Optional course in data science: Fairness in two-sided recommendation

Mauricio Byrd Victorica

February 1, 2022

Introduction

Fairness in machine learning systems has become a highly active area in the past few years, as the ethical concerns surrounding automated decision-making become more prominent and these systems become more commonplace. Recommendation systems are not the exception: they operate in different contexts, from seemingly innocuous content recommendation to more sensitive areas such as targeted job offers. Although most of the literature on recommendation fairness focuses on whether users receive fair recommendations, in the context of platforms that mediate between said users (or consumers) and producers (or creators) whose livelihood is directly impacted by the recommendation system in place, it becomes crucial to consider two-sided fairness. This project focuses on the specific case study of the video-sharing application TikTok and the perceptions of fairness in recommendation from content creators and consumers in the platform. In particular, the aim is to explore insights from online user reviews, not only from TikTok but also from Youtube, another prominent video-sharing application and website; the idea is to gain a broader overview with this comparative study format.

While previous work has displayed effectiveness in eliciting valuable and actionable user perspectives from online reviews, there are particular challenges when an elaborate pre-defined topic such as two-sided recommendation is the focus a priori as in this project. Moreover, the considerable volume of reviews for applications as popular as TikTok and Youtube is another prominent challenge, notably regarding the constraints it implies for performing manual analyses of the data. In this work data analysis was approached first through given metrics in the reviews themselves, i.e. user scores and "thumbs up" or "up-votes" for the reviews themselves. Afterwards more elaborate techniques as sentiment analysis and topic modelling were introduced to gain further insights: sentiment analysis (and emotion classification) is useful to expound on user scores, as the content of the reviews might not align consistently with the provided score. Statistical topic modelling was performed to investigate potential themes present in the data, and manual analysis for some selected reviews was also performed. As discussed ahead, this was helpful in revealing the main problems when eliciting users' perception of fairness in the thematically unstructured setting of self-motivated reviews, and most importantly, the analysis reveals how, under some considerations, TikTok users seem to exhibit a greater preoccupation with recommendation fairness, and how the “two-sided” aspect of the issue seems to prevail most on the TikTok platform as well (in the sense that users embrace the creator/producer role more often than on Youtube, where most reviews come from the point of view of consumers).

In the next section the literature review conducted is covered, followed by the different stages of the analysis of online reviews, from collection and pre-processing to topic modelling. Finally some discussion and concluding remarks are presented.

Literature Review and Related Work

Prior work has addressed fairness in recommendation systems, in particular, the work by Patro et al. [7] was a relevant reference early on for this project as it focuses specifically in the two-sided perspective, considering fairness among both consumers and producers, and between these groups. The authors provide a recommendation algorithm (Fair-Rec) with fairness guarantees for both consumers and producers, in terms of recommendation quality and exposure, respectively. Although algorithm design is beyond the scope of this project, the details of the two-sided recommendation fairness setup are nonetheless quite useful to frame the current work. While ensuring fairness in a consumer-centric or producer-centric way is not trivial, the problem becomes yet more complex when considering both sides jointly, as their utilities are at odds with one another; this was illustrated by their results in that naively improving exposure for producers by recommending the products with least exposure resulted in the worst utility for consumers, and conversely, naively recommending only the most relevant products achieved the best utility for consumers at the direct expense of utility and fairness among producers.

In general, plenty of work has also been devoted to analysing user reviews, of which two important references were covered in the literature review. In the work by Eiband et al. [3], user problem reports for “intelligent everyday applications” are extracted from online reviews through topic modelling and manual coding, and then a complementary user survey is conducted to elaborate on potential solutions to the extracted issues. In the work by Svikhnushina, Placinta, and Pu [9], online reviews for different chatbot applications were analysed using quantitative and qualitative methods in order to extract general user expectations for chatbots. These are both relevant antecedents due to the similarity with the practical setup of this project, although a notable difference is that in this case the issue (fairness in recommendation) is defined beforehand, while in these references the relevant issues addressed are extracted from the reviews themselves.

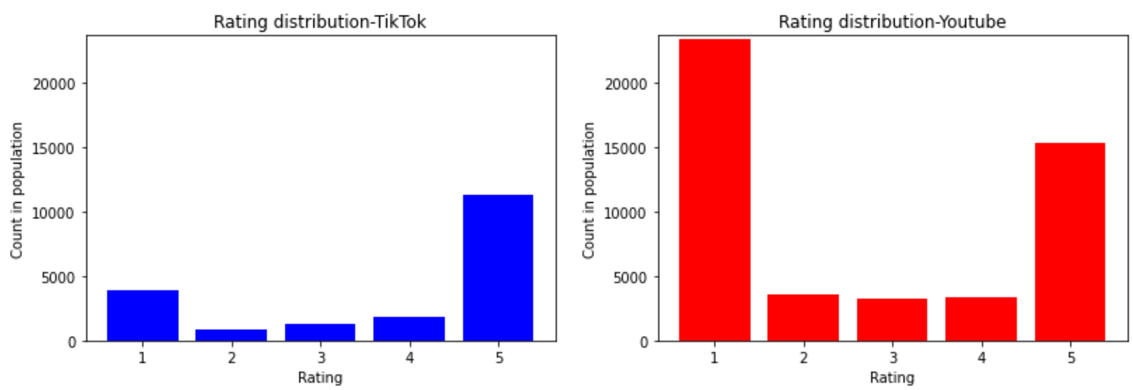
Review analysis

In this section the main results are presented. These consist of analysing the collected reviews in terms of metrics provided by the users regarding the approval of both applications and the support for the reviews themselves by other users. Sentiment analysis, emotion classification and topic modelling are also used to gain further insights into the perception of the users in both platforms. The main goal of the analysis is to uncover the main issues and specially whether or not fairness in recommendation is relevant to users and to what extent.

Data collection and analysis

Reviews from the Google Play Store for both platforms were scraped using the *google_play_scraper* package[1]. In both cases, reviews from November 2021 and corresponding to the US only were collected exhaustively. This resulted in a total of 359566 reviews for Youtube and 90241 for TikTok. To filter out empty or extremely short reviews, all reviews less than 40 characters long were dropped, as well as those not written in English (this last part was done using the *langdetect* package[2]). This resulted in a total of 53511 Youtube and 20937 TikTok reviews after filtering by length, which were further reduced to 48835 Youtube and 19244 TikTok reviews after applying language filtering. Besides the reviews themselves, each review has an associated score indicating the rating provided for the application by the author and a "thumbs up" count, indicating how many users found the review useful.

In terms of scores given by users for each app, the general distributions can be seen in Figure 1. As is common for online user reviews, the distributions are skewed towards extremely positive and negative cases [1]. In particular, the histogram of TikTok reviews clearly reveals the J-shaped distribution discussed by Hu, Zhang, and Pavlou [4], where most reviews are extremely positive, a considerable amount are extremely negative, and the amount for any scores in between is notably lower. However, for Youtube a surprisingly large quantity of extremely negative reviews is present, surpassing the extremely positive reviews while moderate (2-4 stars) reviews remain scarce.



