Executive summary

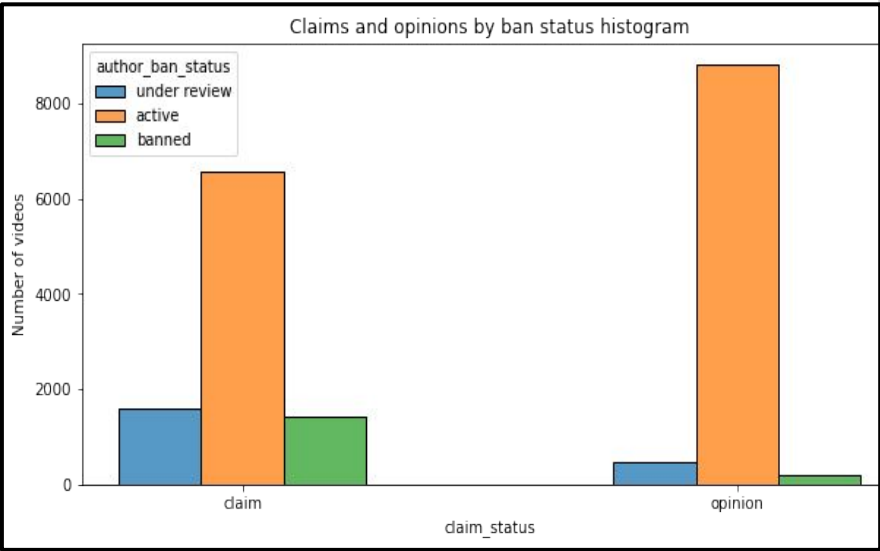Milestone 3 of the TikTok Claims Classification Project

Project Overview

Create a predictive model that can determine whether a video contains a claim or offers an opinion. In this phase, we conducted an exploratory data analysis on the dataset.

Details

Key Insights

- The number of not-verified users is really high on the platform. Claim videos are more likely to come from a not-verified user. Nevertheless, it seems that there's no relation between being a not-verified user and uploading claim videos.

- Active users are the majority on the platform, regardless of the video's nature. Anyway, claim videos are more likely to come from banned or under-review users.

- Banned authors are the most watched on the platform, followed by under-review authors.'

- 99% of the views come from claim videos.

Claims and opinions by ban status histogram

Next Steps

- It's necessary to make a decision about the outliers; my recommendation is to leave them. That will help the machine learning model understand the natural behavior of social media and viral videos.
- Inform the stakeholders about the state of the data.