Kevin Amaya - Junior Data Analyst

TikTok Claims Classification Project

10 May, 2025

Problem Overview

TikTok receives a high volume of reported videos, but not all can be reviewed manually. Videos that make claims are more likely to violate the platform's terms of service compared to opinions.

The goal

- build a machine learning model to classify videos as either claims or opinions.

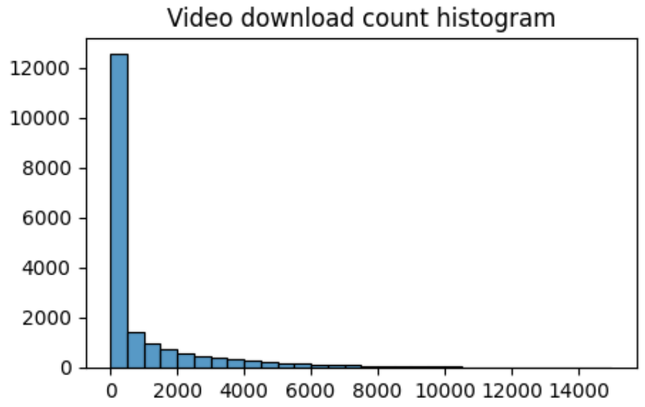
Data Exploration and Key Variables

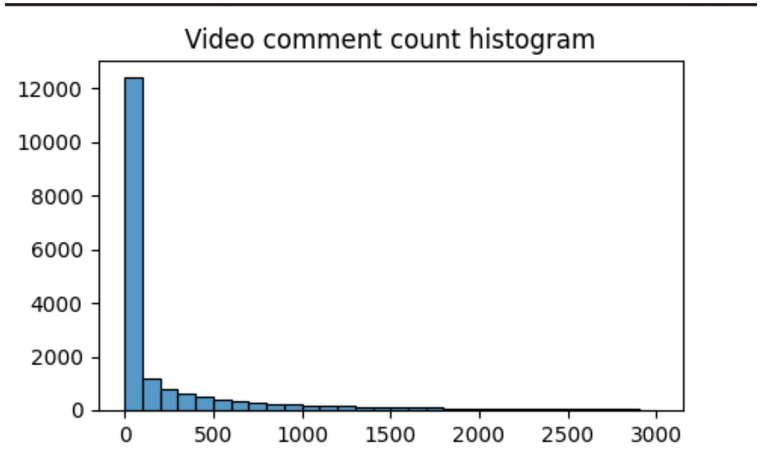
Initial Review

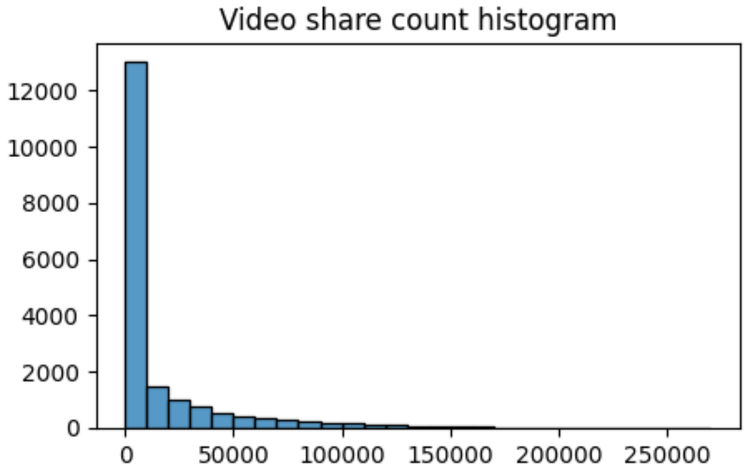
- Dataset was assessed, focusing on the claim_status variable (claim vs. opinion).
- Data was fairly balanced:
  - Claims: 9,608
  - Opinions: 9,476

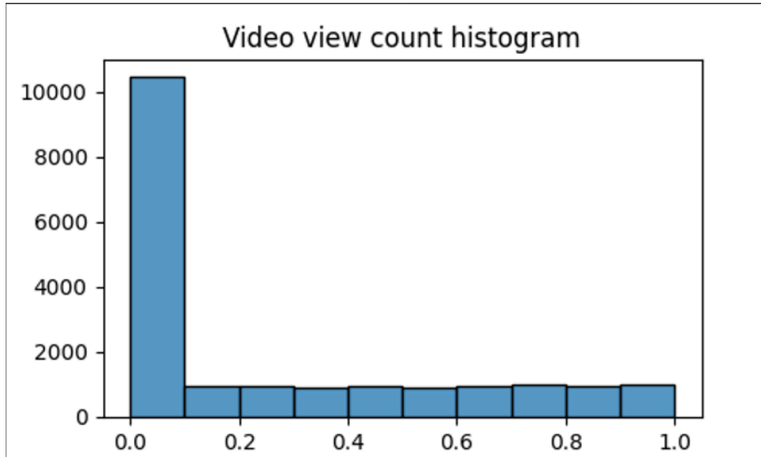
Engagement Metrics

- View counts were significantly different between categories:
  - Mean views (Claims): ~501,000
  - Mean views (Opinions): ~4,956
- Identified important predictors for future modeling:
  - video_duration, video_view_count, like_count, comment_count





