

# Executive Summary

Milestone 4 of the TikTok Claims Classification Project

## Project Overview

TikTok's data team is building an ML model to classify user-generated claims. Current focus is on performing EDA to identify distinguishing features between videos with claims and those with opinions.

## Key Insights

- A statistically significant difference in TikTok video views between verified and unverified accounts was found through analysis.
- This finding has implications for understanding fundamental behavioral differences between these account types.

## Further Investigation

Further investigation into the underlying causes of this behavioral difference is recommended. This investigation could explore the following possibilities:

- Whether unverified accounts exhibit a tendency to post more engaging content, irrespective of its harmfulness.
- Whether unverified accounts are associated with bot activity that artificially inflates video view counts.

## Details

The relationship between `verified_status` and `video_view_count` was examined. Mean `video_view_count` was calculated for each `verified_status` group. Unverified accounts exhibited nearly three times the mean view count of verified accounts.

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

- A two-sample t-test confirmed the prior findings, demonstrating that the observed differences reflect actual population differences and leading to the rejection of the null hypothesis.

## Next Steps

- A regression model using `verified_status` as a key predictor will be developed to investigate the behavioral differences between verified and unverified users and the relationship between verification status and the presence of claims or opinions in their videos.