

Exercise 12: Explaining Pup Inflation

Exercise 2 and 7, explored pup inflation in @dog_rates Twitter account using the provided dog_rate_tweets.csv file. This data was produced using tweet_dumper.py to scrape the @dog_rates feed. First, all the tweets that contained an “n/10” rating were extracted, and its numeric rating was used to produce the following scatter plot with a line of best-fit.

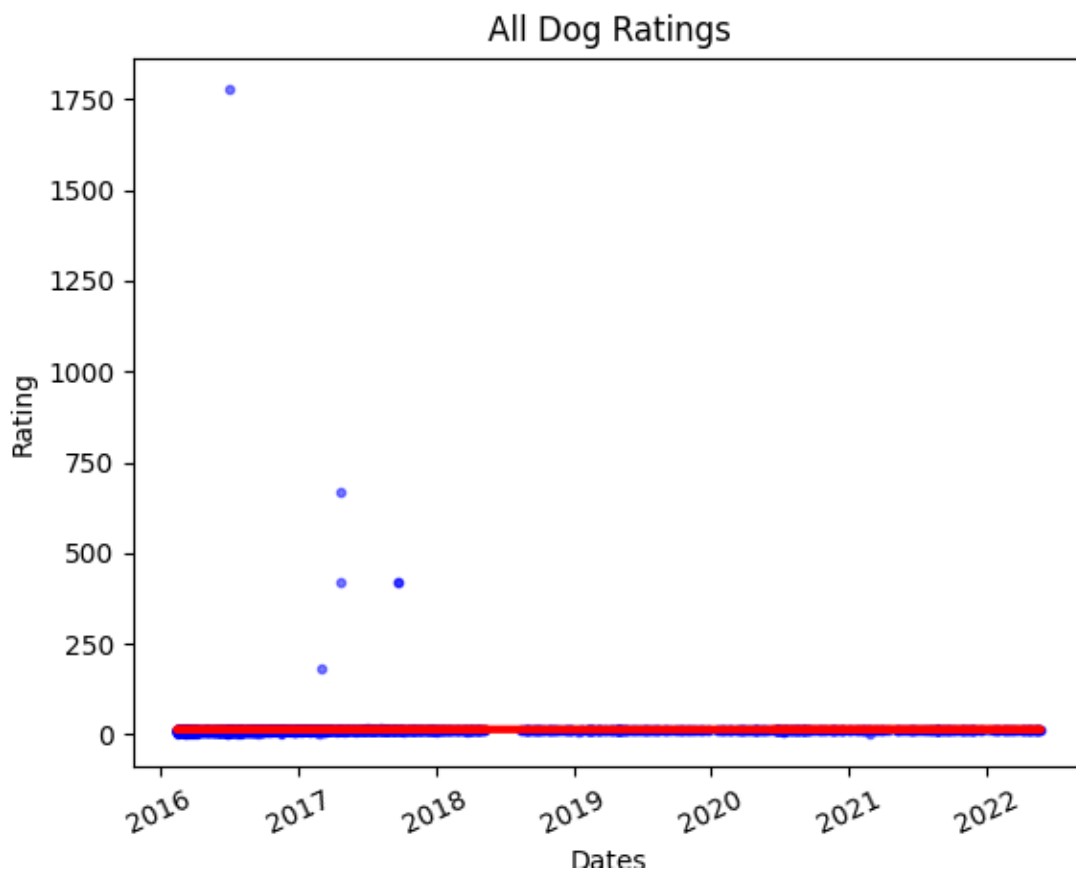


Figure 1: All dog ratings

There are a few outliers that make this visualization meaningless. After removing ratings larger than 25/10, we can see what’s happening with the data a bit better. The result is shown below.

