ANALYSIS AND INSIGHTS

Wrangle and Analyze Data: Disha Mukherjee

INTRODUCTION

This Wrangle and Analyze Data Project is a component of Udacity's Data Analyst Nanodegree. The project involves haggle of information from varied sources related to tweets from the Twitter user @dog_rates, conjointly called WeRateDogs. WeRateDogs rate's photos of people's dogs during a comic manner, most frequently giving ratings beyond 10/10. With of scraping data, quality and tidiness problems were assessed then cleansed. Finally, two visualizations were created, and insights can be found below.

FAVORITE VS RETWEET COUNT

At the time this data was collected, WeRateDogs had over 4 million followers; in this manner, their tweets are likely to urge numerous favorites and retweets. In expansion, there may be a few tweets that are immensely popular in this case they ended up part of international news or gone viral. In Figure 1, it can be seen that favorite and retweet tallies are exceedingly correlated. For about each 4 favorites there's 1 retweet. The tweet's share of the data falls underneath 40000 favorites and 10000 retweets. The foremost well-known tweet has almost 130000 favorites and 80000 retweets.
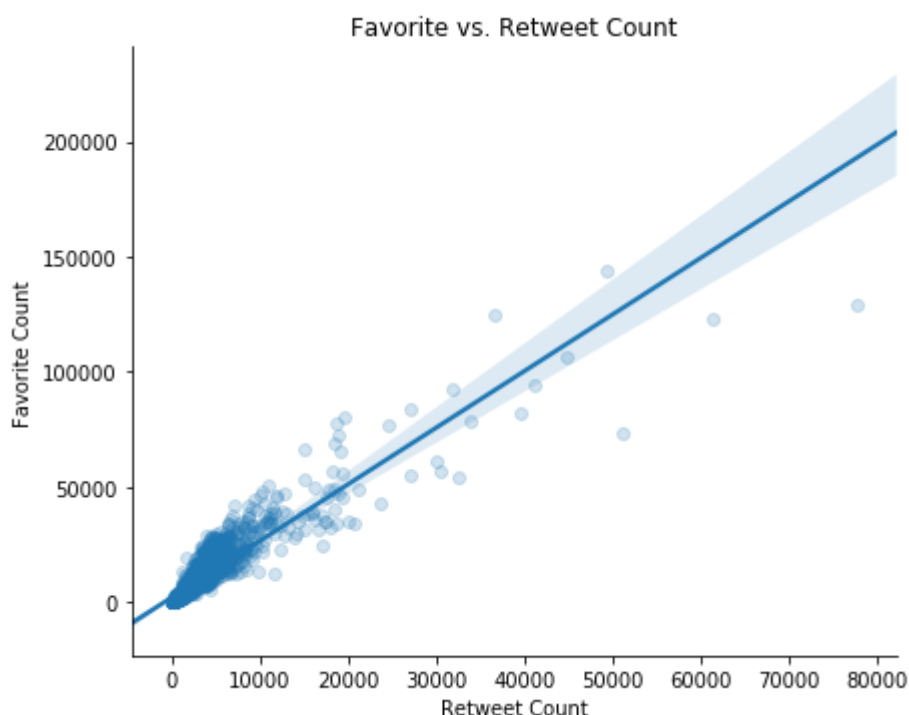

Figure 1. Favorite count and retweet count from 2068 tweets from the Twitter user @dog_rates, also known as WeRateDogs.

STANDARDIZED RATING OVER TIME

The thought behind the WeRateDogs account is that they ask individuals to send them photographs of their dogs, and they will rate them on a scale of 1-10 with funny comments; be that as it may, they are regularly given ratings that is higher than 10. I expected that nearly all the dogs were given a rating higher than 10/10 but I was shocked to take note numerous of evaluations were lower than 10/10. In expansion, numerous evaluations did not have a denominator of 10. Hence, to standardize the evaluations I calculated an standardize numerator separated by denominator. I was most

