Data Visualization Report

WeRateDogs Tweet Analysis

Introduction

The dataset comes from the tweet archive of Twitter user @dog_rates, also known as WeRateDogs.

WeRateDogs is a Twitter account that rates people's dogs with humorous comments about the dog. These ratings almost always have a denominator of 10 though the numerators can be absurdly large. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

I got access to their tweet archive via Udacity. 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 Data Gathering process is supposed to comprise 3 methods of gathering data:

- The WeRateDogs Twitter archive
  to be downloaded manually

- The tweet image predictions
  image_predictions.tsv... To be downloaded.programmatically

- Additional data from the Twitter API
  Each tweet's retweet count and favorite ("like") is to be queried from Twitter's API into a file, tweet_json.txt.

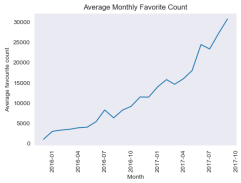
However, due to my Twitter Developer account not being verified in time, I proceeded to download this file manually as Udacity foresaw scenarios like this.

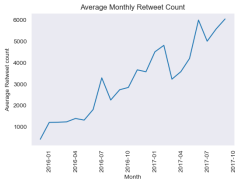
Data Wrangling:

For more info on the data process carried out on this project, click this link.

Data Explorations

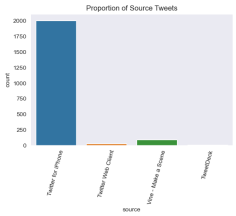
How did the WeRateDogs tweet perform overtime?





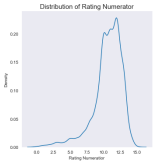
From the above line plots, it appears that the average amount of retweets and likes increased overtime.

Which of the sources Were the WeRateDogs Tweets tweeted from the most?



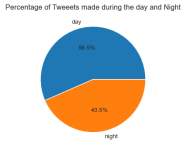
From the countplot we could easily see that iPhone users were represented in the WeRateDogs datasets.

Rating Distribution

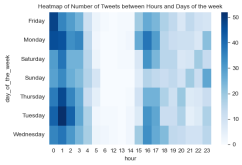


The graph shows that the kdeplot was skewed to the left which means that it clustered the most within 12 and 12.5. This implies that most of the dogs were rated within the cluster region of the kdeplot.

What time of the day do most people tweet?

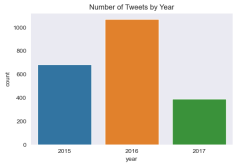


From the observed pie chart,it shows that about 56.5% of the WeRateDogs tweets were made during the day(12 am the morning to 5 pm) while the remaining 43.5% were made at night.



From the heatmap, we could see that most of the WeRateDogs tweets were tweeted in the early hours of the day, from 12 midnight to 3am and rapidly declines from the 4am to 2pm .This could be partly down to their busy schedule during the day as most people are known to work with that period of the day.For much of the evening the tweet activity slightly pickup again as most persons are known to have returned from work during that period. Mondays, Tuesday and Fridays tweet activities are mostly within the period of 12, 1 and 2 in the early hours of the morning.

What year has the most WeRateDogs Tweets?



From the observed barplot, it shows that most of the WeRateDogs Tweets were made in 2016.

Conclusion

After wrangling and carrying out exploratory data analysis. We were able gain the following insights:

- There was steady rise in the amount retweets and likes
- About 56.5% of the WeRateDog tweets were made during the day
- WeRateDogs tweets attracted the highest average number of likes in the month of July.
- Most of the dogs were rated between 12 and 12.5.
- An overwhelming amount of WeRateDog tweets were made from an iphone.
- 2016 was the year with the highest number of WeRateDog tweets.

-
- WeRateDog tweets were tweeted the most in the early hours of morning between 12 am and 2am.
- There is always a spike in tweets activities on Monday.