Arpit Kaur
310367801

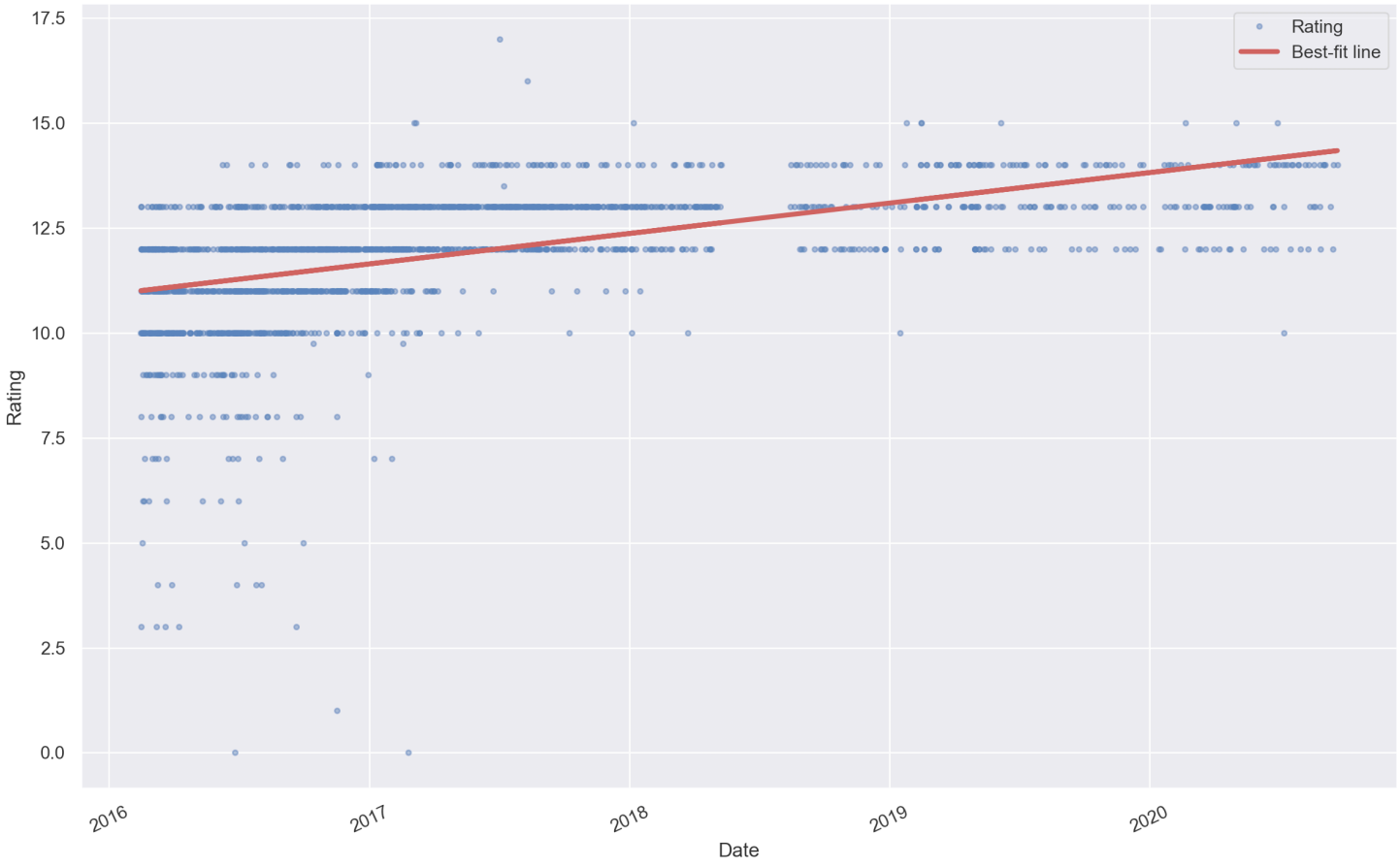
CMPT 353
Exercise-12

Pup Inflation Analysis

The Pup inflation : Analyzing tweet question asks whether there has been a grade inflation on the @dog_rates Twitter, which rates the cuteness of users' dog pictures.We used a pre scraped version of @dog_rates feed.

We perform linear regression to analyze the trend between predicted and actual data.To look for score inflation, we first had to make sense of the data, so we loaded the csv file into a pandas data frame. To make the data usable, we needed to do some preprocessing .After loading the data, we extracted the numeric rating of the tweets that contained "n/10" rating while excluding the tweets that did not contain a rating (as they were not required) with the help of string extract function(regex). The outliers were then removed, i.e, the values that were too large to be a rating.(The ones more than n=25).

After preprocessing, we apply linear regression on date and ratings to get the slope and the intercept of the best fit line for predictions.After that, for visualization, we plot graphs of rating vs date to see the trend and the relation or difference between the actual and the predicted data.We can analyze that from the figure given below.



Arpit Kaur
310367801

We can see from the graph that the ratings have increased over years.The slope of the best fit line is approximately 2.29.

To do further analysis, we can find the p value. Here, the p value comes out to be very less.A smaller p value represents a more significant impact. It means here that the result is what we were hoping to conclude.

To see if the ratings change with time we subtracted the actual ratings from those of the predicted ones to see if a relation arises or not.We can analyze this with the help of the histogram in the figure below.