Act Report by Tyler Sanders

Finding the Winners

At number one with the highest rating is Atticus, earning a 1776 out of 10!

At second place Snoop Dogg comes in at very impressive 420 out of 10.

Removing the Global Outliers for Cleaner Comparison

The two winners above are data points that were accurately gathered, but their ratings are global outliers to the rest of the observations. These global outliers greatly increase the standard deviation of the field 'rating' in our DataFrame.

To simplify this analysis, I decided to remove the two outliers, and take a deeper look into the distribution of the ratings by dog type.
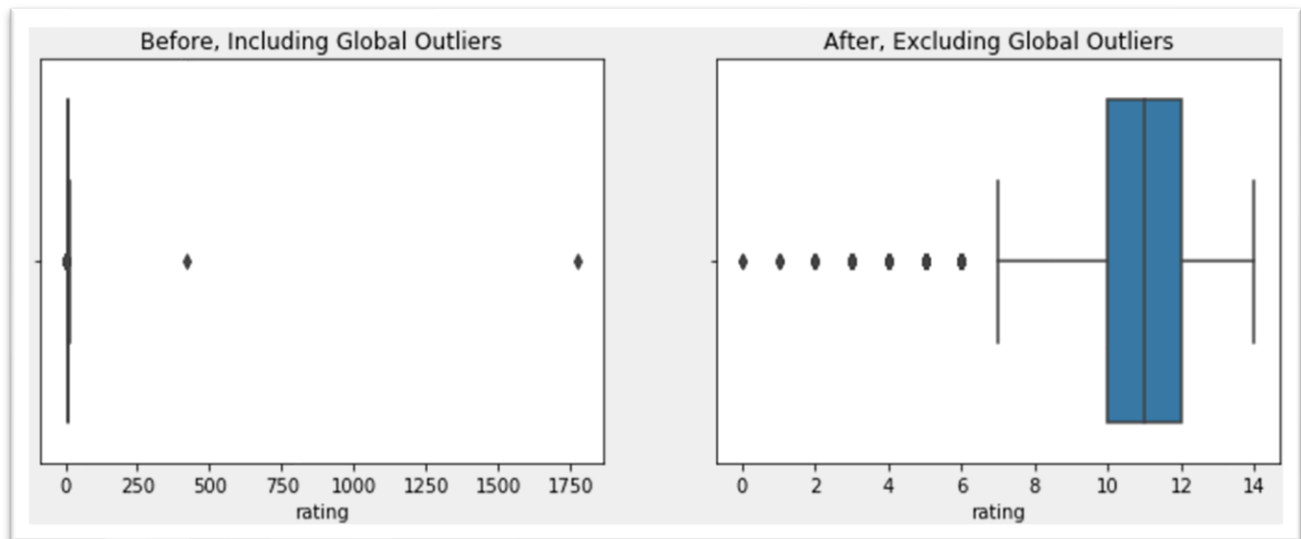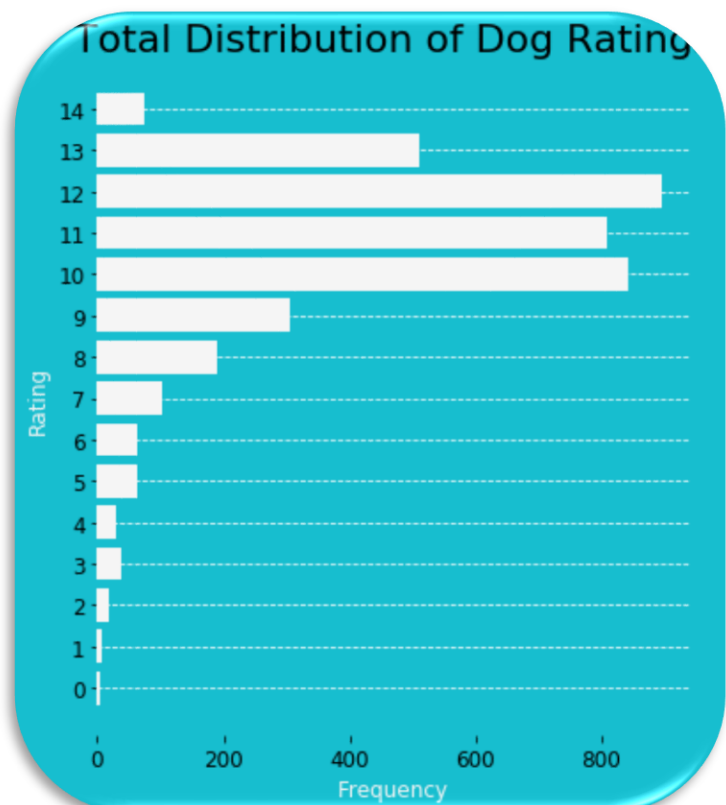

Distribution of Dog Ratings by Dog Types

With the outliers removed you can see that the total distribution of the WeRateDogs is skewed the left (since the ratings are group in descending order).

