

TikTok Insights: Trends in Engagement Activity by Claim Status

A deeper dive in exploratory data analysis

Overview

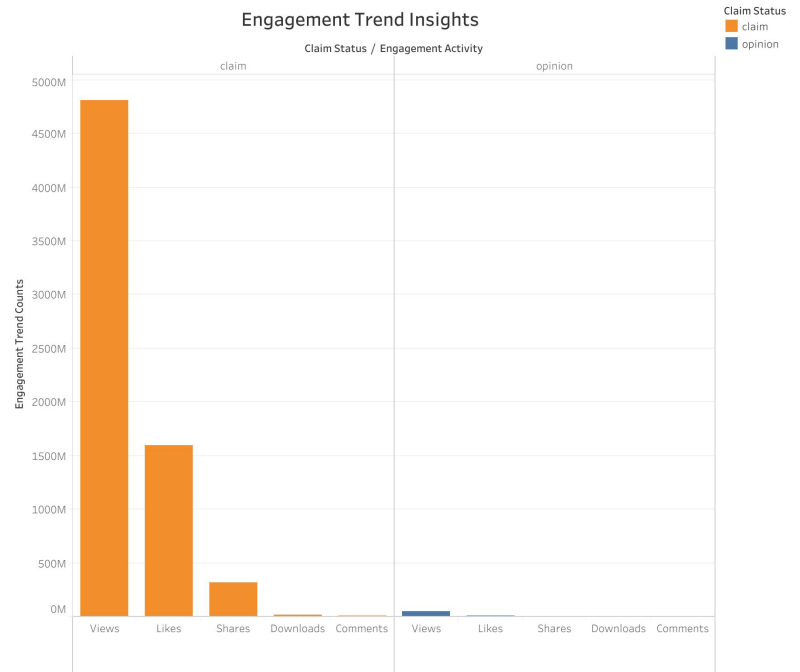
TikTok aims to build a Machine Learning Model to predict when a user is making a claim or an opinion to assist in handling the current backlog that keeps growing. Initial patterns were identified in the previous step and a more thorough analysis of each attribute is necessary.

Objective

Thoroughly review each attribute. Identify and isolate variables that are indicators of claim status and run a statistical analysis on them. Start preparing the dataset for modeling.

Results

- Claims and Opinion statuses are equally represented in the dataset (number of examples), but the actual values of each of the activities within each status are drastically disparate and will affect the model.
- Overall, engagement activity is higher for accounts making a claim.
- Number of Views and Number of Likes were identified to be the mostly likely indicators of Claim Status
- Outliers and Missing Data was Present.
- At this time, missing data will be reassigned, outliers will be walled (and should not affect modeling performance).



Next Steps

- More Analysis is needed to compare engagement activity against author_ban_status and verified status, to uncover any important correlations there.
- Path forward: Prepare the dataset for regression modeling and consider one-hot encoding certing variables.