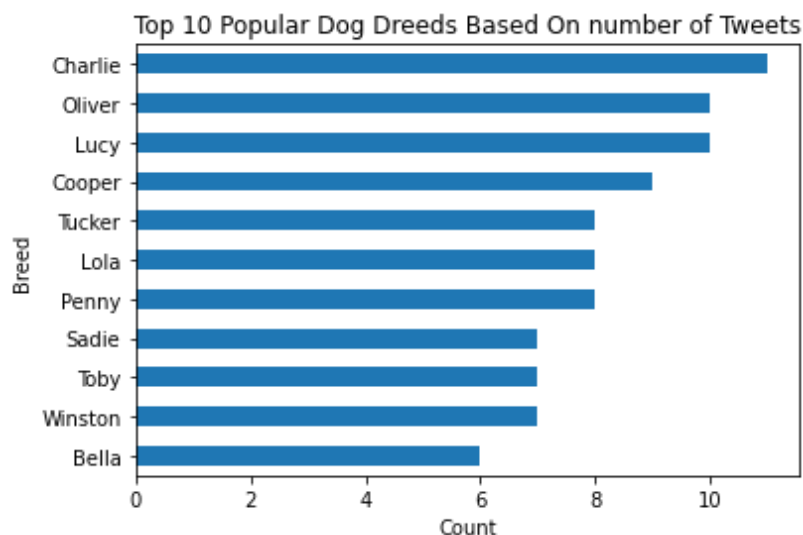


The most common breed is the Golden Retriever which seems obvious since they're everyone's favorite smart & loyal family breed (including me). Followed by the Labrador retriever & Chihuahua.

Insight 4: Among these dogs , what are the most popular dog names ?

There are a lot of dog names in the dataset, so what is the most popular dog name?

I answered this question using the name column by ranking the dog names and visualizing it using an histogram.



The Top 3 Dog Names based on number of tweets are : Charlie, Oliver & Lucy.