The frequency of lower ratings is dwarfed by the sizeable number of ratings that are either 10/10 or greater than 10/10. Ratings of 10, 11, and 12 are the most common.

14/10 happens at a much less frequent manner. Below, I will look more into the specific rating of 14 and 13 to see which breeds were most likely to receive it.

This plot categorizes the distributions by dog breeds, showing subtle disparities between the frequency of ratings.

At the top, Golden Retriever appears to consistently outperform the other breeds, receiving the most 14's, 13, and 12's. The frequency of ratings also shows that the most common dog breed detected in the data set was Golden Retriever, with a count of 138 out of the 1977.

Another observation that can be taken away from this plot was that WeRateDogs had a consistent pattern of rating all dogs at or above 10 out of 10, regardless of breed.

Although, they did like picking on the occasional Pug or Chihuah, who have suspiciously received a few more 7 to 3 ratings than the others.
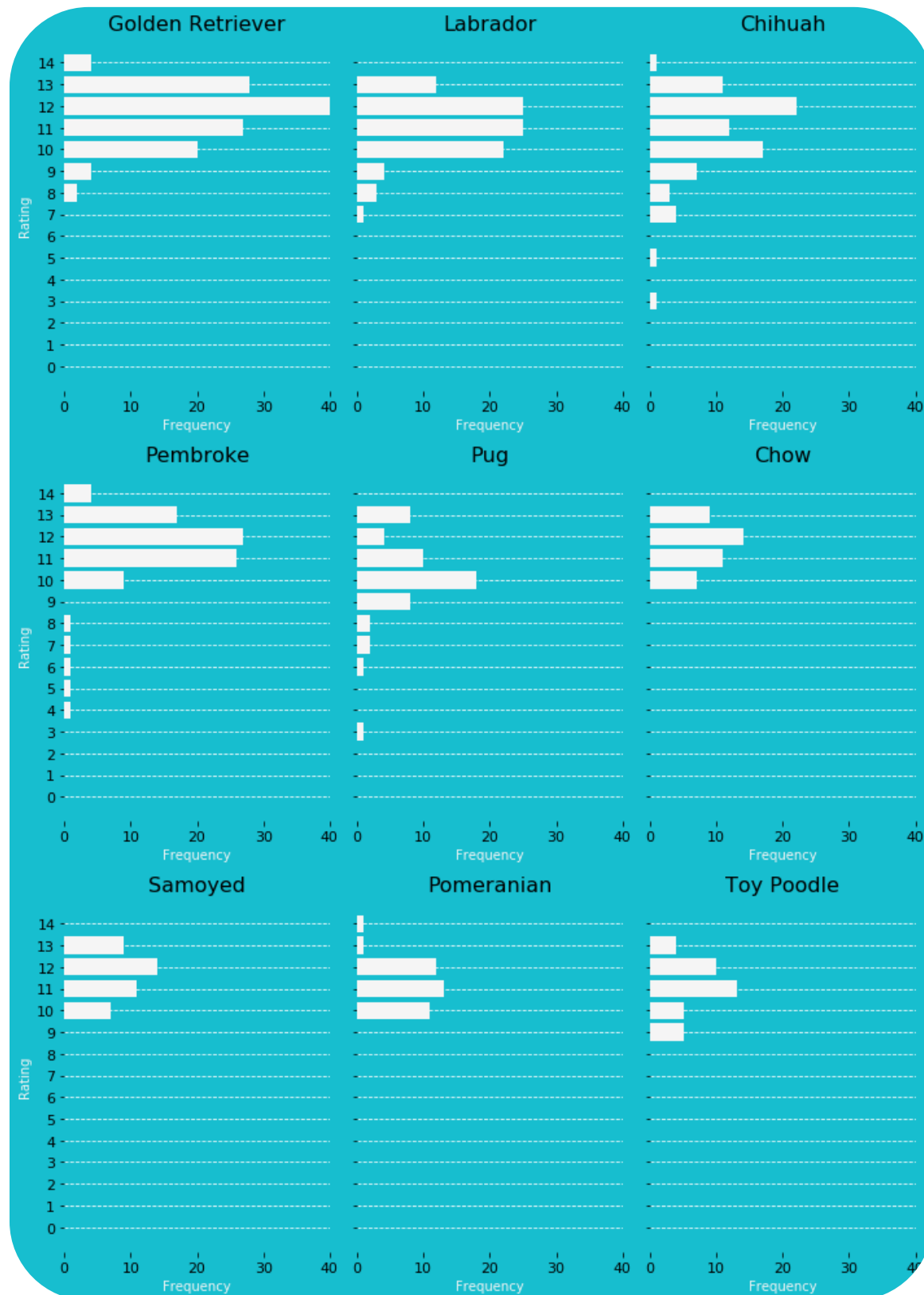
Bayesian Probability of Dog Ratings for the Top 5 Dog Types

