Claim Classification Project

Executive Summary - Statistical Testing

Project Overview

The TikTok data team seeks to develop an accurate predictive model that determines whether a video contains a claim or an opinion. In this part of the project, the team conducted hypothesis testing to analyze the relationship between video_view_count and verified_status.

Key Insights

- There is a statistically significant difference in the video view count between verified and unverified accounts.

- It might be interesting to investigate why and what behavior caused the difference in view count. Why does videos posted by unverified accounts have more views?

For example,
- Does unverified users use spam bots to generate more views?

Details

We analyzed the relationship between verified_status and video_view_count.

	verified_status	average_video_view_count
0	not verified	265663.785339
1	verified	91439.164167

When comparing the average video views between verified and unverified, we can see that there is a significant difference in view count.

The hypothesis testing result also shows that there is a statistically significant difference. This means that the observed difference is due to actually difference and not due to chance.

Next Steps

The next step would be to build a **regression model** on verified_status.

Investigating the relationship between verified_status and other variables will tell us more about the user behaviors that can help us to predict claim status.