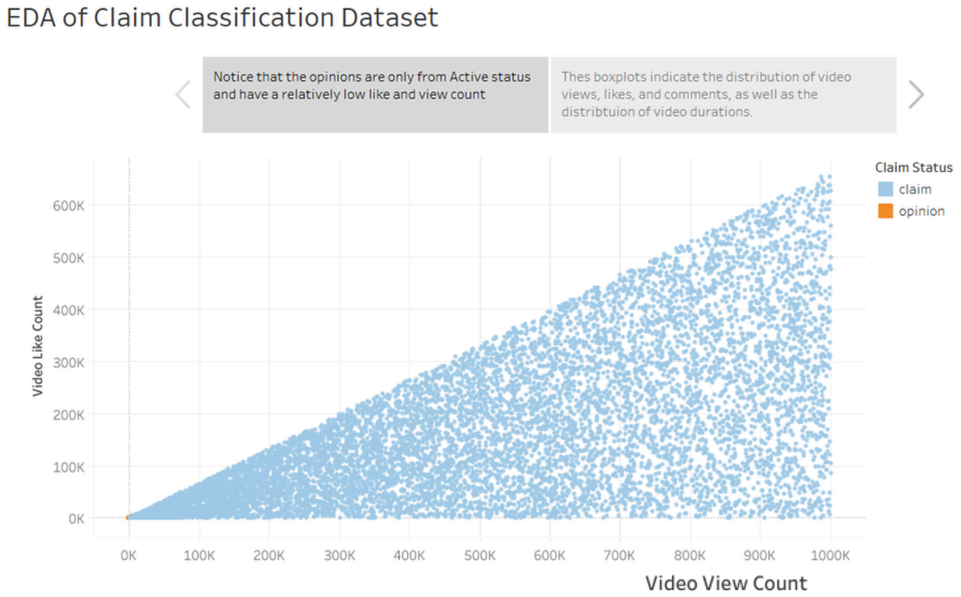


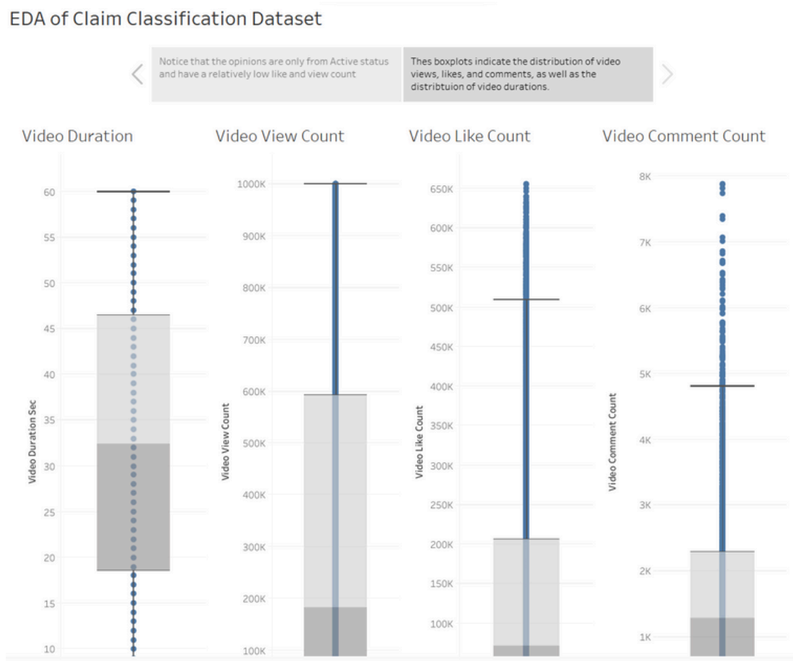


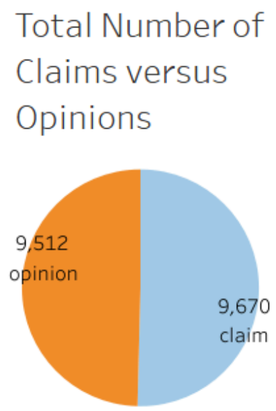
EDA Insights

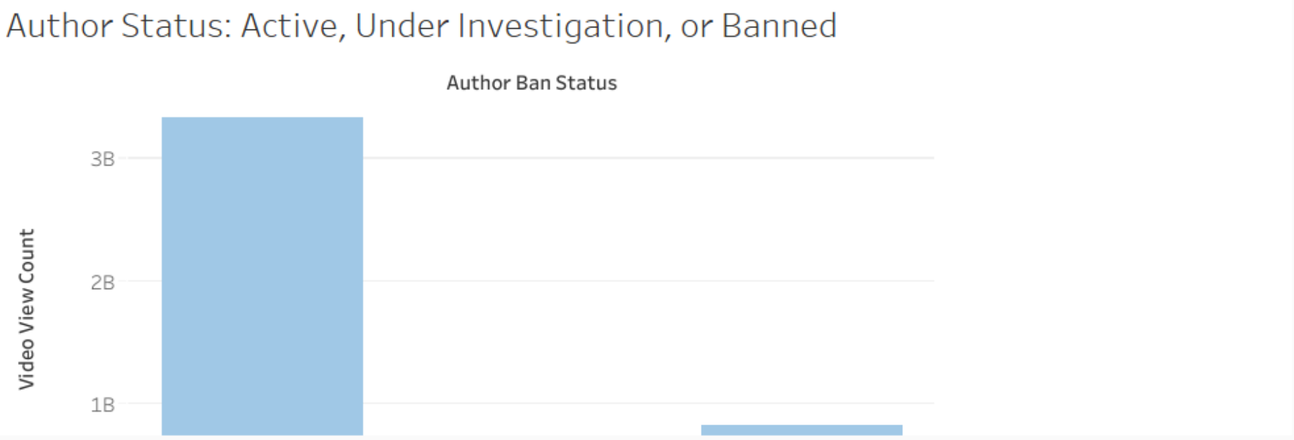
EDA Insights

- Tableau interactive dashboard
- Data was right-skewed: Most videos had low engagement.
- Over 200 null values across 7 columns — to be handled before modeling.









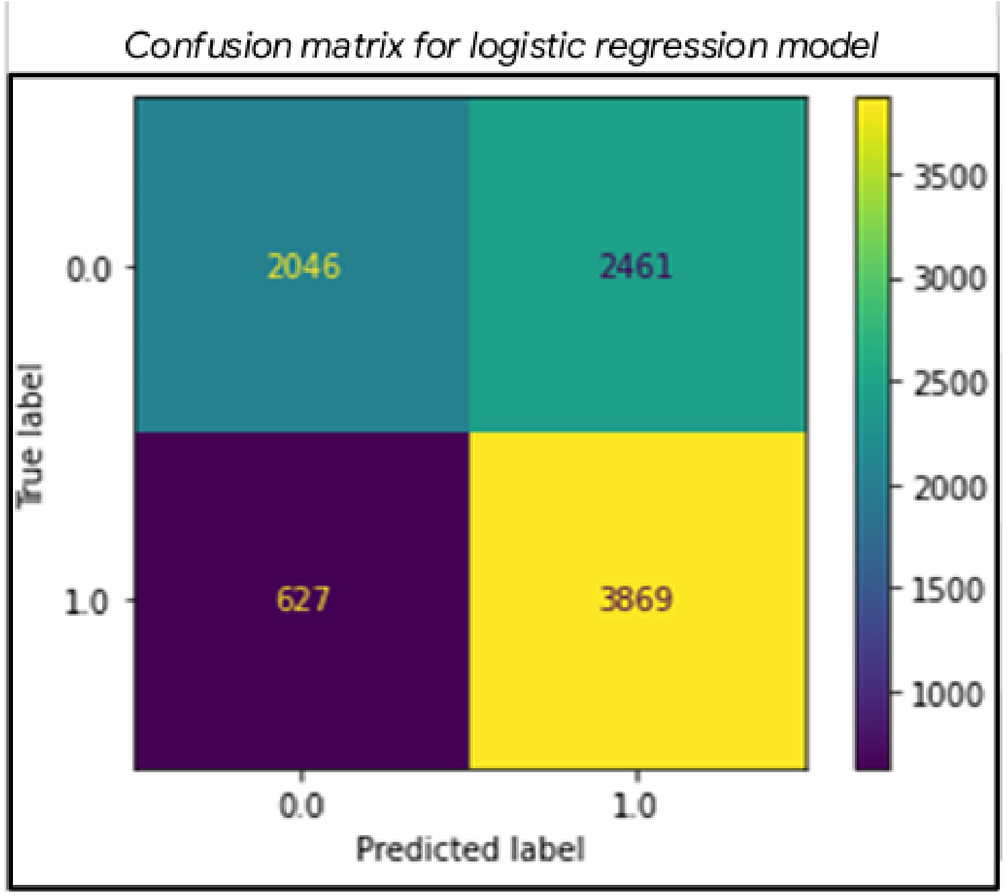
Statistical Testing

Hypothesis Test: verified_status vs. video_view_count

```
verified_status
not verified    265663.785339
verified         91439.164167
Name: video_view_count, dtype: float64
```

- Unverified accounts had higher average views (~265k) than verified ones (~91k).
- Two-sample t-test confirmed a statistically significant difference.
- Hypothesis: Unverified users may use more engaging or misleading content to boost views.