Once the data was discrete and the frequencies were gathered I could then use what I know about how WeRateDogs twitter account rates each dog breed and derive a probability table.
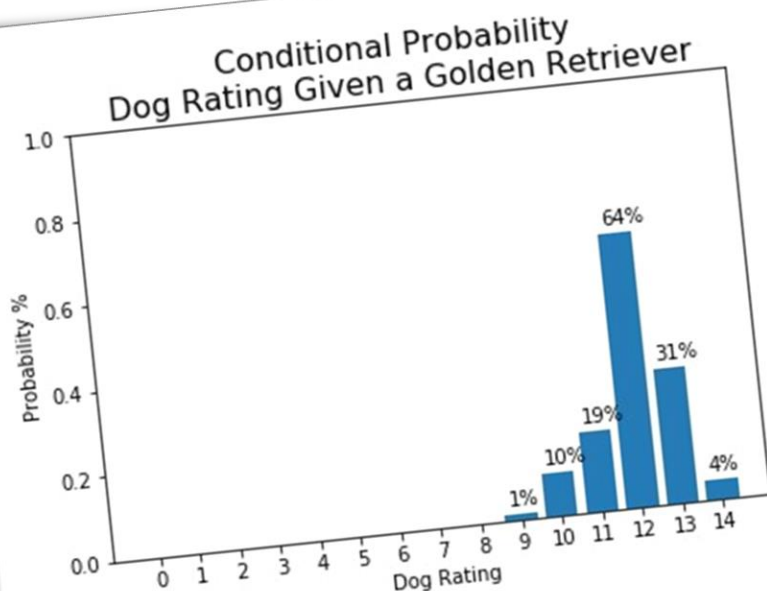
Using the probability table I could then calculate the bayesian conditional probability of how likely each rating was given the dog's breed.

I found that if a Golden Retriever is posted on WeRateDogs it has a 1% chance of receiving a rating of 9, a 10% chance of receiving a 10, a 19% of receiving an 11, but the rating the Golden Retriever is most likely to receive is a 12 out of 10, which has a probability 64%. The likelihood of receiving a rating of 12 is more than twice as likely as the next most likely rating.

