

# Wrangle and analyze data

Act report

5 TH PROJECT

Udacity Data Analysis Nanodegree

## Data Wrangling Act Report

### Introduction

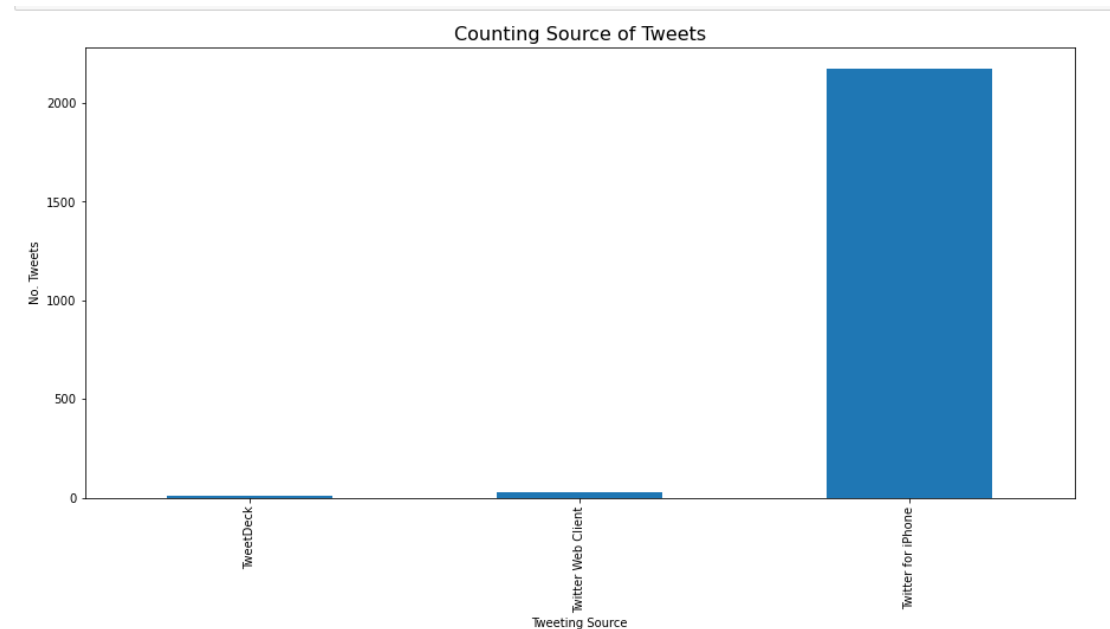
This Project demonstrate the data wrangling process using twitter data on archive called WeRateDogs. In this archive, users are given a dog image and they rate it with their comments. It is worth noting that the setup of the rating system requires that the rating denominator is always higher than 10. In the project, I have used this dataset to practice data wrangling steps through gathering, accessing and cleaning the data inside the archive. In this report, we summarized the data analysis process done after data wrangling.

### Exploratory Data Analysis

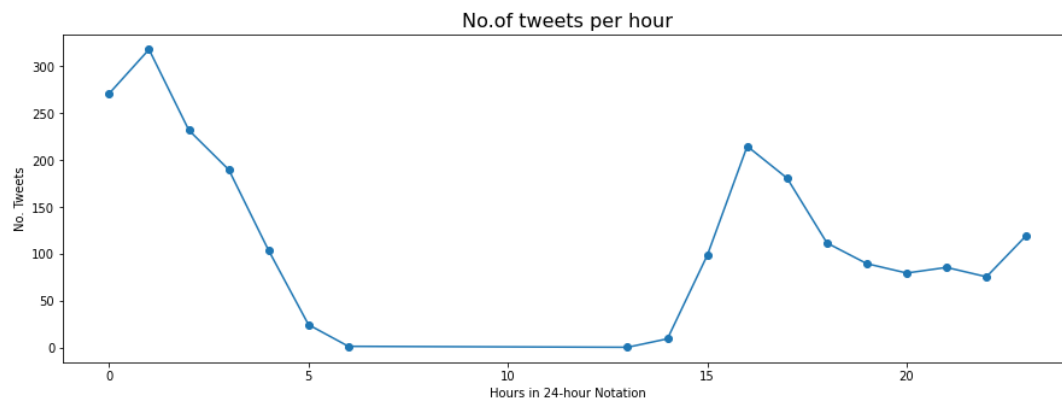
In this report, I aimed to answer some question related to our dataset:

- What are the most used source for rating in weRateDogs?
- What is the most common hour of dogs rating tweets?
- What are the top favorited bread?
- Is there graphically relationship between 'retweet\_count' and 'favourite\_count'?
- What is the image with the highest favorite\_count
- What is the image with the highest rating\_numerator

