Introduction

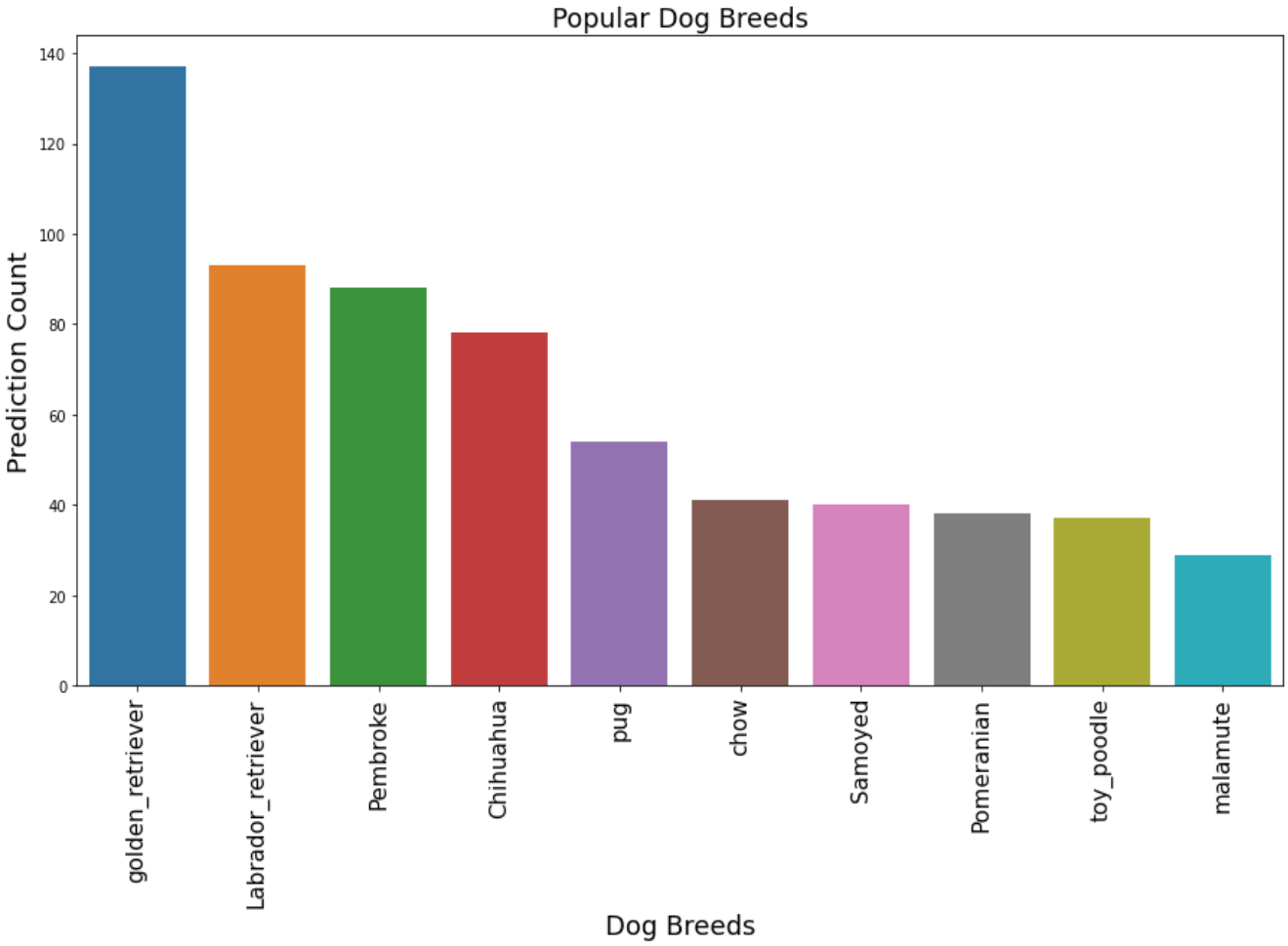
The dataset used was a merger of a CSV document from Udacity provided by WeRateDogs, a TSV file programatically extracted and JSON text document provided by Udacity.

