Analysis of Possible Reddit Rating Inflation

We are given the task to investigate if the rattings in certain reddit channel are inflating over the past year. So, we gathered the comment data from the “CatDog” subreddit over the past year, which contains the comment itself and the time the comment was written.

We extracted the rating score from each comment and performed some data-cleaning against the scores, including removing outlier comment scores which does not make sense. Let’s take a look of the data by visualizing them on a graph.

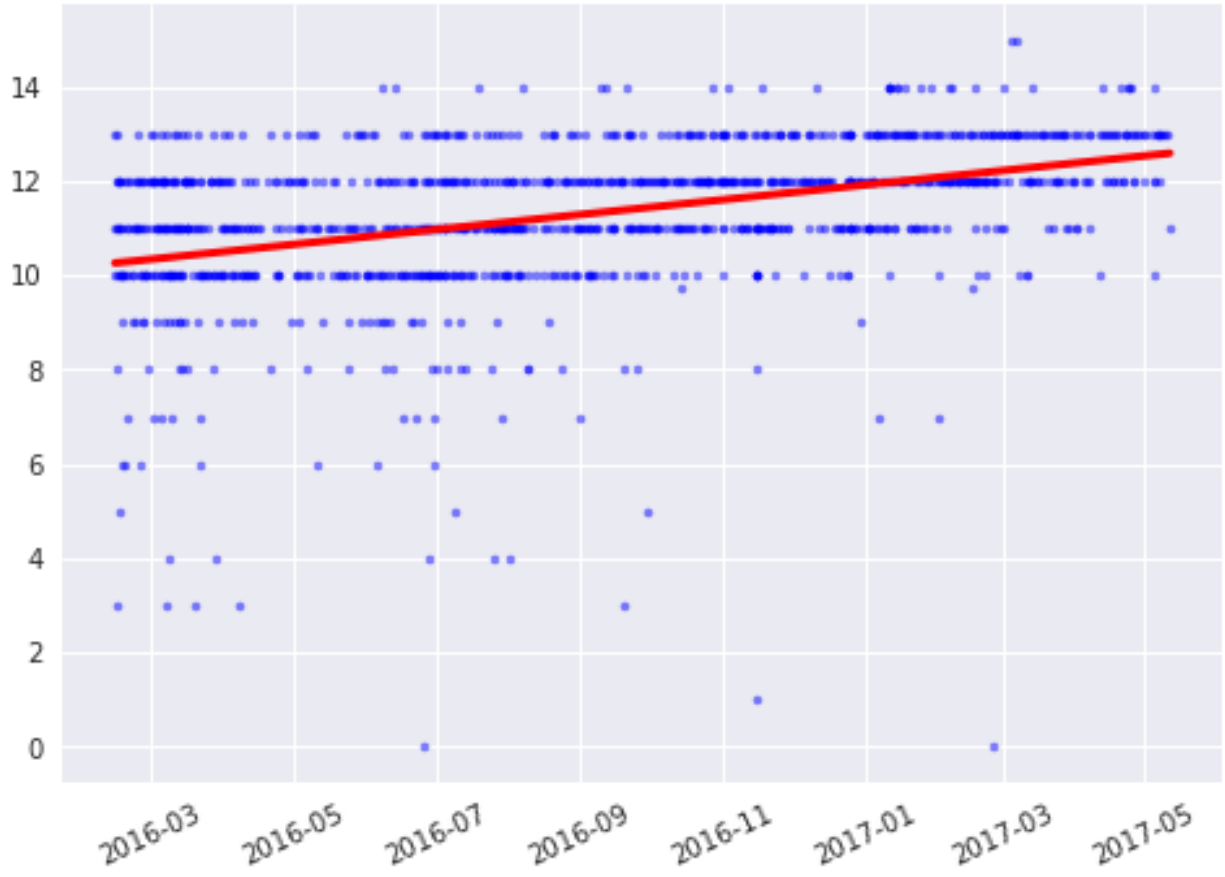

Figure 1-1, Reddit Rating Linear Regression Plot

In the graph, the blue dots represent each comment, with the x-axis representing the time and the y-axis representing the rating score. The interesting part is the red line. We performed a linear regression over the rating score and the red line is the best-fit line plot. As you can see in the graph. The slope of the line is slightly positive.

