

Analysis of @WeRateDogs Twitter Account (Udacity Project) – Report

By Manuel Cabrera

This report is a summary of the most relevant insights on the analysis performed on the @WeRateDogs Twitter account data.

Introduction

@WeRateDogs is a Twitter account that makes humorous comments on pictures and videos of dogs using common internet lingo. He also gives a ranking to the dog based on how a "good boy" the dog is.

The purpose of this project is to gather, assess, clean, and analyse data obtained from Twitter user @WeRateDogs and provide any significant insights.

Datasets and Gathering

There were three main data sets used:

- The Enhanced Twitter Archive – provided by Udacity, it has been enhanced by programmatically extracting variables from the body of the tweet (rating numerator, rating denominator, etc...)
- Twitter API (Likes/Favourites + Retweets) – extracted from Twitter's API, mainly for the likes/favourites and retweets for each of the tweets.
- Image Prediction file – programmatically extracted from Udacity servers, this file is the result of using a neural network on the media (pictures/videos of a dog) from each tweet and identifying the dog breed for each tweet.

Assessment and Cleaning

Database had to be assessed and cleaned prior any further analysis. Main issues were found using grammatical and visual inspections, some of the issues were:

- Variables having an incorrect category.
- Invalid variables.
- Multiple variables combined into a single variable.
- Unnecessary information.
- Single variable split into multiple variables.
- Tables having different observational units.

Analysis and insights

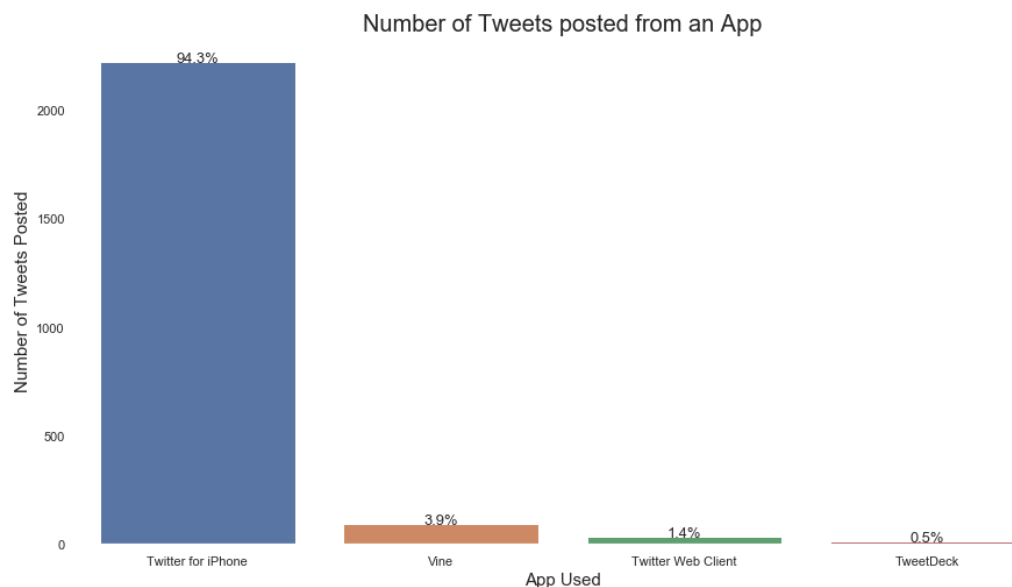
The main questions raised from this project were:

- What is the user's most used platform/device to tweet from?
- What is the @WeRateDogs user preferred time to tweet?
- What is the best time to tweet according to previous user engagement?
- Is there a difference between like and retweet engagement?
- Does the user prefer certain dog breeds (According to rating)?
- Do certain dog breeds promote a higher user engagement?

Using exploratory and explanatory visualisation, it was possible to answer the proposed questions for this project:

What is the user's most used platform/device to tweet from?

