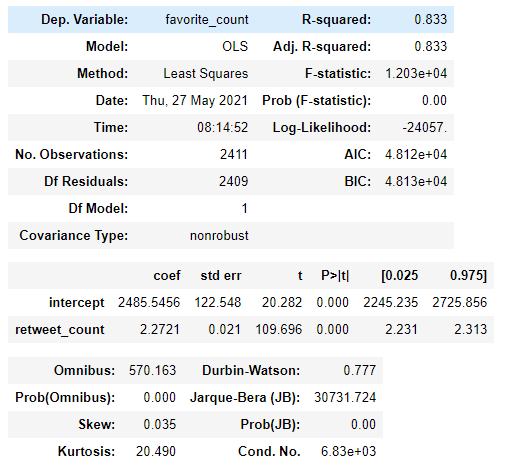
The Analyzing and Visualizing Process

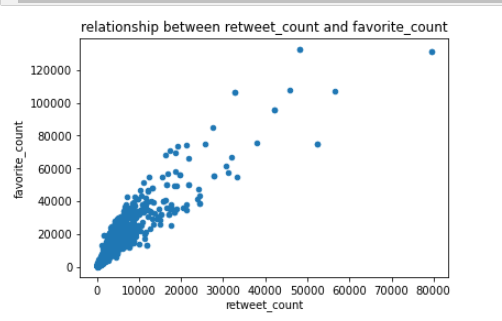
# Introduction

In this project I have done the analyzing and visualizing process in 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. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

# Analyzing and Visualizing Process

## The relationship between **retweet\_count** and **favorite\_count**

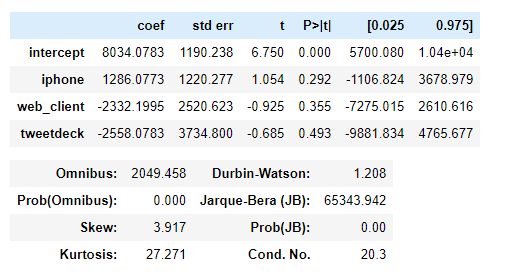
Using the Regression Model and scatterplot I found these interesting conclusions about the relationship between retweet\_count and favorite\_count.x

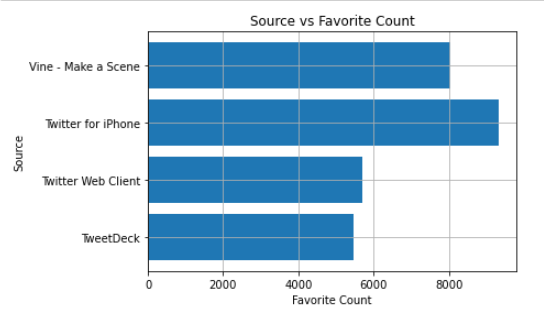


### From the model summary and scatter plot we can conclude the following:

* + The predicated favorite\_count = 2485.5456 + 2.2721 \* retweet\_count -> if the retweet\_count = 0, we predicate that the favorite\_count will be around 2486, and for every one more retweet, we predicate that the favorite\_count would be increase by 2.2721.
  + The p-value for retweet\_count = 0 -> retweet\_count is statistically significant for predicting the favorite\_count.
  + The R-squared (= 0.833) is closer to 1, which mean the better our model fit, and it is suggest that there is a positive strong realtionship between favorite\_count and retweet\_count (as you can see in the scatterplot). Also, we can interpret the value of R-squard as that 83.3% of the variability in favorite\_count is explained by the retweet\_count.

## The relationship between **soure** and **favorite\_count**





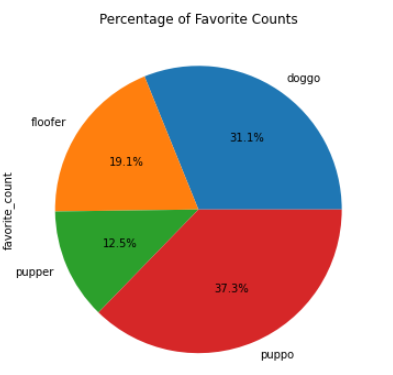
### We can get the following conclusions from the multiple linear regression summary and the plot

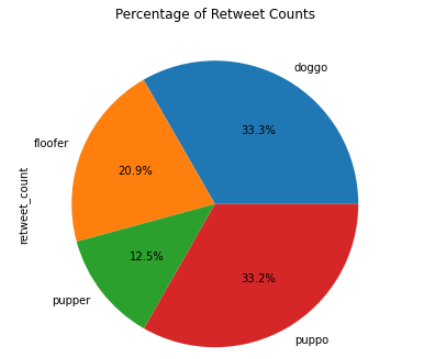
* Based on Multiple Linear Regression : If our tweets are from \*vine\* we predict its favorite\_count to be 8,034, the tweets from \*iPhone\* 1,286 greater than \*vine\*, a \*web\_client\* 2,332 less than \*vine\*, a \*tweetdeck\* 2,558 than \*vine\*.
* Based on plot: we can notice that if we tweets from iPhone the means of favorite\_count will be higher.

## The most popular breed of dogs

* + Based on prediction1 for the algorithm, the most favorite breed is **Golden Retriever** with sum of **1,977,583** favorite\_count.
  + Based on prediction2 for the algorithm, the most favorite breed is **Labrador Retriever** with sum of **1,670,195** favorite\_count.
  + Based on prediction3 for the algorithm, the most favorite breed is **Labrador Retriever** with sum of **776,200** favorite\_count.

## The heighest favorites and retweets means by dog stages





### From the above two pie plot we can conclude the following:

1. The highest favorite\_count for **puppo**
2. The highest retweet\_count for **doggo**

## 5. the relationship between **favorite\_count** and **rating**

### I found these conclusions about the relationship between favorite\_count and rating:

1. When the rating was > (median = 11) the mean of favorite\_count was = 15,936, and the retweet\_count = 4,934
2. When the rating was <= (median = 11) the mean of favorite\_count was = 4,685, and the retweet\_count = 1,625
3. The Higher rating go along with higher favorite\_count and retweet\_count