REPORT

ON

THE ANALYSIS MADE ON

THE WeRateDogs TWITTER

DATA

INTRODUCTION

Real-world data rarely comes clean. The dataset that I had cleaned, analysed and visualized was the tweet archive of Twitter user @dog_rates rates, also known as WeRateDogs. WeRateDogs. is a Twitter account that rates people's dogs with a humorous comment about the dog.

This project works through the data wrangling process, focusing on the gathering, assessing and cleaning of datasets. However, this report presents the results from all analysis and visualizations performed.

ANALYSIS AND VISUALIZATION:

To answer some of my curiosities as to What the most popular dog breed is?, What relationship there is between the number of retweets a tweet has and the number of favorites such tweet gathered?, Through what major source were all tweets made? And what is the most common dog stage?

I have wrangled the WeRateDogs twitter data and have analyzed it to get answers of these questions. Python libraries have helped me with some cool visualizations for my analysis.

Following are the results of my analysis and visualization:

1. Analysis and Visualization of the most popular dog breed

I had realized the data used for this analysis from extracting the correctly predicted dog breeds. Which is a proof of quality for this analysis. Figure 1.0 below shows the visual frequency representation of the ten most popular various dog breeds.

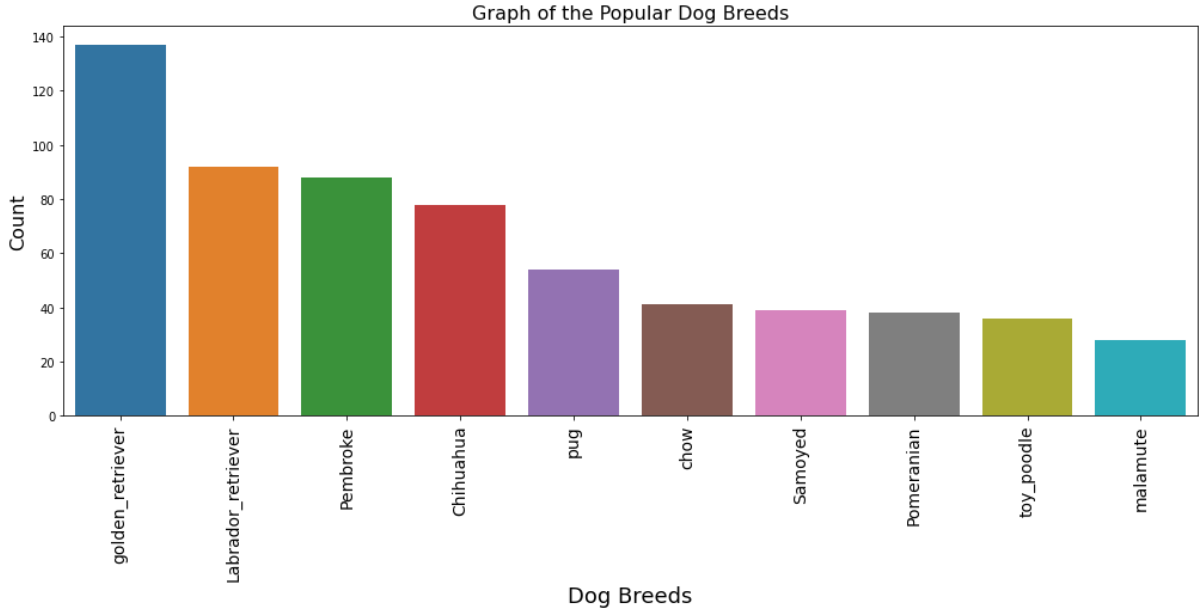

Figure 1.0: *Graph of the ten most popular dog breeds*

As it could be seen from the graph, the most popular Dog breed of all was the 'golden_retriever' which counts 137. What does a typical golden_retriver look like? This is answered in the Figure 1.1 below:

Figure 1.1: *golden_retriever with the highest favorite counts*

Such a very cute breed.

