PROJECT 4: WRANGLE AND ANALYZE DATA
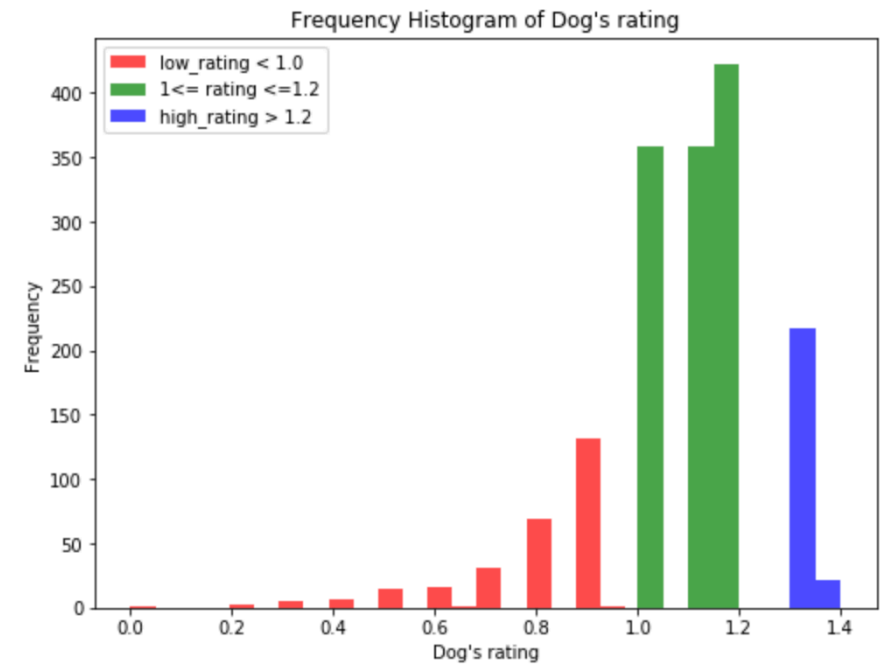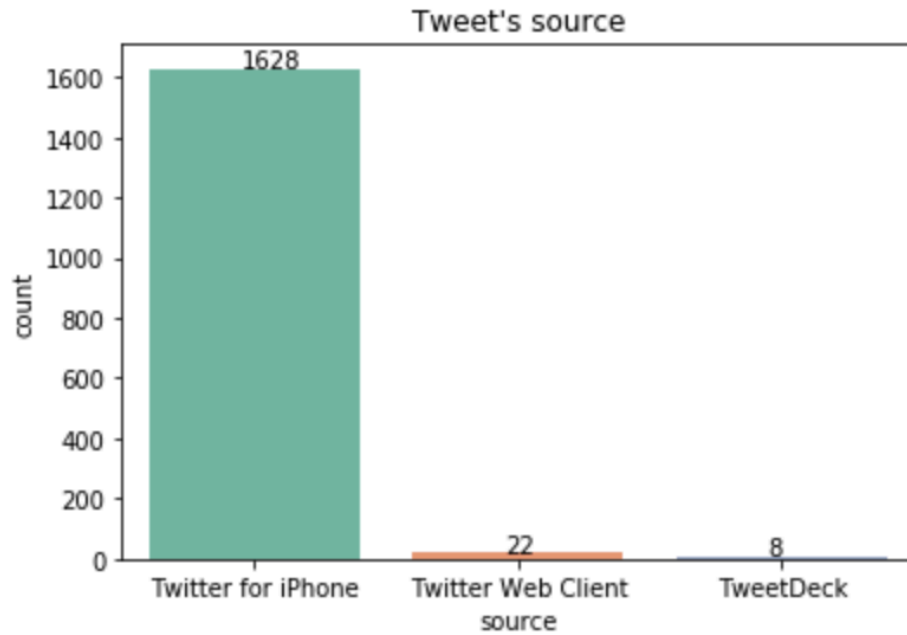
WeRateDogs

INTRODUCTION

Objective:

This project focused on wrangling data from WeRateDogs Twitter account using Python. The project is coded using Jupyter Notebook (wrangle_act.ipynb) and this report will be focusing on communicates the insights and displays the visualizations produced from my wrangled data.

