

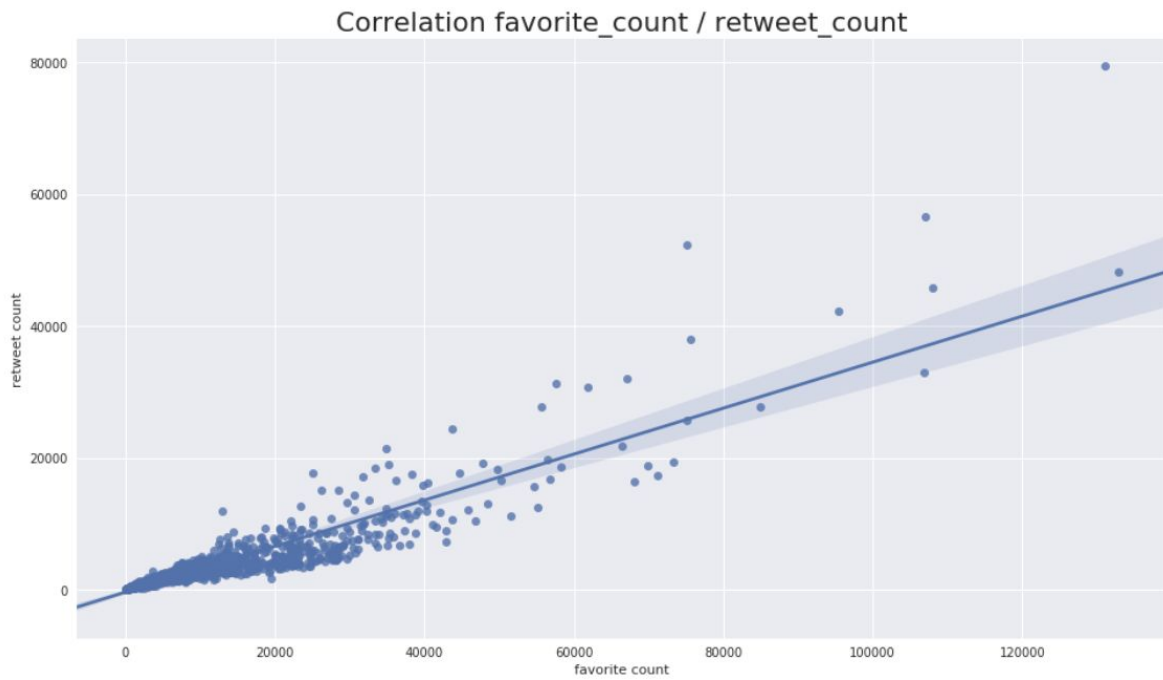
This report tries to summarize the analysis and insights obtained from the Wrangle and Analyze project in the Udacity Data Analyst Nanodegree.

During the project, we've gathered data from different sources, assessed it visually and programmatically, identified quality and tidiness issues, and proceeded to clean it before trying to answer some questions about the data:

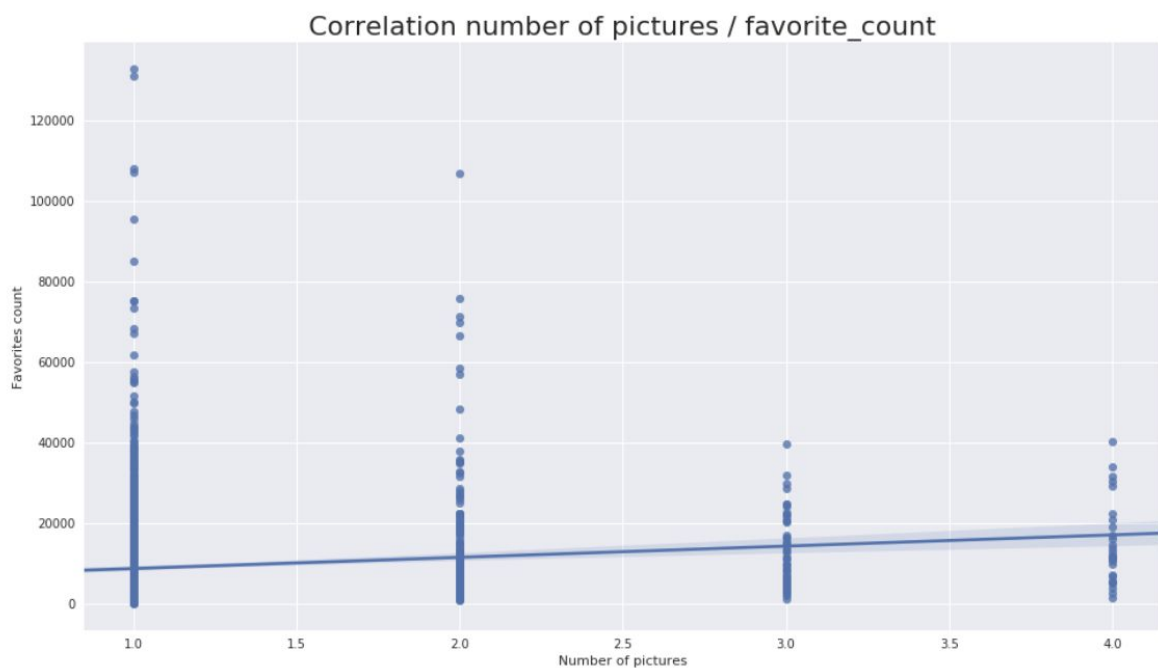
- If there is any correlation between the number of favorites and the number of retweets a tweet get
- If there is any correlation between the number of pictures per tweet and the favorites it receives
- Then, we've created a graph with the most popular dog breeds by number of favorites in absolute terms.
- Finally, we've created another graph to show the average number of favorites per dog breed.

## **Correlations**

We expected to find a strong positive correlation between the number of favorites and the number of retweets a given tweet received, and we confirmed it plotting this graph:

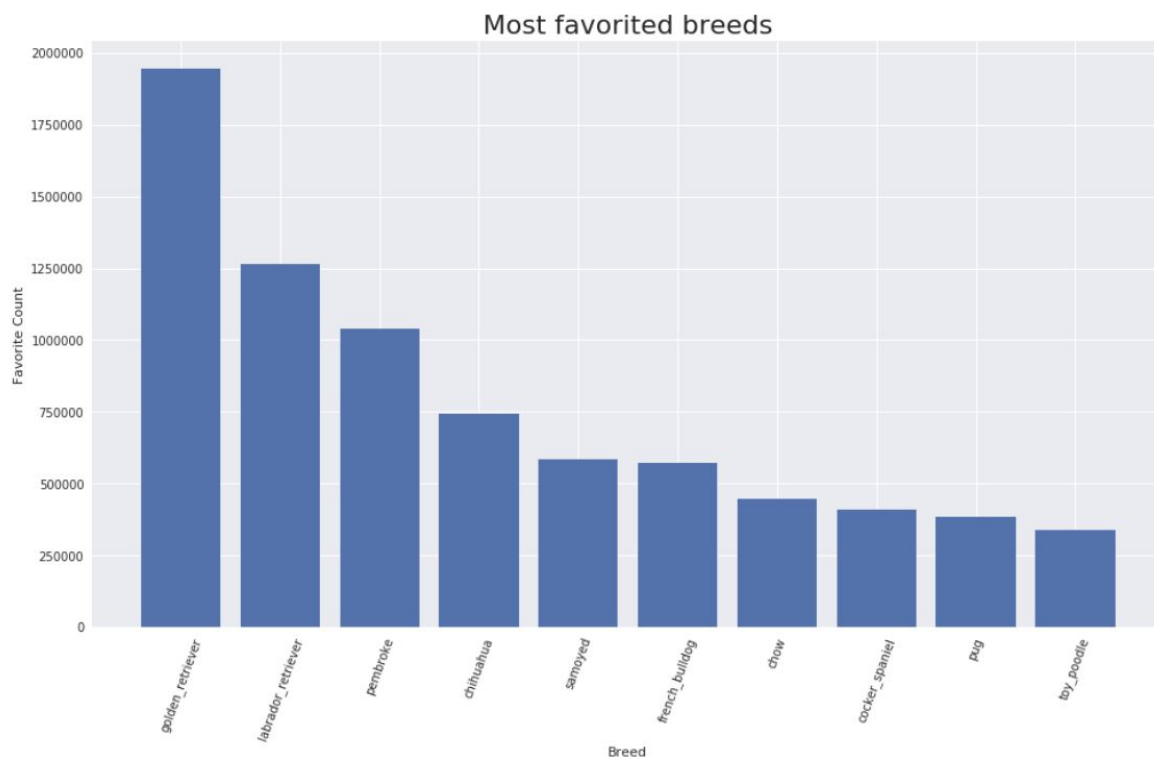


We also plotted the correlation between the number of pictures attached to a tweet and its number of favorites. In this case, there we've found a weak small positive correlation.

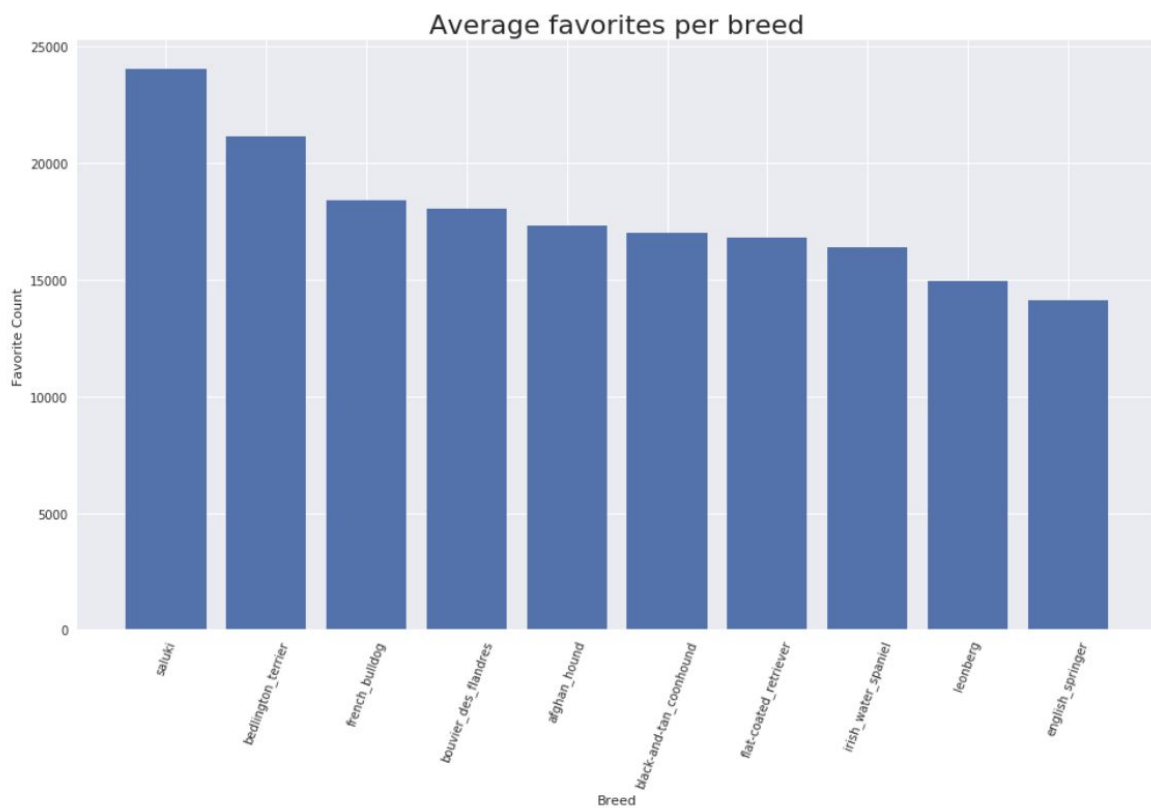


## Most popular breeds

Regarding the most favored dog breeds in absolute terms, Golden Retriever, Labrador Retriever and Pembroke are the ones heading the graph:



But, when we factor in the number of samples of each breed to get the number of favorites that each breed get in average, the scenario changes, and now Saluki, Bedlington Terrier and the French Bulldog are the leading ones.



I can't conclude this small report without posting the picture of the most favorited dog in our data.