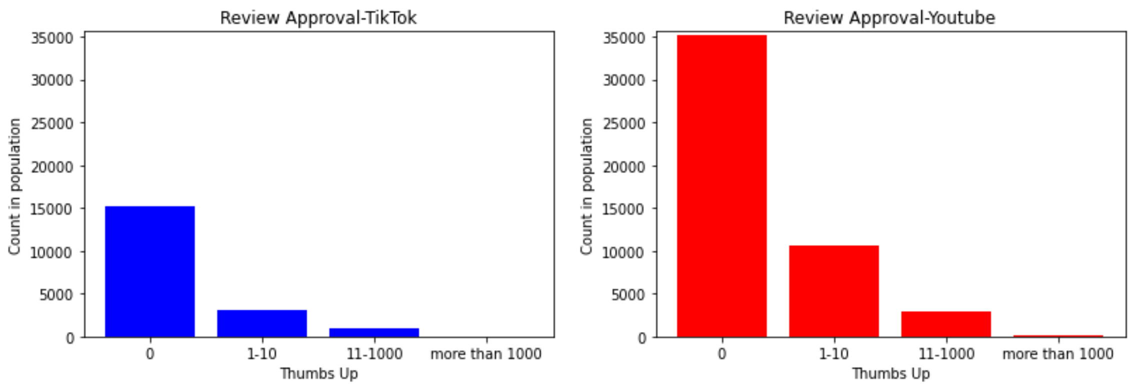
(a) TikTok. (b) Youtube.

Figure 1: Distribution of scores for both platforms.

Regarding approval by other users, both platforms show very similar distributions, which are presented in Figure 2: most reviews receive no attention at all, and those that do usually receive less than ten "thumbs up" or up-votes. Reviews with more than a thousand up-votes are very rare. In both cases, the review with most support is an extreme outlier: for Youtube, the most supported review received 103044 up-votes while next most supported received 13205 up-votes, and for TikTok the most supported review has 13651 up-votes while the second most supported received 3210 up-votes.

Taking a closer look at the reviews which receive the most support from users some details can be noted. For the top 1000 reviews, shown for both platforms in Figure 3, it can be seen

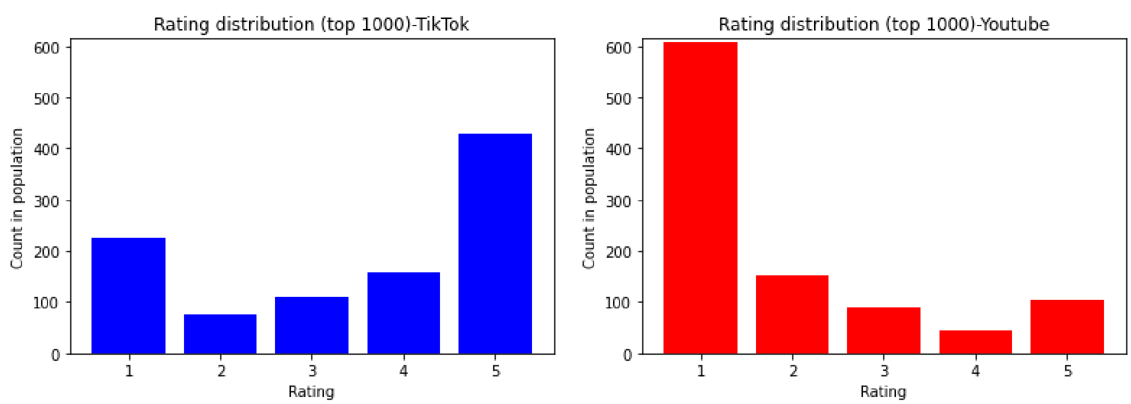
[1] https://github.com/JoMingyu/google-play-scraper
[2] https://pypi.org/project/langdetect/



(a) TikTok. (b) Youtube.

Figure 2: Thumbs up counts for both platforms.

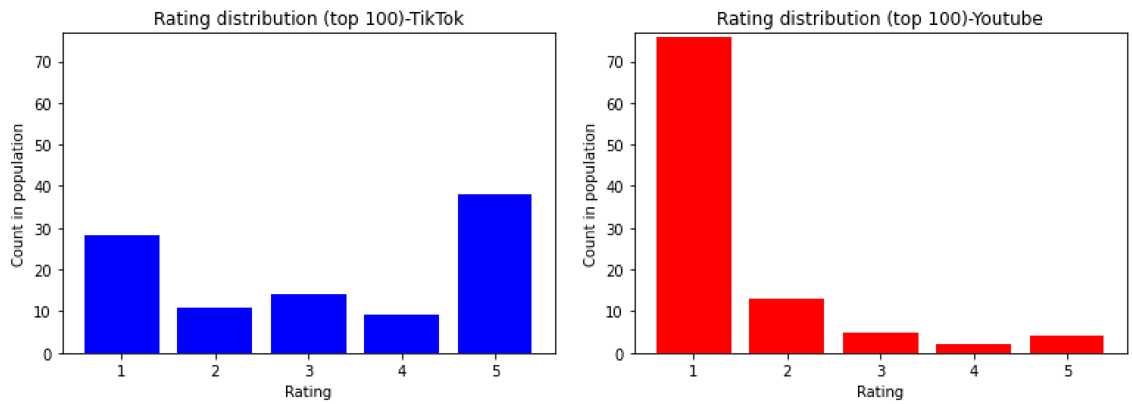
that while the J-shaped distribution persists for TikTok, the differences between each score category are attenuated; however, for Youtube the dominance of extremely negative reviews becomes even sharper, confirming users' agreement with the widespread negative reception during this particular time period. Further honing into the top 100 reviews (in terms of user approval) reveals how these contrasting trends continue, in that scores become yet more even for TikTok while the dominance of extremely negative reviews in Youtube is even more acute, as shown in Figure 4. To further study these trends, up-vote distributions considering only reviews belonging to each of the five scores were obtained for both platforms and are presented in Figures 9 and 10 in the appendix; these results also support what was observed in the rating distributions for most relevant reviews in Figures 3 and 4.



(a) TikTok. (b) Youtube.

Figure 3: Score distributions for the 1000 most approved reviews.

It is also of interest to look into the reviews with most support for both platforms. For Youtube, the most supported review gave the application the lowest rating and reads as follows: "*After that stupid Dislike button update, the YouTube app on all devices has been suggesting these terrible, irrelevant videos in mass quantities. It is actually ridiculous! Is this how it's going to be now? Or is the team at YouTube actually going to do their jobs now and manage suggested videos better? And will the long, repetitive ads be stopped? Or do they want their competitors to inherit their business, like Instagram, Reddit, Twitch, Odysee,*



(a) TikTok.

(b) Youtube.

Figure 4: Score distributions for the 100 most approved reviews.

