Analyzing and Visualizing WeRateDogs

Udacity: Wrangle and Analyze Project

By: Dwiti Shah

Introduction:

As a Udacity Data Analyst Nanodegree candidate I need to wrangle and analyze the WeRatedog data. WeRateDog is an archived tweet from Twitter users which basically rates people's dog by positing comments. To begin with, I gathered data from:

- Enhanced Twitter Archive
- Image Predictions File
- Data via Twitter API

I have used 3 different sources to enhance my finding since Twitter archive would provide only basic information and for more advanced information like breed, retweet count I have included other 2 useful datasets.

After a careful look at the data sources, I have cleaned, transformed and remove unnecessary columns and combined them useful columns in a master dataset to perform visualizations.
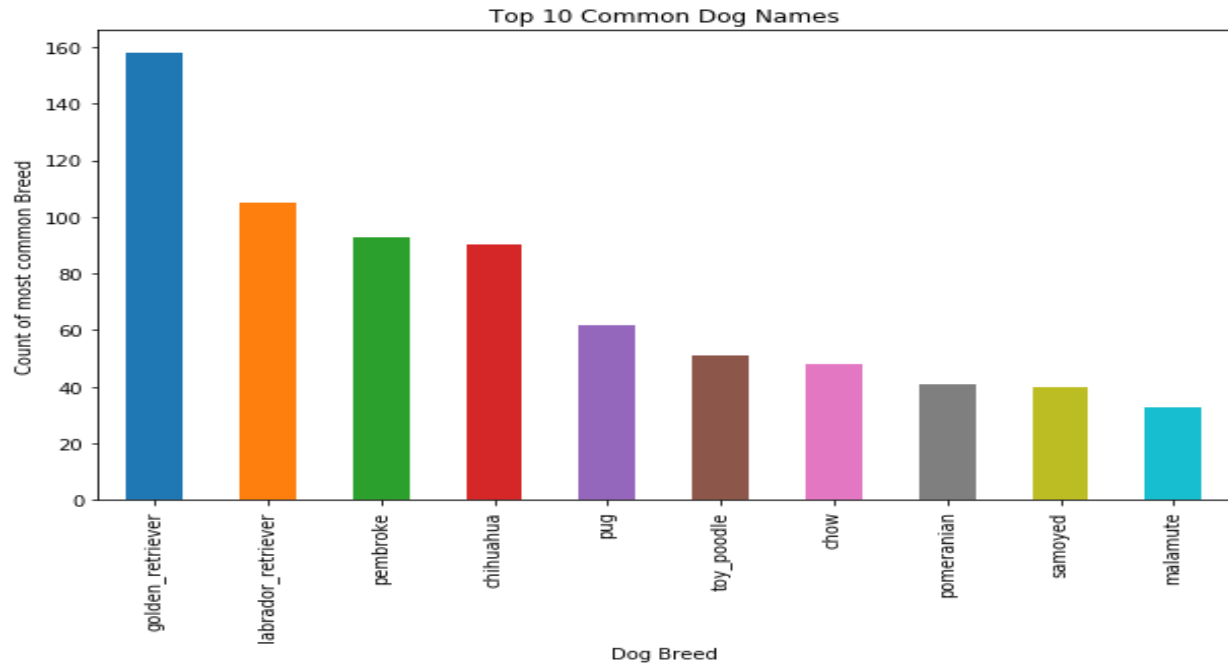
Visualization:

WeRatedogs has received over 5000 tweets, however after accessing and cleaning the dataset we have 1991 tweets to be visualized.

