

# Explaining Pup Inflation

## Introduction:

In this report, I will create an analysis of the @dog\_rates data, which was a Twitter activity that uses upload their dogs' pictures or videos, and all users will rate those pets with a numerical score, out of 10 can be higher than that. This report will show whether there has been grade inflation from @dog\_rates data.

## Step-by-step:

### First step filter data:

From Figure1, the image shows some improper values that we need to filter out for invalid or outlier values for the initial data.



(Figure 1)

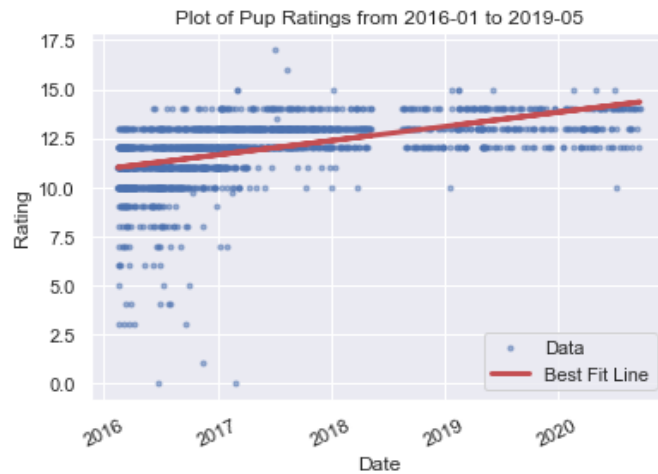
(Figure2)

### Comparing unfiltered and filtered data:

From Figure1, the unfiltered data shows a near consistent line representing the values close to the original "out of 10" rating system. Almost all the data is good to take analysis. We only need to filter a few outliers and remove those data that rating > 25. Now the data is good to analyze. We take a linear regression which generates the best fit line prediction, and a p-value of  $1.51e-106$  based on the data.

### Second step visualize the data:

We generate the best fit line that serves as a predictive model for the dog ratings using Ordinary Least Squares linear regression (OLS). The best fit line trends upwards from 2016 to 2019. (Figure3)

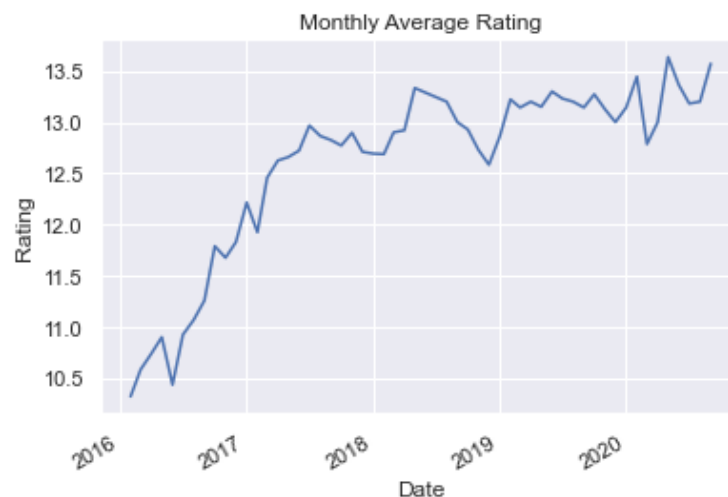


(Figure3)

Then we can make some assumptions:

- a) The sample is representative of the population
- b) The relation between Date and Rating is linear
- c) The residual is close-enough to being normal to look at the OLS p-value

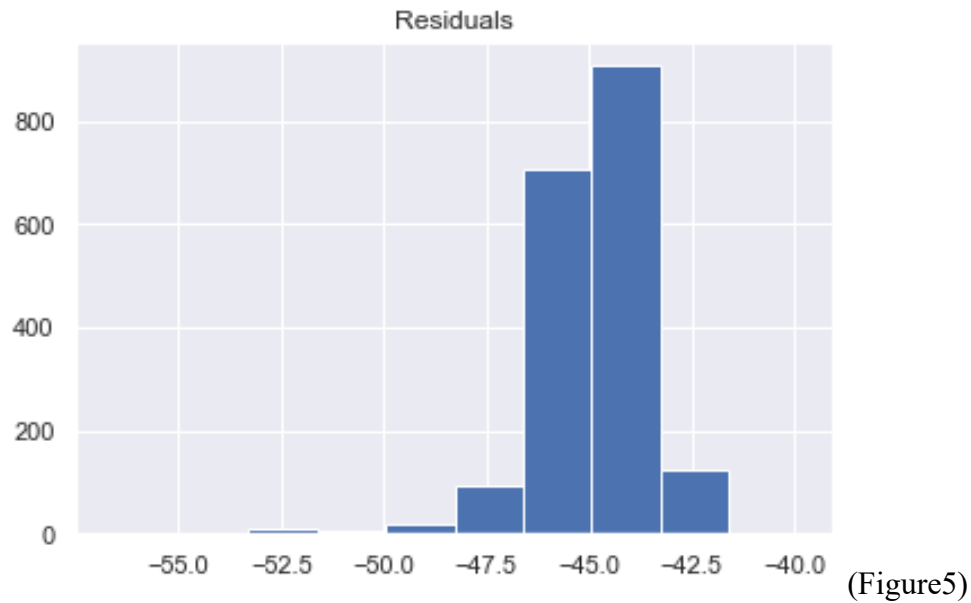
Then we can see the Figure4 which has an increasing distribution of the average rating and date.



(Figure4)

The regression p-value is  $1.5139606492960756 \times 10^{-6}$  so that we can reject the null hypothesis that there is no correlation between date and rating because of  $p < 0.05$ . Then we can conclude that the ratings are increasing and the sample can represent the population.

The histogram (Figure5) shows a left-skewed distribution, which should be enough to consider the OLS p-value. Then we can conclude the residuals look close enough to be normal.



### Conclusion:

There is a grade inflation in the dog ratings given by Twitter user @dog\_rates. Based on the upwards trend of the OLS best fit line and supplemented by the increasing month-to-month average in Figure4. We still need more data and information to know the reason why the dog ratings' data is grade inflation.