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10.

The numerators almost always greater than 10. It has over 4 million followers and has received international media coverage.
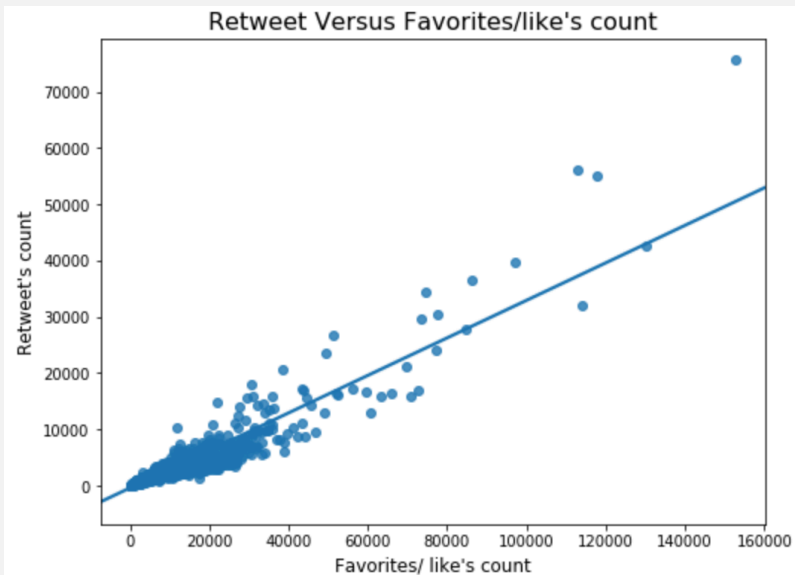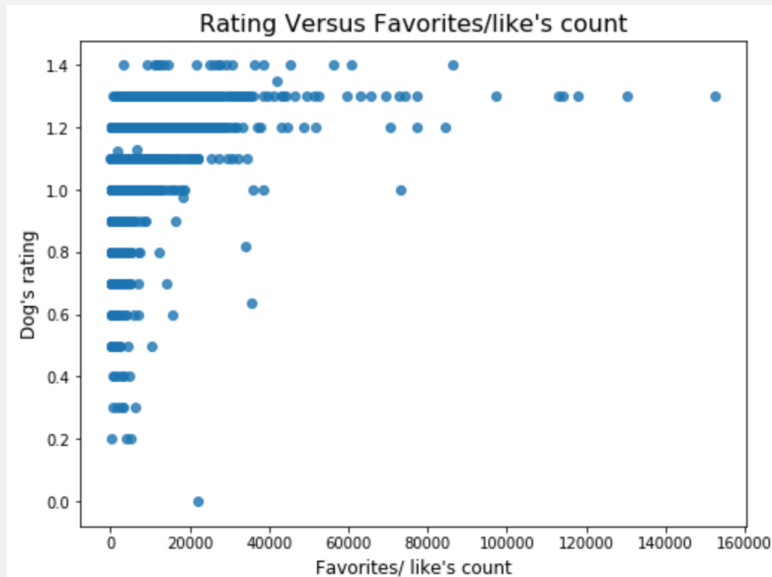
QUESTIONS TO FIND OUT FROM THE DATA

- Which is the common platform user used to push their tweet?

- How's the rating distribution?

- Any correlation between rating with favourites **(like)** counts?

- Any correlation between favourites count and retweet count?

- What dog breed is most popular (high number of favourites, counts)?

- What is the common (high in number of tweet) dog's stage tweet in this account?

- What is the day of week with highest number of tweet post?

- How's the trend of Rating and favourites count over time?

- Which is the top rate dog (highest dog's rating) ?

- Which is the most popular dog (highest favourites count) ?

