Executive Summary

Exploratory Data Analysis (EDA)

Overview

The goal of this EDA is to find what variables are related to the classification of claims. The relevant variables were engagement metrics and the ban status of the author.

Problem

There are many videos that require processing. Data can be used to create a machine learning model that will improve efficiency in the classification of claims.

Solution

Determine relationships between variables that can be used when developing the machine learning model.

Details

As can be seen in the table on the right, videos from banned authors regularly outperform those from active authors and authors under review.

It was also found that videos categorized as "claim" outperformed videos classified as "opinion"

	video_share_count
author_ban_status	
active	437.0
banned	14468.0
under review	9444.0

In this EDA, I built a data frame from the raw CSV file, evaluated the data for missing values, and determined the data type in each column. Lastly, descriptive statistics were gathered to summarize the data.

Next Steps

These findings indicate that the team should utilize engagement metrics such as views, likes, comments, and shares along with the author's banned status to develop the machine learning model.