# ACT REPORT

## 1. INTRODUCTION

Real-world data rarely comes clean. Using Python and its libraries, I will gather data from a variety of sources and in a variety of formats, assess its quality and tidiness, then clean it. This is called data wrangling.

The dataset that I will be wrangling (and analyzing and visualizing) 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. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. WeRateDogs has over 4 million followers and has received international media coverage.

## 2. INSIGHTS

### A. Image Prediction Algorithms

Counting the True Boolean values for the three prediction algorithms to see which algorithm works better.

```
Total True values of the first Prediction Algorithm: 1425
Total True values of the second Prediction Algorithm: 1442
Total True values of the third Prediction Algorithm: 1397
```

The figure illustrated above shows that the second algorithm predicted 1442 dog pictures. Therefore, the second algorithm works best.

### B. The Five Number Summary for the Retweets and Likes of a tweet

i.   Retweets

```
count      1928.000000
mean       2514.632261
std        4187.214895
min          16.000000
25%         611.750000
50%        1299.000000
75%        2911.250000
max       79515.000000
Name: retweet_count, dtype: float64
```

From the figure illustrated above, it was found that the maximum retweet count of a tweet is 79515, while the minimum retweet count of a tweet is 16, and the mean retweet count is about 2515.

ii.    Likes

```
count       1928.000000
mean        8296.298237
std        11371.277140
min           81.000000
25%         1873.750000
50%         3904.000000
75%        10280.000000
max       132810.000000
Name: favorite_count, dtype: float64
```

From the figure illustrated above, it was found that the maximum likes count of a tweet is 132810, while the minimum likes count of a tweet is 81, and the mean likes count is about 8296.

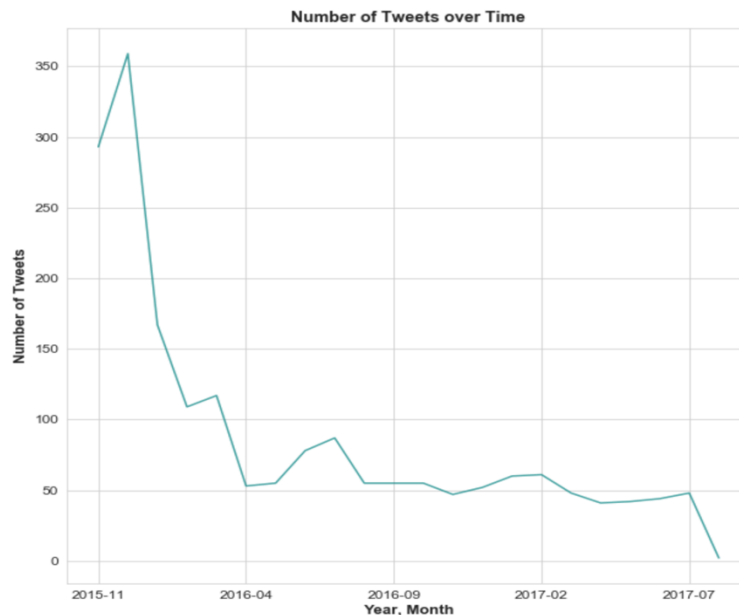**C.  Total Number of Tweets for each Year**

Counting the number of tweets for each year to see which year has the most number of tweets.

```
2015    652
2016    930
2017    346
```

The figure illustrated above shows that the year 2016 has the most number of tweets.

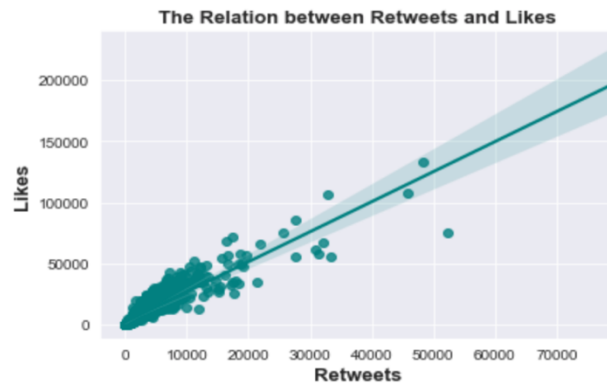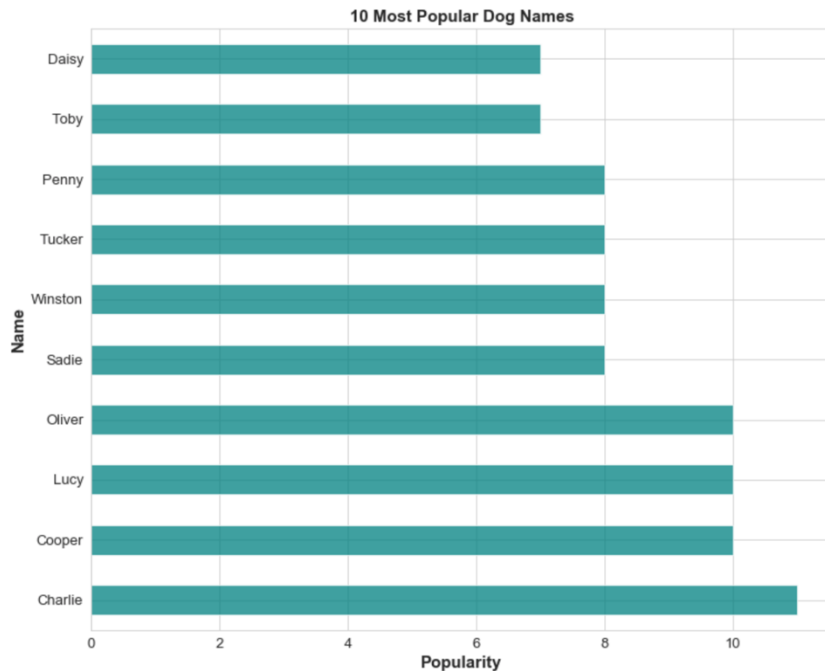## 3. VISUALIZATION

**A.  Number of Tweets over Time**



The figure illustrated above shows that several tweets were posted in December 2015. Afterward, the number of tweets decreased, then started increasing during the year 2016, and generally decreasing from there.

## B. The Relation between Retweets and Likes



From the figure illustrated above, it was found that the like and retweet counts are highly positively correlated. The majority of the data falls below 50000 likes and 20000 retweets, and the most popular tweet has about 135000 likes and 80000 retweets.

## C. 10 Most Popular Dog Names



From the figure shown above, it was found that the most popular name for dogs is Charlie, where it has been repeated more than 10 times, while the names Oliver, Cooper, and Lucy for a tie at 10 times each.

**D. Most Liked Picture**