



CrowdTangle Statistical Summary

在此处插入文本在此处插入文本在此处插入文本在此处插入文本

Yu Wang
2023-07-03

Part Zero - Simple Report Generate

The total posts crawled are around 23000, in order to generate the statistical report, I separated them by the account name by using Pandas 'groupby' function.

| account | posts_num | account | posts_num |
|------------------------|-----------|---------------------------|-----------|
| Marine Le Pen | 2378 | Mr. Peter Obi | 338 |
| NEW WORLD ORDER ☆) | 4235 | Renaud Camus | 391 |
| Sahara Reporters | 17416 | Steve King | 610 |
| Tucker Carlson Tonight | 586 | Tucker Carlson | 59 |
| We Are Igbos | 395 | Asiwaju Bola Ahmed Tinubu | 147 |
| Atiku Abubakar | 602 | נוער הגבעות - להכיר מקרוב | 111 |
| Congressman Matt Gaetz | 2112 | I HATE THE ZOO(NIGERIA) | 45 |
| INEC Nigeria | 556 | | |

As the tables shows, each account's posts num are varied with other accounts. Luckily, we can use the 'actual', 'expected', 'scores' statistical information to evaluate the post.

Part One - Post Statistical Summary

1. Using 'Ydata-Profiling'

For each account, I kept the original post text, 'actual' related statistical summary, 'expected' statistical summary and 'scores'. Using the 'Ydata-Profiling', we can get related heatmap and other summary information.

2. The 'Scores'

According to the official information from the CrowdTangle(<https://help.crowdtangle.com/en/articles/2013937-how-do-you-calculate-overperforming-scores>):

Overperforming $\geq 1.0x$

Every post is compared to a benchmark, which is the expected value of that post. The equation is:

$$\text{score} = \text{actual} / \text{expected}$$

So if a post has 100 interactions and its benchmark was 50, that's $100/50 = 2.0x$. To generate the benchmarks, we take the last 100 posts from a given account and of a given post type (link post, image post, etc.). We drop the top and bottom 25% of those 100 posts, and calculate the mean number of interactions that the middle 50% of the posts have at each age (15 minutes old, 60 minutes old, 5 hours old, etc.).

Underperforming $\leq -1.0x$

If you know math, and I know you do, you'll recognize that if we're doing simple division, then anything overperforming should score ≥ 1.0 , and anything underperforming should be between 0 and 1.

But we liked the idea of making underperforming be negative, because it seemed more instantly obvious. Remember the philosophy above. So how to do that?

If something is expected to have 200 interactions and it has 100, it's at 50% of its expected value, and that seems like $-2x$ underperforming to us. So we flip the equation and make it negative:

$$\text{score} = -1 * (\text{expected} / \text{actual})$$

Or in this example, $-1 * (200 / 100) = -2x$.

However this documentation raised some questions:

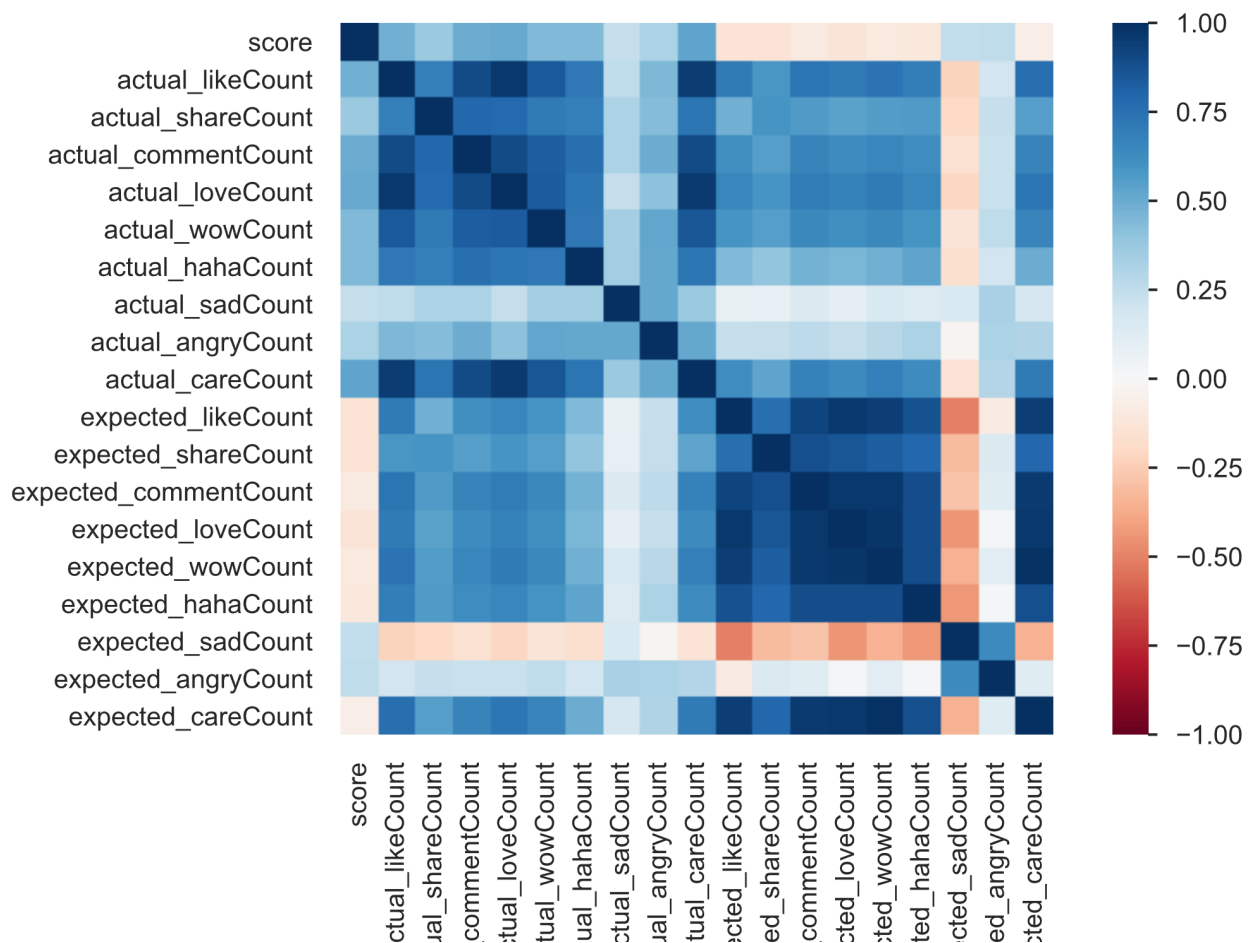
Where did the 'expected scores' come from? What is the benchmark?