inquisitive to see in the event that extra time, as the account got be more prevalent and individuals related the over 10/10 ratings with being humorous attributes had got the higher evaluations would and gotten to be more popular. Undoubtedly, as appeared in Figure 2 and 3, shows up that over the time the recurrence of rating below 1 decreased. Sometime recently 2016-11 there are numerous evaluations underneath 1, whereas after that time there are scarcely any fall. The maximum standardized rating is almost 1.3 but for three exceptions(outliers) including the joke ratings 1776/10
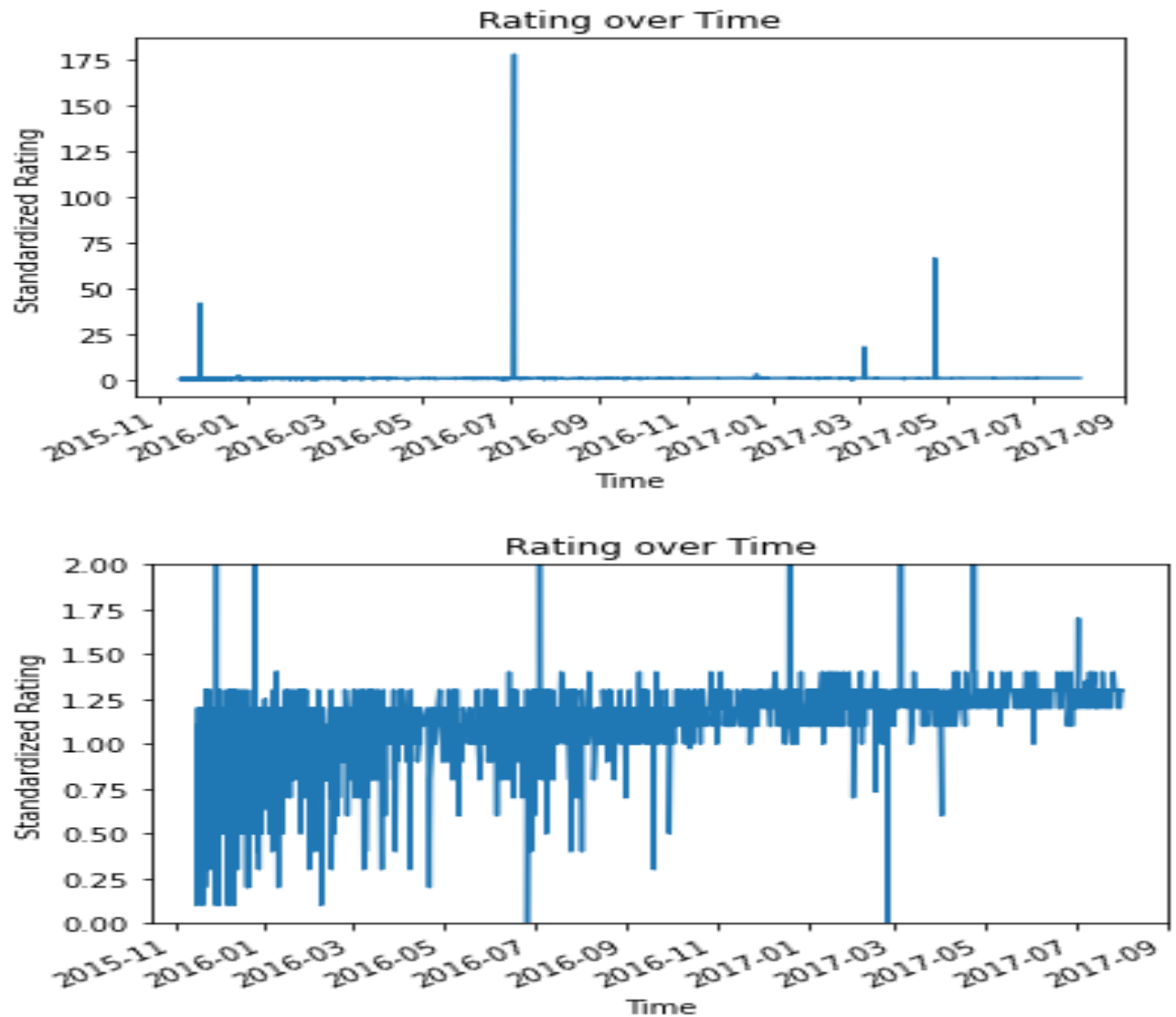

Figure 2. Standardized rating over time from 2068 tweets from the Twitter user @dog_rates, also known as WeRateDogs.