Analysis and Insights

OCTOBER 20

Wrangle and Analyze Data Project
**Authored by: Shady Bartakh**

UDACITY

Datasets

Nowadays, data are everywhere, with various formats and big amounts. Most parameters related to any references can be categorized or tabulated, so forming the data. Recently, it is a trend to analyze these to get valuable insights.

They can be used in different subjects. Data driven from the sensors around a production line and the warehouses can give new indications or form innovative ideas. Also, Internet now is a rich source of data in different fields, whether for scientific reasons or in social media platforms. And the latter is our study here.

Dogs' rates on Twitter

One of the applications is studying the tweets generated from the WeRateDogs page on Twitter. It shares their followers' dogs' pictures and rate them on a scale. People react to their postings, like and retweet the original tweets.
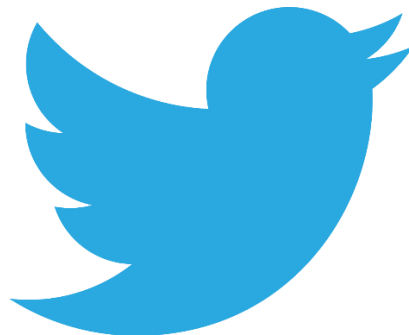
We started gathering our data that includes tweets and predictions of dogs' breeds based on their pictures. They can be held through their different tweets ID number. Then assessing the quality of the data and after that wrangling and cleaning it from arising issues to be tidy data.

Analysis

A data of nearly 2000 tweets have been analyzed and their statistics is calculated as follows

	rating_numerator	rating_denominator	favorite_count	retweet_count
count	1986.000000	1986.000000	1986.000000	1986.000000
mean	12.295065	10.539778	8122.270393	2399.928499
std	41.576848	7.332456	11962.777682	4293.901507
min	0.000000	10.000000	70.000000	11.000000
25%	10.000000	10.000000	1737.500000	537.250000
50%	11.000000	10.000000	3655.000000	1160.000000
75%	12.000000	10.000000	10101.000000	2733.500000
max	1776.000000	170.000000	153185.000000	75908.000000

The above table includes ratings with different values in the form of i.e. 11/10 beside the counts for users likes or favorites and retweeting times. And from these and other coding with Python an Insight tells that; through the entire set of WeRateDogs tweets, only 2 times where they gave nothing (0/10) for nothing also (no dogs). As, one for criticizing a plagiarism of a similar page, and another for disappointing about no dogs are in the near.

In the same set, maximum ratings have been given to the number one 'Atticus' dog which received 1776/10 wearing in America style flag as well as the USA flag in the background. While the number two 'Calvin Cordozar' or 'Dogg' received 420/10 although being a rapper person not a Dog. But a set of dogs received the third place of ratings of 204 even from 170.

Another insight is in the favorites number given for each tweet; so the minimum of 70 likes was for a dog with its twins. And the highest likes were given to a swimmer dog in a pool that reached 154,185 like.

The same above findings are witnessed in the retweeting numbers for the same tweets; as the dog and the 2 little ones received only 11, but the swimmer dog had 75,908 retweets.

In a grouping of dog stages to analyze numbers with the below table, found that elderly dogs received generally the most counts of both favorites and retweets as indicated in the chart below.

```
   dog_stage   favorite_count   retweet_count
0      doggo      18311.507937     6299.714286
1    floofer      12292.875000     4091.625000
2     pupper       6570.565854     2047.512195
3      puppo      21581.565217     6183.043478
```
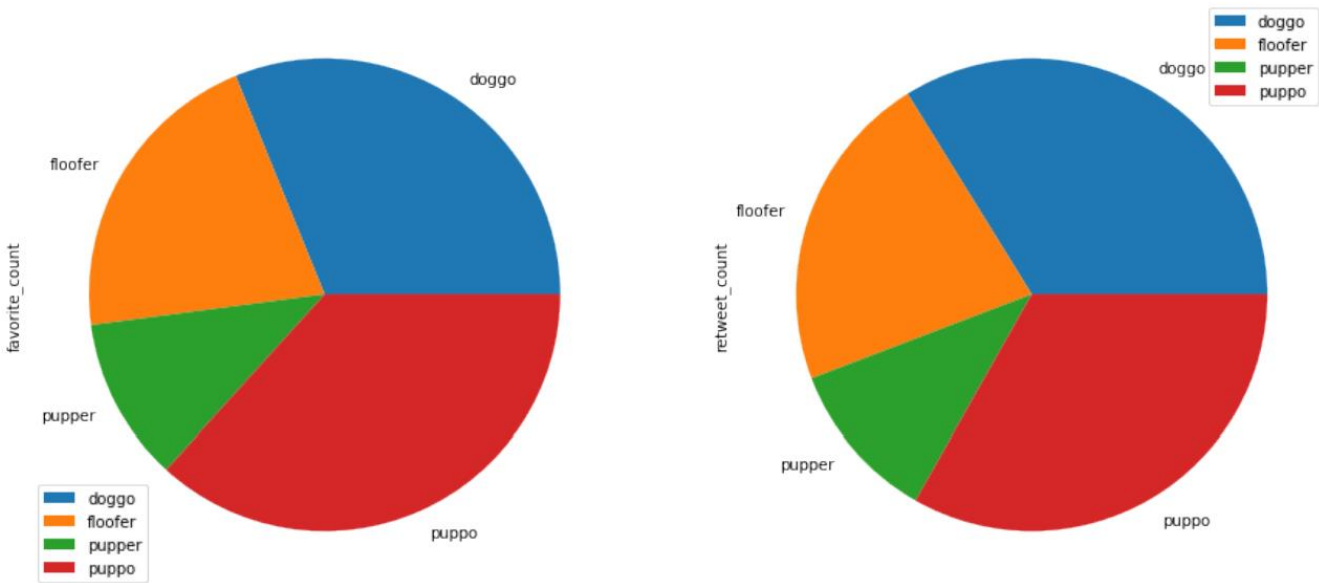
At last, the dog breeds seemed affecting the followers' feedback, so some like the Golden retriever, Labrador retriever and Pembroke Corgi were in the head of likes as depicted below.

Dog breeds favorites reactions