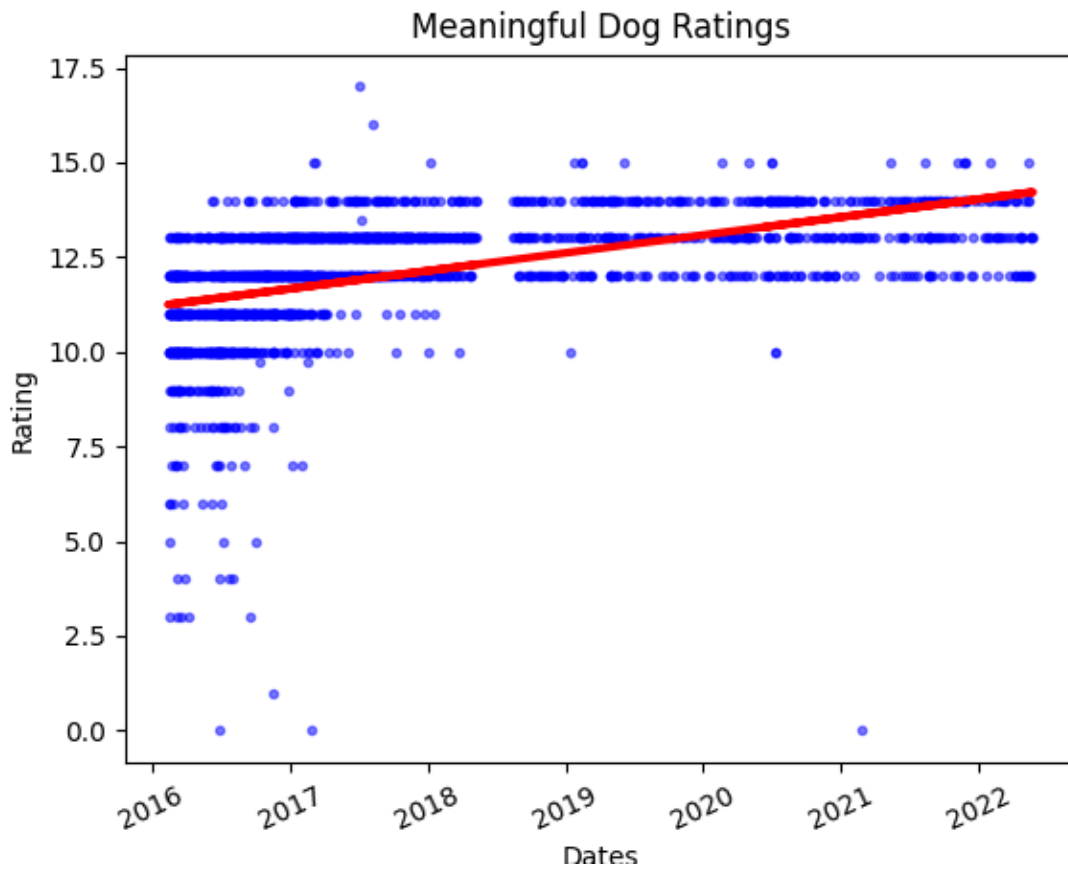


Figure 2: Dog ratings after removing outliers

Upon initial inspection of Figure 2, the rating and the date appear to have a positive relationship, further shown with the line of best fit created using the linear regression model with ordinary least squares. To justify this, we can check p-value from the linear regression model, which is about $2.6e-127$. Since it is less than 0.05, we can conclude that the slope is not zero, and the slope is about $1.50e-08$, which is positive. It seems likely that pup inflation is real.

Let's further explore if the average dog rating is changing monthly. After grouping the ratings by their Year-Month, we calculate the mean of those ratings and get the following line plot.

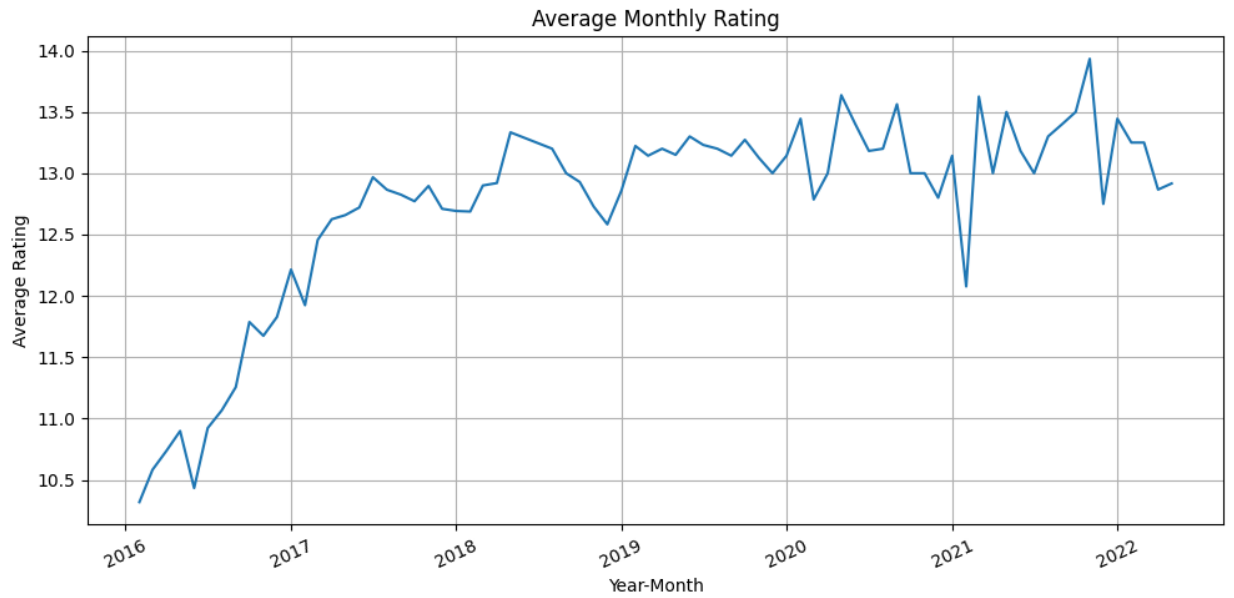


Figure 3: Average monthly dog rating

Although some fluctuation exists, it does seem that the average dog ratings are increasing over time. Therefore it feels reasonable to conclude that pup inflation is real.