WeRateDogs - Analysis and Insights

Introduction :

This Wrangle and Analyze data project was the 4th project of Udacity's Data Analyst Nanodegree program.It involved wrangling and analyzing the tweet archive of Twitter user @dog_rates, also known as WeRateDogs. 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, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 4 million followers and has received international media coverage.

Gathering Data :

This project involved gathering data from 3 sources :
1.**Enhanced Twitter Archive -** Provided by the Udacity team
2.**Image Predictions File -** The tweet image predictions, i.e., what breed of dog (or other object, animal, etc.) is present in each tweet according to a neural network. This file (`image_predictions.tsv`) is hosted on Udacity's servers and was downloaded programmatically using the request library
3.**Additional Data via the Twitter API -** Each tweet's retweet count and favorite ("like") count at minimum, and any additional data you find interesting. Using the tweet IDs in the WeRateDogs Twitter archive, queried the Twitter API for each tweet's JSON data using Python's tweepy library.

Assessing Data :

After gathering data, the assessment on the data was done using following methods :
- head
- Info
- Describe
- value_counts()

Tidiness issue that were solved :

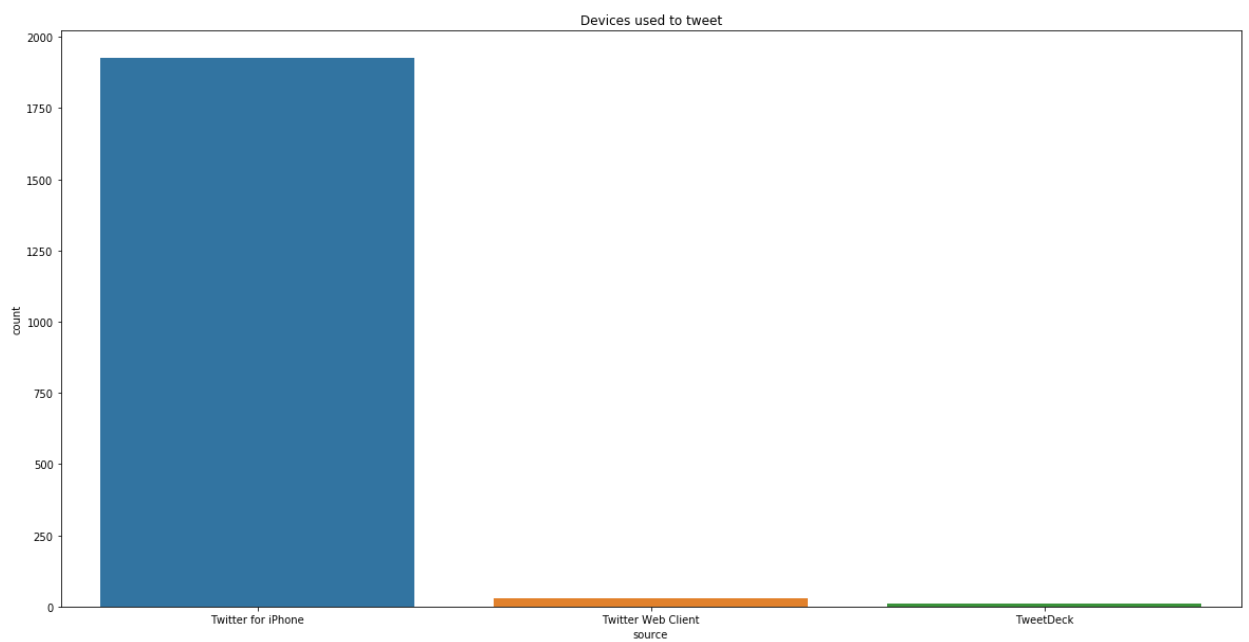- Combining data from all 3 sources into 1 final table

