Project: Wrangle and Analyze WeRateDogs Tweet

*By: Redi Sunarta - Udacity Student (2020)*

Exploratory Data Analysis Report

WeRateDogs is an account that designated to rating dogs on their appearences and stories in twitter since 2015. The account has grown tremendously in popularity with many users sharing its content and requesting their dogs being rated as well. Even though, their own unusual rating system which majority dog is rated above 100% because ['They are good dogs, Brant' (https://knowyourmeme.com/memes/theyre-good-dogs-brent)](https://knowyourmeme.com/memes/theyre-good-dogs-brent). They also driving force the development of the dog culture with it's new dog terms such as pupper, floofer, and doggo. My job is to create interesting and trustworthy analyses and visualizations based on the cleaned dataset before which contain WeRateDogs tweet archive for 2015-11-15 to 2017-08-01. I can raise a few questions and know a several things using the information available in the dataset about tweet pattern over time, what kind of characteristic most get like, or what kind of animals, outside dog, are commonly detected in models. I'll use favorites (likes) count and retweet count as our main indicator for proxy of netizen rating. In fact, `rating column` can't be used because its humor purposely.

Research Question

> The more spesific questions as follow: **How WeRateDogs account behave or netizen response the his tweet? What kind of dog stage that undetected as dog? What kind of dog stage most likely to be favorites? Which breed of dog have highest median favorites?**
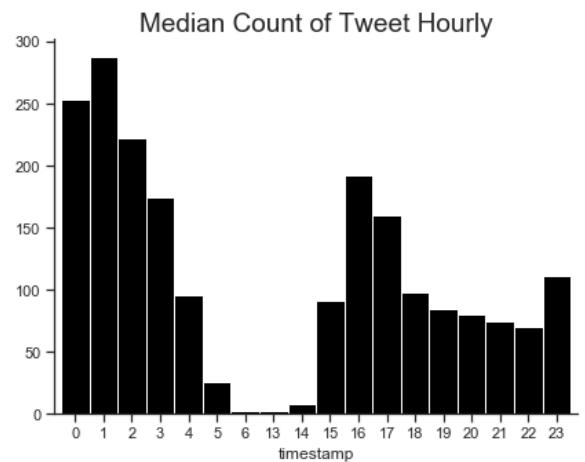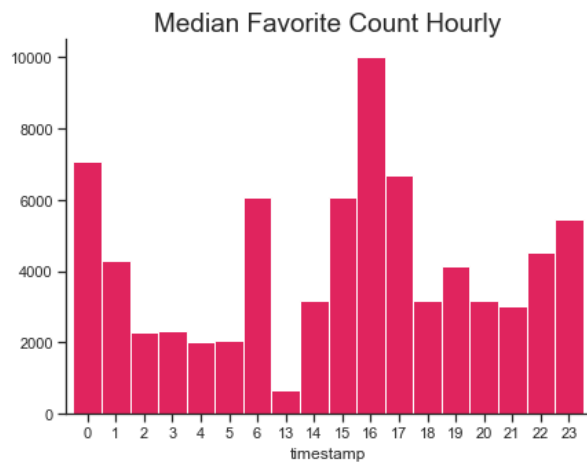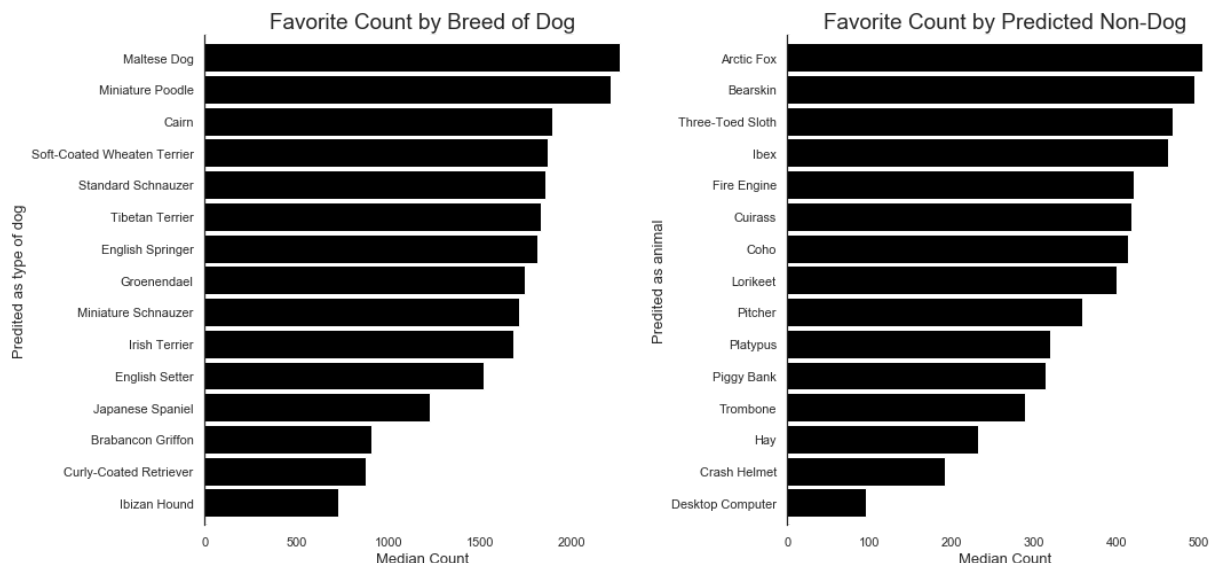
Key Insight