3.The Conclusions From The Heatmap

3.1 The relationship between 'actual' and 'expected' are different

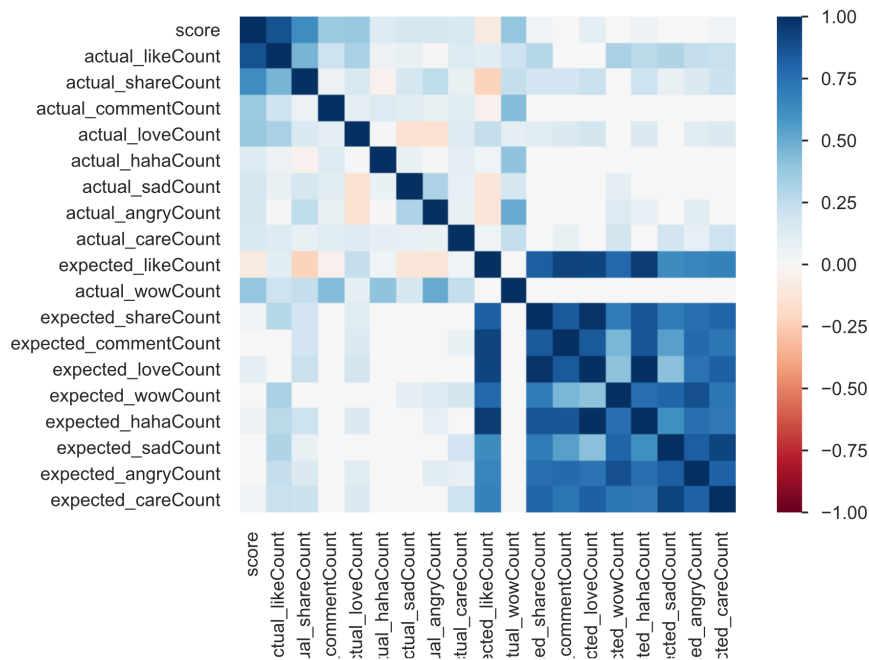
Generally, we may assume a post which has a high 'expected' related scores tends to also have a relative high 'actual' scores. However from the heatmap, the trends are not the same.

