# Executive Summary (Task 4)

## Overview

The TikTok data team seeks to develop a machine learning model for content classification. In this part of the project, the clients are interested in whether there is a difference between verified and not verified accounts. To tackle this, I conduct a 2 sample T hypothesis test to verify whether not verified accounts have more mean view count than verified accounts.

|  |  |
| --- | --- |
| **Key Insights**   * The hypothesis test results reveal that not verified accounts have much higher mean views than verified accounts. * The results are statistically significant (P value < 0.05) * I believe there are might be fundamental behavioral differences between the two groups and it would be interesting to investigate the roots of the differences:  1. What make unverified videos more engaging than verified ones? 2. Why verified accounts have lower views? | **Details**   * Distribution of views for verified and not verified accounts are highly skewed, yet we have large sample size enough to proceed the hypothesis test. * Not verified accounts have a mean view of **265.6k**   Verified accounts have a mean view of **91.4k**   * Null (HO): Mean view count (not verified) <= Mean view count (verified) * Alternative (H1): Mean view count (not verified) > Mean view count (verified) * 2 sample Welch’s T test results: T statistics (25.49) > T critical value (1.65), P value <0.05 * Conclusion: reject the null hypothesis (Ho) |
| **The next steps**  The team can move forward and build a predictive regression model on verified status. A technical note is that data sample is highly skewed and should be considered to choose appropriate models. | |