Rumble, and SO many others?".

For TikTok, the most supported review also gave the lowest score and reads as follows: "*Love the app, have over 50k followers, but my post that went viral and promoted my small business was suddenly pulled for no reason and there's no way to contact the team about it*".

Something quite interesting is how the Youtube review is from a consumer's perspective, and actually has some connection to recommendation, since it mainly complains about how video dislikes are a useful indicator for content quality and its removal resulted in widespread bad recommendations. Moreover, this closely relates to the trade-off between consumers and producers in two-sided recommendation, since the removal of the dislike counts was not only to prevent harassment against creators, but also to "ensure small creators can thrive"[10] thus this relates directly to sacrificing recommendation quality for consumers as an effect of improving producer exposure. At any rate, this feature is external to the recommendation algorithm, so strictly speaking one can talk about recommendation fairness for consumers only in terms of the alleged "terrible, irrelevant" recommendations after the modification. On the other hand, the most supported TikTok review doesn't address recommendation, but interestingly it switches to the creator or producer perspective, which is reaffirmed as the reviews reveal this to be one of the main differences between the two platforms.

Sentiment Analysis

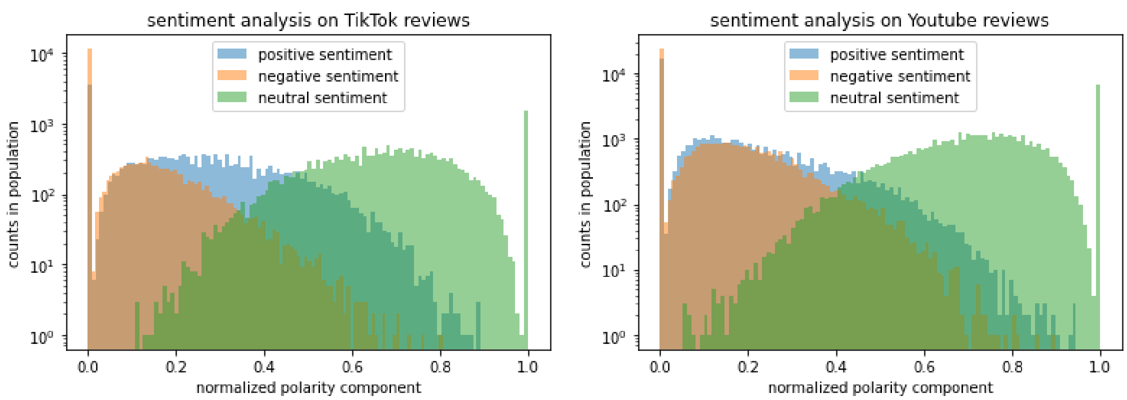
Although review scores are the primary feature regarding user approval (it is not uncommon for users to rate an application without elaborating on their decision through text), in this study more emphasis should be given to the information contained in the actual reviews, which does not always match the corresponding score intuitively; for example, one review for Youtube with the lowest score reads: "*I think that it is a very good app, it is one of the best video watching apps in the world*".

Using the VADER lexicon for sentiment analysis [5] through the *NLTK* package[3], the sentiment of the reviews can be analysed in terms of negative, neutral and positive normalized sentiment components, indicating with a real valued number between 0 and 1 the extent to

[3]https://www.nltk.org/api/nltk.html#module-nltk

which each review corresponds to each of the three considered sentiment types (plots using the more common single compound sentiment value in the range from -1 to 1 are included in the appendix). While VADER was not designed to deal with online reviews explicitly, its original focus was to process content from social media. Therefore, slang, abbreviations, contextual sparseness and text shortness were all design considerations for the VADER lexicon, which are an adequate match for the online review setting. A histogram for sentiment values in reviews of both platforms is provided in Figure 5, and the one corresponding to the 100 most supported reviews is provided in figure 6.

Sentiment analysis supports what was reflected by the score or rating distributions, as the ones for negative and positive sentiment components show similar trends: for both platforms the distributions for negative and positive components seem to decay gradually as their magnitude increases, however for Youtube the postive component has a sharp peak for values close to zero and then decays quickly, indicating that most reviews have low positive sentiment, while this component seems to have a more even distribution among TikTok reviews, indicating that most of them have at least some non-negligible positive component. The opposite can be observed for the negative component. Thus the trend for score ratings persists in terms of sentiment analysis, although the contrasts are not as stark as the ratings alone would suggest. The neutral component, in both platforms, seems to be distributed approximately normally around high values.



(a) TikTok.

(b) Youtube.

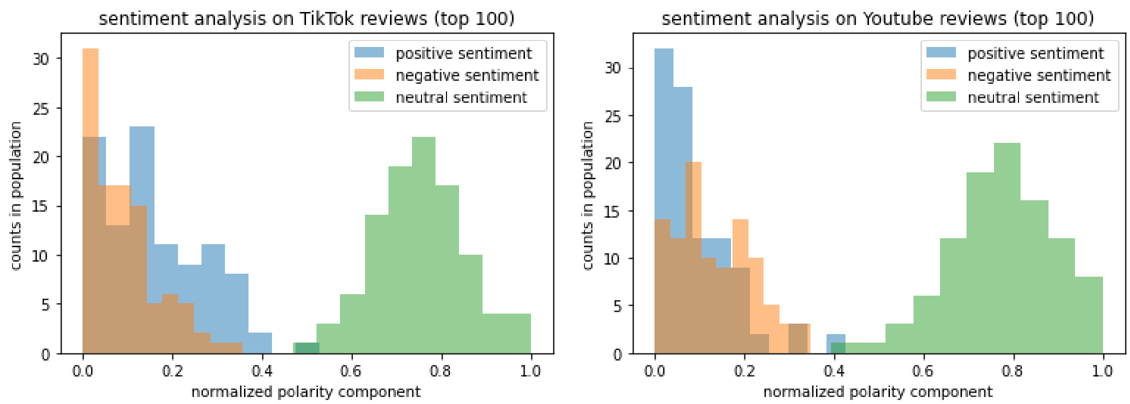
Figure 5: Sentiment components for all reviews.

