

Pup Inflation

Ruixuan Liu
301383988

Introduction:

This report was created to analyze data on how people rate @dog_rates on Twitter. Users will rate the dogs that participate in this activity by tweeting their photos on a scale of 0-10, but it can go beyond ten. So I'll be analyzing the data to see if it's inflated.

Analysis:

Figure 1 shows all of the data, but we can see that there are indeed some outliers, so we filtered the data and kept only those less than 25 points or less for analysis. After removing the outliers, our data becomes Figure 2.

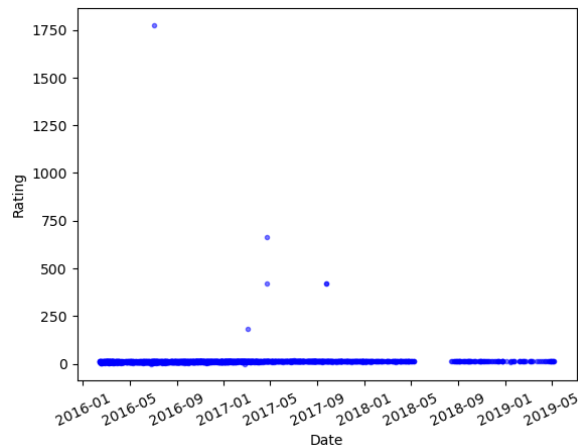


Figure 1

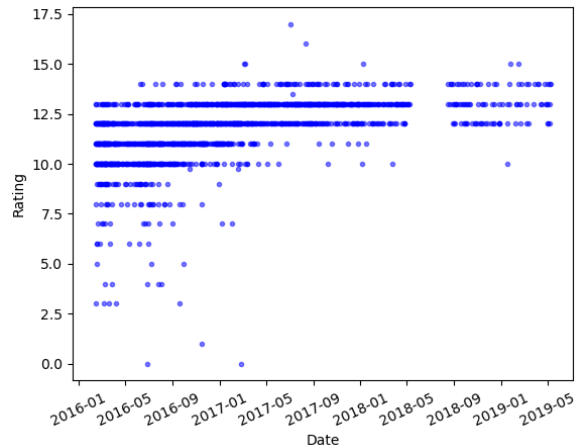


Figure 2

We first performed a linear regression on the data to calculate the slope and intercept, and then applied it to that data to draw a straight line that best fits the red color. (as shown in Figure 3)

Based on this red herring, we can now assume that there is a positive correlation between rating and year. Based on my analysis, I found $p\text{-value} = 2.536058725135737e-127$, which is much less than 0.05, so we can reject this hypothesis. That is, there is no positive correlation between ratings and years.

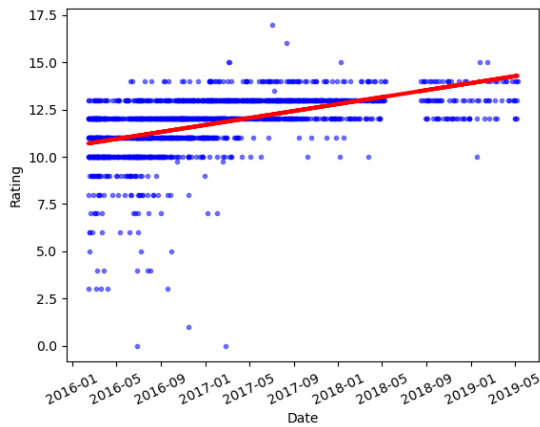


Figure 3

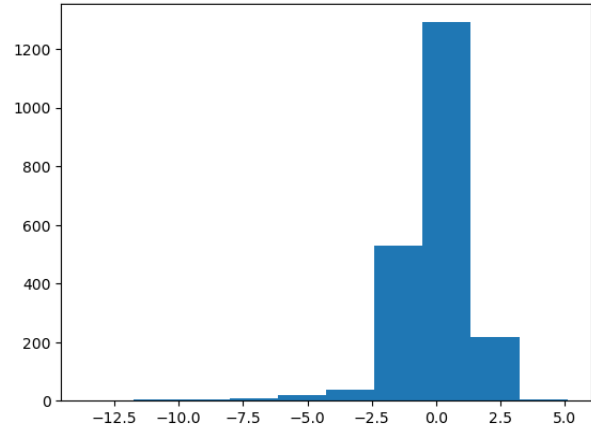


Figure 4

Finally, we also did a plot of the residual distribution and found that it is left-skewed.

Conclusion:

Based on our analysis of the linear regression and residual distribution of the data, there is inflation in these scores for rating dogs.