Claim Classification Project

Executive Summary - Exploratory Data Analysis (EDA)

Project Overview

TikTok seeks to develop an accurate predictive model that determines whether a video contains a claim or an opinion. In this part of the project, the team explored, structured, cleaned, analyzed, and visualized the data to help us better understand the data.

Tableau Public Story: Link

Key Insights

- Almost 300 null values in 7 columns.

- Dataset is heavily right skewed.

- Verified users are more likely to post opinion.

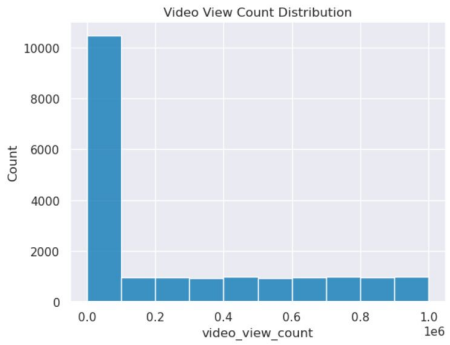
- Users who post claim are more likely to be under review or get banned.

- View count can be a good indicator of claim status.

- 98-99% of views are dominated by claim videos even though the number of claim and opinion video are about the same.

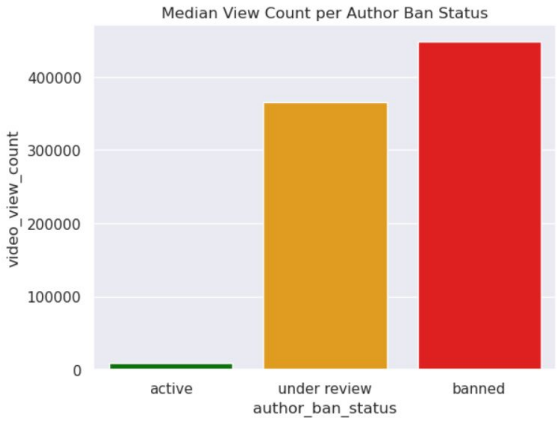
- At least 2000 outliers in the video statistics columns (views, like, share, download, and comment count).

Details



The distribution for view count is heavily right skewed. This is the case for like, share, download, and comment counts too.

The median view count for active authors is significantly less than non-active authors.



Next Steps

After exploratory data analysis (EDA), the next step is to test the relationship between important variables.