TikTok Claims Classification Project – Exploratory Data Analysis

Executive Summary Report II

Milestone 3

ISSUE / PROBLEM

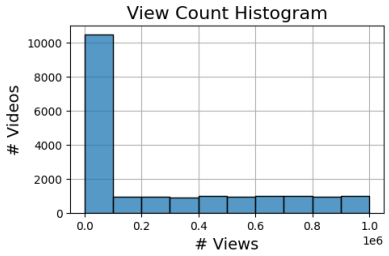
The goal of the TikTok data team is to build a machine learning model to classify user-submitted claims. At this stage of the project, the data needs to be explored, cleaned and structured before moving forward with model development.

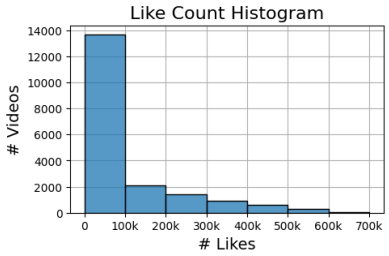
IMPACT

The analysis revealed several key factors that will need to be considered in the classification model, including missing data and an imbalance between the number of views and likes for claims and opinions. Additionally, the majority of videos fall at the lower end of the range for key engagement metrics like view count, like count and comment count.

RESPONSE

Exploratory data analysis was conducted to understand how TikTok videos impact user engagement. This involves analyzing variables like view counts, like counts and comment counts to identify user interactions with the content.





KEY INSIGHTS

- Further analysis is needed to determine the cause and impact of null values.
- Most videos have fewer than 100,000 views, with a uniform distribution for videos that exceed that number (left graph above). Most videos also receive fewer than 100,000 likes and 100 comments, with a heavy right skew in the data (right graph above).
- View and like counts for opinion videos are heavily skewed, with most values clustering below 1,000 (graph below). This right-skewed distribution will inform the model selection.