DATA ANALYSIS AND VISUALIZATIONS

Aisulu Raganina

Udacity Project: Wrangle and analyze Data

Introduction

The report includes the basic data analysis of WeRateDogs twitter account data and provides 4 insights from the analysis and visualization results.

1.The most popular dog stage

According to the data, the distribution of dog stages are as follows:

| Dog stage	Count
Pupper	710
Doggo	259
Puppo	97
Floofer	38
No specification	26

The most popular dog stage is Pupper with 64,3%. The second most popular is doggo with 23,5%, followed by puppo – 8,8%. The least popular is floofer with 3,4%

It also should be mentioned that 26 positions have no specified dog stage.
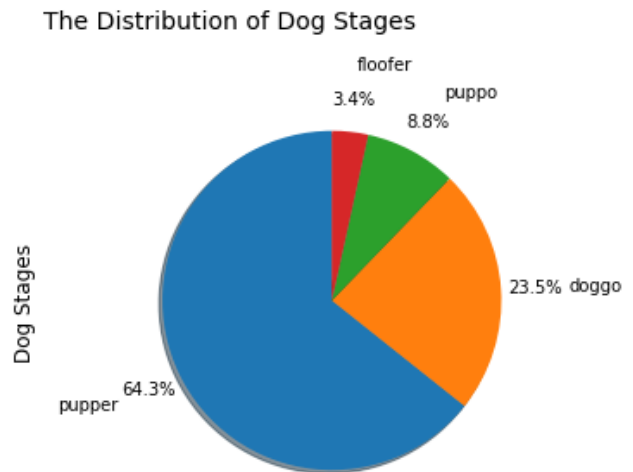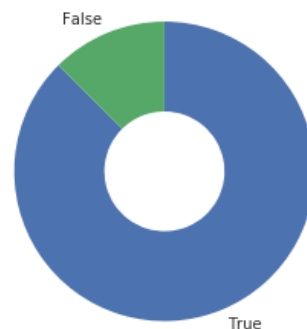
2.The most populat breed of a dog

After comparing the three predictions for the dog breed on the photos from p1, p2 and p3, we made the conclusion that the p1 column can give us better prediction of the dog breed with the most of the cases being a 100% prediction.
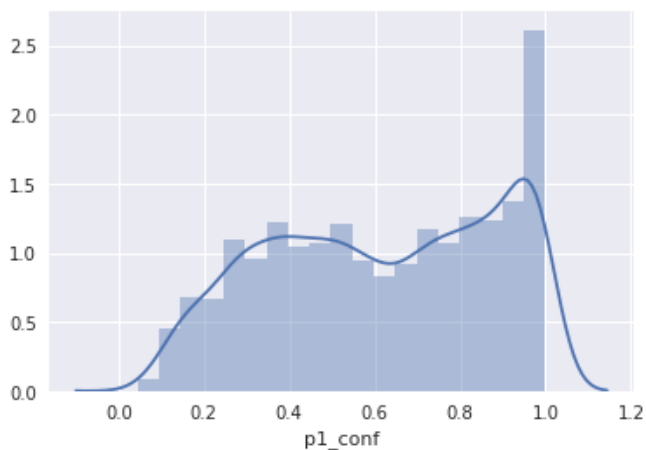
Dog Breed	Count
Golden_Retriever	597
Pembroke	385
Labrador_Retriever	384
Chihuahua	291
Pug	207

The Distribution of Dog Breeds

The breeds above are the top 5 dog breeds this model predicted. The Golden Retriever, Pembroke and Labrador are top 3 with all over 350 predictions.

