Data analysis and visualizations

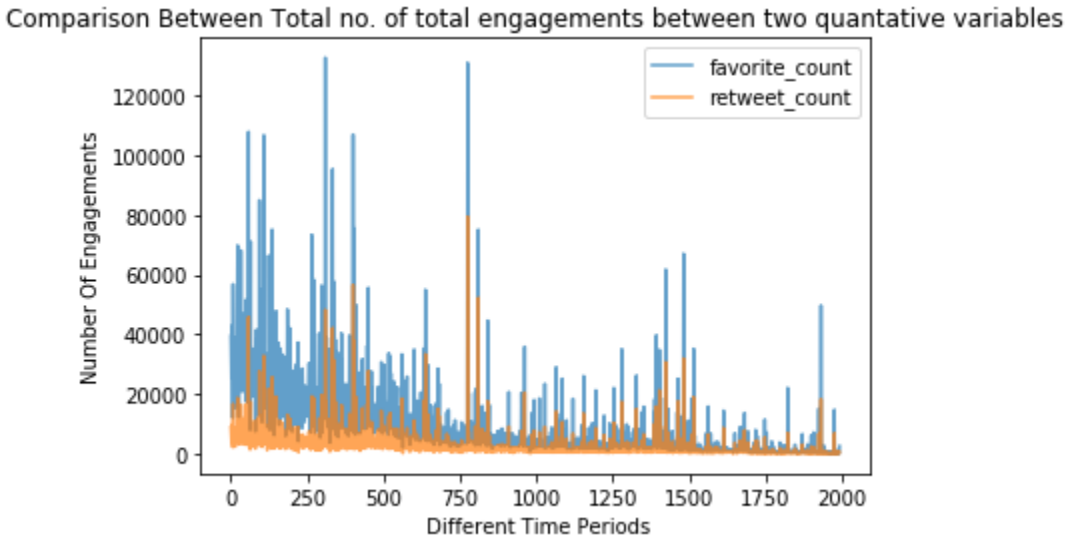
By: Mahmoud Esmail

Introduction:

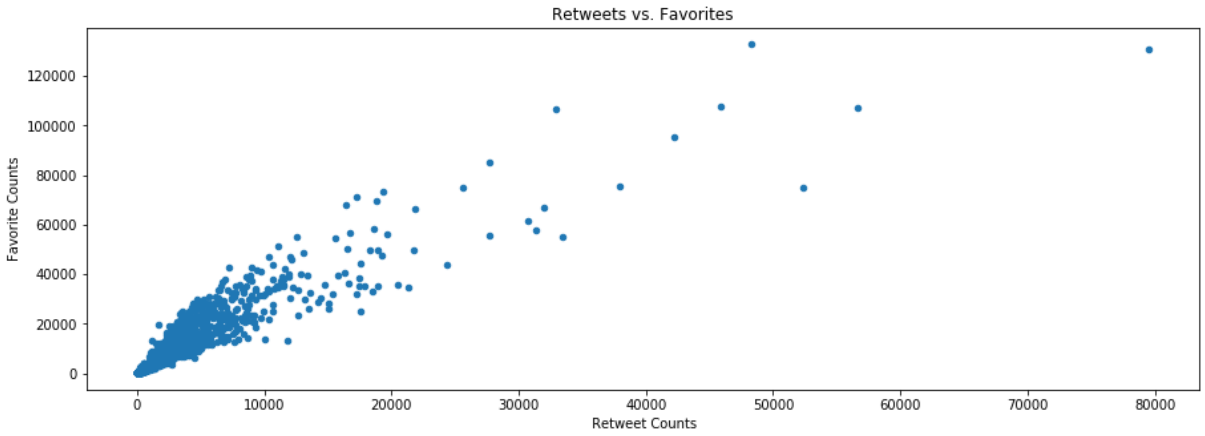
- Real-world data rarely comes clean, we wrangled a dataset to be able to analyze it

- The dataset we're analyzing and visualizing in this report 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. WeRateDogs has over 4 million followers and has received international media coverage, and they're known for a famous meme "They're All Dogs Brent

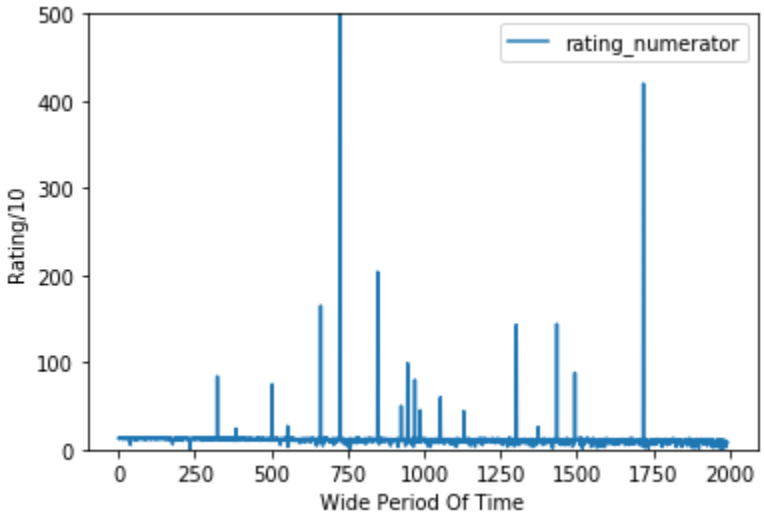
Visualizations & Some Insights on it:



