

Executive Summary

Milestone 2

ISSUE / PROBLEM

Tiktok seeks to use Machine learning model to pickout ideas within tiktok videos and comments and identify them either as claim or opinion

RESPONSE

The data team performed a preliminary analysis on the claim investigation tiktok dataset with goal of understanding underlying relationship between claims, opinions and the factors that can be possibly influencing them

IMPACT

The result of this preliminary analysis can be helpful later when performing exploratory data analysis.

UNDERSTANDING THE DATA

Below attached screenshot shows the number of claims and opinions in the given data set

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

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

It shows approximately the contain 50% of both the values

To understand the user engagement with videos marked claims and opinions , an analysis was done and it revealed following information

Claim

Mean view counts : 501029.45

Median view counts : 501555.0

Opinions

Mean view counts : 4956.43

Median view counts : 4953.0

KEY INSIGHTS

- There is a near equal balance between claims and opinions in the given data set. This can help in our future analysis since we know our data set has equal shares of both claims and opinions
- Now since we have identified our key variable **claim_status** we can proceed with exploratory data analysis