

# TikTok Claims Classification Exploratory Data Analysis (EDA)

## Executive Summary

### ➤ ISSUE / PROBLEM

An EDA needs to be conducted on a large amount of TikTok data to help prepare for potential future machine learning models to classify claims from TikTok users. The data needs to be analyzed, cleaned, structured and explored thoroughly prior to the creation of any models.

### ➤ RESPONSE

The exploratory data analysis was completed on the dataset with the purpose of having a deeper understanding of the effect TikTok videos were having on the viewers (TikTok users). Seeing as this was the main purpose of the EDA the variables that were analyzed included those that can easily be linked to TikTok user engagement.

The analyzed variables:

- Views
- Likes
- Comments

### ➤ KEY INSIGHTS

Upon completion of the EDA on the dataset we were able to identify two key insights.

#### Skewed Data:

The data is right-skewed as can be easily seen in the graphics provided. This will have to be considered when choosing which models to build.

#### Null Values:

There are 298 null values present in 7 of the columns. These values need to be analyzed further to determine why they are null as well as their potential impact on future models.

### ➤ IMPACT

According to the completed EDA of the TikTok dataset any claim classification models that will be built will need to account for a few things. Firstly, the very right-skewed data and finally the imbalance in opinion video counts, secondly, the null values that were present in the dataset.