Emotion classification

A more fine-grained approach to analyse the reviews is emotion classification. Using the NRC Lexicon [4] through the *NRCLex* package[5], words can be associated to 8 different emotions, and then a score for each of these emotions within a review can be computed according to the emotions of the words contained in it. It should be noted that this emotion lexicon might not be as good of a match as the VADER lexicon used for sentiment analysis: rather than

[4]https://nrc.canada.ca/en/research-development/products-services/technical-advisory-services/sentiment-emotion-lexicons

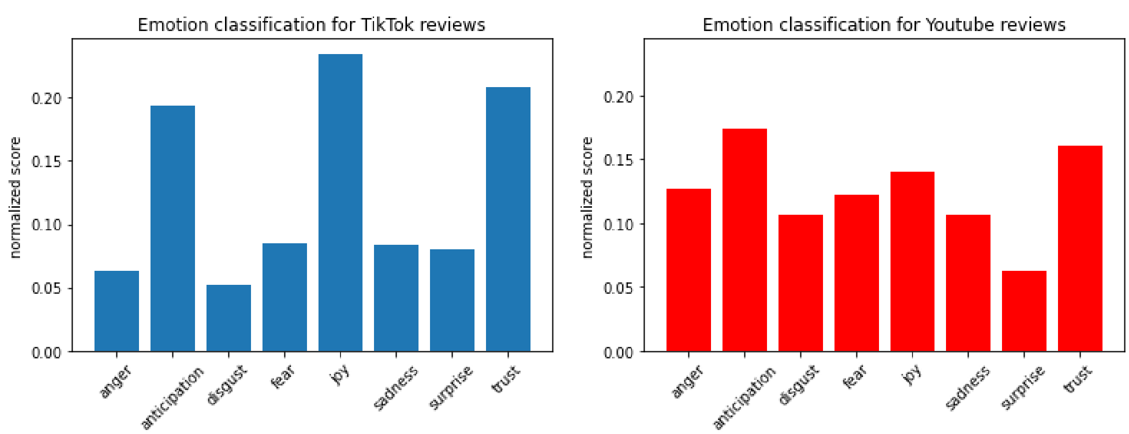
[5]https://pypi.org/project/NRCLex/



(a) TikTok. (b) Youtube.

Figure 6: Sentiment components for top 100 approved reviews.

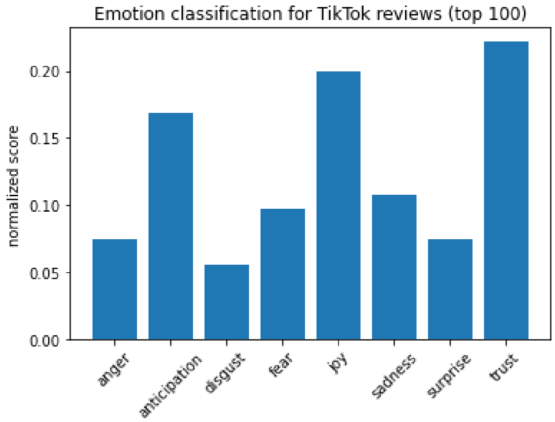
using data in a specific context like social media posts, this lexicon is based on terms from the Macquarie Thesaurus, and is thus centered in a vast collection of singular words and everyday expressions which do not capture the peculiarities of the online review setting [6]. The normalized frequencies for each of the eight emotions on reviews from both platforms is presented in Figure 7 for all reviews and in Figure 8 for the top 100 most supported reviews; In both cases, the scores for each emotion were summed across reviews, and then the eight resulting total scores were normalized by their sum.



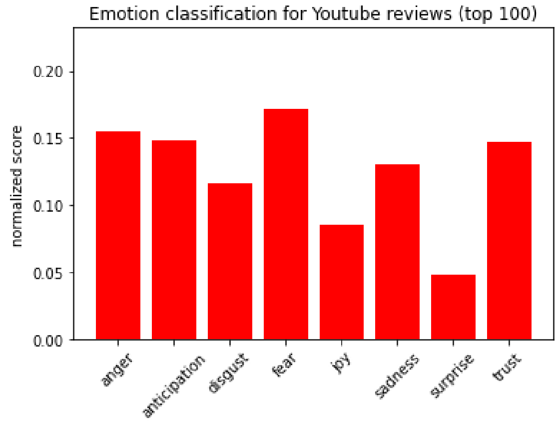
(a) TikTok. (b) Youtube.

Figure 7: Normalized emotion scores for all reviews.

Although these discrete emotions are not as straightforward to interpret as user ratings or even continuous sentiment components, it is clear that positive emotions like joy and trust dominate the TikTok reviews, while emotions like anger, disgust, fear and sadness become more prominent in the Youtube case (even though surprisingly anticipation, joy and trust are still the ones with larger scores). Honing into the top 100 reviews shows no major changes for TikTok, but for Youtube negative emotions seem to take over, such that fear and anger become the highest scored and joy drops sharply, becoming the second lowest scored emotion. So it would seem that this simple emotion classification also supports what was



(a) TikTok.



(b) Youtube.

Figure 8: Normalized emotion scores for top 100 most approved reviews.

seen in user ratings, including the previously observed increment in negativity for Youtube reviews when the focus is solely on the most up-voted ones.

Themes

Finally the themes emerging in the reviews were studied. Using both an automated approach using Latent Dirichlet Allocation (LDA) [2] and manual analysis of select reviews, the aim was to identify whether one or more groups of reviews addressed recommendation fairness. First, using the *gensim* library [8], LDA models were fitted on the whole body of reviews for each platform, varying the number of topics from 2 through 50 and evaluating coherence to determine the adequate value of this hyper-parameter. After settling on 31 and 28 topics for Youtube and Tiktok, respectively, topics were determined according to the ten most relevant reviews for each one.

The topics extracted showed significant overlaps: for Youtube 5 different topics focused on the removal of the dislike counts and another 5 all just revolved around praising the app; in general, none of these topics showed tendencies towards fairness in recommendation besides potential connections to the removal of the dislike counts as mentioned previously when analysing the most supported Youtube review. For TikTok, the situation didn't improve, having 6 different topics focusing on general praise for the app, multiple topics focusing on similar sets of technical issues, and so on. However, an interesting topic emerged, where reviewers were soliciting greater exposure or asking for their accounts to be "unfrozen", but in any case these all seemed to follow very closely a dull short template suggesting they are probably automated.

After hypothesising that the poor outcome for LDA topic modelling on the whole dataset might be due to the fact that most of the top reviews per topic were short and received little to no support, the same process was repeated taking into account only the 1000 reviews with most up-votes. In this scenario the range for the number of topics was taken from 2 through 20, and 13 and 11 topics resulted in the highest coherence for Youtube and Tiktok, respectively.

For Youtube, most topics overlapped even more than before, and the reason seems to be that many reviews with a lot of support cover simultaneously different aspects that matter to the population of users at large: all topics involved reviews covering different combinations of complaints about ads, complaints about the dislike count removal, and certain technical issues. Once again no topic seems to focus significantly on recommendation. For TikTok a similar trend can be observed as certain technical issues, complaints about the moderation policy, and general praise for the app are elements present for most topics; although recommendation does seem to be a larger issue in these topics, no comment beyond the quality of the recommendations is made, so it would seem that fairness in recommendation doesn't emerge as a relevant topic in either platform even after honing into the most up-voted 1000 reviews. All the topics and their most representative review are included in Tables 2 and 3 in the appendix.

Finally, the most up-voted 100 reviews of both platforms were explored manually. For Youtube, only 3 of the top reviews had strong links to recommendation, and for TikTok the number rose to 19. Moreover, the Youtube reviews are focused on the consumer side and only one addressed fairness. For TikTok, there are reviews from both sides: some discussing recommendation quality from the consumers' perspective and some addressing the exposure received from the creators' point of view. Interestingly, of these TikTok reviews on recommendation, the five of them which addressed fairness were from creators, claiming that the exposure they received was not fair compared to the exposure given to other creators. Some representative reviews are shown in Table 1.