EXPLORATORY DATA ANALYSIS AND VISUALISATIONS

- Vast majority, **98.2%** of user uses iPhone to tweet.
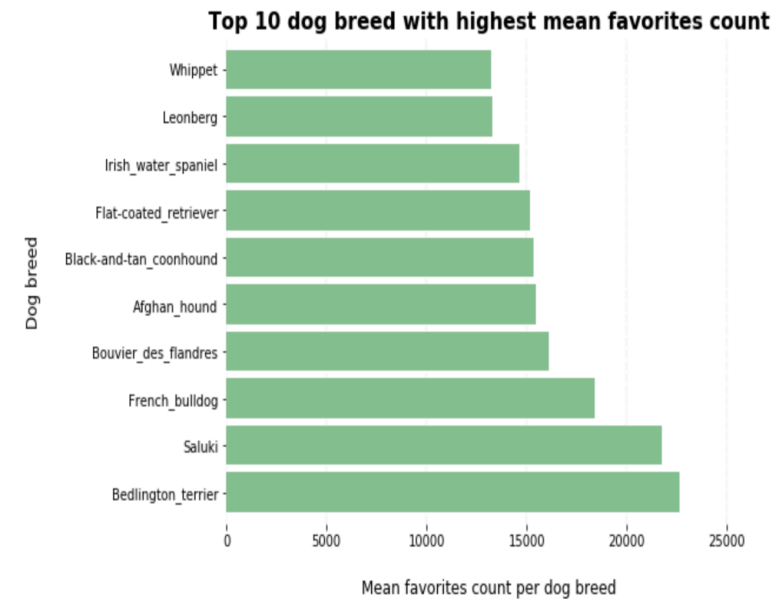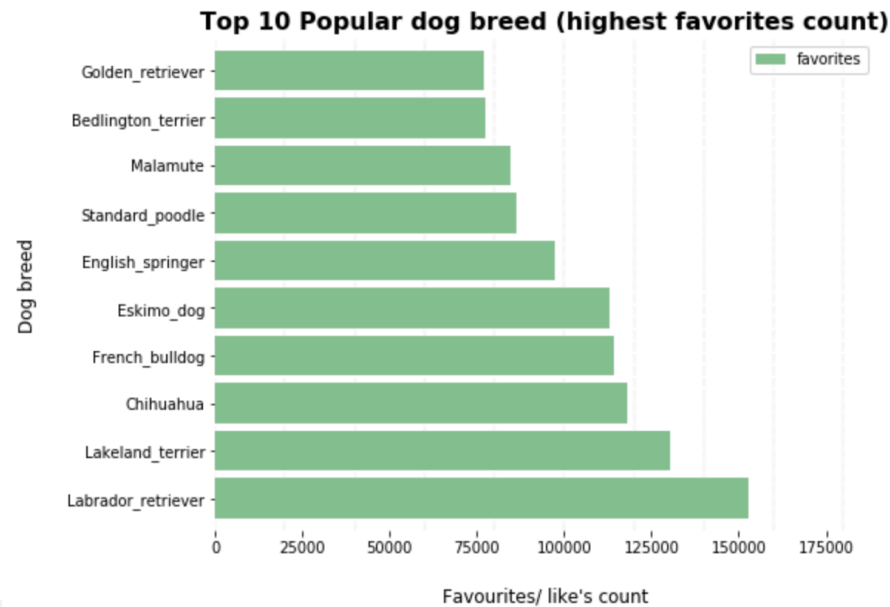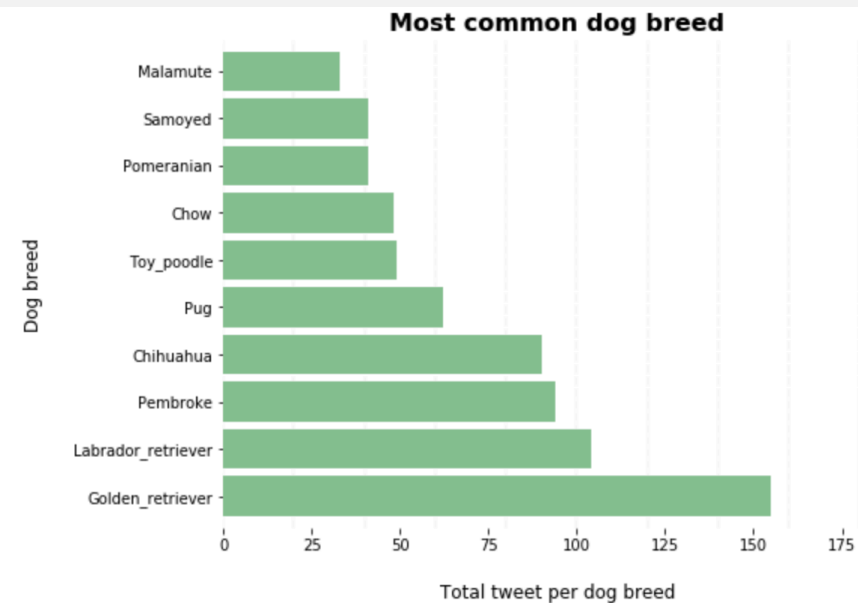- Majority of rating is between 1.0 to 1.2.