- Based on median favorite count and retweet count by hourly aggregation, WeRateDogs never tweet among 7-12 clock. I dont know what if the account use tweeter bot, have more than one admin, or other possibility. But, if Matt Nelson (https://en.wikipedia.org/wiki/WeRateDogs) manually made tweet alone, that time must to be his schedule for bedtime or doing his 'real-job' every day. Moreover, the count of tweet on clock 5,6, and 14 is lowest point, even less than 3% of total tweet. Moreover, early morning is the time most often WeRateDogs tweet although the afternoon is a time to get lots of likes/retweets.
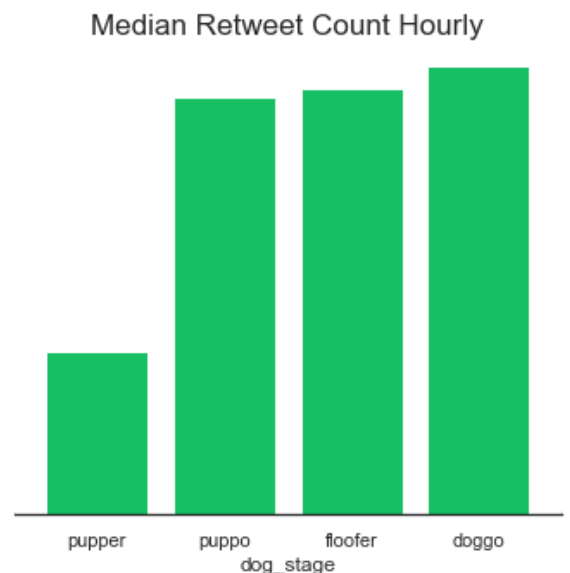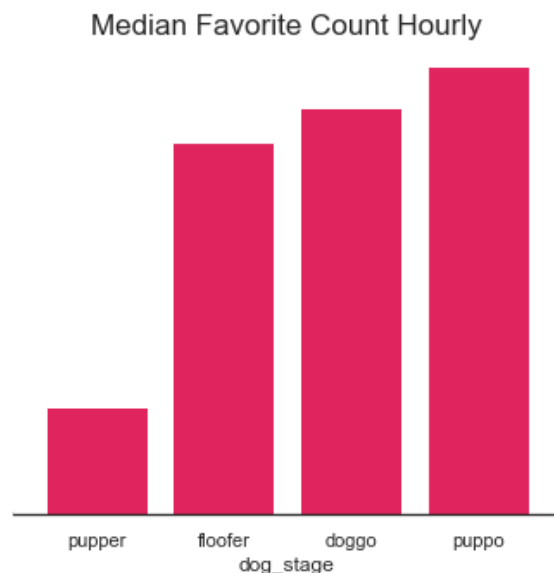
Key Insight

- Stage of dog, pupper, have least of likes/retweets. Also, pupper have highest count, 60 times, detected in model as another animal compare to Doggo (16 times), puppo (4 times), and floofer (0 times). While, based on median of like and retweet count, doggo is most favorite stage.

Favorite Count by Breed of Dog

Favorite Count by Predicted Non-Dog

Key Insight

- If the predicted photo show another animal, the tweet have low like count, even compare 1th another animal with the 15th most favorite bread of dog. Those group even almost doesn't reach 500 likes. This is interesting to empasize and can indicate that netizen/follower want to WeRateDogs focus on only dog content. If the account tweet doesn't related with dog content nor don't look like a dog, the response (interaction intensity) will be decrease.

> **Disclaimer**: This exploration only utilize descriptive statistics. I can't be certain that this relationship has a causal relationship. Moreover, I just use few columns in analysis process. Indeed, I can't discuss so much about dog because my culture has no close relationship with dogs. Therefore, I doesn't have good prior knowledge about dog.