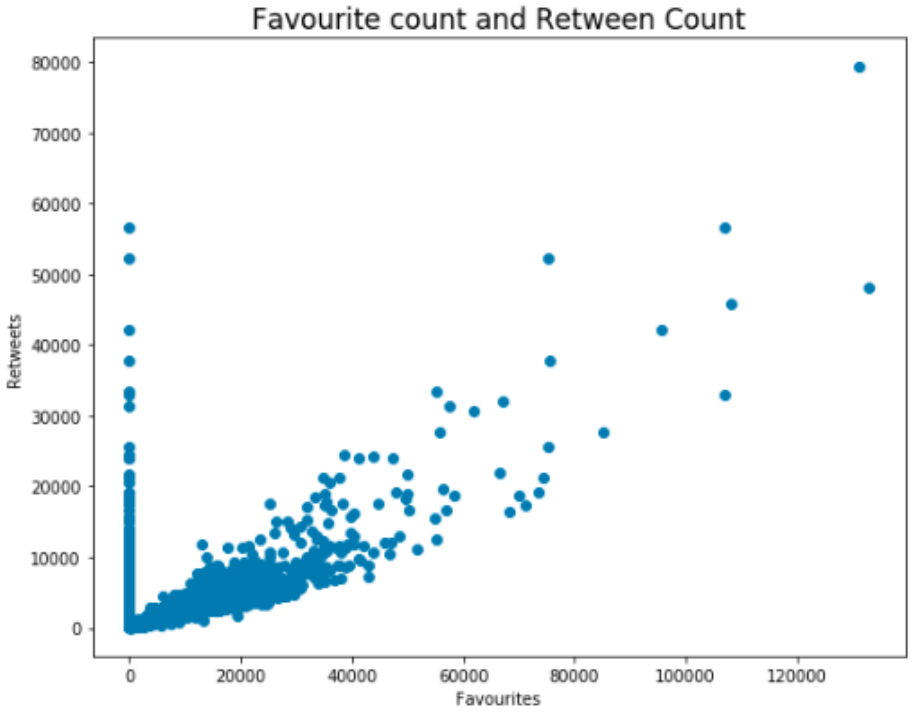
Analysis and Insights

After the data cleaning phase, we attempted to visualize components of the dataset so that we can perform some analysis. In general, we tried to answer three key questions:

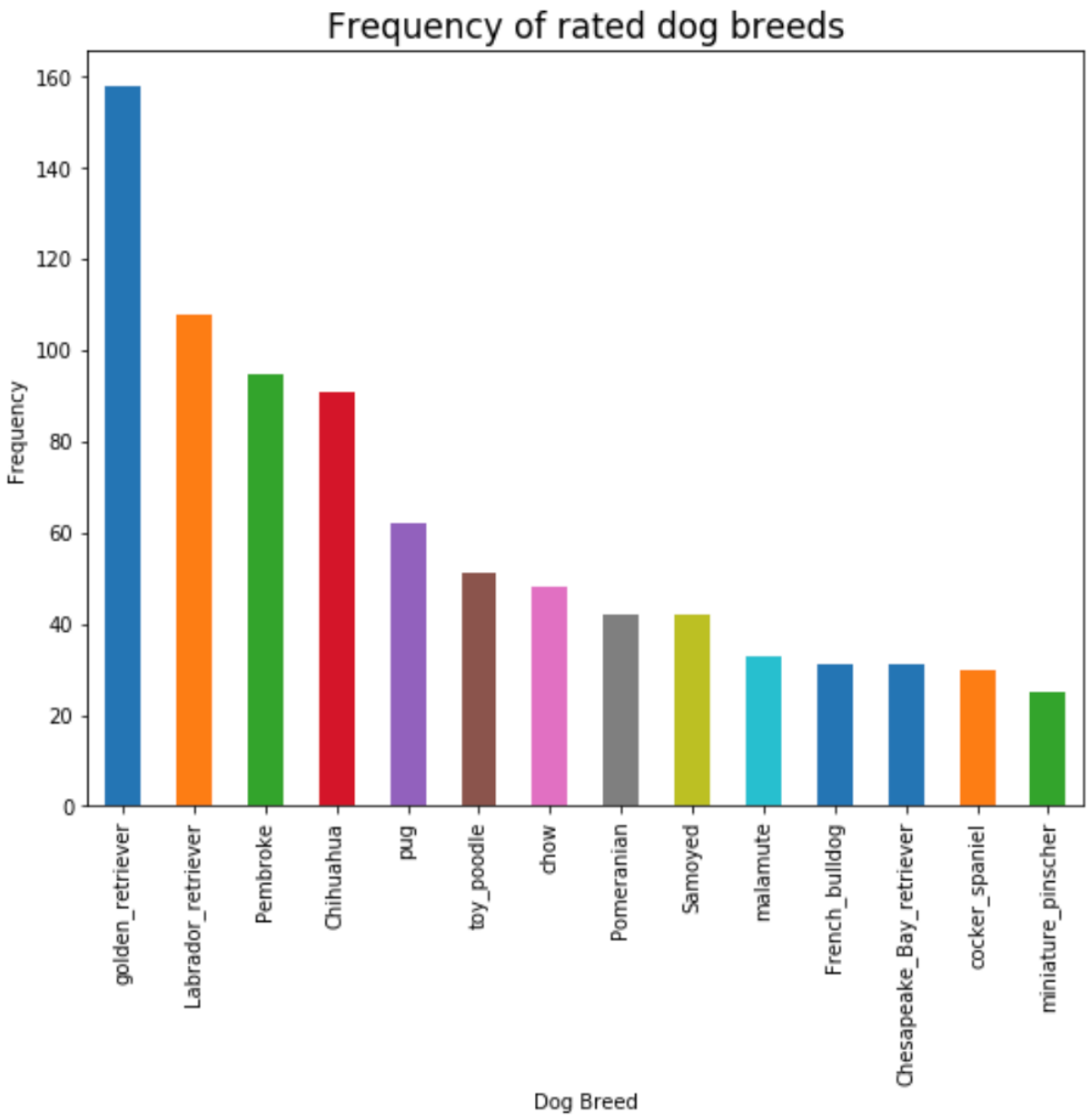
1. What is the correlation between a tweet's favourite and retweet count?
2. What are some of the most rated dog breeds?
3. What are the most common dog names in our dataset?