Rating Versus Favorites/like's count

Retweet Versus Favorites/like's count

CORRELATION BETWEEN RATING VS FAVORITES, RETWEET VS FAVORITES

Weak positive correlation between dog's rating and favorites count (top chart), with correlation coefficient around r = 0.4

From the scatter plot (top), it was observed that low favorites count spread across all the dog's rating (0 – 1.4). However, higher rating tends to have relatively higher number of favorites count.

Strong positive correlation between retweet and favorites count (r = 0.93). This is logical as retweet tends to reach out to more audience. In addition, people likely to retweet post that they like.

Most common dog breed — Total tweet per dog breed

Top 10 Popular dog breed (highest favorites count) — Favourites/ like's count

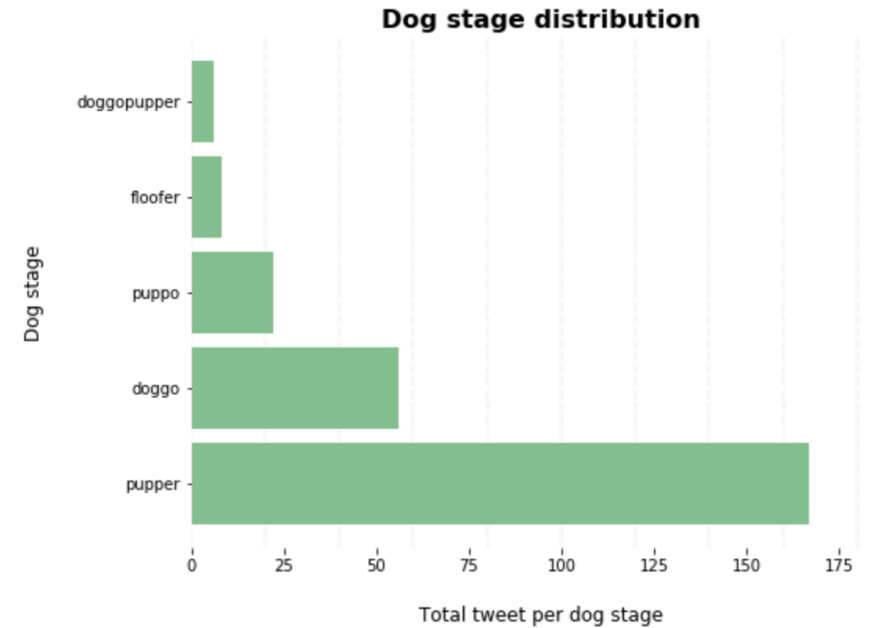Top 10 dog breed with highest mean favorites count — Mean favorites count per dog breed
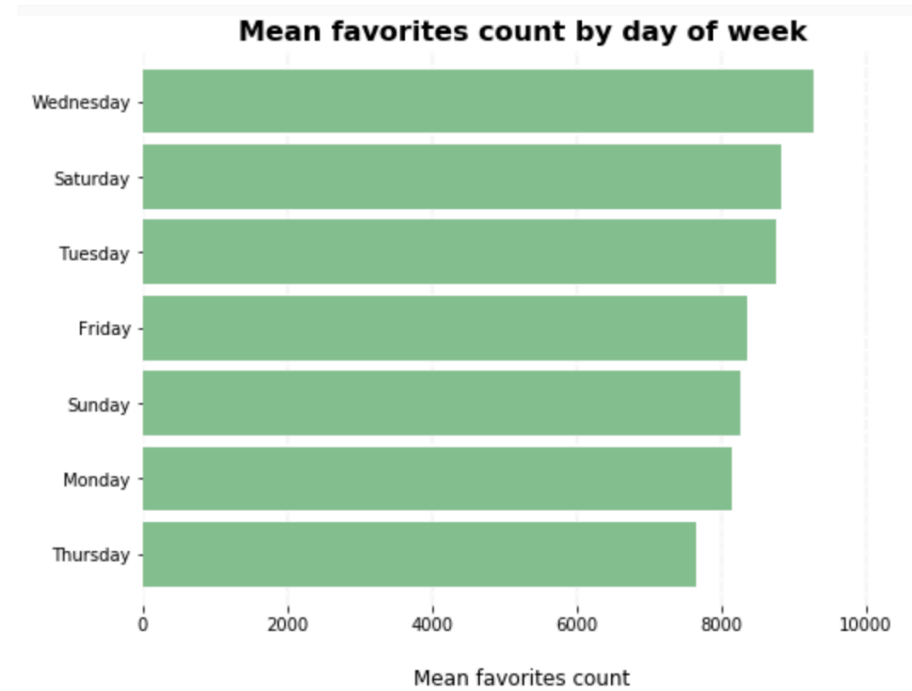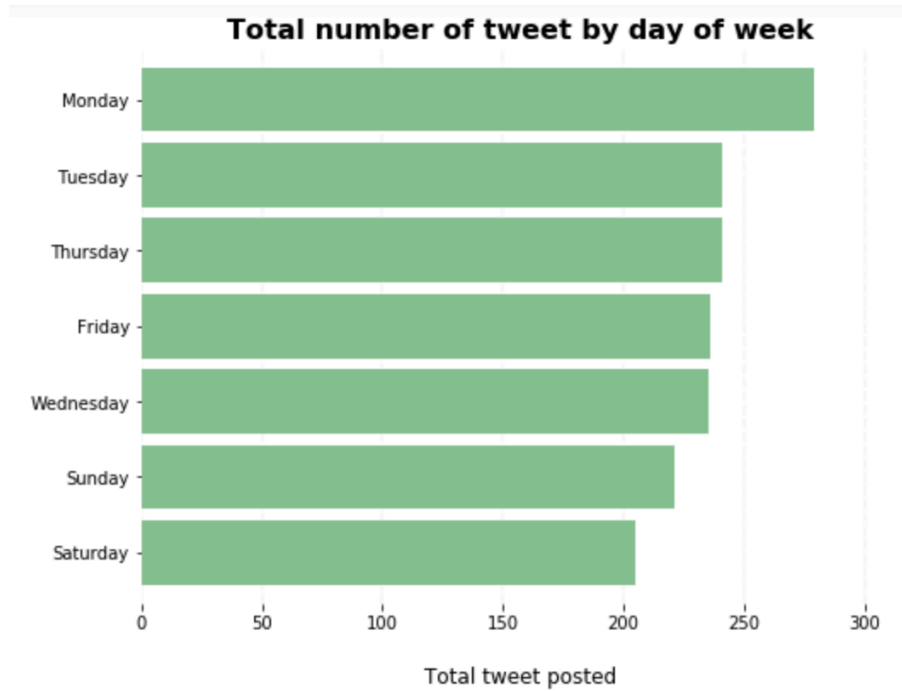
POPULAR DOG BREED

- There are total of 113 unique dog breed.
- Above bar chart showed 10 most common dog breed, which has the highest number of tweet. **Golden retriever** top the list!

- Popular dog breed showed the top 10 refers to tweet that received highest number of favorites count.
- **Labrador retriever** top the list!

- Chart above showed the dog breed that has highest MEAN favorites count.
- They are typically less common dog breed. This means, common dog breed on average has lower favorites count.

COMMON DOG STAGE

Common dog stage appeared in the tweet is pupper, about 85% of all the known dog's stage.

