WeRateDogs Analysis

There were several steps taken in order to complete this project. To kick the project off, I had to gather three data sources and read them into separate dataframes using Pandas. The first source was the twitter-archive-enhanced.csv file that was kindly provided with the course materials. The second source was collected by using the tweepy and json libraries in order to call upon the WeRateDogs Twitter page using the Twitter API. This process produced the tweet_json.txt file which stored all Tweets pulled from the WeRateDogs account. The tweets were then read into a dataframe for analysis. For the third and final data source, I used the requests library to download the image_predictions information and wrote the contents of that file locally to a file titled image_predictions.tsv.

Once my data was gathered, I then visually and programmatically assessed the data from quality and tidiness issues. I documented 8 cleanliness issues and 2 tidiness issues which I would address during cleaning. After my work had been cut out for me, I went through each of the 10 issues identified and fixed them programmatically.

After the data was clean, I set out to create several visualizations which would lead to uncovering a few key insights from the data. The first insight uncovered is that the highest rated dog stage by the WeRateTwitter account, as well as the dog stage with the highest average number of favorites by Twitter users is the Puppo. However, the dog stage that gets the most retweets is the Floofer. This was demonstrated by taking the average rating, number of favorites, and retweets, and comparing the totals for each type of dog. Another insight gained from this visualization is that the Pupper type of dog is the least popular in all categories. For my final visualization, I created a line plot demonstrating the fluctuation in daily tweets from 2017 -2017. It is interesting to note that the WeRateDogs account was tweeting 20+ times daily towards the end of 2015, and had slowed to ~ 5 tweets per day by July 2017.

Overall – it was very interesting to observe the WeRateDogs tweet info.