Visualization 1: Correlation between favourite and retweet count



To explore a bivariate relationship between two variables, we used the scatterplot. This is a strong, positive correlation between a favourite count and retweet count, which is not very surprising given that the level of engagement on one variable should intuitively also be present in the other variable.

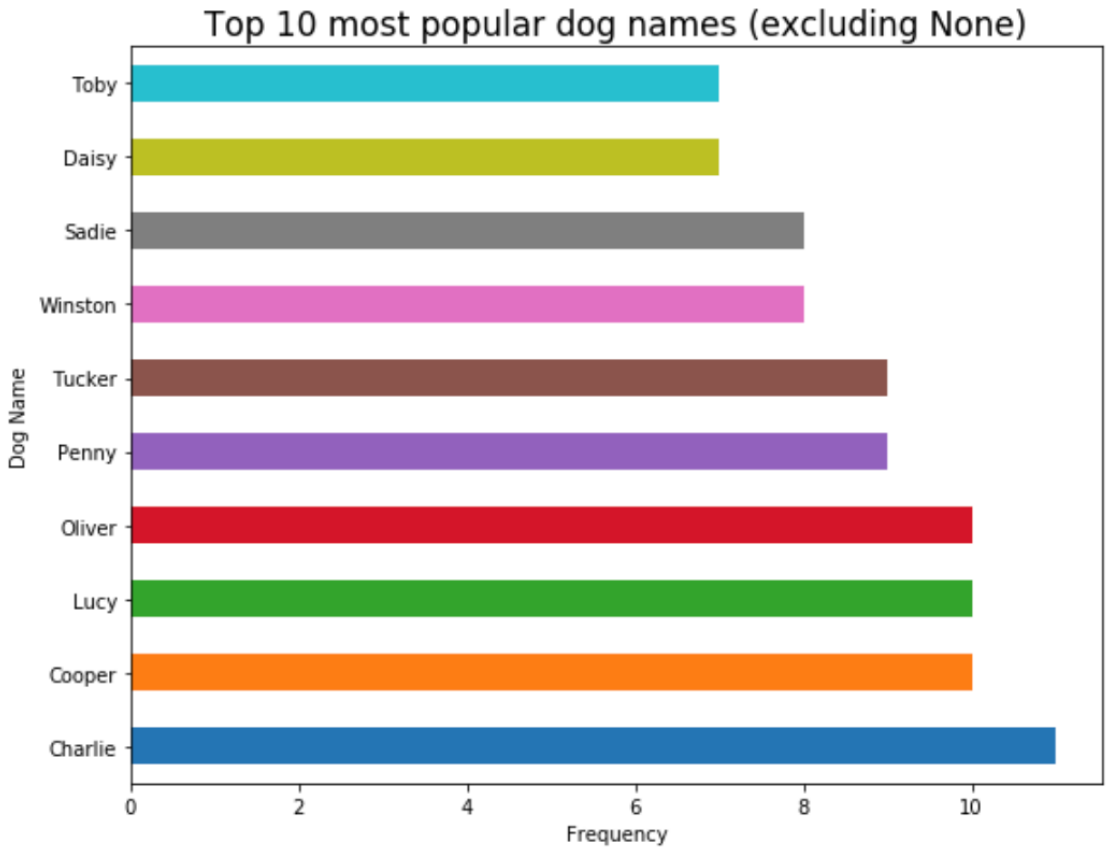
Visualization 2: Popularity of dog breeds



In the second visualization, we explored the popularity of dog breeds. Ignoring the datapoints with “None” assigned, the Golden retriever was the most popular breed, followed closely behind by the Labrador Retriever!

One of the usefulness of such insights is that pet retailers can utilize the data to prepare merchandizes featuring such dog breeds due to their popularity. Similarly, they will be able to identify which breeds are the most popular and thus can focus on carrying these breeds in their shops for sale to prospective owners.

Visualization 3: Most popular dog names



In the final visualization, we attempted to uncover the most popular dog names. Interestingly from our dataset – Charlier was the most popular name, followed by Cooper, Tucker and Lucy tied in second place.

Conclusion

These are the following analyses and visualization performed for our data exploration project.