3.The most favoured/retweeted tweets

Even though the Labrador_retriever is the third most popular breed in the top 5 Dog breeds, it is on the top 5 most favourite/retweeted tweets with corresponding favorite_count maximum = 156359 and retweet_count maximum = 77882.
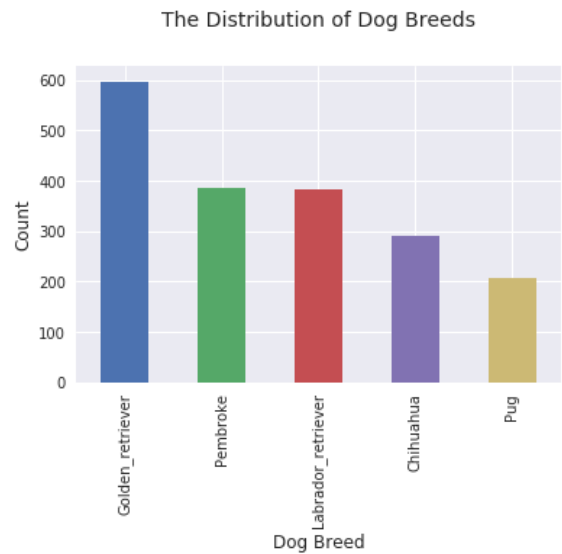
4. The correlation between 'retweet_count' and 'favorite_count'.

The hypothesis is that most popular tweets usually get a large number of retweets and favorite counts. According to the test, the correlation between 'retweet_count' and 'favorite_count' has a strong positive relationship with the pearson r^2 is 0.928436. The plot confirms the hypothesis.

Favorite Count vs. Retweet Count
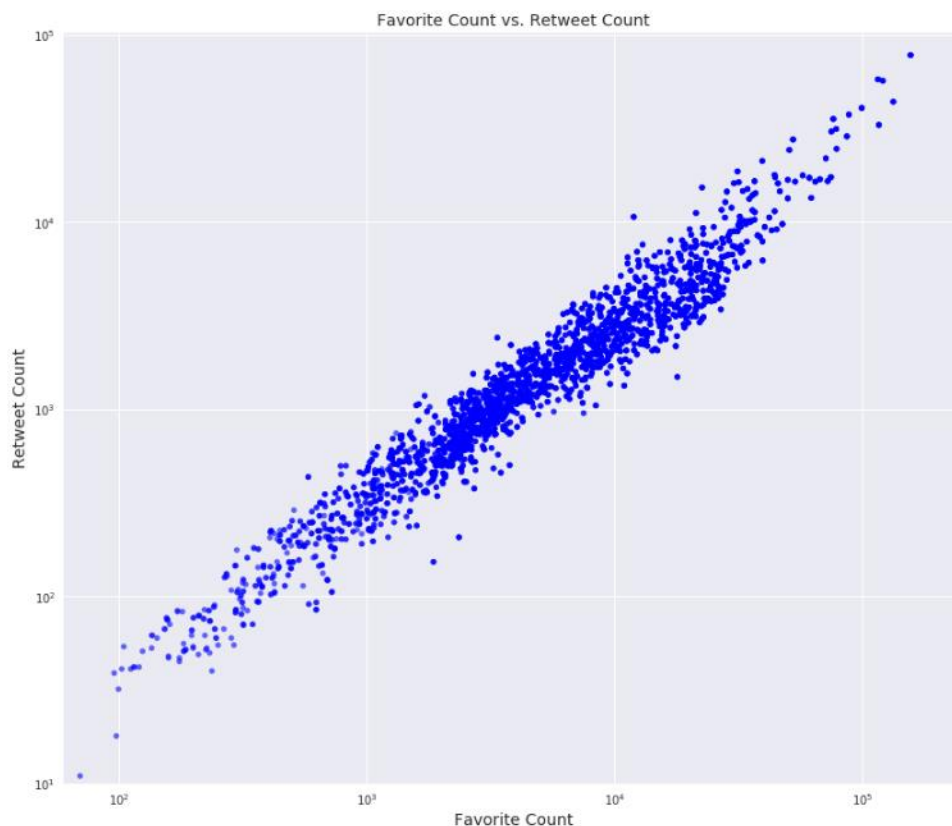
Conclusion

Based on the data, the Retrievers at the pupper and doggo stages are likely to get the better ratings.