

# Claim Classification Project

## Executive Summary - Preliminary Data Investigation

### Project Overview

TikTok seeks to develop an accurate predictive model that determines whether a video contains a claim or an opinion. The data team started by inspecting, organizing, and preparing the data for analysis. The following included some insights from the preliminary investigation.

### Key Insights

- The number of claims videos and opinion videos are fairly balanced.
- Claim videos in general generates more views than opinion videos.
- Videos from banned authors and those under review generates more views, likes, and shares than active authors.
- Engagement level seems to be strongly correlated with claim status and somewhat with author ban status.
- Relevant columns: claim\_status, video\_id, author\_ban\_status, video\_view\_count, video\_like\_count, video\_share\_count, video\_download\_count, video\_comment\_count

### Details

#### Number of Claim/Opinion Videos:

```
data['claim_status'].value_counts()
```

claim	9608
opinion	9476

#### Engagement Level per Claim Status:

	Average Views	Median Views
Claim	501029	501555
Opinion	4956	4953

#### Video Share Count per Ban Status:

author_ban_status	video_share_count_median
active	437.0
banned	14468.0
under review	9444.0

### Next Steps

After the preliminary investigation, the dataset is ready for exploratory data analysis.