Data Analysis

This data analysis project will attempt to answer these questions: -

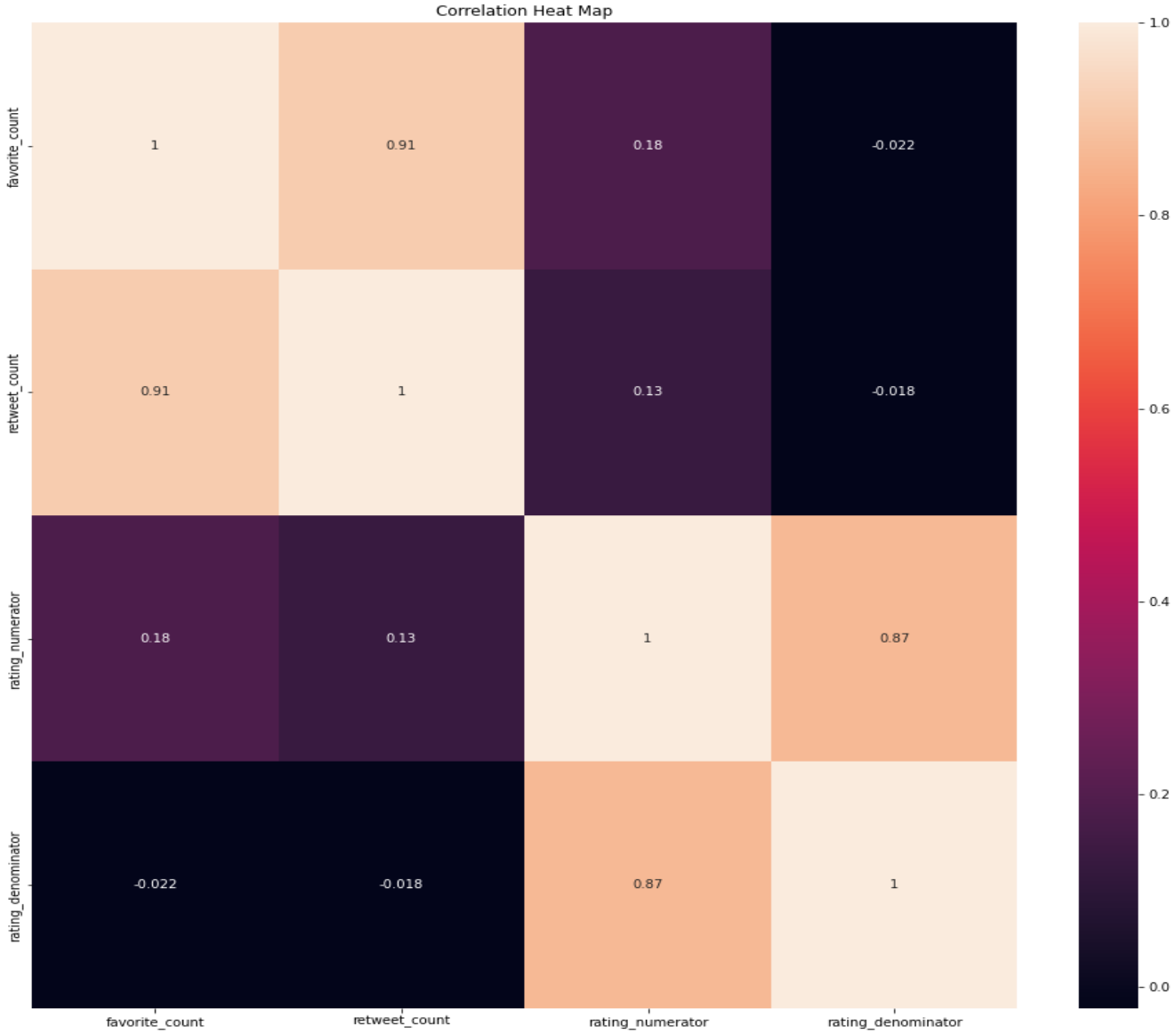
- What type of dogs are in the tweets (Depicting top 10 dogs)?
- Is there a relationship between ratings, favourite counts and retweets?
- What is the highest user tweet source?
- What dog image had the highest favourite count?
- What dog image had the highest retweet count?

What type of dogs are in the tweets (Depicting top 10 dogs)?

