

Act Report 2019

In analyzing the data that I gathered, assessed and cleaned, I first stored the final merged dataset into a CSV file so I could have the final dataset without needing to redo my entire process. I was first interested in finding out which tweets received the most retweets as well as the most favorites. To do so, I sorted the `retweet_count` column so that it would show me the top `tweet_id` in terms of retweets. I found out that the post received 79573 retweets! When sorting the data based on the `favorite_count` I found that the top post received 159484 favorites! Not only that but from looking at the results and the `tweet_id`, they were both from the same post! This was a really interesting find.

```
master_merged_df.sort_values('retweet_count', ascending=False)[['tweet_id', 'retweet_count', 'favorite_count', 'name', 'text', 'rating_numerator', 'rating_denominator', 'dog_stage', 'p1']]
```

	tweet_id	retweet_count	favorite_count	name	text	rating_numerator	rating_denominator	dog_stage	p1
836	744234799360020481	79573	159484	None	Here's a doggo realizing you can stand in a po...	13	10	doggo	Labrador_retriever
870	739238157791694849	58978	117769	None	Here's a doggo blowing bubbles. It's downright...	13	10	doggo	Eskimo_dog

After finding the top posts, I wanted to understand more about the dogs. I pulled out the counts of the name column to see which dog name may have been the most popular from the tweets gathered. I found that Tucker, Cooper, Penny, Charlie, and Oliver had the most with 10 mentions each! This is very interesting to me, since I am a dog owner and my dog's name is Penny!

```
master_merged_df.name.value_counts().head(20)
```

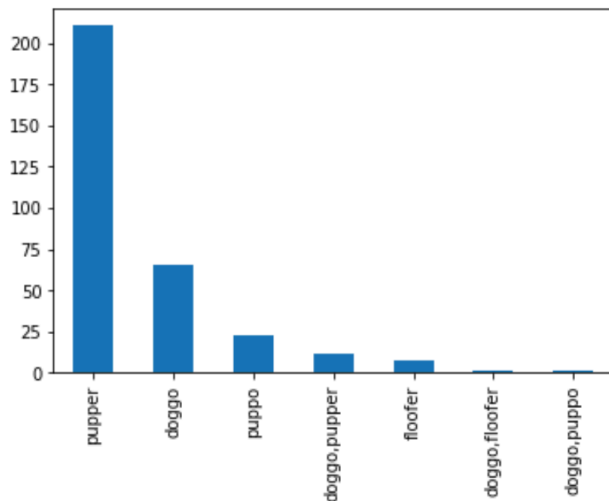
None	562
a	54
Tucker	10
Cooper	10
Penny	10
Charlie	10
Oliver	10
Lucy	9
Bo	8
Sadie	8
Lola	8
Winston	8
the	7
Toby	7

To look at the classification data, I analyzed the `p1` column which provided me with the top prediction from the neural network created and found that the majority of the predictions were resulting in golden retrievers. I am not completely sure if the neural network was 100% accurate but from some of the images I saw it certainly made sense. I saw that some predictions made by the neural network resulted in a classification of seat belt and website, which I found pretty funny.

```
master_merged_df.pl.value_counts().head(20)
```

golden_retriever	146
Labrador_retriever	95
Pembroke	88
Chihuahua	81
pug	57
chow	44
Samoyed	42
Pomeranian	38
toy_poodle	38
cocker_spaniel	30
malamute	30
French_bulldog	25
Chesapeake_Bay_retriever	23
miniature_pinscher	23
seat_belt	22
German_shepherd	20
Siberian_husky	19
Cardigan	19
web_site	19
.	.

Finally. I wanted to create a visualization for the dog stages to see how often WeRateDogs would classify a dog as pupper, doggo, puppo, or floofer. To see this in a visualization, I used a histogram for dog stages and I found that pupper was the most used stage at over 200 mentions, versus doggo at just under 75 mentions.



It was great working with this data and I was able to understand a lot from working with various sources and format as well as finding cool insights on dog data from twitter!