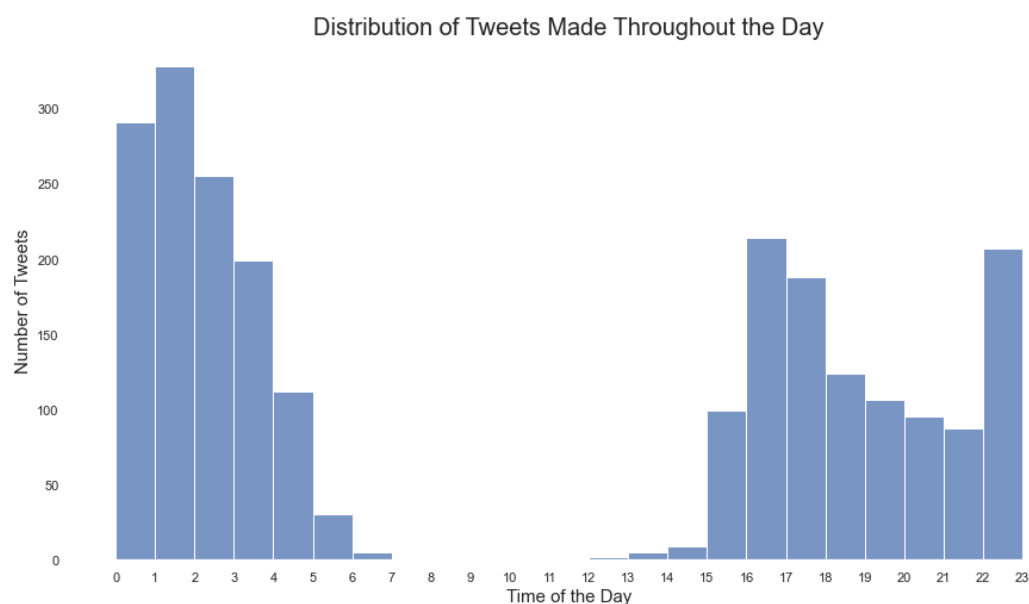
@WeRateDogs user posts most of its tweets from the Twitter app for iPhone(94.3%), followed by vine (3.9%) (vine was a social app which as of 2019 was discontinued).



What is the @WeRateDogs user preferred time to tweet?

@WeRateDogs user made most of his tweets during late hours; most of his interactions were between 22:00 and 03:00. Surprisingly, the user did not make any tweets between 07:00 and 12:00; there could be possible explanations for this:

