**Executive Summary**

TikTok Claims Classification Project | milestone 2

**ISSUE / PROBLEM**

**UNDERSTANDING THE DATA**

After reviewing the provided dataset, the variable ‘claim\_status’ seemed particularly useful, give. The client’s proposed project. The following screenshots show important points of analysis required to understand the ‘claim\_status’ variable.

Note: The counts of each claim status are quite balanced. There are 9,608 claims and 9476 opinions.

We are developing a machine learning model to assist in the classification of claims for user submissions on behalf of TikTok. To begin, the data team needs to organize the raw dataset and prepare it for future exploratory data analysis.

A screenshot of a computer

Description automatically generated

**RESPONSE**

The data team performed a preliminary investigation of the claims classification dataset with the aim of learning important relationships between variables.

Given the ask for a classification of user claims, the data team looked at the counts of claims and opinions in order to understand the count of each type of video content.

**ENGAGEMENT TRENDS**

The data team considered viewer engagement with each video in the claim and opinion categories. In order to understand viewer engagement, the data team considered the view count. The mean and median view count show the impact of each category of video; specifically, the mean and median view counts for both categories show the association between content (claim or opinion) and the video views.

Claims:

Opinions:

**IMPACT**

The impact of this preliminary analysis will be evident in the next steps. In order to understand the impact of user videos, the data team identified two important variables to consider. The variables ‘video\_duration’ (in seconds) and ‘video\_view\_count’ are both important factors to consider for future prediction models.

A black text on a white background

Description automatically generated



A pie chart with text and a blue circle

Description automatically generated

**KEY INSIGHTS**

* There is a near equal balance of opinions versus claims. With this understanding, we can proceed with our future analysis knowing that there is a fairly balanced number of claims and opinions for the videos included with this dataset.
* With the key variables identified and the initial investigation of the claim’s classification dataset, the process of exploratory data analysis can begin.