Executive Summary: Hypothesis Testing on Video Views by Verified and Unverified Accounts

TikTok Claims Classification Project

Project Overview

The TikTok data analytics team is working on developing a machine learning model to support the classification of user-submitted claims. In this phase of the project, a hypothesis test was performed to examine the association between verified_status and video_view_count.

Key Insights

- The analysis indicates a statistically significant difference in the number of video views between TikTok videos from verified and unverified accounts. This result suggests underlying behavioral differences between these two groups of users.
- Further investigation into the root causes of these differences would be valuable. Questions to explore include whether unverified accounts tend to post more engaging or provocative content, and whether such content is more claim-driven or opinion-based. Another possibility worth examining is whether some unverified accounts are linked to automated behavior, such as spam bots artificially inflating view counts.

Details

The TikTok data team investigated the relationship between verified_status and video_view_count. Initially, descriptive statistics were computed, revealing that unverified accounts had an average of 265,664 views, compared to 91,439 views for verified accounts. To validate these findings, a two-sample hypothesis test (Welch's t-test) was conducted. The results confirmed that the observed difference in mean view counts is statistically significant and unlikely to be explained by random variation, indicating a true difference between the two account types at the population level.

Next Steps

The next step is to develop a regression model focused on verified_status to further explore behavioral patterns among verified and unverified users. Insights gained from this model will provide important context for the future claim classification model, enhancing its predictive capabilities and improving decision-making based on user verification status.