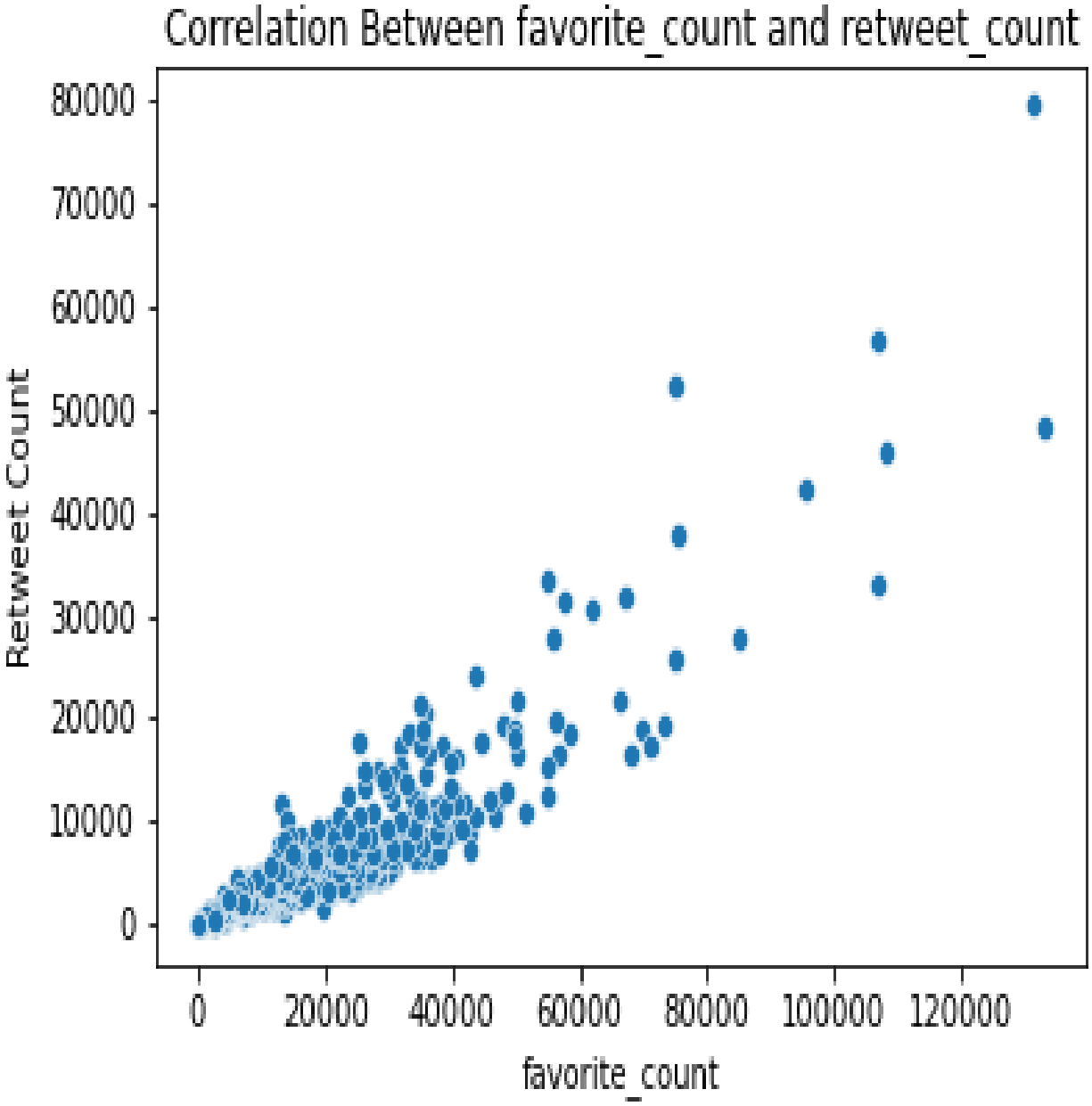


Observation: This plot shows that the Golder retriever is the most popular dog breed, followed by Labrador retriever and Pembroke.

Is there a relationship between favourite_count and retweets?

