**Executive Summary**

**Overview**

This project analyzed a dataset of over 19,000 TikTok videos to explore the relationship between video content type—specifically whether a video makes a claim or shares an opinion—and various engagement and account status metrics such as views, likes, shares, comments, and ban status. The goal was to identify patterns that could inform moderation practices and support future machine learning efforts to classify content automatically.

**Problem**

TikTok aims to better understand what differentiates videos classified as claims from those categorized as opinions, especially in terms of engagement behavior and content moderation outcomes (e.g., author bans). This is critical for improving platform safety and ensuring that controversial or misleading content is handled effectively.

**Solution**

Using Python and Pandas, we:

* Cleaned and summarized the dataset.
* Examined the distribution of videos by claim status.
* Analyzed engagement metrics (views, likes, shares, comments).
* Investigated how author ban status and claim status are correlated.
* Created new features (likes per view, comments per view, shares per view) to better understand engagement rates.

**Key Insights**

* The dataset is nearly evenly split between claim (49.6%) and opinion (49.0%) videos, with ~1.5% missing data.
* Claim videos receive significantly higher engagement (views, likes, shares) than opinion videos.
* Banned authors are more likely to post claim videos and show much higher engagement per view, suggesting that their content may be more controversial or sensational.
* Engagement rate (e.g., shares per view) is consistently higher for claim videos regardless of author ban status.

**Next Steps**

* Incorporate natural language processing (NLP) to analyze video transcription text for claim detection.
* Use these insights to build a classification model to automatically distinguish between claims and opinions.
* Investigate whether highly engaging claim videos violate community guidelines or contribute to misinformation.
* Collect additional metadata (e.g., publish date, content topic) to refine analysis.

**Impact**

This analysis helps TikTok:

* Better understand how different types of content drive engagement.
* Identify which content may require greater moderation oversight.
* Lay the groundwork for automated claim classification using machine learning.
* Inform policy decisions on how to handle high-engagement content from banned or flagged users.