2. Analysis and Visualizaation of the relationship between retweet counts, and favorite counts.

From my experience, twitter users most times tends to retweet a tweet they admire or like. So, base on this it was expected there would be a relationship between the retweet_counts and the favorite_counts. The resulting visualization is as shown Figure 2.0 below:

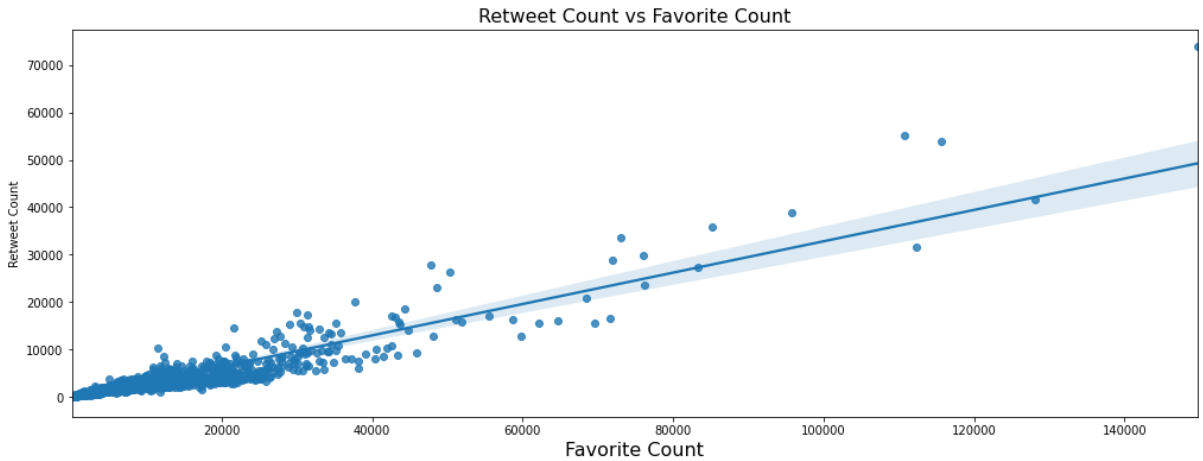

Figure 2.0: *Graph of retweet count against favorite count*

As it would be thought logical, a positive linear correlation is seen between number of retweets and the number of favorites. We could however say, a high number of retweets should be expected for a post with a high number of favorites. To further be sure about this correlation, I found out the Pearson correlation between this two columns and the result is as seen in Table 2.0.

Table 2.0: *Pearson correlation of favorite_count and retweet_count*

	favorite_count	retweet_count
favorite_count	1.000000	0.928452
retweet_count	0.928452	1.000000

YES! The inference made earlier from the graph prove correct. A 92% correlation was found which a very strong positive correlation.