Keywords	Review	Rating	Up-votes
Unfair exposure (Youtube)	*Search function is dishonestly skewed to show content that has revenue earning advertising. Users are not able to see results sorted by date and so content is usually too old to be relevant.*	1	1069
Unfair exposure (TikTok)	*TikTok is the best app,it keeps you busy when you are lonely I don't even know how to explain it. The only problem is when you put in all your efforts to make a video you hardly get likes or followers, but the videos that are lame are the ones that get likes and followers.*	5	516
Bad recommendations (TikTok)	*I like it... but recently there's been a lot of problems with my algorithm, no matter how many times I select not interested on a genre of videos, it continues to show similar videos. Another problem is that it will glitch out and I can't block people.*	3	317

Table 1: Some representative reviews related to recommendation fairness issues on the top 100 most up-voted reviews for both platforms. Note that some grammatical corrections were made to the original texts to make them clearer.

Discussion

There are relevant limitations in this analysis. Starting with the data itself, it was particularly clear that the recent update removing the dislike count in Youtube set the tone for the reviews at large, possibly obscuring other long-standing issues; although there are no indications of such a prominent recent issue for the TikTok reviews, the timescale limitation should always be kept in mind.The description of the data through user ratings, sentiment analysis and emotion classification didn't show large high-level discrepancies, and overall supported the dominance of negative reviews for Youtube and positive ones for TikTok, and also the strengthening of these trends in the pool of reviews with most user support. However, the potential mismatch between the NRC Affect lexicon and the online review content should not be disregarded, and it could be interesting to use better suited models, such as the NRC Hashtag Emotion based on Twitter data or perhaps the latent semantic analysis (LSA) approach proposed by Zhang et al. [12] in the context of online movie reviews.

Automated topic modelling through LDA was not successful in providing clear separations for the issues on both platforms, as a single set of relevant separate issues seem to spread across topics in both cases. Perhaps different approaches to pre-processing would be helpful, like the filtering out of positive reviews done by Eiband et al. [3] to avoid unhelpful reviews focused on praising the app (in this case it might also be helpful to filter out negative reviews considering their dominance in the Youtube review population) or doing the modelling with a technique better suited than LDA for this particular setting, such as GSDMM [11] which is intended for short text data like these online reviews. In any case, it can be said that recommendation was mostly absent from the most representative reviews for the majority of topics in both platforms, and fairness in recommendation was not present in any of them. However, analysing the most up-voted reviews for each application reveals that users are not indifferent to this topic (as shown before even Youtube's most up-voted review addresses recommendation quality).

What these top reviews also show is that TikTok reviews seem more concerned with recommendation than Youtube reviews, and some of these top reviews even address recommendation fairness, although only from the producers' point of view. It makes sense that fairness among producers is more prominent, since it is not as trivial for users to compare their received recommendations as it is for producers to compare their received exposure, in any case, analysing the precise statements is out of the scope of the current project. It might be useful to combine the automated topic modelling and manual analysis performed by using techniques like LDA or GSDMM while also taking into account the support of each review, which might provide a more complete set of themes that reflect the weight of these most up-voted reviews in terms of user perception.

Overall, without disregarding the aforementioned limitations, this comparative study suggests that fairness in recommendation is a relatively relevant topic for TikTok, and that online reviews reflect the two-sided nature of the issue by including the producer's point of view to a considerable extent in comparison to Youtube, where the most relevant reviews addressing recommendation tend to focus on the quality of recommendations received by consumers. These reviews might not be well suited to study fairness in recommendation for consumers on an individual scale, since consumers tend to approve or disapprove of the recommended content, but not in relation to what others receive (i.e. how fair are these recommendations, as good or as bad as they might be); however it could be interesting to explore consumer fairness in general by identifying the distribution of consumers approving and disapproving their personalized content feed.

A general analysis regarding the differences between consumers and producers is complicated because separating reviews from both populations is not trivial, and in the particular case of TikTok it is not uncommon for single users to take on both roles simultaneously. However, an interesting next step for Youtube would be analysing reviews from the Youtube Studio platform[6], which allows users to manage their channels, hence one could expect to get a population of reviews heavily skewed towards creators, in contrast to the reviews for the general platform used throughout the current study.

Conclusion

In this work the US Google Play Store reviews for video-sharing platforms TikTok and Youtube from November 2021 were analysed to gain insights into the issue of recommendation fairness as perceived by TikTok users and creators. After characterizing reviews in terms of user ratings, user support, sentiment analysis and emotion classification, reviews were further studied through topic modelling and manual analysis. While content recommendation did emerge as a perceived user problem in both platforms (albeit not a particularly dominant one), fairness in recommendation was more rarely addressed by the reviews in general. Overall, the topic was found to be more strongly perceived by TikTok users: interestingly, fairness in terms of exposure for producers was a noticeable topic in the most up-voted TikTok reviews, and it might be worth investigating further why this part of the two-sided fairness setup has such prominence in the platform compared to fairness among consumers or between consumers and producers. It is also worth noting that the results suggest that online reviews might not be a suitable source for important aspects of this problematic, such as fairness among consumers on an individual scale.

References

[1] Georgios Askalidis, Su Jung Kim, and Edward C. Malthouse. "Understanding and Overcoming Biases in Online Review Systems". In: *Decis. Support Syst.* 97.C (May 2017), pp. 23–30. ISSN: 0167-9236. DOI: `10.1016/j.dss.2017.03.002`. URL: `https://doi.org/10.1016/j.dss.2017.03.002`.

[2] David M. Blei, A. Ng, and Michael I. Jordan. "Latent Dirichlet Allocation". In: *J. Mach. Learn. Res.* 3 (2003), pp. 993–1022.

[3] Malin Eiband et al. "A Method and Analysis to Elicit User-Reported Problems in Intelligent Everyday Applications". In: *ACM Trans. Interact. Intell. Syst.* 10.4 (Nov. 2020). ISSN: 2160-6455. DOI: `10.1145/3370927`. URL: `https://doi.org/10.1145/3370927`.

[4] Nan Hu, Jie Zhang, and Paul A. Pavlou. "Overcoming the J-Shaped Distribution of Product Reviews". In: *Commun. ACM* 52.10 (Oct. 2009), pp. 144–147. ISSN: 0001-0782. DOI: `10.1145/1562764.1562800`. URL: `https://doi.org/10.1145/1562764.1562800`.

[6] https://play.google.com/store/apps/details?id=com.google.android.apps.youtube.creator&hl=en&gl=US

[5] C.J. Hutto and Eric Gilbert. "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text". In: Jan. 2015.