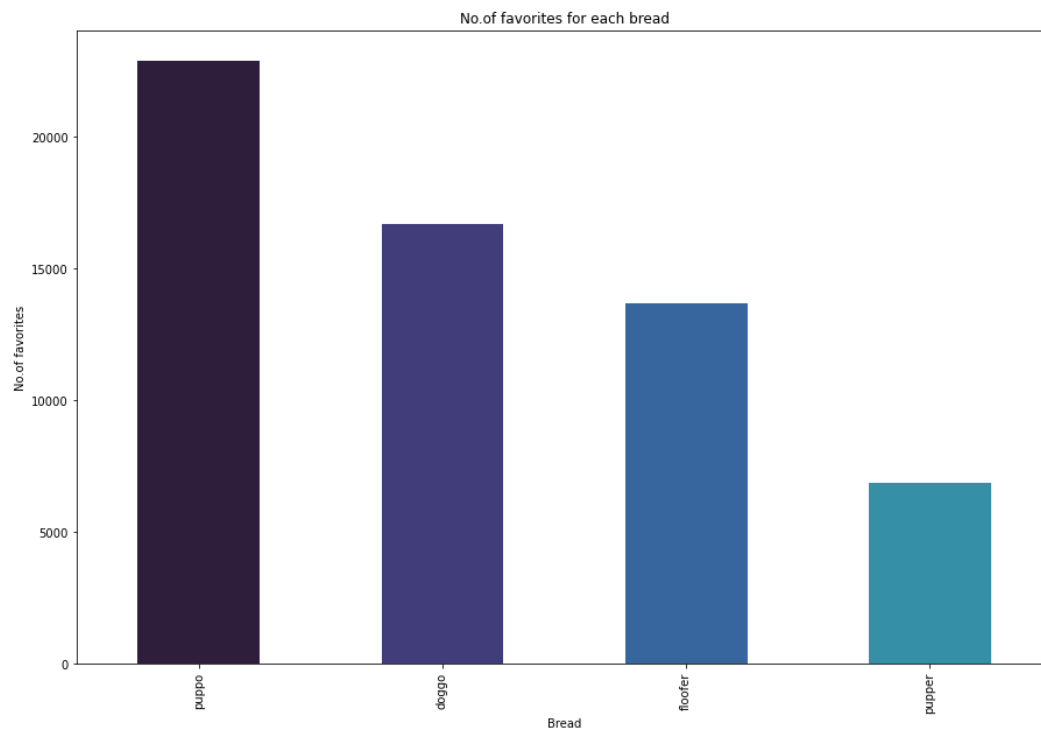
### Results:



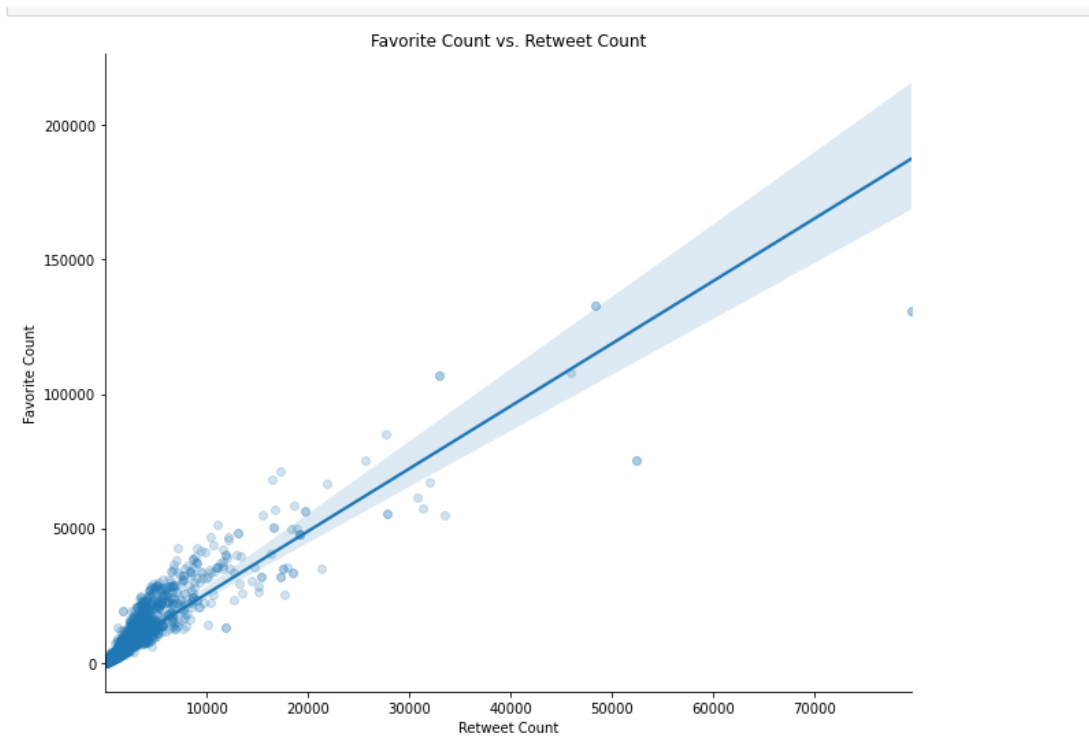
from the figure above, it is clear that twitter for iPhone is the most used application for tweeting in this WeRateDogs followed by twitter web client .



we can see from the above figure that the most common tweeting hour is between 1 am - 2 am



from the figure, we can see that the stages in average favorites. On the other hand, the top bread was puppo followed by doggo.



The scatterplot revealed that favorite counts and retweet counts are positively correlated, with the majority of the data falling below 50000 favorites and 10000 retweets.

the image with the highest favorite\_count :



the image with the highest rating\_numerator:



Resources :

<https://stackoverflow.com/questions/32370281/how-to-embed-image-or-picture-in-jupyter-notebook-either-from-a-local-machine-o>

[https://seaborn.pydata.org/tutorial/color\\_palettes.html](https://seaborn.pydata.org/tutorial/color_palettes.html)