TikTok Submission Classification

Machine Learning Model Development

Project Overview

This project is to produce a machine learning model to classify all submissions on TikTok as 'claim' or 'opinion', currently we are exploring the data and beginning the exploratory data analysis .
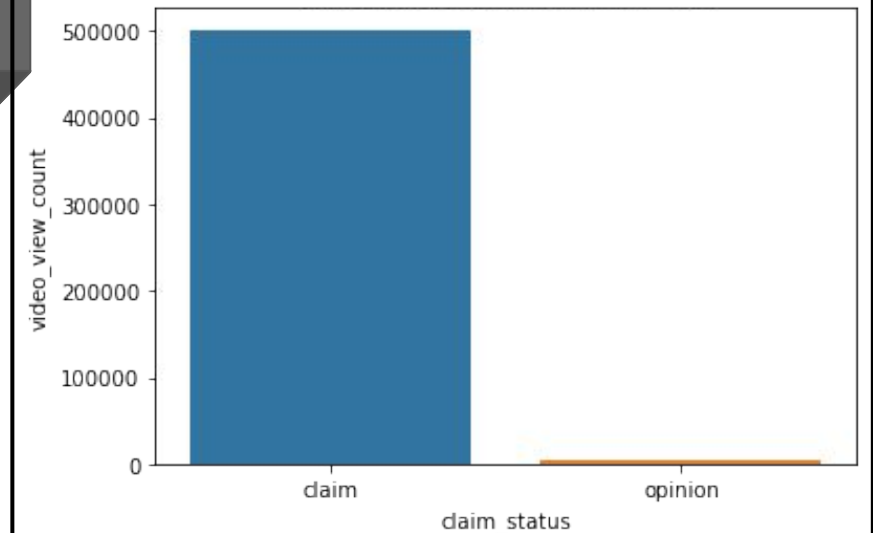
Key Insights

Details

Data Exploration

- Approximately 50% are claims and 50% are opinions, with less than 200 more claims vs opinions out of over 19000 rows of data.
- The claim category they are within a small deviation from each other. this is the same as with the opinion variable as well.
- most submitted videos for banned accounts are a claim,
- The author_ban_status shows likely correlation with the claim status. Additionally, view, shares, likes, and comments were all skewed towards videos with 'claim' and a 'banned' author, potentially indicating inflammatory content.
- Active/banned/under review all seemed to be very similar when looking at the video engagement for both a claim or opinion.

There is a clear skewing of view on videos that have been marked as being a legitimate claim, possibly indicating issues with the content.

Next Steps

- Further explore statistical values of data to determine correlation
- Create visualizations for clear decision making
- Collect more data for better training and testing
- Begin developing a machine learning model