TikTok Claims Classification Project – Statistical Review

Executive Summary Report III

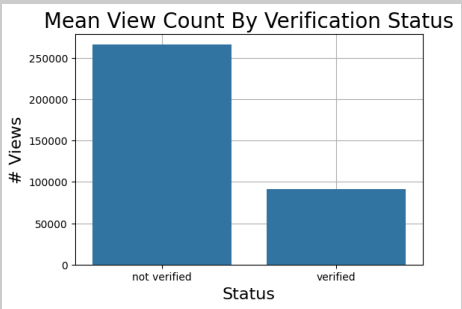
Milestone 4

ISSUE / PROBLEM

The TikTok analytics team aims to build a machine learning model that categorizes user-submitted claims. In this phase, the focus is on hypothesis testing to explore the link between a user's verified status and the number of video views.

IMPACT

The next recommended step is to develop a regression model based on verified status. This model can provide deeper insights into user behaviors among verified users and serve as a foundation for the subsequent claim classification model.



not verified	265663.785339
verified	91439.164167

RESPONSE

The analysis examined the connection between verified status and video view counts. Initially, the average view counts for videos from verified and unverified accounts were compared. A two-sample hypothesis test was also performed to confirm that the differences observed in the sample data are likely reflective of real differences in the population means.

KEY INSIGHTS

- Videos from unverified accounts averaged 265,663 views, while those from verified accounts averaged 91,439 views. This analysis aligns with the results of the hypothesis test.
- The findings indicate a variation in the average view counts between videos posted by verified and unverified TikTok users (**statistically significant** difference).
- There may be inherent differences in behavior between these two groups of accounts (verified, unverified). This disparity warrants further investigation into its causes.
- The following research questions remain open:
  - ❑ Are unverified accounts more likely to post content that draws higher engagement?
  - ❑ Is engaging content posted by unverified users more likely to be opinion-based?
  - ❑ Could unverified accounts be tied to spam bots that artificially boost view counts?