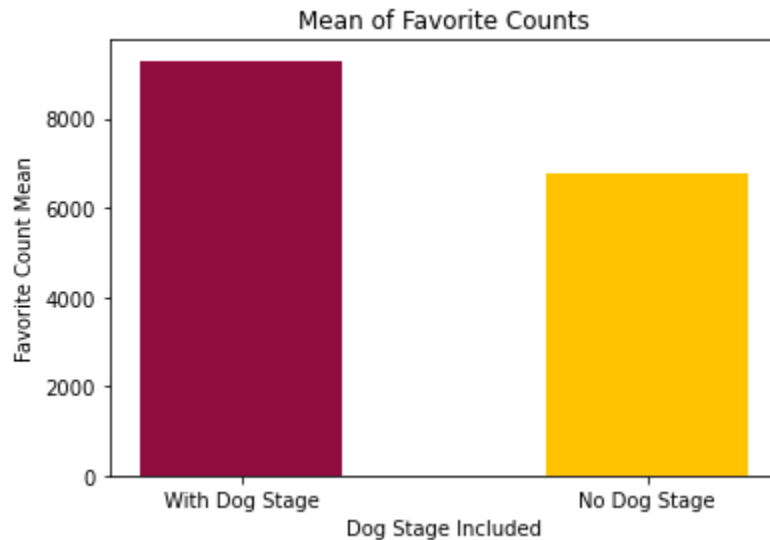


Actions Report

Question 1: "Do people who list the dog_stage in their tweet receive a significant number of favorites in comparison to people who do not list their dog's dog_stage?"

To try and answer this question I queried the dog_stage column for those rows that were not null, meaning they contained a dog stage and performed a separate query for those rows that were null (no dog stage listed). I then calculated the favorite count mean for each of these groups. I chose to run a Mann-Whitney U test on the means because the data for each group was not normally distributed (both datasets were right-skewed). The alpha value for my test was 5%. The results from the Mann-Whitney U Test show that the pvalue for this test was .017. This is below my threshold of 5%. I consider this to be enough evidence to reject my null hypothesis that there is no significant difference between the means of these two datasets. Furthermore, because the test was set to an alternative parameter of "greater" I can say that according to the data tweets with dog stages listed in the text have a significantly greater amount of favorites than those that do not list the dog stage in the text.



Question 2: Do people who post pictures of dogs with objects or just objects get more retweets than those who post pictures of just dogs?

Every image in the WeRateDogs Twitter archive has been run through a neural network that can classify breeds of dogs. The table "img_predict" holds the image predictions. We can see that in some of these predictions of breed, something other than "a dog" was identified. This is most likely because there was an object in the picture with the dog or the picture contains only an object and no dog. I will use these predictions to answer the above question. My null hypothesis assumes that there is no significant difference in the number of retweets a tweet will get if an object is in the picture(s) with the dog (or if there is only an object in the picture(s)) and tweet with a picture(s) of a dog only. The alternative hypothesis assumes that there is a greater number of retweets for tweets containing dogs and an object or just objects than tweets with pictures of dogs only. The data in the dog_objects and dog_only dataframes do not follow a normal distribution and do not have equal variances. Both sets of data do

have the same shape (right-skewed). I performed a Mann-Whitney U test on this data as this is a nonparametric test so it does not make assumptions about the distribution of the data (ie. the test does not assume normal distribution). Again, I set my alpha (threshold) to .05 or 5%. The pvalue from the Mann-Whitney U test was .98 or 98%. This is well above my threshold of .05 or 5%. This means that I do not have enough evidence to reject my null hypothesis.

Question 3: Which day has the highest average retweet count?

I used the gathered data to determine which day of the week had the highest favorite count. The first thing I did was merge the `twitter_arc_clean` and `df_rt_fav` data frames together on tweet id. This way the information that I needed was in one dataframe. I then used the timestamp information and the `datetime` method to determine which day of the week went with each time stamp. I grouped the table by day of the week and then found the favorite count mean for each day. I used this information to create a new data frame holding the days of the week and their mean favorite counts and to plot the bar chart below. A median line was included in the bar chart to show the mean for all retweet counts for the data set (average favorite count for the week). The bar chart shows that Wednesday and Saturday have on average more retweets than the other days of the week. Future opportunities for analysis may be to combine and average Saturday and Wednesday retweet counts and compare them to the other days of the week retweet count averages. Statistical tests could be run to see if the differences are significant or not. This could be used to tell WeRateDogs if there are days that they have better engagement from their audience or not.

