TikTok User Behavior: Verification and View Counts

Executive Summary 4: Statistical Analysis for Claim Classification Model

Project Overview

The TikTok data team is developing a machine learning model to classify user-submitted claims. As part of this effort, we conducted a hypothesis test to analyze the relationship between `verified_status` and `video_view_count`, aiming to understand user behavior and inform model development.

Key Insights

A two-sample t-test revealed a statistically significant difference in video view counts between verified and non-verified accounts.
Specifically, non-verified accounts exhibit significantly higher average video view counts (mean: 265,663) compared to verified accounts (mean: 91,439).
This suggests potential fundamental behavioral differences between these two groups.
Possible explanations include:

- Unverified accounts may post more engaging content, potentially including claims or opinions that resonate strongly with viewers.
- Unverified accounts may be associated with spam bots that artificially inflate view counts.

The p-value resulting from the t-test was 2.6088823687177823e-120, which is extremely small, and therefore we reject the null hypothesis.

Details

Data Preparation: Loaded data, removed rows with missing values.

Descriptive Analysis: Observed significantly higher average view counts for unverified accounts.

Hypothesis Testing:

- Performed a two-sample t-test.
- Resulting p-value: 2.6088823687177823e-120.
- Confirmed a statistically significant difference in view counts.
- This low p value means that the difference in views between the two groups is not due to random chance, but is a real difference between the two populations.

Next Steps

Build a regression model specifically for `verified_status` to analyze user behavior within this group, providing valuable context for the subsequent claim classification model.
Conduct further investigation to determine the root cause of the observed behavioral differences. Specifically:

- Analyze the content of highly viewed unverified accounts to assess engagement and content type.
- Investigate potential spam bot activity associated with unverified accounts.

Investigate the algorithms that are used to promote or display videos, to better understand why non verified accounts are receiving more views.