TikTok Data Project - Questions & Considerations

- What are you trying to solve or accomplish?

  The main goal is to reduce the backlog of user reports and prioritize them more effectively.

  Note. An user report is a submission by a TikTok user identifying a video/comment containing user claims. All these claims need to be reviewed by moderators.

- Who are your external stakeholders that I will be presenting for this project?

  The main teams working on this project are the data team and the Cross-functional team. I will report to them both. My explanation will be more technical as long as the data team is involved. Instead, to the cross-functional team I will present the final insights and model's highlights from a more business-oriented perspective.

- What resources do you find yourself using as you complete this stage?

  I will employ different Machine Learning models along with a regression model to classify data about claims.

  While developing the model, I'll turn to standard model developing techniques, as EDA, feature engineering, validation and model development and evaluation.

- Do you have any ethical considerations at this stage?

  The main ethical concern is to keep users making claims anonymous.

  Furthermore, claims coming from/to banned/ban-risk users will be examined thoroughly, since this can be the requests raising more concerns.

- Is my data reliable?

  Data is coming from TikTok itself. Therefore, It appears to be reliable.

  Note. Probably, there may be the necessity to clean data.

- What data do I need/would like to see in a perfect world to answer this question?

  Ideally, I would like to have data about 4 main things:

  The claim status: such as claim or opinion

  The video itself: such as the id, duration, view, comments, download, likes, posting hour, verification

  The video's topic: like the main topic, description, keywords

  The Author: unique id for the author, ban status, previous ban, amount of reported claims

- What data do I have/can I get?

  I can get different information about the claim, such as: status, video id, duration, transcription, views, like, comments and download and the author's ban and verification status.

- What metric should I use to evaluate success of my business/organizational objective? Why?

  The major metrics I should use to assess the business objective success/failure is the amount of claim that are actually recognized as claim. Indeed, the major risk is that some videos that are actually claims may pass as opinion and still be displayed on the platform.