- Combining 4 different variables(columns) into 1 named dog_type

Quality issues were cleaned :
- Completeness of Data
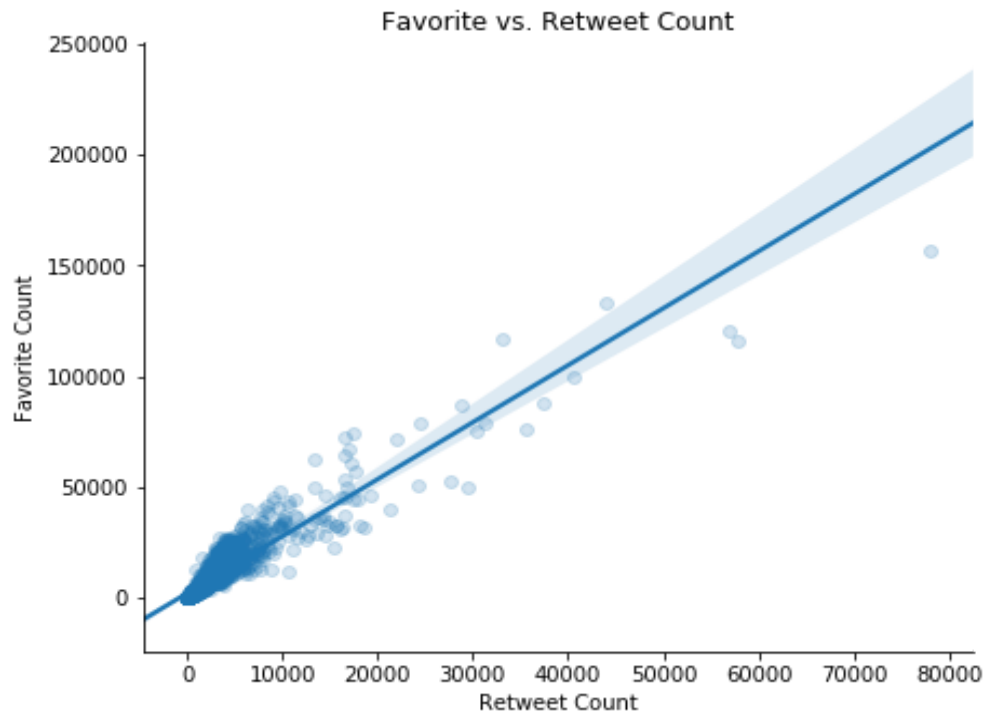- Validation of Data
- Accuracy of Data
- Consistency of Data

<u>**Visualizations :**</u>

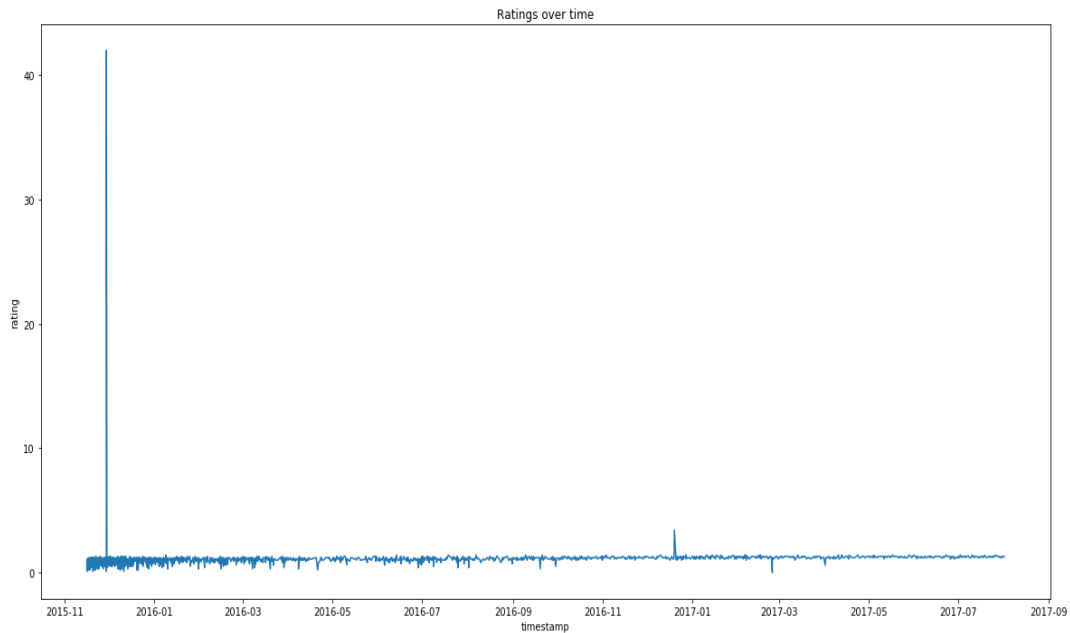- **Checking the source of all the tweets**

