

TikTok Claim Classification EDA Report

Findings from Exploratory Data Analysis

Overview

We found that the difference between claims and opinions is almost a 50/50 split, and that most opinion claims come from lower viewed videos while claim claims come from a wide range of popularity.

Problem

TikTok tasked us with creating a machine learning model that can distinguish between opinions and genuine claims in their report system.

Solution

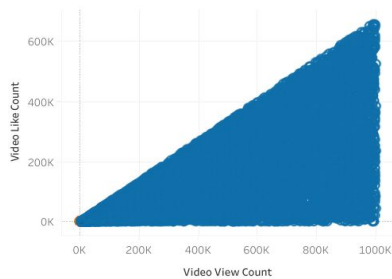
To do this we explored the data and have an idea of which data is the most relevant into distinguishing between the 2 and what to train the model on.

Details

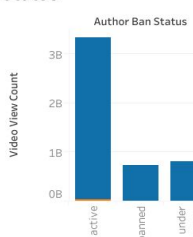
While most data is right skewed, there is a clear difference between opinion and claim claims when it comes to engagement with most engagement coming from 'claim' status. Length distribution is the only one without skew and outliers.

TikTok Engagement Metrics

TikTok Views vs Likes



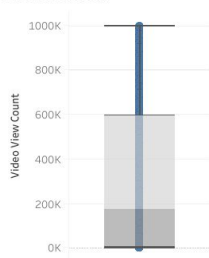
TikTok Author Account Status



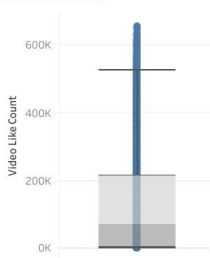
Claim vs Opinion



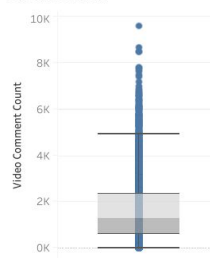
TikTok Video View Distribution



TikTok Video Like Distribution



TikTok Video Comment Distribution



TikTok Video Length Distribution

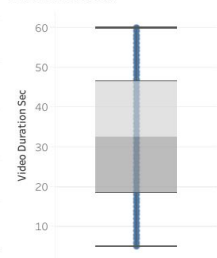


Image Alt-Text Here

Next Steps

Create a model using these insights that can accurately predict between claim and opinion