- I see over the plot here, people tend to interact by using favorite button more than interacting with retweets
- the interaction with retweets was more than favorites only one time in the period before value 250
- this page went viral and then the engagement is decaying in the graph
- Note: this data archive originally contains 5000 tweets as they stood on August 1, 2017

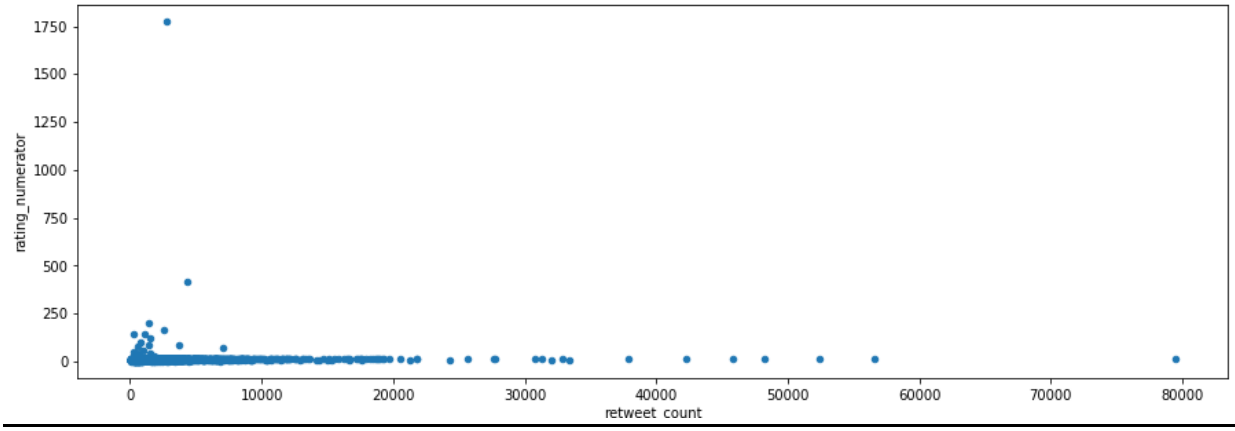


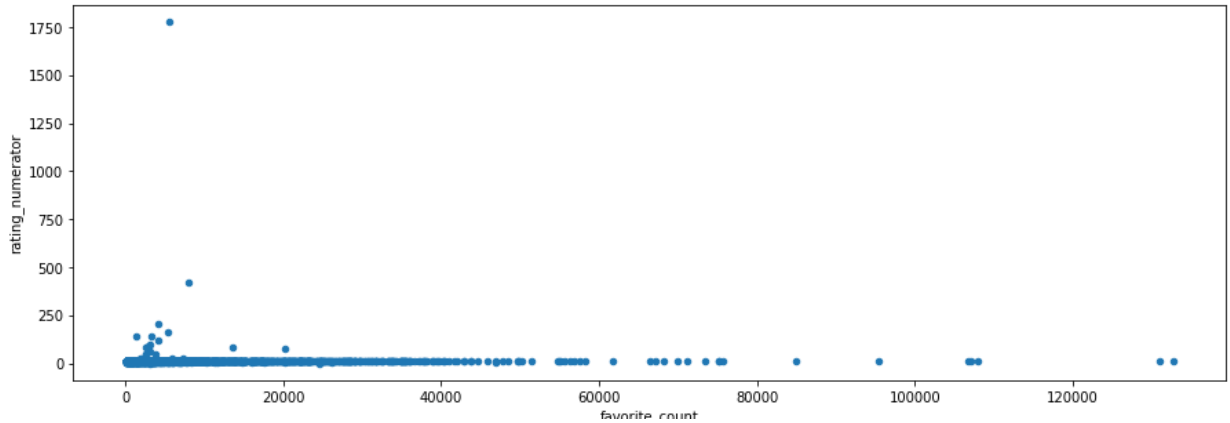
- though, the favorites engagements is always higher than retweets, there's a strong correlation between them
  as we see here in the period from 0 to 1000 on x-axis and from 0 to 20,000 on y-axis, also there's a small outliers
  in the other regions in the plot



- We see here, that the people's engagements on WeRateDogs account was very small in the beginnigs and started to grow and
  it's clearly that was the time that account created,
- there's a few outliers as the standarized rating is preferable not to exceed extremly high values

- extremely high ratings could be a sign of biasing towards certain type of dog or could be
  a sign that the people are in love with content of the account





- as we see when we look at those two figures we see that the higher ratings tends to have a smaller retweets engagements than that of favorites engagements