At the highest end of the spectrum, a Golden Retriever has a 31% chance of receiving a 13 and a 4% chance of receiving a rating of 14. Next up I will take a look at the inverse conditional probabilities, the liklihood of the dog breed given that the rating is a 13 as well as the likelihood of the dog breed given the rating is a 14.
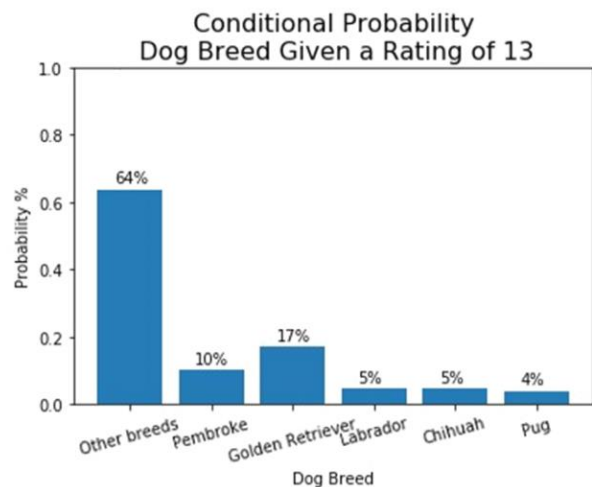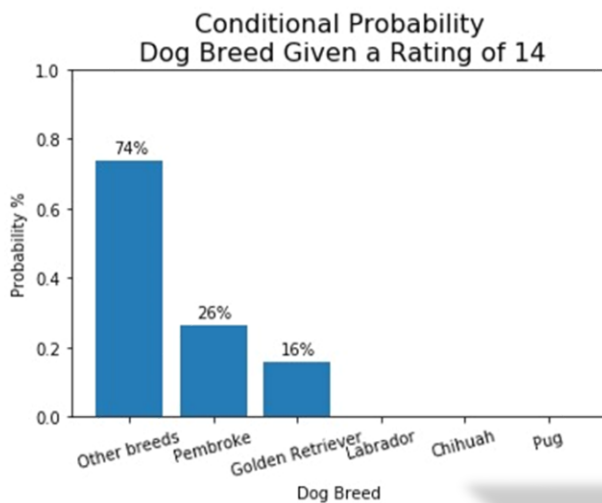
rating	other_count	pem_count	gold_count	lab_count	chih_count	pug_count	Total
0	2.0	0.0	0.0	0.0	0.0	0.0	2.0
1	4.0	0.0	0.0	0.0	0.0	0.0	4.0
2	9.0	0.0	0.0	0.0	0.0	0.0	9.0
3	17.0	0.0	0.0	0.0	1.0	1.0	19.0
4	14.0	1.0	0.0	0.0	0.0	0.0	15.0
5	30.0	1.0	0.0	0.0	1.0	0.0	32.0
6	30.0	1.0	0.0	0.0	0.0	1.0	32.0
7	43.0	1.0	0.0	1.0	4.0	2.0	51.0
8	84.0	1.0	2.0	3.0	3.0	2.0	95.0
9	129.0	0.0	4.0	4.0	7.0	8.0	152.0
10	335.0	9.0	20.0	22.0	17.0	18.0	421.0
11	304.0	26.0	27.0	25.0	12.0	10.0	404.0
12	318.0	27.0	53.0	25.0	22.0	4.0	449.0
13	179.0	17.0	28.0	12.0	11.0	8.0	255.0
14	28.0	4.0	4.0	0.0	1.0	0.0	37.0
Total	1526.0	88.0	138.0	92.0	79.0	54.0	1977.0

rating	other_count	pem_count	gold_count	lab_count	chih_count	pug_count	Total
0	0.001	0.000	0.000	0.000	0.000	0.000	0.001
1	0.003	0.000	0.000	0.000	0.000	0.000	0.002
2	0.006	0.000	0.000	0.000	0.000	0.000	0.005
3	0.010	0.000	0.000	0.000	0.001	0.001	0.010
4	0.009	0.001	0.000	0.000	0.000	0.000	0.008
5	0.018	0.000	0.000	0.000	0.000	0.000	0.016
6	0.018	0.000	0.000	0.000	0.000	0.001	0.016
7	0.024	0.000	0.000	0.000	0.004	0.001	0.026
8	0.049	0.000	0.000	0.001	0.001	0.001	0.048
9	0.072	0.000	0.001	0.001	0.004	0.008	0.077
	0.175	0.002	0.007	0.012	0.009	0.014	0.213
	0.150	0.019	0.013	0.017	0.005	0.005	0.204
	0.148	0.018	0.045	0.015	0.014	0.001	0.227
	0.082	0.013	0.022	0.006	0.006	0.005	0.129
	0.014	0.005	0.003	0.000	0.000	0.000	0.019
	0.772	0.045	0.070	0.047	0.040	0.027	1.000