Let’s also take a look of a distribution residuals rating score.

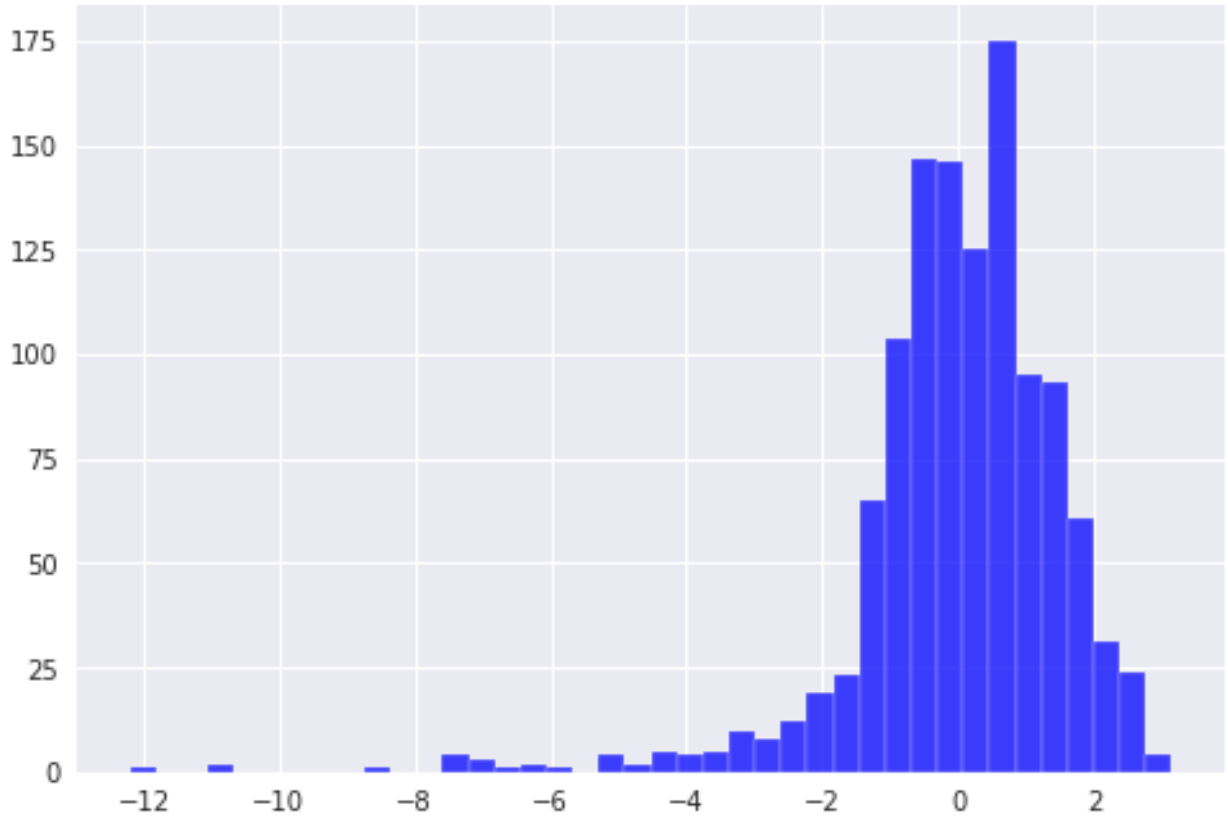

Figure 1-2, Reddit Residuals Rating Score Distribution

As you can see, the scores are in normal shape, and the p-value of the linear regression is.

```
P-value for the fit : 2.62541672742e-44
```

As the p-values is < 0.05 and the red line has a positive slope, it implies the trend of the rating score: The rating seems to be growing over the year!