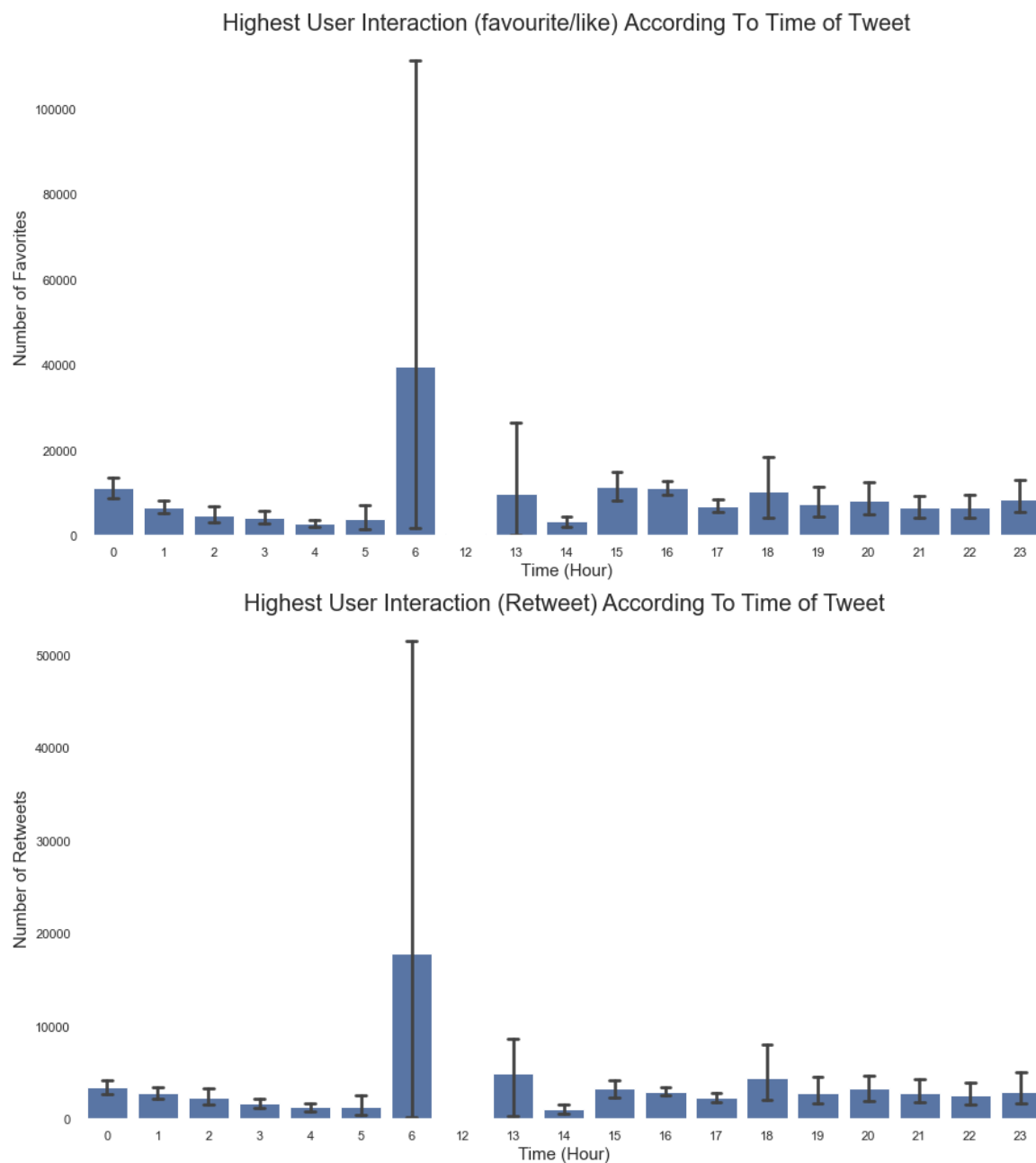
- Incomplete data
- Incorrect timestamp extracted (time was from a different time zone the user was in).
- The user restrains using his phone during the morning and midday and performs his tweets during his own leisure time, to avoid any impacts on his work or studies.



What is the best time to tweet according to previous user engagement? And is there a difference between like and retweet engagement?