10 most common Dog Breed

Golden Retriever breed is most popular among people and received over 150 count. There could be multiple reasons as why these breeds are most common:
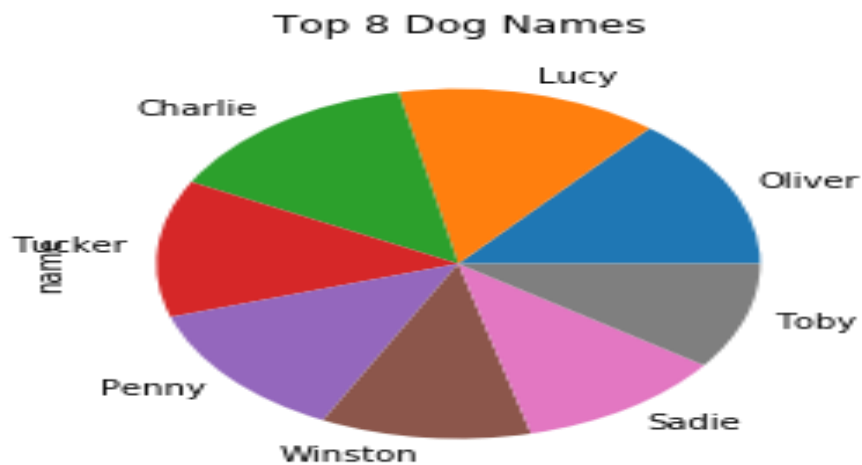- Commonly identified by people or machine
- Maximum people own these dog breeds

Top 10 Common Dog Names

Most common Dog Names

Top 3 common names are:
- Lucy
- Oliver
- Charlie

Top 8 Dog Names

Correlation between Retweet Count and Favorite Count

From the visualization, we can make out that there is strong correlation between Retweet and Favorite data. This makes sense because popular tweets tend to have higher retweet and favorite count.

Favorite vs Retweet Scatter plot
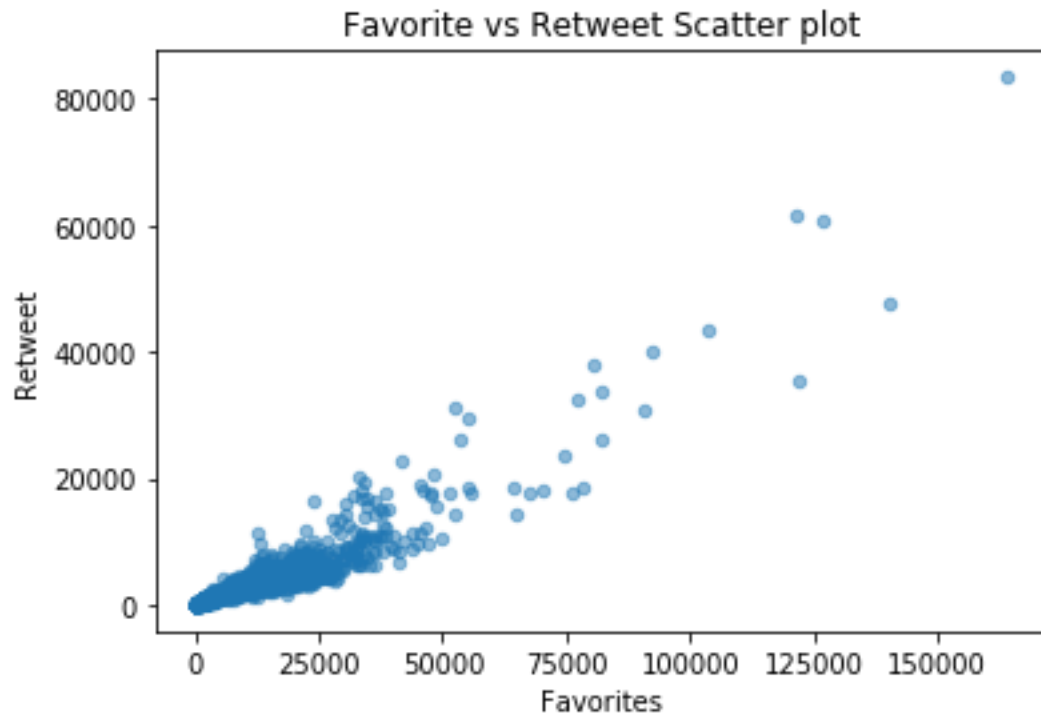
Retweet and Favorites trend over time

We see an upward trend in the number of retweets and favorite since 2015.However, favorite count has increased 7 times more compared to retweet count.

Retweeting and Favoriting trend over time