Highly Correlated : The 'actual' and 'expected' that have high correlations

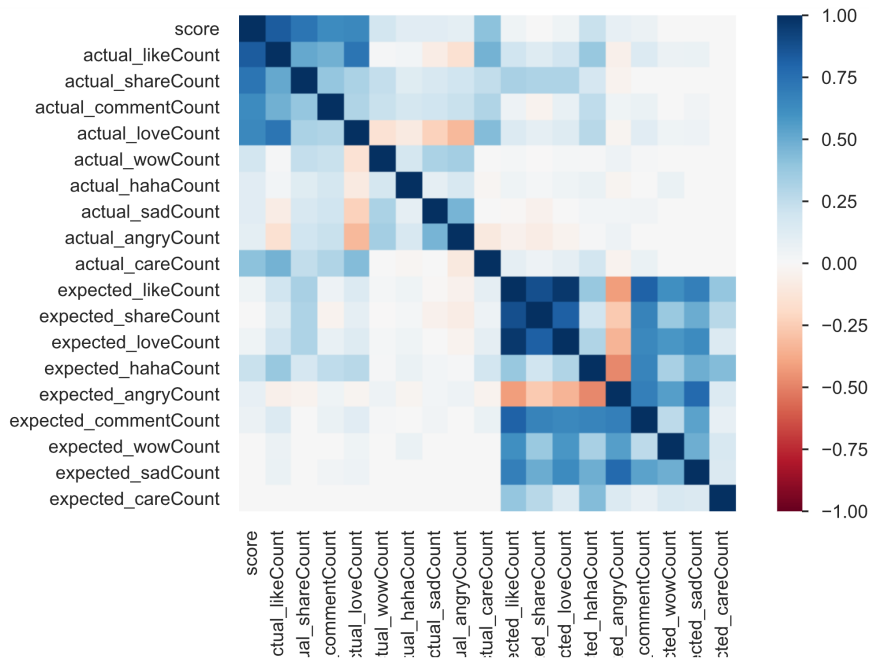


This is the heatmap from the 'Mr. Peter Obi' account.

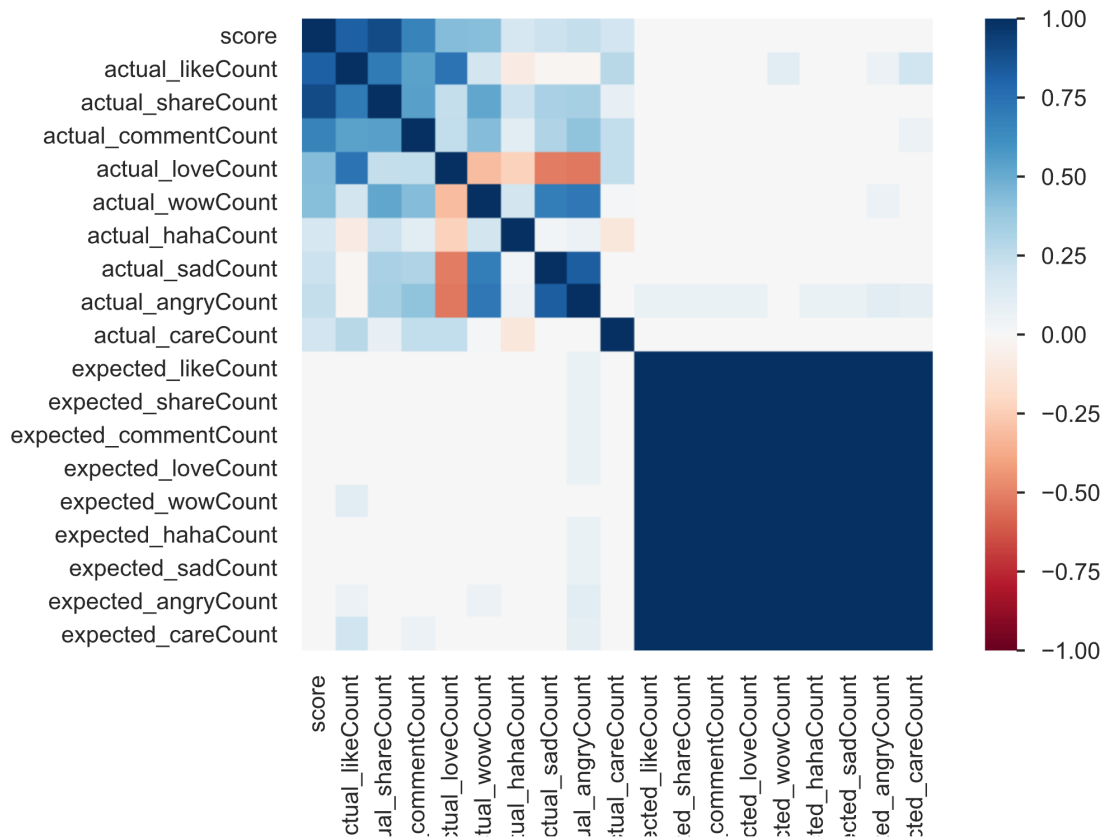
Lowly Correlated Example : The 'actual' and 'expected' that have low correlations



This is the heatmap from the 'Renaud Camus' account.



This is the heatmap from the 'NEW WORLD ORDER ☆)' account.



This is the heatmap from the 'Steve King' account.

Another important information from this Heatmap:

As you can see, the relationships among the 'expected' information are highly correlated, however in the 'actual' information the correlation relationship tends to be different from each other.

3.2 The score is highly more correlated with positive emotions than the negative emotions

This conclusion is drawn from all the correlation score tables. Also, people tend to 'like' and 'love' the posts than feeling 'sad' or 'angry' about the posts.

3.3 The scores with positive emotions

The scores are highly correlated with 'like' or 'love' emotions.