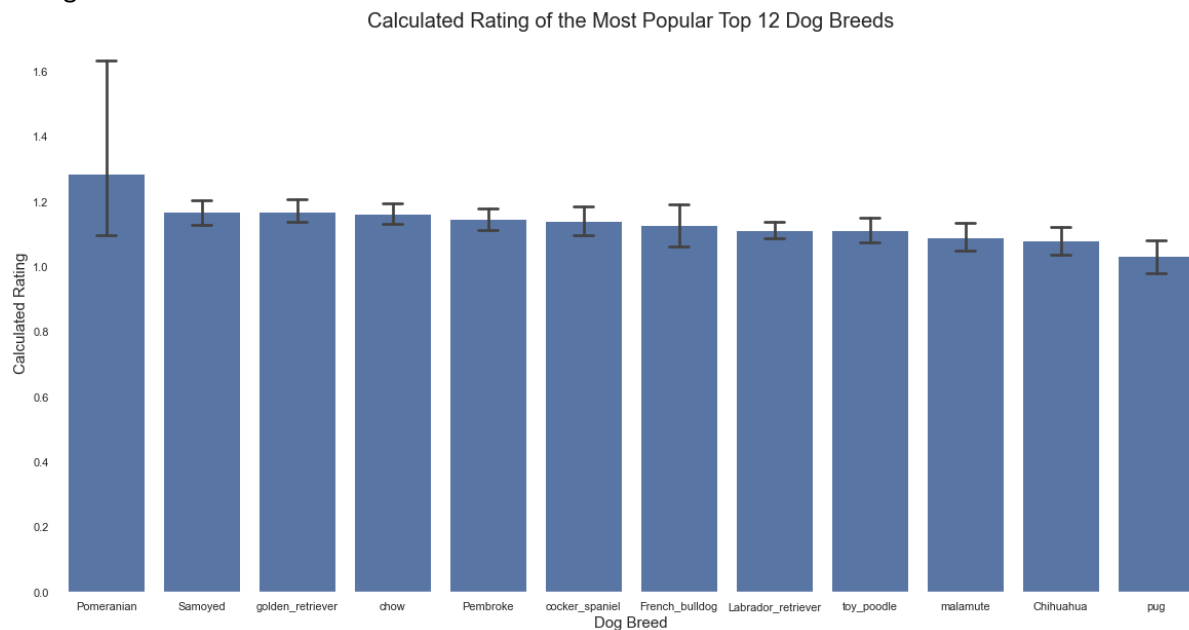
User engagement can be categorised as likes/favourites and retweets. Surprisingly, the preferred engagement time for these is slightly different. For better like/favourite engagement, it is recommended to tweet between 15:00 and 18:00. The engagement throughout the evening and late night seems to be consistent with a big drop at 01:00.

Whilst looking at the bar charts, the largest surge in favourites/likes and retweets occurred at 06:00, however, it is important to point out that the error bar for this timeframe is exceedingly high and it would be best to ignore this timeframe.



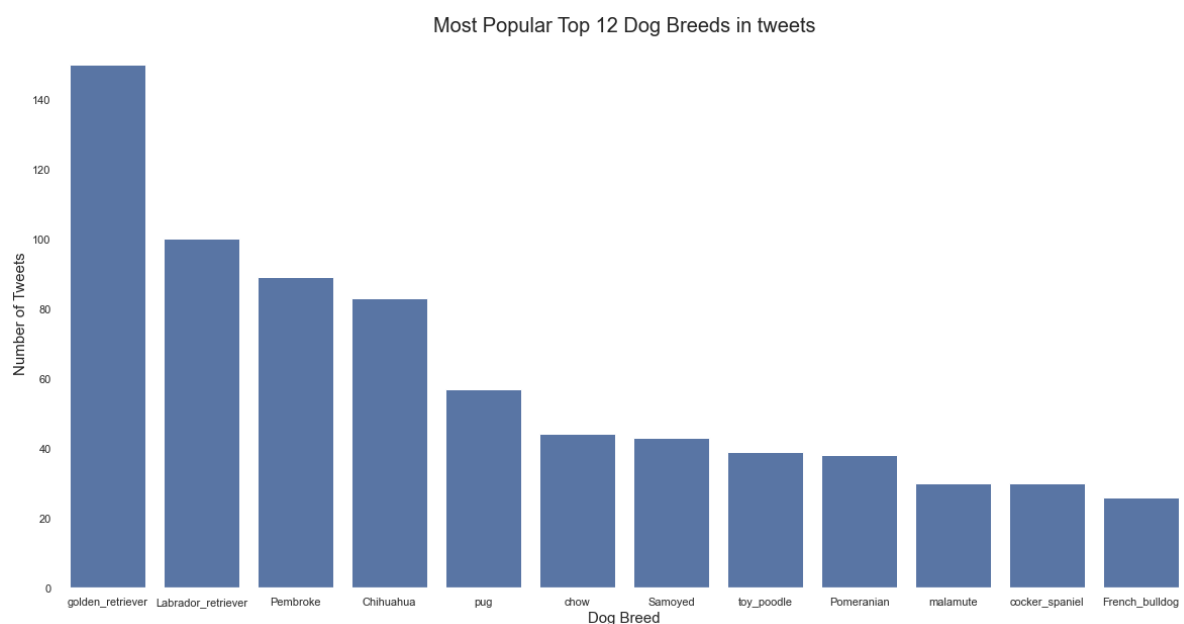
Does the @WeRateDogs user prefer certain dog breeds (According to rating)?

