Executive Summary Report

TikTok Claims Classification Project

ISSUE / PROBLEM

The TikTok data team aims to develop a predictive model to determine whether the video contains claims or offers an opinion . To begin, the data team has organised raw data for future EDA.

RESPONSE

The initial data inspection of data was done to understand the relationship between variables.

As to classify any video, the count of 'claim' and 'opinion' category are taken into consideration to understand the type of video content in the data.

IMPACT

The preliminary inspection helped to understand the impact of user videos . Variables like 'video_view_count ' and 'claim _ status' can be considered in the future when making a classification model.

UNDERSTANDING THE DATA

After reviewing the dataset, the variable claim_status seemed particularly useful, given the client's proposed project . The following ScreenShot shows an important point of analysis to understand the claim_status variable .

```
tiktok_df['claim_status'].value_counts()
```

```
claim_status
claim      9608
opinion    9476
Name: count, dtype: int64
```

NOTE: The count of each claim status is quite balanced. There are 9608 claims and 9476 opinions.

ENGAGEMENT TRENDS

The data team considered the viewer engagement with each video in claim and opinion category . In order to understand the viewer engagement, the data teams considered the view count . The mean and median view count of each category highlighted the association between videos and claim counts .

CLAIMS

	video_view_count
mean	501029.452748
median	501555.000000

OPINIONS

	video_view_count
mean	4956.43225
median	4953.00000

KEY INSIGHTS

There is an equal balance of opinion and claim categories . With this understanding, we can proceed further with our analysis, knowing that the dataset has a fair balance of both categories .

NOTE: The ScreenShot provides count of claim_status category.

```
claim_status
claim      9608
opinion    9476
Name: count, dtype: int64
```