Insights / Displays

The insights of this project show various different components of the datasets. To begin I thought it would be useful to only show dogs from the datasets with the filtered twitter_data.dog == True being the way to show that the analysis was being made on actual dogs. I realise that this could be a mistake in the model in actually identifying dogs but for scrutiny, I thought it best to rely on the data.

Insight 1 gives an interesting output and we are able to see that images that have been predicted as dogs tend to be on average more favourited by nearly 20% over images that have not been predicted as dogs. However, retweets are roughly the same. Also, The model tends to be more confident (roughly 10%) in identifying images that are dogs over images that are not. What may be somewhat contradictory is that the average rating is roughly 20% higher for non-dogs than dogs.

Insight 2 shows various ways that dogs can be categorised as ‘popular’. For example, when looking at favourite counts the saluki is almost 40% more popular than the next most popular French bulldog which conforms to my personal knowledge of this dog in social media trends. (popular dog within parts of Asia especially Japan) On rating, the most popular dog was recognised as the climber, which could be a personal favourite of admins of the page. (responsible for issuing ratings) Finally, the most popular breed on retweets would be the poodle, possibly a result of this dog is one of the most common breeds across the world.

Insight 3 would go over the most popular dog names. Stephen would be considerably higher than the second most popular name of Duddles, and this could be recognised as a relation to the popularity of this name

among humans. Next interestingly saw a favourite count of the name Stephen also being consistent with the retweet count, something that was not seen as a correlation with dog breeds. And finally, looking at ratings the most common name was Logan which could also relate to social media trends/influencers and the popularity of this name in North America.

Insight 4 shows some statistics of the average counts of ratings, favourites and retweet counts as well as confidence percentages. An interesting min rating of 0 and max rating of 177 shows a large range in ratings from the admins of the page.

Finally, the visualisation shows some of the most reoccurring dog breeds in the form of a pie chart. From this, we can see that the retriever (labrador and golden) make up over 30% of the most common dog breeds across all submitted to the page with some of the less common being the German shepherd and miniature pinscher.

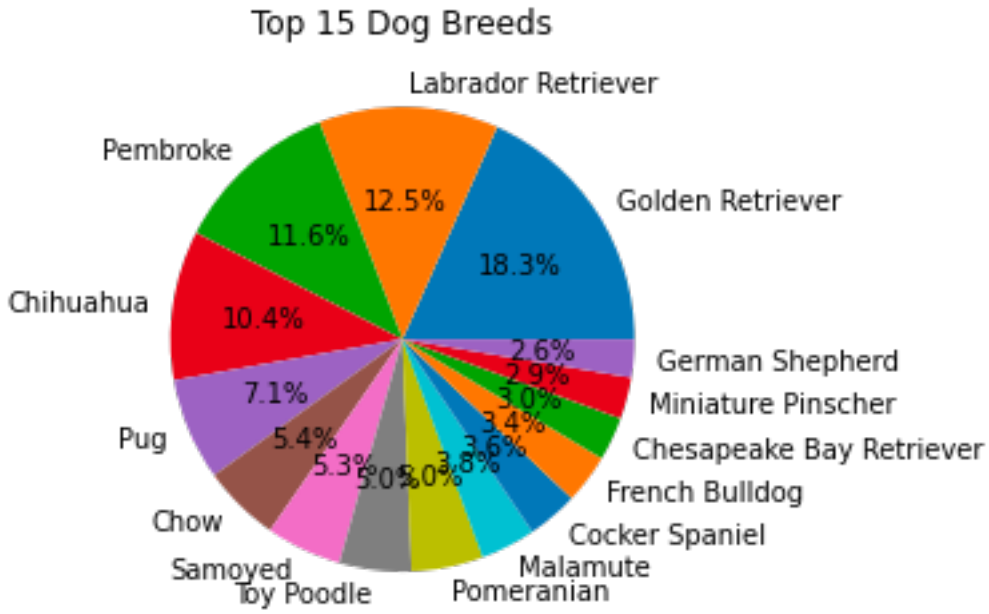

Image 1 - Pie Chart showing the top 15 most common dog breeds from the entire datasets given.