TikTok Claims Classification Project

Stage 4 - Executive summary

ISSUE / PROBLEM

The TikTok data team seeks to develop a machine learning model to assist in the classification of claims for user submissions. In this part of the project, it will be analyzed if there is a relationship between the verified status of the accounts and the number of views.

IMPACT

Now, we know there is a statistical difference between the two types of verified statuses in terms of the number of views. This brings to the table many questions about the platform and how the ML model should be conducted in the upcoming stages. Another possibility is to adapt the model to unverified users and make it more accurate in predicting when these types of accounts are more likely to upload a claim video.

RESPONSE

A two-sample test with a significance level of 5% was conducted in order to check whether there is a statistically significant difference between verified and unverified accounts in terms of views.

KEY INSIGHTS

1. The two-samples hypothesis test conducted shows that most of the views come from unverified accounts. As a matter of fact, there is a significant statistical difference between the two types of accounts.

2. Another approach was to get the means for the two types of accounts. The unverified accounts have 265663.78 views on average. Meanwhile, the verified accounts have a 91439.16 average (these means were obtained after missing values).

As we can see, both approaches confirm the same, unverified accounts generate a higher number of views than verified accounts.