

Pup Inflation Analysis

Alec Pershick
301190478
CMPT 318

By scraping the @dog_rates Twitter account, we extracted all the data on dog cuteness ratings by random Twitter users. Our goal is to figure out if the ratings in that twitter account are changing with time.

Firstly, to analyze the data presented in the @dog_rates Twitter account, we needed to do a fair amount of cleaning to make sure we are getting somewhat valid data to work with. We first filtered out anything not containing an '10' rating from the data. Next we removed any outliers, we chose anything greater than '25/10' even though anything over 10 shouldn't really be considered valid. Finally we created a scatterplot of the data to see if we could spot a trend visually within the data to support our hypothesis (bottom left plot).

Looking at the scatterplot, we decided to plot a best-fit line onto the data, to further analyze it visually. The red best-fit line in the plot seems to be trending upwards over the past year and a half worth of data.

We decided to also plot a histogram of the residuals of the plot, to see if a linear regression analysis might be worth conducting on the data. We can see that the residuals seem to be fairly equally dispersed, making a linear regression analysis a viable option for testing it. (bottom right plot)

This is enough evidence to conduct a hypothesis test.

H_0 : Slope of the data is exactly 0

H_A : Slope of the data differs from 0

To approach this problem, we tested the slope using a linear regression analysis to see if the slope is different from zero. We came out with a p-value of $4.68648499391e-74$. We hadn't set a confidence level at the beginning of our test, but even if we had set it to 99%, which is very high for typical hypothesis testing, this p-value still would have been enough evidence to reject our null hypothesis.

We can therefore reasonably conclude that the slope of the graph is not 0, and even though we can't guarantee it, there is most likely an increasing trend of dog ratings on the @dog_ratings Twitter account.

