

TikTok claims vs opinions project

Exploratory data analysis (EDA)

Overview

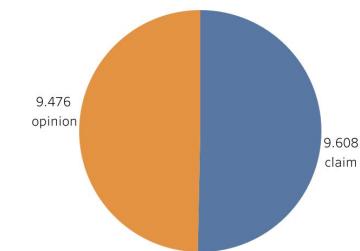
The objective of this project is to develop a machine learning model to classify the videos uploaded in the platform into claims and opinions.

Objective

In this stage of the project, we performed EDA to explore, clean, structure and analyze the dataset containing different metrics (columns) for each video (rows). The results from this analysis will provide insights which will be considered during the development of the machine learning model.

Results

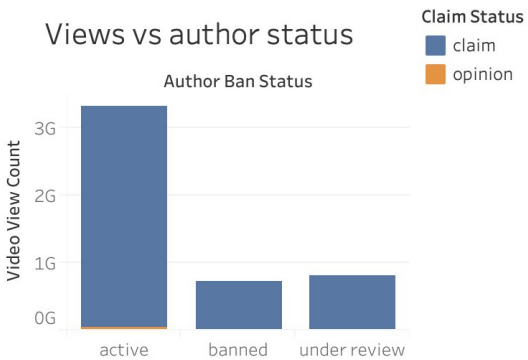
Number of claims vs number of opinions



We found null values in the dataset.

To sum up, the analysis of the distribution of the engagement metrics (views, likes and comments), reveals most of the videos in the platform have reduced engagement, whilst a smaller group of videos (viral videos), achieve much higher engagement.

Furthermore, most of the views generated by opinion videos belong to active users, while claim videos are distributed between active (1st), under-review (2nd) and banned (3rd) users.



Next Steps

According to the evidence gathered via EDA, the machine learning model will have to account for the null values and for the imbalance between claims and opinions with respect to the user status to produce accurate predictions.