

# Executive Summary: Statistical Testing Results

TikTok Claims Classification Project

## Project Overview

The TikTok data team seeks to develop a machine learning model to assist in the classification of claims for user submissions. In this part of the project, the data team will conduct a hypothesis test to analyze the relationship between `verified_status` and `video_view_count`.

## Key Insights

- The analysis shows that there is a difference in number of views between TikTok videos posted by verified accounts and TikTok videos posted by unverified accounts.
- As a result, these findings suggest there might be fundamental behavioral differences between these two groups of accounts: verified and unverified.
- It would be interesting to investigate the root cause of this behavioral difference. For example, consider:
  - Do unverified accounts tend to post more engaging videos? Is that engaging content a claim or opinion?
  - Or, are unverified accounts associated with spam bots that help inflate view counts?

## Details

The TikTok data team considered the relationship between `verified_status` and `video_view_count`.

One approach conducted was to examine the mean values of `video_view_count` for each group of `verified_status` in the sample data. The findings showed that most accounts were unverified. 265,663 accounts were not verified and 91,439 accounts were verified.

```
verified_status
not verified    265663.785339
verified        91439.164167
Name: video_view_count, dtype: float64
```

The second approach was a two-sample hypothesis test. Aligned with preliminary findings from the mean values, this statistical analysis shows that any observed difference in the sample data is due to an actual difference in the corresponding population means.

## Next Steps

The team suggests moving forward and building a **regression model** on verified status.

A regression model for `verified_status` can help analyze user behavior in this group of verified users. Then, this context can be used to consider results from a claim classification model that will be created afterwards.