DATA WRANGLING PROJECT

Analyzing and Visualizing Data

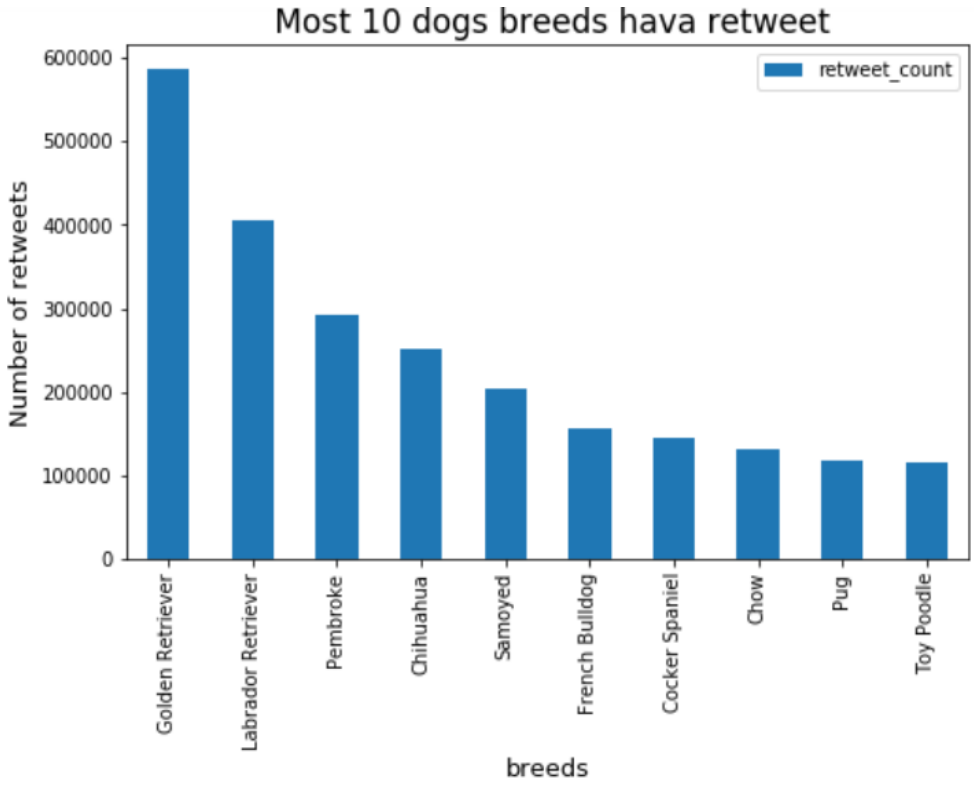
Introduction

In this project, I will be wrangling the WeRateDogs Twitter archive containing that data is a Twitter account that rates people's dogs. At first, I should gather the data and I have three data tables with three different ways to gather it. second, I will also assess Data also I should assess data in different ways. The last thing, I will be cleaning data and Storing data to analyze and visualize data.

I have three insights:

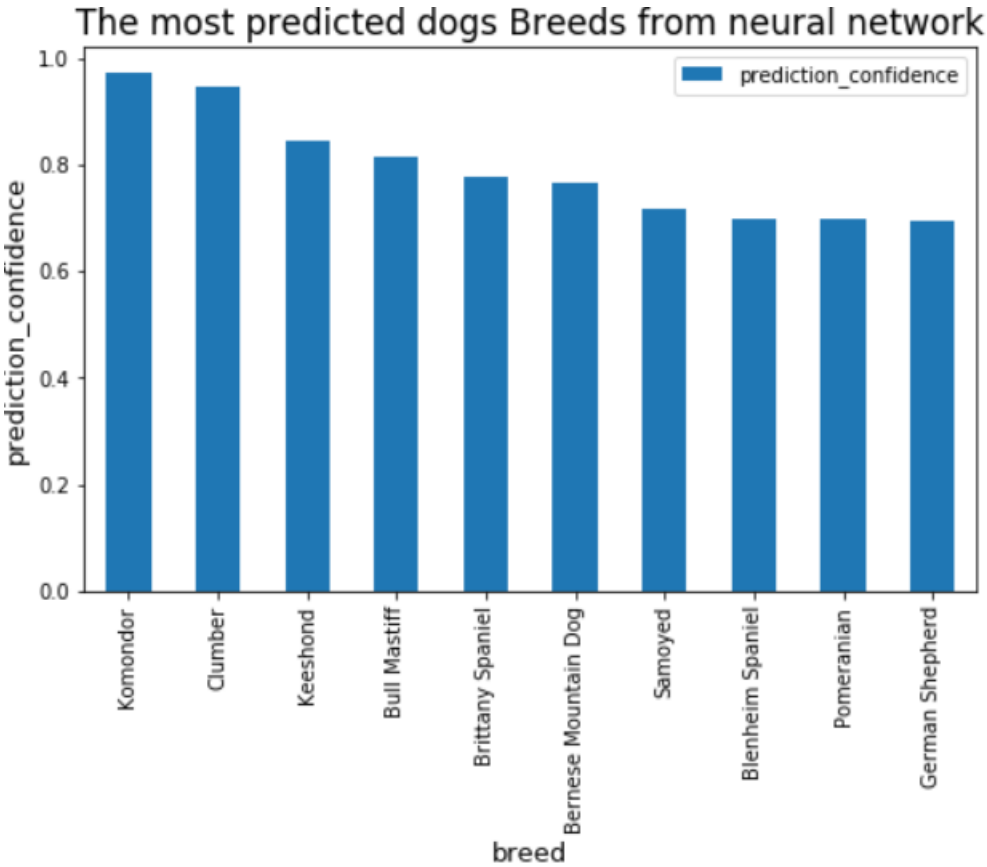
1. What are the Most 10 dogs breeds that have been retweeted in the dataset?
2. What are The most predicted dogs' breeds from the neural network?
3. How do the Favorites and Retweets change over time?

