

# Wrangle and Analyze WeRateDogs dataset

*By Mohammed Essam Rashad*

## **Introduction**

The purpose of this project is use what I learned in data wrangling lesson from Udacity Data Analysis Nanodegree program. The dataset which will be wrangled is the tweets archive of **Twitter user @dog\_rates, also known as WeRateDogs**.

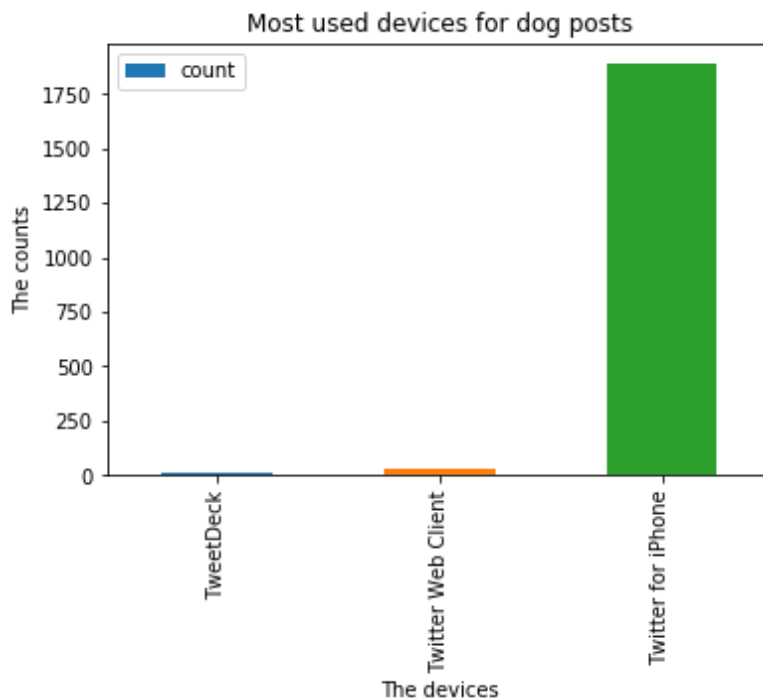
WeRateDogs downloaded their Twitter archive and sent it to Udacity via email exclusively for us to use in this project. This archive contains basic tweet data (tweet ID, timestamp, text, etc.) for all 5000+ of their tweets as they stood on August 1, 2017.

The goal of this project is to wrangle the WeRateDogs Twitter data to create interesting and trustworthy analyses and visualizations. The challenge lies in the fact that the Twitter archive is amazing, but it only contains very basic tweet information that comes in JSON format. So, I need to gather, asses and clean the Twitter data for a worthy analysis and visualization.

## **Most common device used to post Dogs tweets**

WeRateDogs has over 6000+ tweets. I was able to analyzed around 1500+ tweets. The most used device to post tweets on dogs related to WeRateDogs Account is **IPHONE** through **Twitter Iphone Application**.

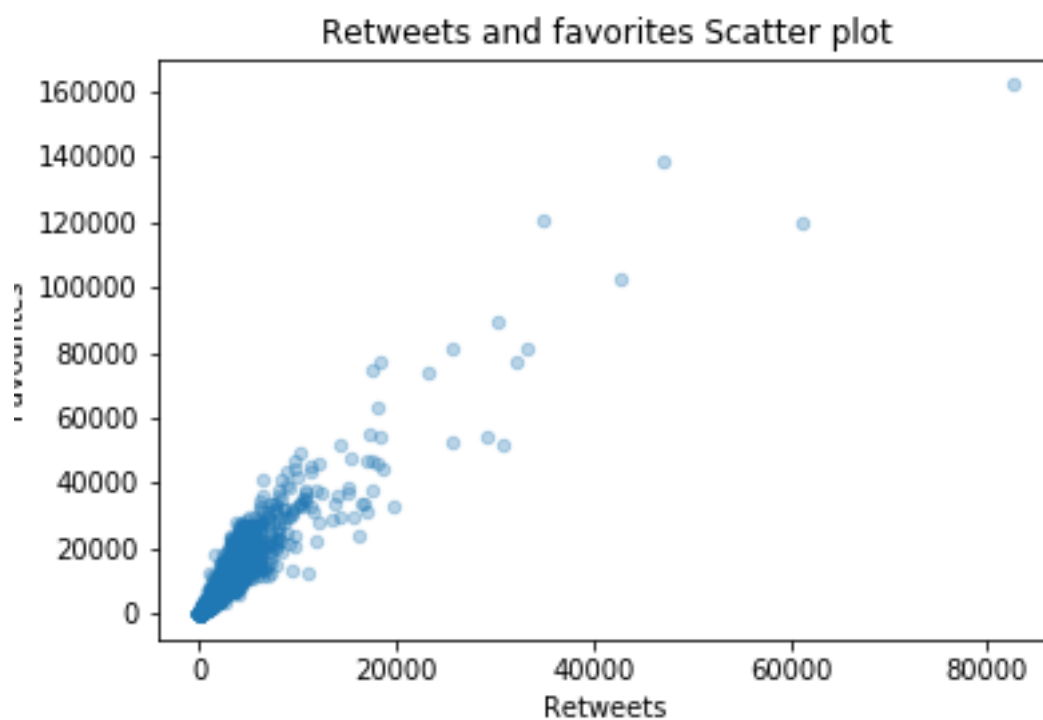
The below bar chart shows more info about this insight:



## Retweets vs Favourites(likes)

The blue dot on the below scatter plot represents the counts of retweets in respect of the counts of favourites(likes). This **plot shows that the retweets is high correlated with favourites(likes)**

The below Scatter Plot shows more info about this insight:



## Retweet counts

Amazingly many tweets have been retweeted more than **1000** times some more than even **50000** times. **Hence, there is not a clear relationship between the ratings and the retweets.**

