

# TikTok Claims Classification Project

Data Discovery: An initial deep dive into drawing parallels and identifying patterns

## Overview

The TikTok data team aims to develop a machine learning model capable of classifying claims versus opinions for accounts reported for violating community standards.

## Problem

The backlog of user accounts waiting review exceeds the capacity for manual processing.

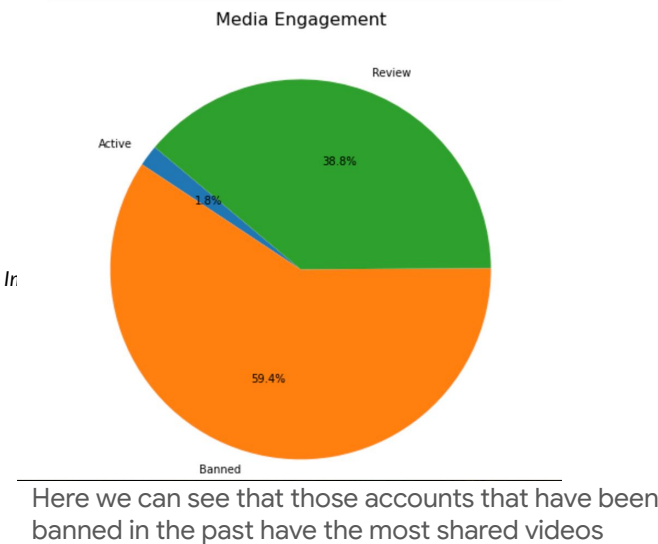
## Solution

The team is conducting an initial discovery phase to uncover insights and patterns within the data, which will be used to curate a dataset to train a ML model.. Concurrently, efforts are underway to prepare the dataset for detailed exploratory analysis.

## Details

- The variable ‘claim\_status’ seems to be the most useful to serve as our positive and negative examples
- The dataset is balanced giving us nearly a 50/50 split between claims and opinions
- Video engagement (likes, number of comments, number of shares) and author ‘banned status’, are other features correlated highly with claim status for classification

claims vs opinions:  
claim 50.35%  
opinion 49.65%



## Next Steps

The data team will take these initial insights to conduct a more thorough analysis of the data and prepare it for statistical analysis..