

WeRateDogs Twitter archive – Act Report

Introduction

WeRateDogs dataset is archived from Twitter user @dog_rates. This account allows followers to rate dogs highly than the usual rating (more than 10 which is the maximum). Currently, the account has more than 9million followers and over 16,000 tweets though we will focus on the archive data for the period up to August 1, 2017.



Source: @dog_rates Twitter

Data Wrangling

This is the process of gathering your data, assessing its quality and structure, and cleaning it before you do things like analysis, visualization, or build predictive models using machine learning.

Data was gathered from 3 different sources i.e. Manually from Udacity Classroom, Using Request Library and also using API.

Insights

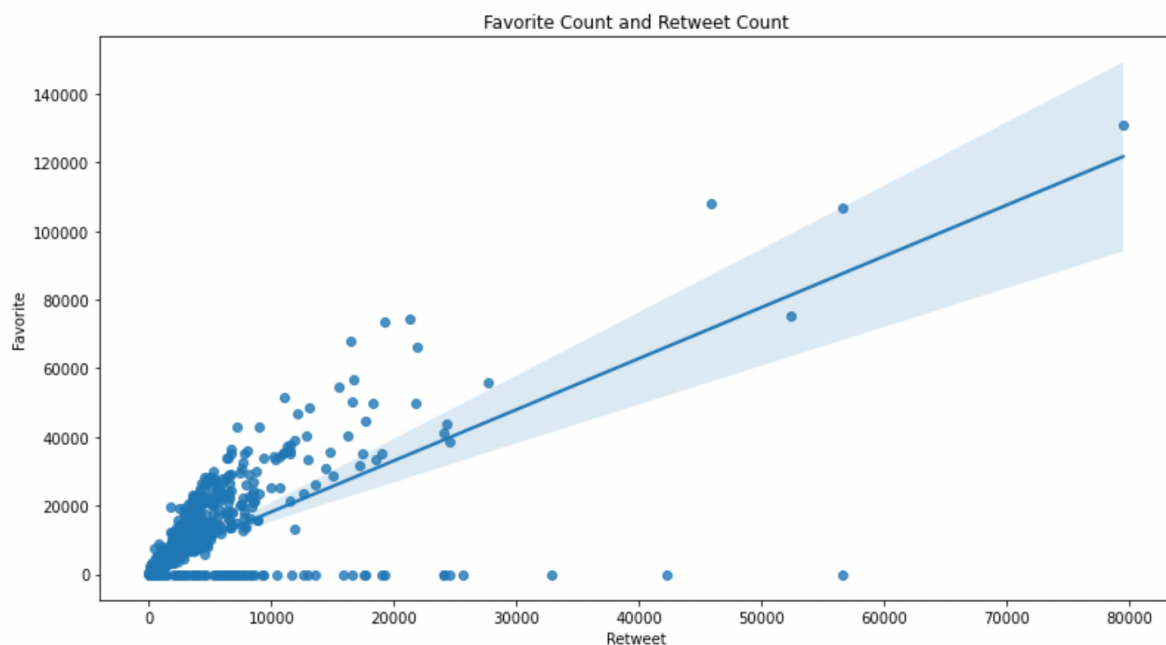
1. Favorite Count is More compared to Retweet Count

Most of the followers preferred clicking on 'favorite/like' icon on a post. This is mostly because it doesn't post in profile as compared to retweet and most of the social users would prefer their accounts clean with minimum posts.

	breed_name	retweet_count	favorite_count
0	afghan hound	5312.0	17305.0
1	airedale	11716.0	45507.0
2	american staffordshire terrier	22392.0	84648.0
3	appenzeller	758.0	2594.0
4	australian terrier	700.0	2244.0
...
108	welsh springer spaniel	3976.0	18811.0
109	west highland white terrier	6613.0	31115.0
110	whippet	19986.0	69958.0
111	wire-haired fox terrier	0.0	0.0
112	yorkshire terrier	2744.0	8650.0

2. Correlation for Retweet and Favorite Count.

There is a positive correlation between our variables (Favorite Count and Retweet Count), this means that clicking on favorite and retweet icon is directly related only that more people click on favorite icon.



3. Where do we get most tweets?

From the dataset, more users use iPhone to access twitter as compared to using tweetdeck, web or vine. This is true because world population most uses phones to access the internet hence the high numbers.

```
df_all_copy.source.value_counts()
```

```
Twitter for iPhone      2221
Vine - Make a Scene      91
Twitter Web Client      33
Tweetdeck               11
Name: source, dtype: int64
```

Visualization

Comparison between various sources of data in a Pie Chart

