Executive Summary: Statistical Testing Results

TikTok Claims Classification Project

Project Overview

The TikTok team wants to develop a machine learning model to help classify the claim_status of a specific video, as this factor is important for the TikTok team to decide whether to manually review a video. In this part of the project, the data team will conduct hypothesis testing to analyze the relationship between verified_status and video_view_count.

Key Insights

- The analysis revealed a difference in the number of views for TikTok videos posted by verified accounts and those posted by unverified accounts.
- There could be a fundamental user behavioral difference between verified status of users.
- There could be two possible reasons for this relationship between the user verified status and video engagement levels:
    - Unverified accounts tend to post more attractive videos.
    - There is a high chance that unverified accounts are associated with spam bots that help increase pageviews.

Details

The TikTok data team considered the relationship between verified_status and video_view_count.

One approach taken was to examine the mean of video_view_count for each group of verified_status in the sample data. The results showed that the average number of views for unverified accounts was 265,663, while the average number of views for verified accounts was 91,439. The second approach was a two-sample hypothesis test. Consistent with the initial results on the means, this statistical analysis showed that any differences observed in the sample data were due to actual differences in the corresponding population means.

Next Steps

The team plans to develop a classification model for predicting verified status.

A classification model for verified_status can help analyze the relationship between user behavior and verified status. This information can then be used to consider the results of a subsequently created classification models for predicting claim_status of videos.