TikTok Dataset: EDA Summary

Analysing Viewership, Engagement and Content Claim Trends

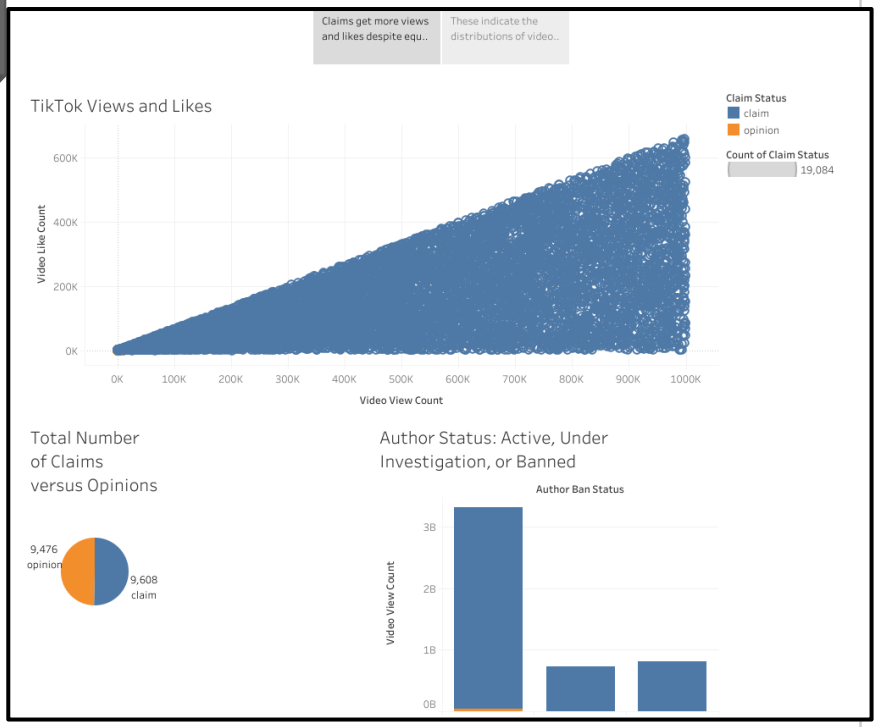
Project Overview

This exploratory data analysis used Python and Tableau to uncover trends in TikTok content based on claim status, user activity and video engagement metrics (views, likes, comments and duration). Visualisations such as histograms, boxplots, scatter plots and pie charts revealed patterns and outliers that inform how content performs and who produces it.

Key Insights

- **Claim vs Opinion:** Although claims and opinions are nearly equal in number, claim videos contribute over 99% of total views, highlighting their dominance in visibility.
- **Engagement Skewed:** Histograms and boxplots in Python and Tableau show high right-skew in views, likes and comments with extreme outliers.
- **Author Ban Status:** Banned and under-review users had higher median view counts than active users, which is counterintuitive and suggests potentially viral but problematic content.
- **Scatter Plot Patterns:** There's a clear positive relationship between video views and likes, especially for claim content. Opinion videos show lower engagement.
- **Duration Distribution:** Most videos are under 60 seconds, with a tight clustering around the max duration, especially evident in Tableau boxplots.

Details



Next Steps

- Apply log transformation or cap outliers to normalise engagement metrics for further analysis
- Dive deeper into why non-active users (banned/under-review) attract high view counts.
- Present visuals tailored to non-technical audiences (via Tableau Dashboard) and technical teams (via Python Notebook).