Regression Analysis



Upper-left: the number of videos posted by unverified accounts. Upper-right: the number of videos posted by unverified accounts. Lower-left: the number of videos posted by verified accounts. Lower-right: the number of videos posted by verified accounts

Objective

- To understand which factors predict verified_status.

Model Used

- Logistic regression chosen for binary classification.

Key Results

- F1 score: 66%, Precision: 69%, Recall: 66%
- Video duration was the strongest positive predictor.
- Other features had low impact on verification status.

Final Classification Model: Claim vs. Opinion

Models Evaluated

- Random Forest (RF)
- XGBoost

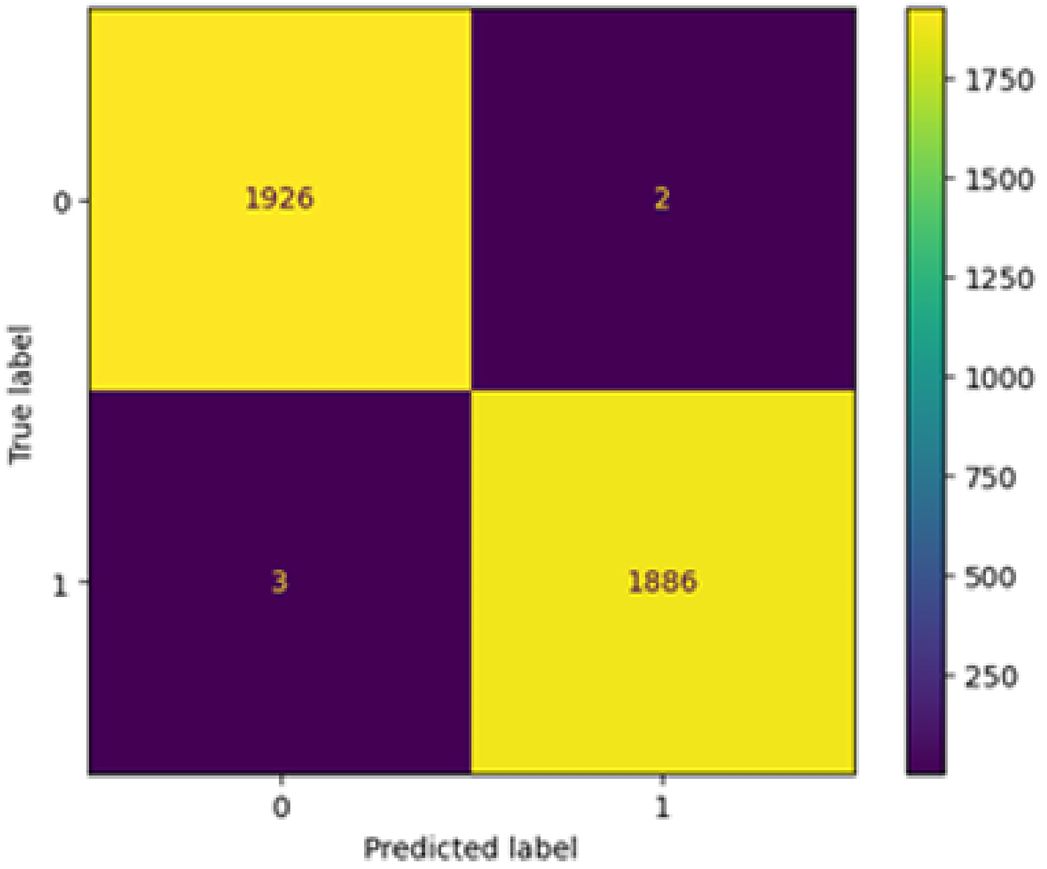
Model Selection

- RF model selected due to superior recall (0.995) on validation set.

Test Performance

- Only 5 misclassified out of 3,817 samples.
- Primary predictors:
- view_count, like_count, share_count, download_count
- Strong engagement correlates with videos being claims.

Confusion matrix for the champion RF model on test holdout data shows only five misclassified samples out of 3,817.



Impact & Next Steps

- The final model enables automatic prioritization of harmful content.
- Moderation can now focus on high-risk videos with greater confidence.

Thank You

- 587-893-7400
- kevin.amayav@outlook.com
- https://www.linkedin.com/in/kevin-amayav/