[6] Saif Mohammad and Peter D. Turney. "Crowdsourcing a Word-Emotion Association Lexicon". In: *CoRR* abs/1308.6297 (2013). arXiv: 1308.6297. URL: http://arxiv.org/abs/1308.6297.

[7] Gourab K. Patro et al. "FairRec: Two-Sided Fairness for Personalized Recommendations in Two-Sided Platforms". In: *CoRR* abs/2002.10764 (2020). arXiv: 2002.10764. URL: https://arxiv.org/abs/2002.10764.

[8] Radim Řehůřek and Petr Sojka. "Software Framework for Topic Modelling with Large Corpora". English. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. http://is.muni.cz/publication/884893/en. Valletta, Malta: ELRA, May 2010, pp. 45–50.

[9] Ekaterina Svikhnushina, Alexandru Placinta, and Pearl Pu. "User Expectations of Conversational Chatbots Based on Online Reviews". In: *Designing Interactive Systems Conference 2021*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 1481–1491. ISBN: 9781450384766. URL: https://doi.org/10.1145/3461778.3462125.

[10] *Testing new designs for the like and dislike buttons - youtube community*. URL: https://support.google.com/youtube/thread/104325801/testing-new-designs-for-the-like-and-dislike-buttons?hl=en.

[11] Jianhua Yin and Jianyong Wang. "A Dirichlet Multinomial Mixture Model-Based Approach for Short Text Clustering". In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '14. New York, New York, USA: Association for Computing Machinery, 2014, pp. 233–242. ISBN: 9781450329569. DOI: 10.1145/2623330.2623715. URL: https://doi.org/10.1145/2623330.2623715.

[12] Wei Zhang et al. "Sentiment Classification and computing for online reviews by a hybrid SVM and LSA based approach". In: *Cluster Computing* 22.S5 (2018), pp. 12619–12632. DOI: 10.1007/s10586-017-1693-7.

Appendix

Additional complementary plots and results removed from the main body of the report are presented in this section.

To further analyse the relationship between the most up-voted approvals and the score their authors assigned to the platform, the up-vote distributions for each score were computed. They are shown in Figure 9 for TikTok and in Figure 10 for Youtube.

For TikTok, the relatively more balanced rating distributions observed when focusing on the most relevant reviews is not straightforwardly revealed when observing the up-vote distributions for each score in Figure 9. However, the more moderate reviews (with scores between two and four) seem to have similar fractions of irrelevant reviews which are slightly lower than the corresponding fraction for extremely negative reviews and clearly lower than said fraction for extremely positive reviews. Overall, the rapid decline in population frequency as user approval increases, which is seen for all ratings in general, is slightly (but noticeably) less steep for the reviews whcih assigned more moderate scores; it should still be noted that the lowest rating group is the one with the highest frequency of reviews among the most relevant ones (those with more than a thousand up-votes).

The trend observed for Youtube where reviews became increasingly negative as they became more relevant is supported again in Figure 10. While there is some variability regarding the portion of reviews that get between one and ten or between eleven and a thousand up-votes, the overall trend, supported most by the portion of reviews that get no up-votes at all, is that as the score increases reviews become less relevant overall.

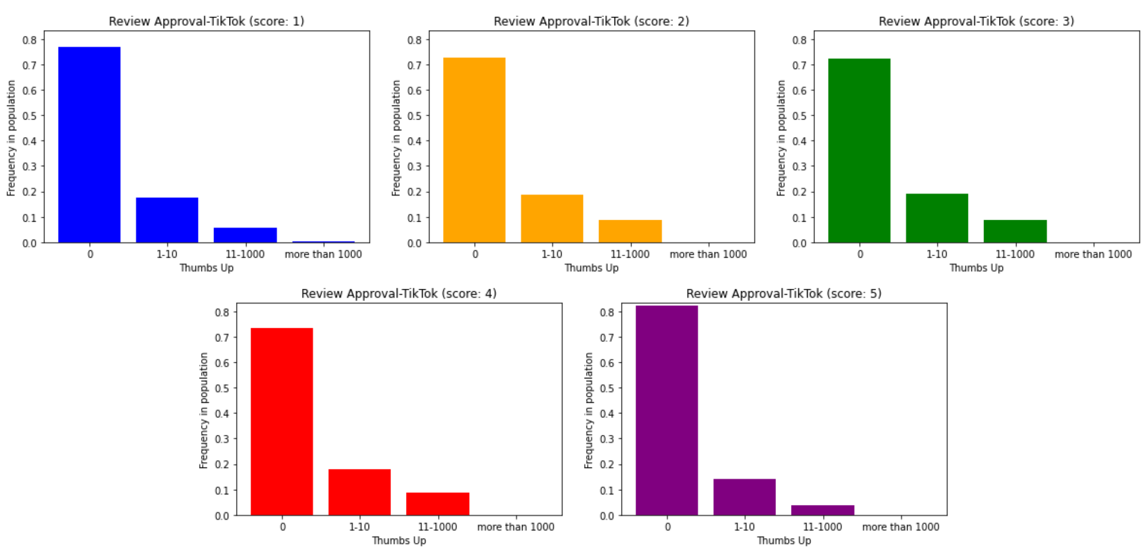


Figure 9: Up-vote counts for TikTok reviews on each of the five scores.

In Figures 11 and 12, the compound sentiment (not broken down into negative, neutral and positive components) is plotted for reviews on both platforms, for all reviews and using only the top-rated 100 reviews, respectively. This compound value goes from -1 to 1 (lower values indicate more negative sentiments and higher ones more positive ones). These plots reflect the dominance of neutral sentiments mentioned in the main report, and the dominance of positive sentiments for TikTok reviews. Interestingly, in Figure 11 the total

