Video View Count for Verified/Unverified Authors

Showing the statistical significance of their view counts

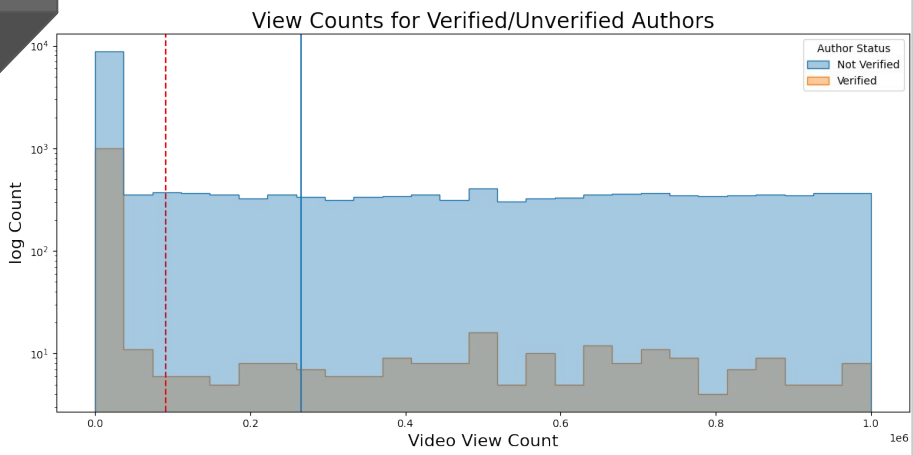
Project Overview

The prior work done on this database indicated that unverified authors get more views than verified authors. Here we want to prove if this difference is real or if it's subject to the variability of views in the platform.

Key Insights

- We will conduct hypothesis testing on the two observed populations
- We need to remove rows with missing data
- We need to remove data points that are considered outliers using 1.5*IQR rule
    - This will remove biases in our analysis.
- The results of our analysis indicate that this difference is statistically significant and that it is not due to chance.
- We also see that the underlying distribution for the view count of both author status is roughly uniform

Details



Histogram of video view count for each each author status. Vertical lines indicate the mean views for that author status category. Solid blue represent the mean views for unverified authors and dashed red line represents the mean views for unverified authors.

Next Steps

It's hard to say anything meaningful about the underlying distribution of the views because of their uniformity. I could prove to be beneficial to randomly sample from this population and compute means of these samples and later on use the central limit theorem results for hypothesis testing.