WeRateDogs Data Analysis

Bin Mohaya, Turki

June 9, 2019

INTRODUCTION

In this project we analyze a data set related to a popular twitter account that rate dogs called WeRateDogs. Dog owners send pictures of their dogs to this twitter account and, thus, they are evaluated in a single tweet. The aim of this project is to understand several factors such as which dogs' type are most popular based on the number of retweets and favorite count for the evaluation tweet.
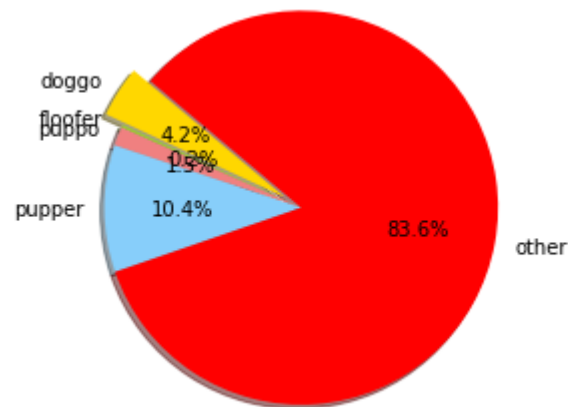
PROCEDURE

Three files are collected from three different sources to start analyzing and understand several facts about WeRateDogs. In these files several columns exist such as retweet count, favorite count, dog name, dog type, and the tweet context. Initially, we started assessing those files to detect any problems in structure or validity. After identifying those problems, we began cleaning in the since that important data are preserved. A plethora of issues have been cleaned such as removing columns, merging, and changing data type.

After cleaning all the inspected problems the analysis stage started to answer the following three questions:

1- What are the percentages of each dog type in WeRateDogs account?
2- What is the highest retweet count mean among based on dog types?
3- What is the highest favorite count mean among based on dog types?

For answering the first question, there are four types of dogs in our collected data set which are Doggo, Floofer, Pupper, and Puppo. the percentages of each type are as follows:

As seen above, the highest known percentage is 10.4% which represent Pupper-type dogs.

To answer the second question, we have used the query command in pandas to isolate each type of dogs. The following table answers both question two and three:

Dog Type	Retweet Count Mean	Favorite Count Mean
Doggo	7597.08	17460.2
Floofer	7606	19531
Puppo	4975.6	17050.7
Pupper	2604.6	7614.3

From the above table, we see that the Floofer type has the highest retweet count as well as favorite count.

CONCLUSION

In this project we have collected data about a popular twitter account that rate dogs and we tried to answer several questions. We concluded that Pupper-type dogs constitute the highest known percentage among evaluations, and the highest retweet count mean and favorite count mean are related to Floofer-type dogs.