TikTok Executive Summary

Milestone 2 for TikTok Classification project

Overview

TikTok requires a machine learning solution that will assist in the classification of user submitted claims.

Objective

The objective of this section is to prepare the data for EDA and gain some simple insights into things such as structure of the data and the type of data that is available currently.

Results

The given data has shown that there is a similar proportion of claims and opinions with there being 50.3% claims and 49.7% opinion.
Further inspection highlighted a long right tailed distribution for video engagement counts indicating presence of outliers on the high end.
Some important columns are video_duration, video_view_count and video_transcript_text which can be useful to consider for the model.
There is some correlation between author_ban_status and the type of claim as under review authors are only assigned to claim under the given data.

Next Steps

For the next steps the team should being Exploratory Data Analysis on the data to find any meaningful trends and perform some statistical inference.