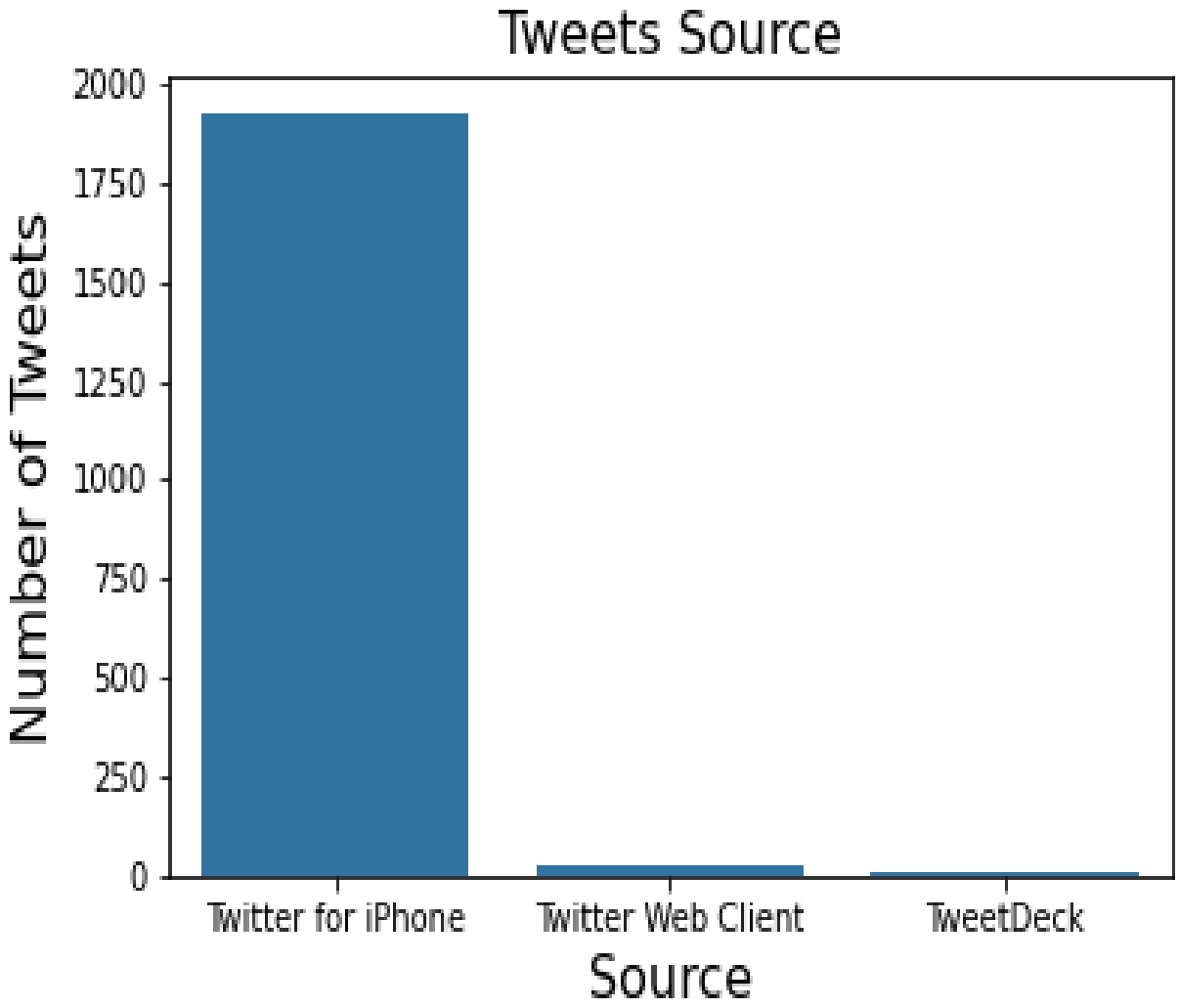


The heatmap shows our correlation relationships, showing a strong relationship between favourites and retweets with a correlation coefficient of approximately (r = 0.9).



Observations: There is a strong linear relationship between Favourites and Retweets counts. This is not a strange thing because most people will favourite what they retweet and vice versa.

What is the highest user tweet source?



Observation: This shows that Iphone users had the highest tweets.

Based on favourite count, This dog image had the highest rating.

Based on Retweet count, this dog had the highest number of retweets.

Conclusion

- During our analysis, there is a **strong linear relationship between the number of Favourites and the number of Retweets** of a given Tweet. The regression coefficient for this relationship is 0.91.
- This plot shows that **Golder retriever** is the most popular dog breed, followed by Labrador retriever and Pembroke.
- This shows that Iphone users had the highest tweets.

Reference

http://freedomcrossroads.com/2017/05/31/weratedogs-will-donate-merchandise-profits-planned-parenthood-updated/