Executive Summary: Statistical Testing Results

TikTok Claims Classification Project

Project Overview

The data team is tasked with making a ML model to predict videos whether they are claims or opinions. In this section we conduct hypothesis testing to check relationship between verified status and view count.

Key Insights

Initial exploration of dataset shows that there is a difference between the number of views for videos posted by verified and unverified accounts.

This does suggest that there might be a behavioural difference between the two groups.

This difference might arise due to either the videos being more engaging from the unverified account or their probable usage of spam accounts to increase views.

Details

Above shown are the mean video counts for each group of account types. The readings show that videos posted by unverified accounts are getting more views compared to the verified counterpart.

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

Then we conducted a hypothesis testing on two samples. It revealed a p-value of 2.5e-120 which is extremely small compared to the significance level of 0.5 thus helping us reject the null hypothesis.

Next Steps

The next step is to create a regression model on verified status column. This will help us analyze user behaviour according to status. After that the claim classification model can be created.