Using the data from the top 12 dog breeds ranked in his tweets (provided by Udacity's Machine Learning team) there seems to be that Pomeranian dogs normally get a higher rating compared to other breeds. Nevertheless, most of the remaining top 12 dog breeds have a similar or close overall rating.

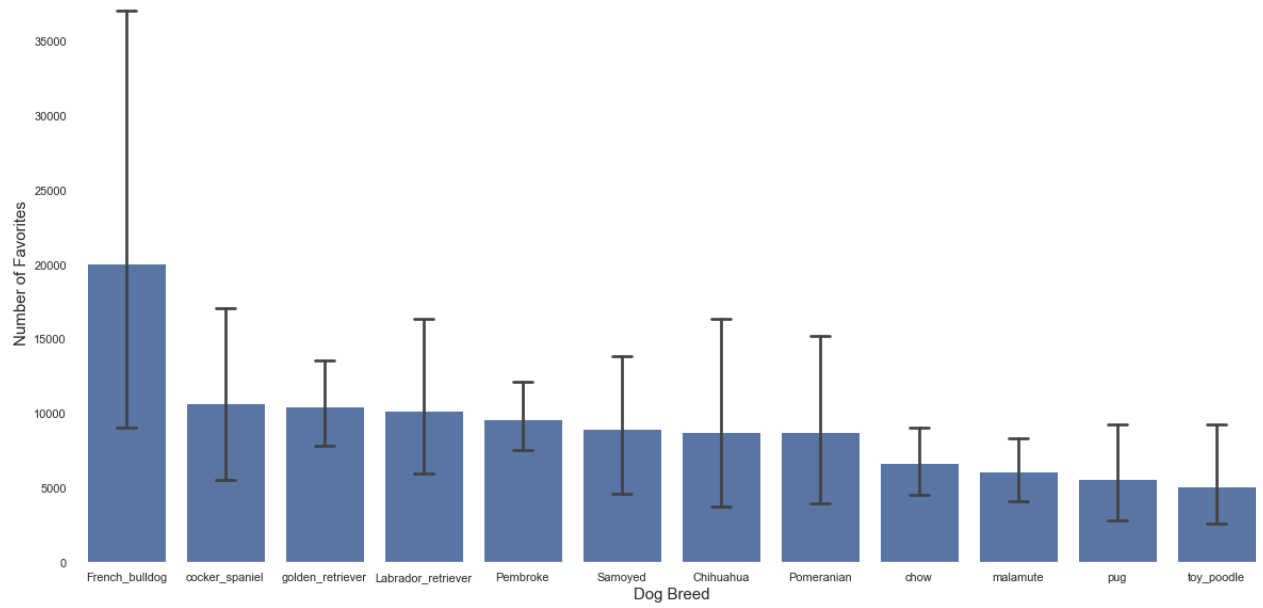


Do certain dog breeds promote a higher user engagement?

Surprisingly, despite the Pomeranian being @WeRateDogs user preferred breed according to rating and the Golden Retriever being the most common dog breed for the user to rate. The French Bulldog and the Cocker Spaniel are the top 2 dog breeds that have the highest user interaction, followed by the golden retrievers; both the French Bulldog and Cocker Spaniel are the least common dog breeds in @WeRateDogs tweets (the least from the top 12 dog breeds).



Number of Favorites/Likes according to Most Popular Top 12 Dog Breeds



Number of Retweets according to Most Popular Top 12 Dog Breeds

