Insights and Data Visualizations from WeRateDogs Twitter Application

For this analysis, 3 tables are cleaned and merged. The first dataset, 'twitter-archive-enhanced.csv,' contains information about the tweets such as timestamp, url, text, ratings, and dog stage. It also contains rows related to retweets, which are removed because they conflict with information in tweet_json.txt. tweet_json.txt is the table that contains the number of retweets per tweet and the number of times that a dog was labeled a favorite. image-predictions.tsv contains the url of the picture of the dog and an analysis by a neural network of the dog breeds contained in the images. Some of the images are predicted not to contain a dog, and those rows are removed from the table. The three tables are merged around the tweet id. The name of the id in tweet_json.txt is changed to tweet_id to match the other tables, and all three of the tweet_id columns are changed to strings.

The first insight from this WeRateDogs analysis is the names of the five dogs most often marked as favorites by users. The names of these dogs, in order, are 'Duddles,' 'Stephan,' 'Jamesy,' 'Bo,' and 'Zoey.'

The second insight gained from the WeRateDogs analysis is the names of the five dogs most often retweeted. The name of these dogs, in order, are: 'Stephan,' 'Duddles,' 'Bo,' 'Jamesy,' and 'Buddy.'

The third insight is the ten dog breeds most frequently pictured in the images uploaded by users on WeRateDogs. A neural network is used to make a prediction about what breed is in the picture. I used the first prediction, rather than the second prediction, since it has a 95% confidence interval. The ten most frequent breeds pictured on this Twitter application, in order, are golden retriever, pembroke, labrador retriever, Chihuahua, pug, toy poodle, pomeranian, chow, malamute, and French bulldog.

The visualization is a bar chart depicting the ten most frequent breeds pictured in images uploaded by users and predicted by the neural network.