TikTok Claims Classification Project

Executive Summary: Statistical Testing

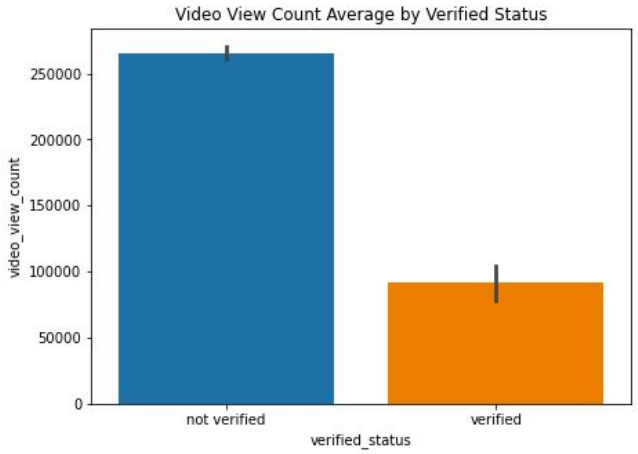
Project Overview

A hypothesis test will be conducted to determine if there is a statistical difference between verified and not verified user data. We will be testing the relationship between the verified status and video view count variables. This is statistical testing will aid in the future development of a machine learning model to assist in the classification of TikTok user claims.

Key Insights

- The analysis done on the variables shows that the average video view count of a unverified user is much larger than that of an verified user
- These findings may be the result of how these two groups of users are actually engaging with TikTok.
    - Are unverified users bot accounts that spam videos to increase video view count?
    - Are users who go through the verification process just making less engaging videos than those without verification?
- Investigating the difference between the posts made by both groups of users may lead to more interesting discoveries as to why the discrepancy is so large

Details



Bar graph showcasing the discrepancy between the average video view count of verified and unverified users

After observing the mean video view counts for both groups of TikTok users we conducted a two-sample hypothesis test. This hypothesis test resulted in the rejection of the null hypothesis meaning that any difference observed between the two groups is statistically significant.

Next Steps

The next steps would to move forward with the claims classification project by building a regression model

A regression model for the verification status will allow for better analysis of the unique difference between the way both groups of users engage with TikTok.

The resulting data will in turn lead to the creation of a claims classification model.