

Analysis and Insights Findings

By Ikechukwu Ogbuchi

Looking at the combined data analytically, I was able to find some interesting insights about the most predicted dog breeds across the three predictions from the models. The insights are as follows:

1. For prediction1 , the golden retriever was most predicted dog followed by Pembroke
2. For prediction 2, the labrador retriever was most predicted, followed by Cardigan
3. For prediction 3, the labrador retriever followed by Chihuahua were most predicted dogs

The following plot shows this at a glance, The different predictions across dog breeds and their frequencies

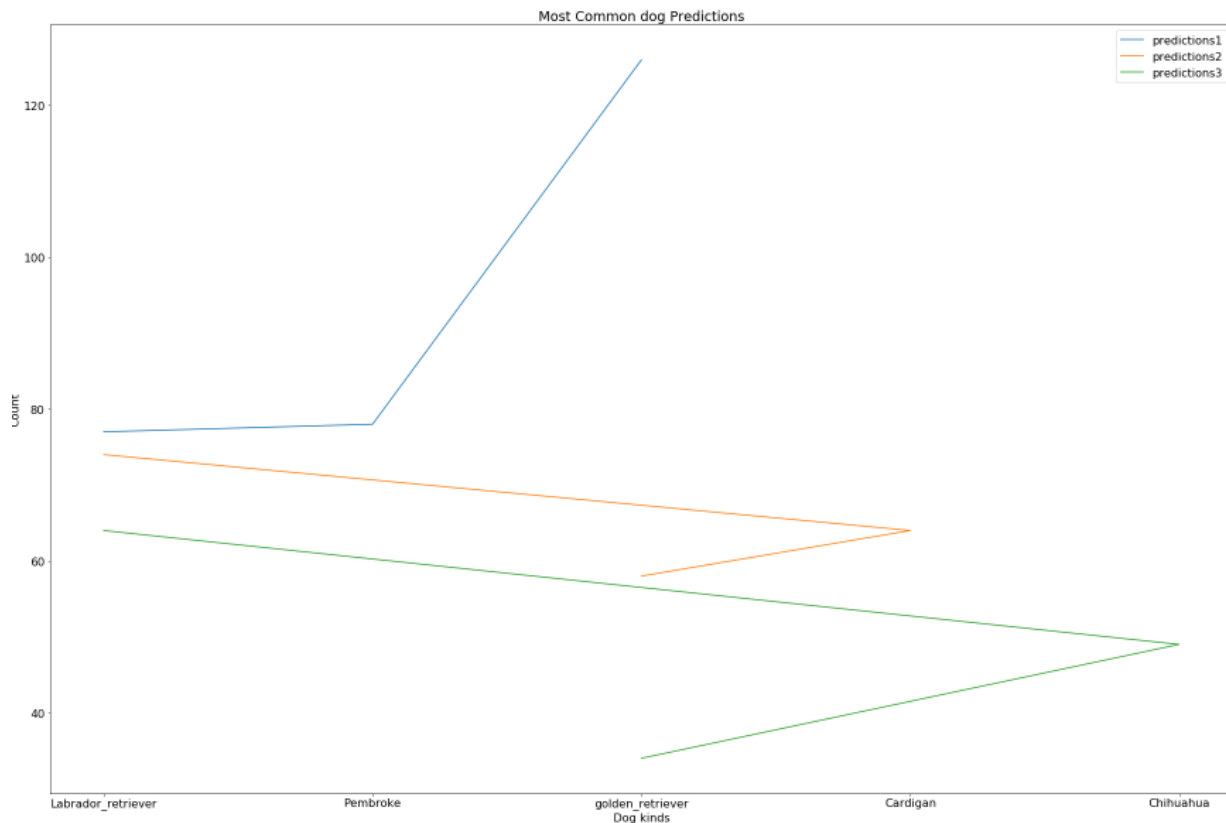


Fig 1 Dog Breed popularity across predictions

To derive this insight, I had to sort each of the prediction categories by their value_counts for the unique dogs using the value_counts() function of pandas. Also looking at the summary statistics of the joined dataframe, I was able to establish that the mean number of retweet counts was 2,911 and that the upper quartile of that was at 3,430 while the lower quartile was at 664 retweets.

I also considered the dog languages mentioned of floofer, puppo, doggo and pupper and by exploring each of their value counts I realized that the pupper was the most mentioned term with 130 mentions , followed by the doggo which had 50 counts and then the puppo with 19 counts followed by the floofer with only 7 mentions.

A histogram plot was made from the prediction 1 column with all of the dog breeds and made it easier to see the most common dog breeds at a glance.

I found this project analyzing tweets using tweepy interesting because it shows how easy it is to get public sentiments from dog lovers online and get useful insights if you are into the pet business or you just want to get popular opinion for marketing reasons.