

Executive Summary: TikTok Claims Classification Data Analysis

1. Overview

This report presents an analysis of TikTok's dataset for claims classification, focusing on understanding the data structure, engagement trends, and key factors that influence claim classification and user interactions.

2. Data Summary

- **Dataset Size:** 19,382 entries
- **Missing Values:** Present in `claim_status`, `video_transcription_text`, and engagement metrics.
- **Claim Distribution:**
 - **Claim:** 9,608 entries (49.6%)
 - **Opinion:** 9,476 entries (48.9%)

3. Key Findings

- **Claims receive significantly higher engagement** (views, likes, shares) than opinions.
- **Banned users tend to have higher share counts**, indicating possible virality concerns.
- **Engagement metrics like likes/view and shares/view are higher for claims.**

4. Summary & Next Steps

- Investigate why **banned authors have high engagement**.
- Perform **sentiment analysis** on video comments.
- Explore **predictive modeling** to automate claim vs. opinion classification.

This summary provides foundational insights for further analysis and model development.

