Statistical Analysis of View Counts Between Verified and Unverified TikTok Accounts

Hypothesis Testing of TikTok Dataset

Project Overview

This project investigates whether there is a statistically significant difference in average video view counts between verified and unverified TikTok accounts. Using a Two-Sample t-test, we explore the impact of verification status on user engagement measured by view counts.

Details

Key Insights

- Descriptive Statistics show wide variability in view counts across the dataset (**mean = 254,708**; **std = 322,893**).
- The average view count for **unverified accounts** is significantly higher (**265,664**) than for **verified accounts** (**91,439**).
- A Two-Sample t-test returned a **p-value = 2.61 × 10^{-120}** , well below the 0.05 significance level.

- **Statistical Method Used:** Independent Two-Sample t-test with unequal Variances (equal_var=False).
- **Null Hypothesis (H_0):** No statistically significant differences in view counts between the two groups.
- **Alternative Hypothesis(H_a):** There is a statistically significant difference in view counts between the two groups.
- **Result:** With a t-statistic of **-25.50** and an extremely small **p-value**, we **reject the null hypothesis.**
- **Conclusion:** Verification status is significantly associated with differences in video view counts.

Next Steps

- Explore **why unverified accounts** have higher average view counts despite lacking verification – consider factors like content type, algorithm bias or posting frequency.
- Develop a **predictive model**, ideally a **logistic regression model**, to estimate future outcomes (e.g. claim status) using variables such as video duration, likes, shares and verification status. This is appropriate given the skewed data and binary nature of some variables.
- Consider additional hypothesis tests to assess interaction effects between other variables (e.g. ban status, comment count, etc).