3. Analysis and Visualization of the most popular source for tweets

See Figure 3.0 below

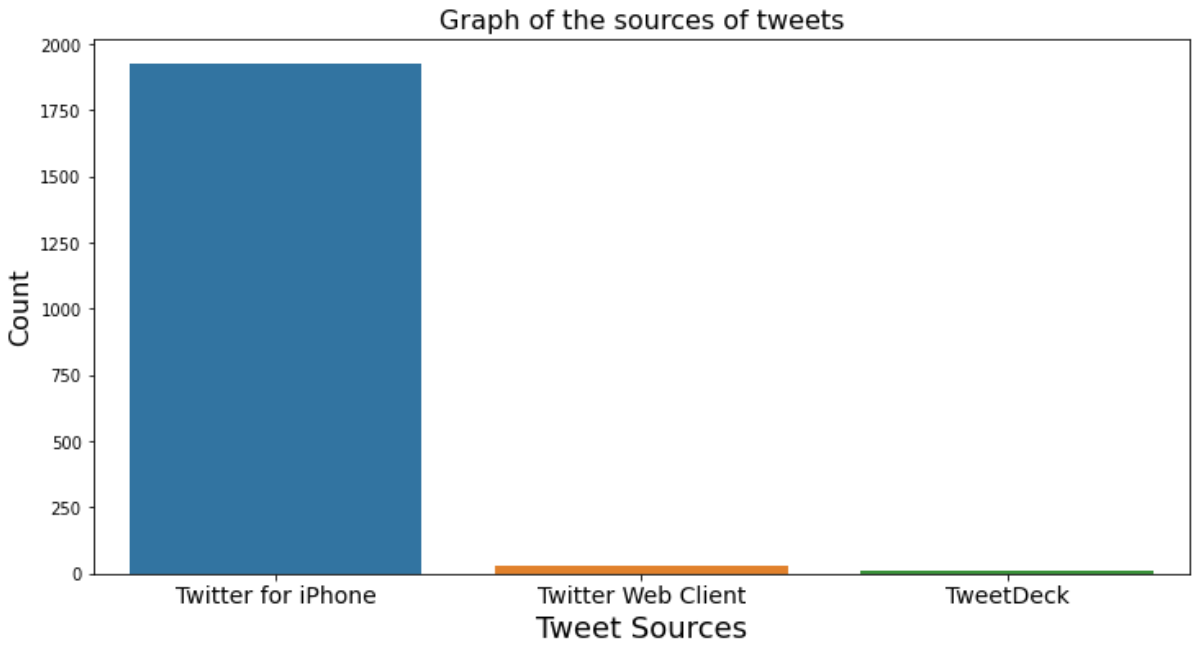

Figure 3.0: *Graph of the sources of tweets*

As seen above, majorly all tweets has been made from an Iphone. Tweets were made with “Twitter for iPhone” for about 1925 times, “Twitter Web Client” for 28 and with TweetDeck for 10 tweets.

4. Analysis and Visualization of the most common dog stage.

As shown in Figure 4.0 below, the most common stage of all dogs is ‘pupper’ followed by ‘doggo’.

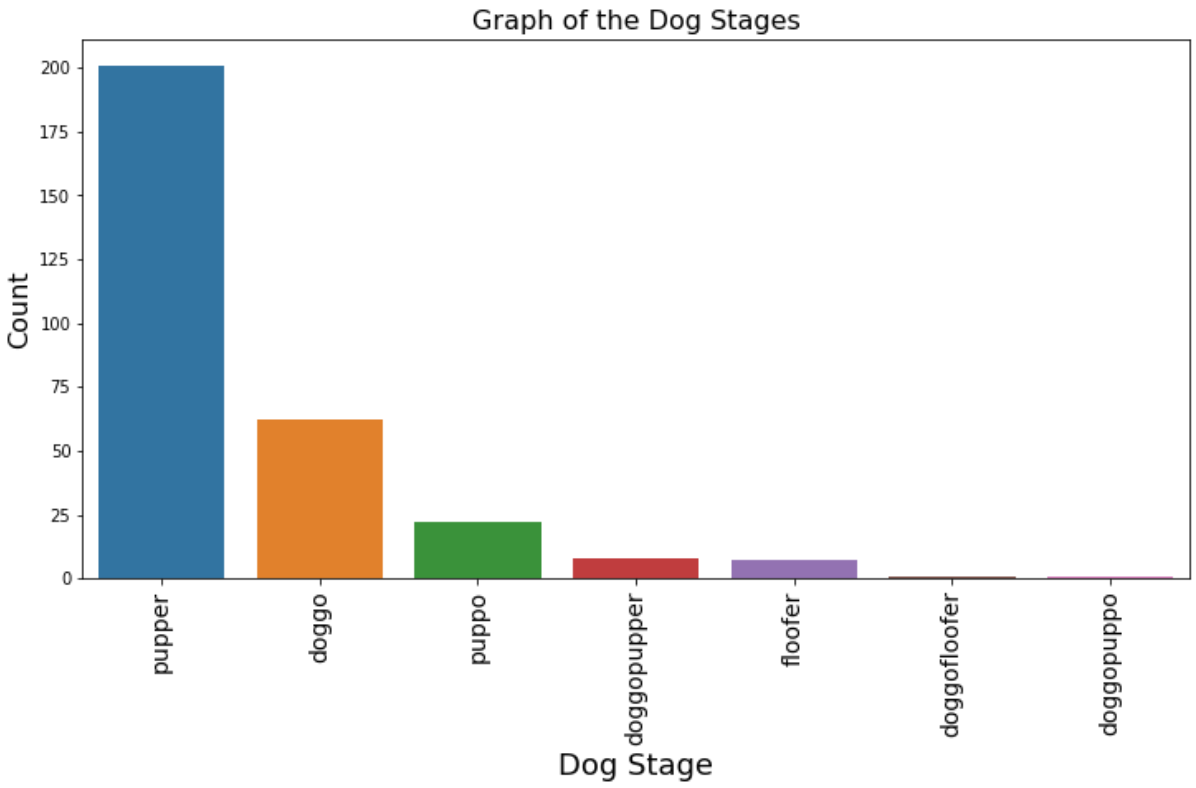

Figure 4.0: *Graph of the counts of different dog stages*

201 of all tweeted dogs are pupper, doggo – 62, puppo – 22, doggopupper – 8, floofer – 7 and only two tweeted dogs of one each are of the stage doggofloofer and doggopuppo.

CONCLUSSION

This is not all insights that could be gotten from the dataset. In fact, there are numerous information to be discovered but this project would only cover just those that has been presented.