Conditional Probability Dog Rating Given a Golden Retriever

rating	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
other_count	2.0	4.0	9.0	17.0	14.0	30.0	30.0	43.0	84.0	129.0	335.0	304.0	318.0	179.0	28.0	1526.0
pem_count	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	0.0	9.0	26.0	27.0	17.0	4.0	88.0
gold_count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	4.0	20.0	27.0	53.0	28.0	4.0	138.0
lab_count	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	3.0	4.0	22.0	25.0	25.0	12.0	0.0	92.0
chih_count	0.0	0.0	0.0	1.0	0.0	1.0	0.0	4.0	3.0	7.0	17.0	12.0	22.0	11.0	1.0	79.0
pug_count	0.0	0.0	0.0	1.0	0.0	0.0	1.0	2.0	2.0	8.0	18.0	10.0	4.0	8.0	0.0	54.0
Total	2.0	4.0	9.0	19.0	15.0	32.0	32.0	51.0	95.0	152.0	421.0	404.0	449.0	255.0	37.0	1977.0

rating	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
other_count	0.001	0.003	0.006	0.010	0.009	0.018	0.018	0.024	0.049	0.072	0.175	0.150	0.148	0.082	0.014	0.772
pem_count	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.002	0.019	0.018	0.013	0.005	0.045
gold_count	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.013	0.045	0.022	0.003	0.070
lab_count	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.012	0.017	0.015	0.006	0.000	0.047
chih_count	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.004	0.001	0.004	0.009	0.005	0.014	0.006	0.000	0.040
pug_count	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.001	0.001	0.008	0.014	0.005	0.001	0.005	0.000	0.027
Total	0.001	0.002	0.005	0.010	0.008	0.016	0.016	0.026	0.048	0.077	0.213	0.204	0.227	0.129	0.019	1.000

Conditional Probability Dog Breed Given a Rating of 14

Conditional Probability Dog Breed Given a Rating of 13

No matter how likely one of the top 5 breeds were to get a 14 or 13, it is more likely to be part of the Other breeds category, once the cumulative probabilities are all added together. The likelihood a rating of 14 or 13 was for a dog who was not part of the top 5 breeds was 74% and 64%, respectfully.
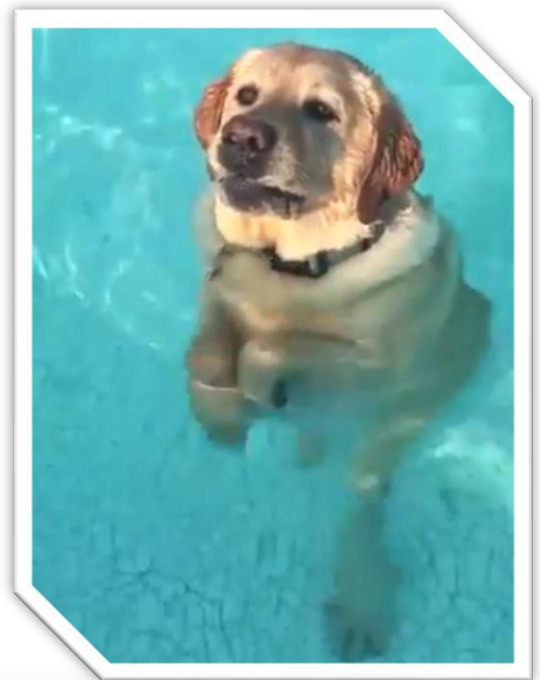
Given that the rating was a 14, the liklihood that the dog is a Golden Retriever is 16% and the liklihood that the dog's breed is Pemboke is 26%. Even though both breeds had 4 ratings of 14, the Pembroke had just over half as many observations in the dataset, 88, compared to Golden Retrievers 138 count.

Given that the rating was 13, the liklihood that a dog is a Golden Retriever is 17% and the liklihood of that the dog is a Pembroke is 10%. Labradors and Chihuahs are both 5% likely. Of the top 5 breeds, it was the least likely of receiving a rating of 13 would be a Pug, with only a 4% chance.

Time Series Analysis of Twitter Activity

From 2015 to Spring of 2016, WeRateDogs had relatively modest twitter traffic. The page did make a few jumps over the 80,000 favorite threshold in December of 2015, but it took until the end of 2016 for the account to consistently hit their over 50,000 favorite count per tweet that they are now accustom.

This doggo to the right is named Cody (Cody is his real name if you watch his now famous video). Cody learned how to stand in the pool and earned himself the most favorites & retweets of all the dogs on WeRateDogs within this dataset. Cody's tweet really took off when it was tweeted out on June 18th of 2016, long before it was common for WeRateDogs tweets to receive this level of activity, receiving 80, 263 retweets and 160,445 favorites. Good boy, Cody.