# Executive Summary (Task 3)

## Overview

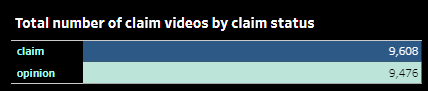
The data team seeks to develop a machine learning model to assist in the classification of claims for user submissions. In this part of the project, I need to analyze, explore, clean, structure prior to any model building.

The purpose of the exploratory data analysis (EDA) is to understand how different video engagement metrics vary based on video claim status and author ban status.

Tableau visualization: [click here](https://public.tableau.com/views/TikTokEDA_17223647820890/Dashboard1?:language=en-GB&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link)

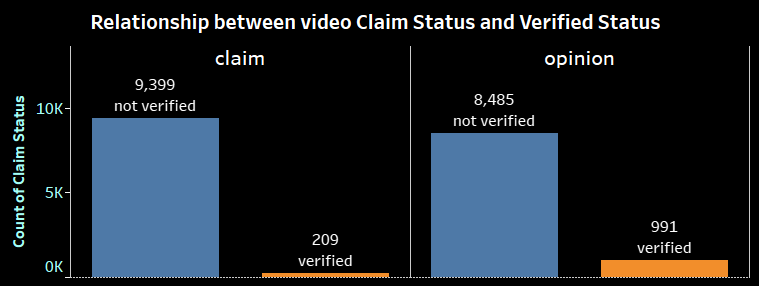
## Key findings

1. **There are more claim related videos than opinion related videos**



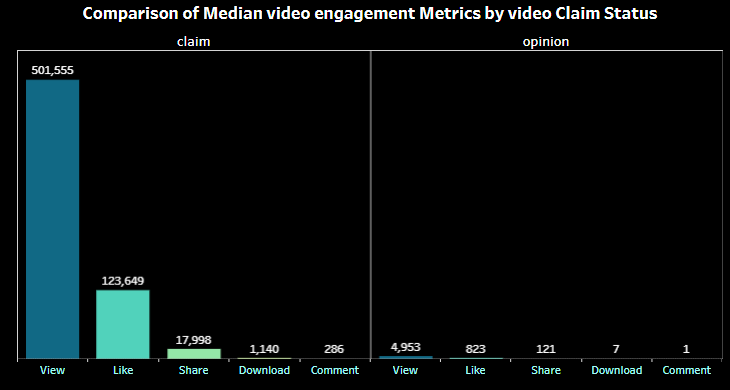
Impact: The future model will need to account for the imbalance in video counts as well as the null values.

1. **Claim related videos tend to have less verified authors than opinion related videos**



Impact: Most of the video authors not verified in both categories. However, claim related videos tend to have less verified authors than opinion related. This insight may help to determine our independent variables for the predictive model later on.

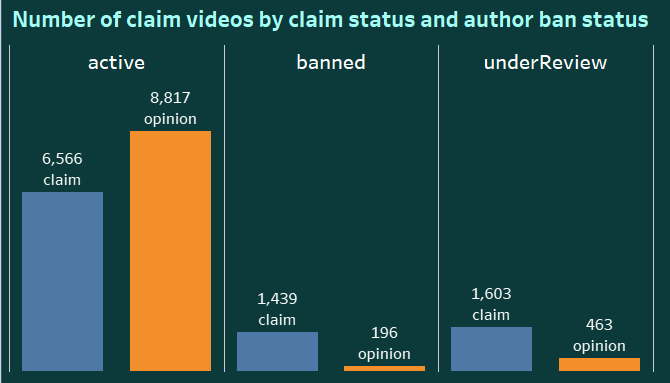
1. **Claim related videos have higher video engagement metrics than opinion related**



Impact: Claim related videos have higher median view, like, share, download, comment count than videos about author’s opinion. This insight may accurately consider higher views, likes, shares, comments associate with claim related videos.

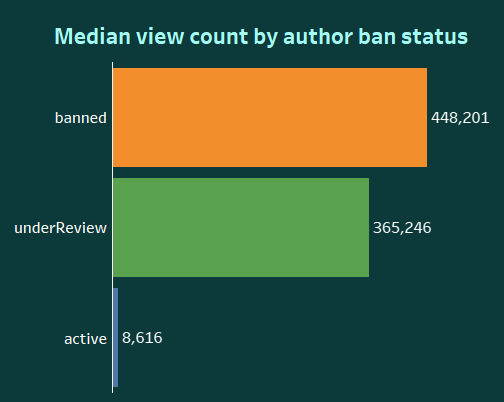
**About the author ban status**

After investigating the dataset, I find author ban status is one of the key independent variables to determine the video classification. By comparing various metrics grouping on author ban status, I found these insights:

1. **There are more claim related videos authors are being banned or under review than opinion related authors**

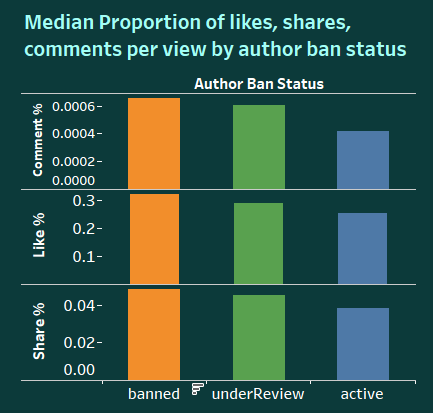
Impact: There are more claim related videos authors are being under review or banned than opinions authors. Surprisingly, there are more opinion related video authors are active than the claim category. The reasons I believe are because claim related videos can have consequences of user’s real life. Claims related to health, safety, or other critical issues are often under stricter scrutiny because they can have real-world consequences if they mislead users.

1. **Banned and under review authors have higher median view counts**



Impact: We can see that banned and under review authors both gain a significant number of views than active authors. I believe Banned or under review authors might have produced more controversial or sensational content that attracts a higher number of views. Controversial content tends to garner more attention, leading to higher view counts.

1. **Banned and under review authors have higher median proportion of likes, shares, comments per view**

Impact: Contents from banned and under review authors have higher users engagement than active authors. The rationales I believe are authors who are banned or under review may discuss polarizing topics that tend to divide opinions. This polarization can drive higher engagement as users like and share the content to express their views. This insight lays the foundation for content analysis, themes of the content posted by banned and under review authors. Identify any common traits that might explain the higher engagement rates.