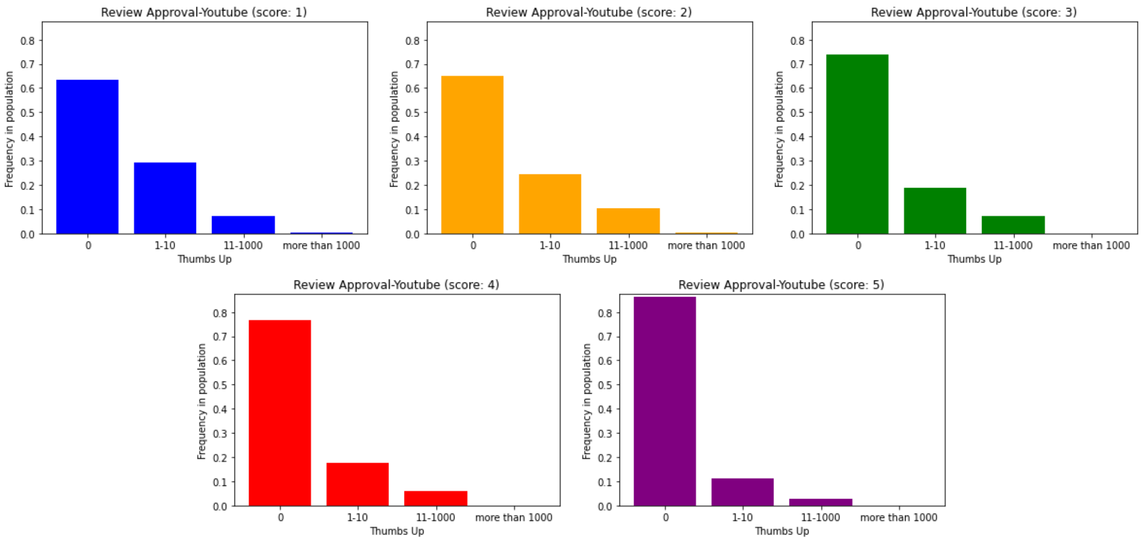
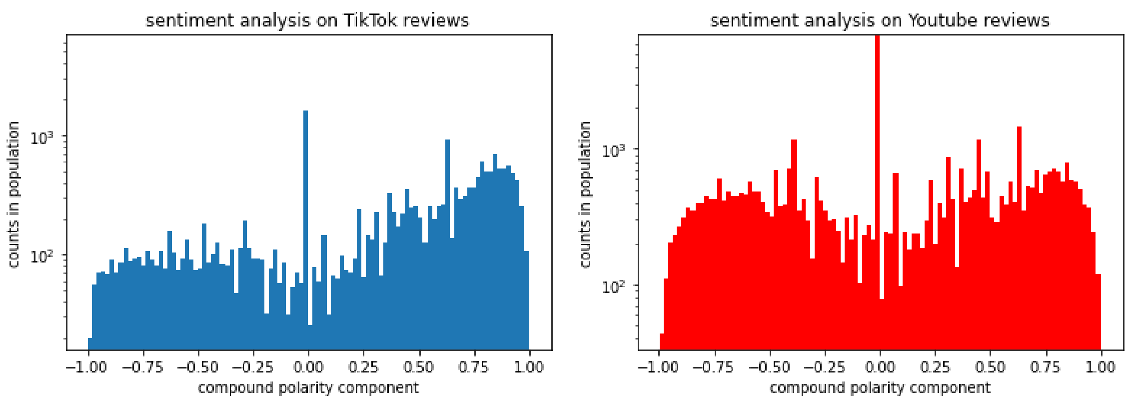


Figure 10: Up-vote counts for Youtube reviews on each of the five scores.

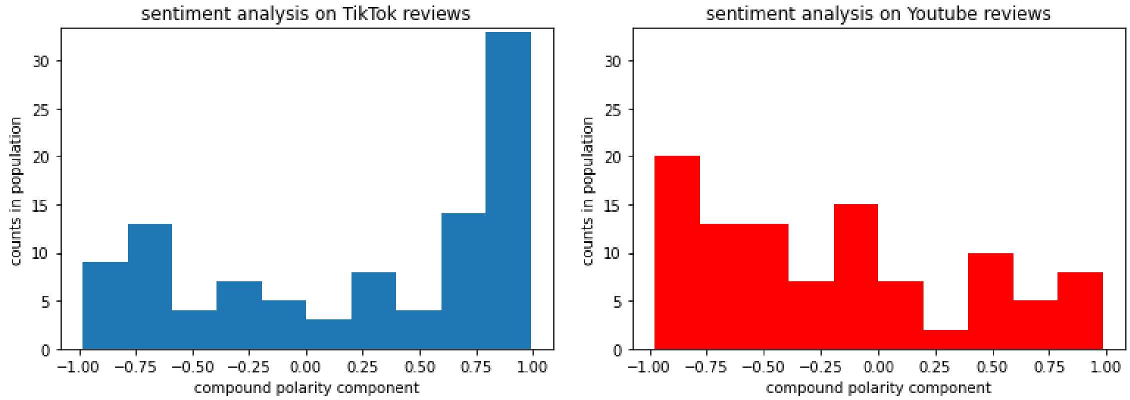
Youtube reviews don't show the clear dominance of negative reviews reflected in the rating distributions, but once again, honing into the top rated 100 reviews augments the dominance of negative reviews as reflected in Figure 12.



(a) TikTok.

(b) Youtube.

Figure 11: Compound sentiment values for all reviews.



(a) TikTok. (b) Youtube.

Figure 12: Compound sentiment values for top 100 approved reviews.

The topics extracted using LDA on the 1000 most up-voted reviews from Youtube and TikTok are presented with their most representative reviews in Tables 2 and 3, respectively.

Topic keywords	Top sample	Rating	Up-votes
Recommendation, dislike button, video quality	*Video recommendations work okish until content solely based on your language and location starts to show up. Streaming quality is ok but the app removes your control over it constantly. Ads cause glitches that ruin the streaming quality and are becoming increasingly intrusive by the day to the point that you could literally end up watching more ad time than content. Content creators abandon this app regularly because of illegal copyright claims by third parties allowed by YouTube's rules.*	1	187
ads, recommendations, technical issues, dislike button	*It's constantly one step forward, several steps backwards. Old quality selection was fine, now I have to mess with settings. Removing public dislike numbers only thinking of YouTube rewind and companies, not content creators or their users. Their broken as hell copyright system that companies and disgruntled exes use to falsely file a DMCA, risking a creator's channel and livelihood. Things need to change or people will start leaving for competitors.*	1	215

ads, video quality, comment deletion, technical issues	*SOS SOS SOS Youtube is HANGING EVERY HALF an HOUR since its last/latest update!!! Youtube App has quietly changed its font from Arial to Roboto and it's a stark difference once you notice it. It's a pain for many people like me with visual problems. Shame on you, after bombarding with ads you even change the font, creating a pain in the neck for visually impaired people. Stopped using the app and running it on the web broswer.*	2	336
Dislike button, ads	*Bring back the dislike counter, or implement a rating system. "Shorts" sucks, lacks most of common video player features. New default quality presets are annoying. Allow advanced features like picture-in-picture for all regions. "Miniplayer not allowed for kid-friendly videos" was another disgusting move, standalone Youtube Kids app exists for a reason. Add advanced video player functionalities like brightness/volume control with gestures, 3rd party subtitles, audio/video synchronization.*	1	496
ads, dislike button, technical issues	*Basically useless now since they took away background play. Absolutely no one wants to sit on the app alone and watch a video when they used to have the option to have a small screen of the YT video play in the background while doing things on other apps. This is a given, but there's waaaay too many ads. It's like we asked for less ads and y'all were like "You want more ads and no dislike button?" Do you guys even listen to your users? Free YT is a joke and premium isn't worth it.*	2	464
technical issues, ads, dislike button	*Year after year, Google finds new ways to provide a less useful product to consumers. I want a dislike counter in YouTube so I know when I'm being lied to, or if my time is going to be wasted. I paid for this service and would like to be treated as a customer, not as a product. I would prefer that my user experience mattered more than advertisers', since again, I'm paying Google for the time on YouTube, not their advertising partners. It frustrates me that my user experience isn't important.*	1	268

