

Executive Summary: Statistical Testing Results

TikTok Classification Project

Overview

The TikTok data team wants to develop a machine learning model to assist in the classification of claims for user submissions.

Problem

In this part of the project, the data team wants to determine if there is a statistically significant difference in the average number of views for TikTok videos posted by verified accounts versus unverified accounts.

Solution

It was concluded at the 5% significance level that non-verified users published more TikTok videos in which a claim/opinion has been made.

Details

The TikTok data team analyzed how verified status relates to video view counts. First, the average view counts for each verification group were compared, and it was found that unverified accounts averaged 265,664 views, while verified accounts averaged 91,439 views.

```
verified_status
not verified    265663.785339
verified        91439.164167
Name: video_view_count, dtype: float64
```

Next, we conducted a two -sample hypothesis test, which confirmed that the difference observed in the sample reflects a real difference in the population means, consistent with the initial comparison of averages.

Next Steps

The next step will be to build a regression model on `verified_status` . A regression model is the natural next step because the end goal is to make predictions on claim status. A regression model for `verified_status` can help analyze user behavior in this group of verified users.