Most 10 dogs breed have retweet



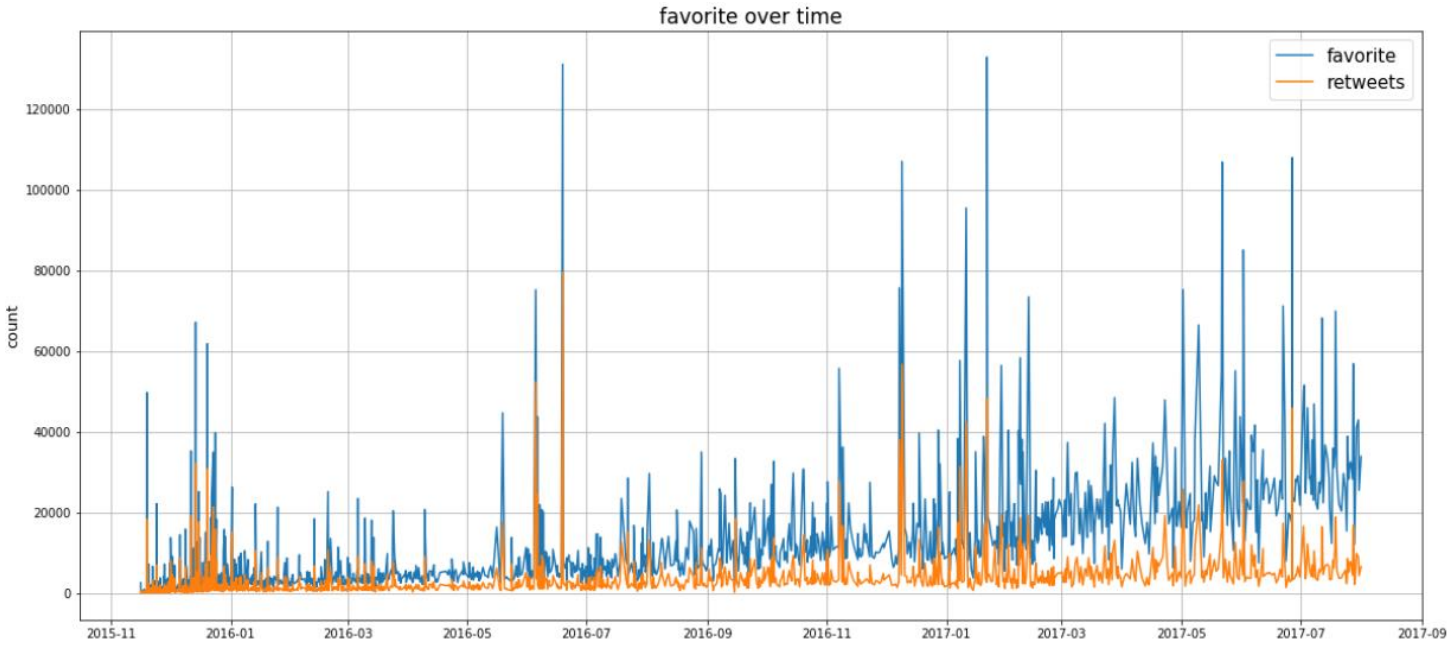
We have 113 number dog's breed but we show most 10 dogs breed have retweeted, as we can see in the bar chart we can get the most breed is Golden Retriever by 586657 number of a retweet with big gab with Labrador Retriever by 405404 after that Pembroke with 291771 and Chihuahua with 250722.

The most predicted dogs Breed from the neural network.



Here we can see what type of dogs the neural network can be predicted. We have 113 dogs typed in the dataset but we display only 10 in a chart. We take the mean of predicted with range 0 to 1 and here the most type of dog that the neural network can predict from the dataset is Komondor with 0.972531 rates, Clumber comes after it with 0.946718 this number close with the first one but the third one is Keeshond by 0.844431 with little gap between the second and third one. But the must 10 rates are more than 0.5.

Popularity over Time.



We can realize that there is an increase in popularity sometimes, but when looking at favorites, we can say that there is a huge increase in popularity in general, but when looking at retweets, the situation is a little different, we can say, there is a slight increase and less of a favorite