Based on the explanation from *Dogtionary*, pupper is a small doggo, usually younger. Can be equally, if not more mature than some doggos.
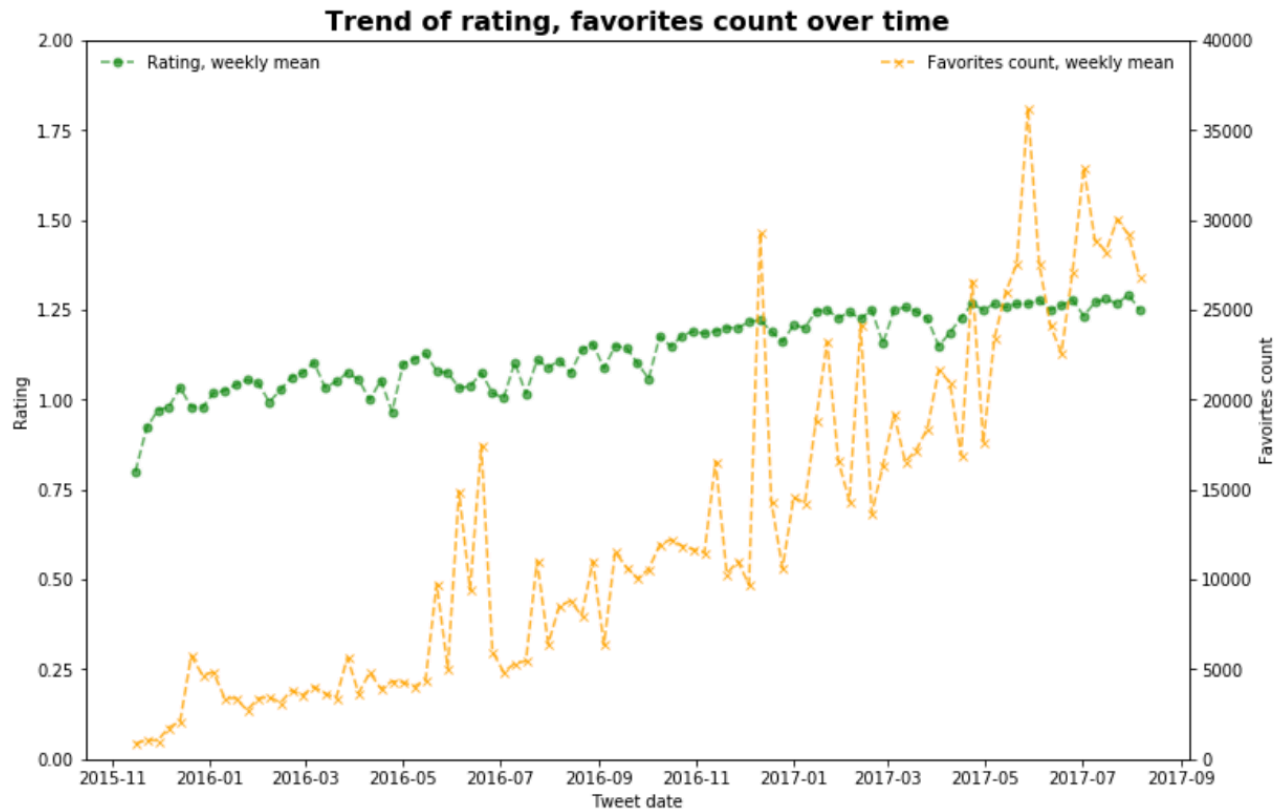

Dog stage distribution

Total number of tweet by day of week / Mean favorites count by day of week

TWEET DAY

- Monday appeared to be the day with highest number of of tweet.
- Wednesday is the day with tweet that has highest MEAN favorites count.

TRENDS

- Dog's rating is in the upward trend over the year. In year 2017, tweet will get higher rating (>1.0) compared to year 2015.

- Favorites count increases significantly by year, which indicating the twitter account is gaining popularity over the year.

- Most Popular | Here's a doggo realizing you can stand in a pool
- Breed | Labrador retriever
- Rating: 1.3
- Favorites | 152629
- Retweet count | 75580
- Stage | doggo

- Top rating dog | This is Bo. He was a very good First Doggo
- Breed | Standard poodle
- Rating | 1.4
- Favorites | 86342
- Retweet count | 36520
- Stage | doggo

MOST POPULAR DOG | TOP RATING DOG