Almost all the tweets by the user were made from iPhone

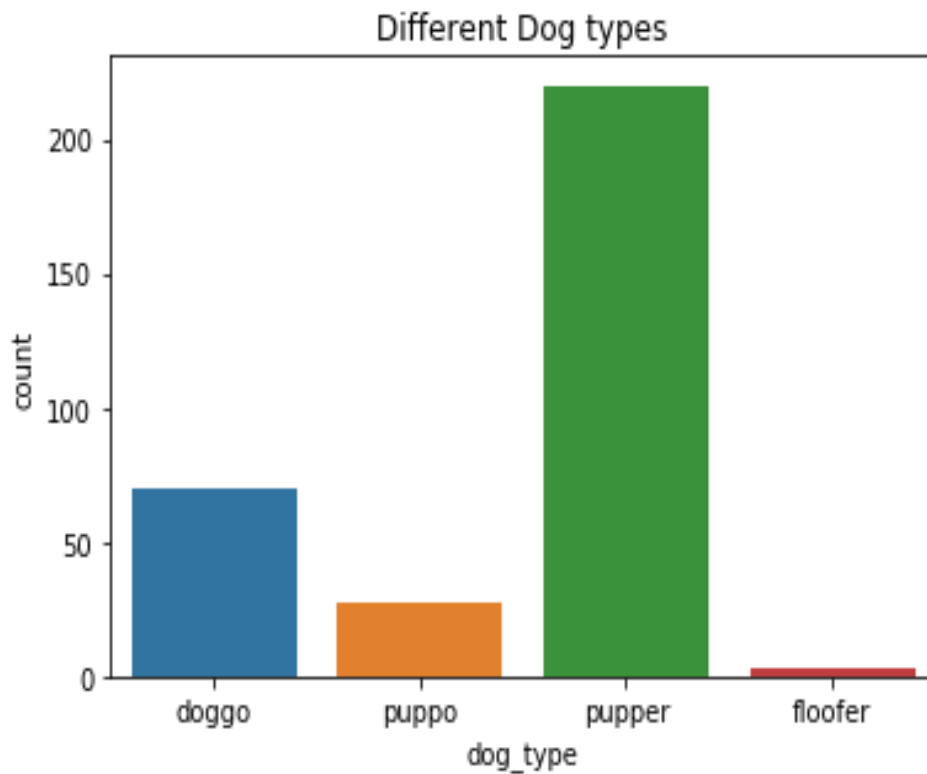- **Checking the relationship between favourites and retweets**

Favourites and Retweets are positively correlated and have a linear relationship.Most of the data falls below 4000 favourite and 2000 retweets.
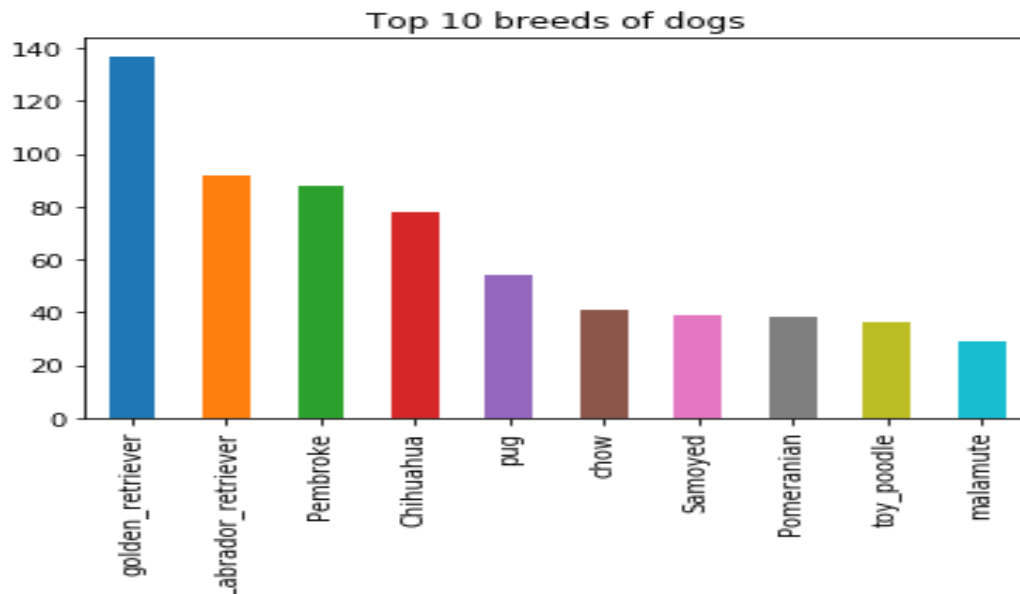
- **Plotting a line chart for rating over the time**

Most of the rating are below 3 , but there is a sudden and unusual increase in rating in 2015-12

- **Checking different types of dog types**

Different Dog types

Pupper has the highest count among all the dog types

- **Checking Top 10 breeds of dog**

Golden Retriever is the most owned dog breed followed by labrador