ads, praise, paid subscription, technical issues	*The controls are SO FRUSTRATING. They're impossible for many disabled people and people with laggy phones. It's painful trying to exit full screen every 3 minutes because I accidentally touched my phone. It's cool that they are hiding brigade dislikes, but most of the time dislikes indicate a low quality vid, which is a useful tool, and they took it away. Looking forward to the additional nasty comments because of this. I wish someone else were in charge of YouTube, it had so much potential.*	2	291
technical issues, interface, ads	*Older YouTube is far better than the present one. Shorts option is in no way useful, instead the explore option from the older version was good. Always old is gold. Soon YouTube is going to lose its presence, be ready to accept it. Ads was unlimited, also there are ad posters in the home page while scrolling. Hire some better thinking developers. Creativity is nowhere nowadays please do think about it. Replace YouTube with the older version with better interface.*	1	469
ads, dislike button, technical issues	*Each time it gets worse. You need to do 3 steps to change video quality, and then it forces you back to the lowest one always, no matter your internet speed. And now you can't see the dislikes, how I'm supposed to detect good tutorials or guides if I can't check the like/dislike ratio? This is starting to get ridiculous, additionally the comment section is flooded with porn bots, unfortunately for the content creators, as a user I'm starting to use less and less this platform, ain't hard hear users.*	1	629
ads, ban policy, dislike button, technical issues	*Way too many ads. They know interfere every time you try to look at anything, plus now after your recent update every time I try to use your app I'm told to update so I press to update and it says it's already updated but won't work, so know I have to uninstall then reinstall to then get bombarded with irrelevant ads... very annoying. come on YouTube sort it out surely you don't need all this damn advertising to keep the app going. Your part of the Google empire you greedy little git.*	2	144

dislike button, ads, video quality	*Simply put: New video resolution system is clunky, the original did the job better. Demonetization and flagging has gotten out of hand, yet ads aren't held by similar standards. Advertising has gotten relentless (in time and frequency) including constant push of premium and other services. Political ads on domain so pushed to cater to kids by YouTube themselves. And the removal of showing the dislike count is another red flag. It's ok I guess, my only reason for using it is due to lack of sufficient alternatives.*	1	146
dislike button, ads, technical issues	*Voice search function is the worst I've ever used, Youtube isn't fixing the simplest problems, like sex chat bots, scammers, spammers and similar things in the comment section. Yet they can spend their time making advertisements about YouTube shorts, all of the black creators on YouTube, and advertisements about transing our kids. Oh yeah, about the advertisements, there is an incredible amount of mrbeast scams and free item scams. Can you at least deal with one of these problems? Uninstalling*	1	255
technical issues, ads, dislike	*Love YouTube. Smoothest of the batch of video players out there. A societal staple, it'd be 5 stars easily for me but recently the videos stopped full screening. Not in the typical sense, but as in the screen filling zoom sense. I have a Samsung S10 plus, and I noticed 2days ago that I couldn't pinch and expand the screen to fill, which is super annoying. Especially while watching hing vids in the exact dimension of my phone. Also the FB player hasn't shown video in months, just audio.*	4	337

Table 2: Identified topics through LDA on the top 1000 most up-voted reviews for Youtube and their most representative samples. Note that some grammatical corrections were made to the original texts to make them clearer.

Topic	Top sample	Rating	Up-votes
general praise, technical issues	*It's good but.. my uploads sometimes stop at 61 percent, can you fix that? I like how you can upload, make friends, follow, like, share, comment, and watch people's uploads in here. This app is amazing, except for the glitch where the uploading stops at 61 percent.*	4	58
good recommendations, varied content, general praise	*I like the app, it's funny to see videos and it offers videos that you might like just by what videos you have liked or saved, plus it keeps me busy when I'm bored .*	5	11

technical issues, general praise	*I'm feeling very good for downloading this application. I've been posting my own voice videos for more than six months. Thanks TikTok.*	5	21
general praise, technical issues	*So much fun! I love this app! It's my most favorite and definitely will pass the hours before you realize it! 5 stars all day long*	5	16
technical issues	*My TikTok age is incorrect. Please update my age in TikTok, my age is 1.1.2001*	1	254
technical issues, general praise	*Please add dark mode it's getting ridiculous at this point.*	1	15
bad recommendations, good exposure, technical issues	*How many times do I have to say "not interested" before I actually stop seeing ASMR. They are literally the only videos that appear on my feed. That's not an exaggeration either, it's literally the only thing!! Update: I still only get ASMR lives on my feed. "We'll show less of these lives" clearly means nothing.*	1	86
bad recommendations, praise, technical issues	*I love this app! I have used it for years! But the problem is I don't have enough followers... and I cant go live! Please fix it even though people don't have 1k followers let them go live thank you!*	5	16
technical issues, banning	*PLEASE let me be able to block audios. I don't wanna have to block more people, just let me block the audios. Some trends just last too long for my comfort and they can get annoying to hear over and over again.*	4	265
mixed reactions, policy	*I hate the content. Hate the company. Hate that half of it is minors showing off how "cute" they are. Hate how people will do anything... Literally ANYTHING to get more followers. Hate the obnoxious sketches. Hate that even when there is a good joke in one of them, they drag it out until the joke is beyond dead. I hate this app.*	1	18
policy, general praise	*So much fun, lots of good content in little bite sized chunks.*	5	18

Table 3: Identified topics through LDA on the top 1000 most up-voted reviews for TikTok and their most representative samples. Note that some grammatical corrections were made to the original texts to make them clearer.

These samples reveal a trend for longer reviews for the Youtube platform. To investigate this more thoroughly, the difference in average review length (both in terms of words and in terms of sentences) was explored for all reviews and the most up-voted 1000 and 100 reviews on each platform. The results are shown in Table 4. Although the differences in review length are not particularly strong when considering all reviews or just the 100 most up-voted, when it comes to the most up-voted 1000 there is a noticeably larger average length for Youtube reviews, in terms of both words and sentences. Although in general review length has a high standard deviation w.r.t. the average, the reviews used for LDA did support the intuition that, among the samples considered for topic modelling, Youtube reviews tend to be more elaborate (or at the very least longer).

Platform	All (words)	Top 1K (words)	Top 100 (words)
Youtube	22.58 ± 18.95	60.63 ± 24.39	54.75 ± 30.66
TikTok	21.86 ± 17.07	43.67 ± 21.49	58.72 ± 22.96
Platform	All (sentences)	Top 1k (sentences)	Top 100 (sentences)
Youtube	1.669 ± 1.381	3.954 ± 2.196	3.94 ± 2.167
TikTok	1.478 ± 1.071	2.587 ± 1.692	3.58 ± 2.272

Table 4: Average review lengths in terms of words (top) and sentences (bottom) for both platforms, reported for all reviews and for the most up-voted 1000 and 100 reviews. The standard deviations are also reported.