Assignment 3

reddit r/spotify

Data Capture/NLP Project

The Company I have decided to serve in this assignment is Spotify a music streaming company. This dataset has been captured from multiple review subreddits on Spotify from Reddit . The review data has 1135 review records which is mainly unstructured textual data reviewing the recent updates on the application from which I look forward to gather actionable insights from the customer's feedback to help the company improve in a step by step question answer manner.

What are the most frequently used words or phrases used in the review?

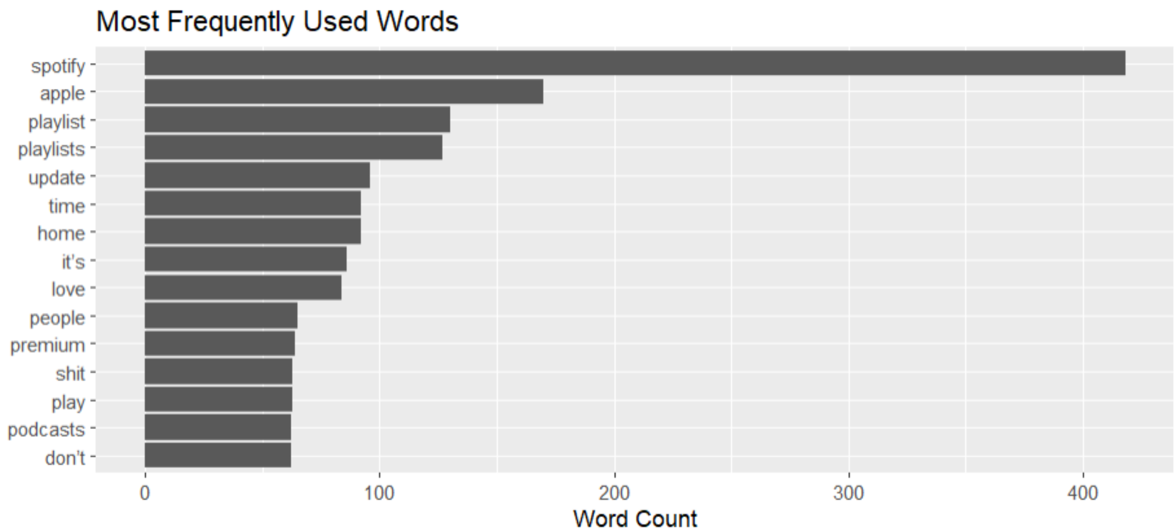
In order to identify what the reviewers are trying to say we first have to break down the reviews into individual words and eliminating the redundant words (like stop words) to get an understanding of what the reviews are talking about the most.

From the word cloud below we can see that the number one word is clearly the application itself i.e. Spotify, but we can also see other competitor names such as YouTube and Apple Music being referenced in the Spotify reviews, which could potentially mean there were comparisons drawn there.

Furthermore, we can see that there are words like Podcasts, Music, Streaming, Premium which are the features that the company offers.

And finally we can see the words being used to describe these services like Love, Hate, Annoying, Terrible, Pretty and how it looks there are more of negative words than positive ones from which we can say that many updates were not received well by the customers.

Fig : WordCloud



What is the overall sentiment of the reviews?

To get a statistical value for the previous assumption and to prove our hypothesis we needed to get the overall sentiment of the reviews by applying lexicon based word level sentiment analysis on the distinct words in the reviews. From figures below we can clearly conclude that the reviews have an overall negative sentiment. We can also see how the sentiments are distributed on a scale of -5 to 5 (-ve value being negative sentiment and vice versa).

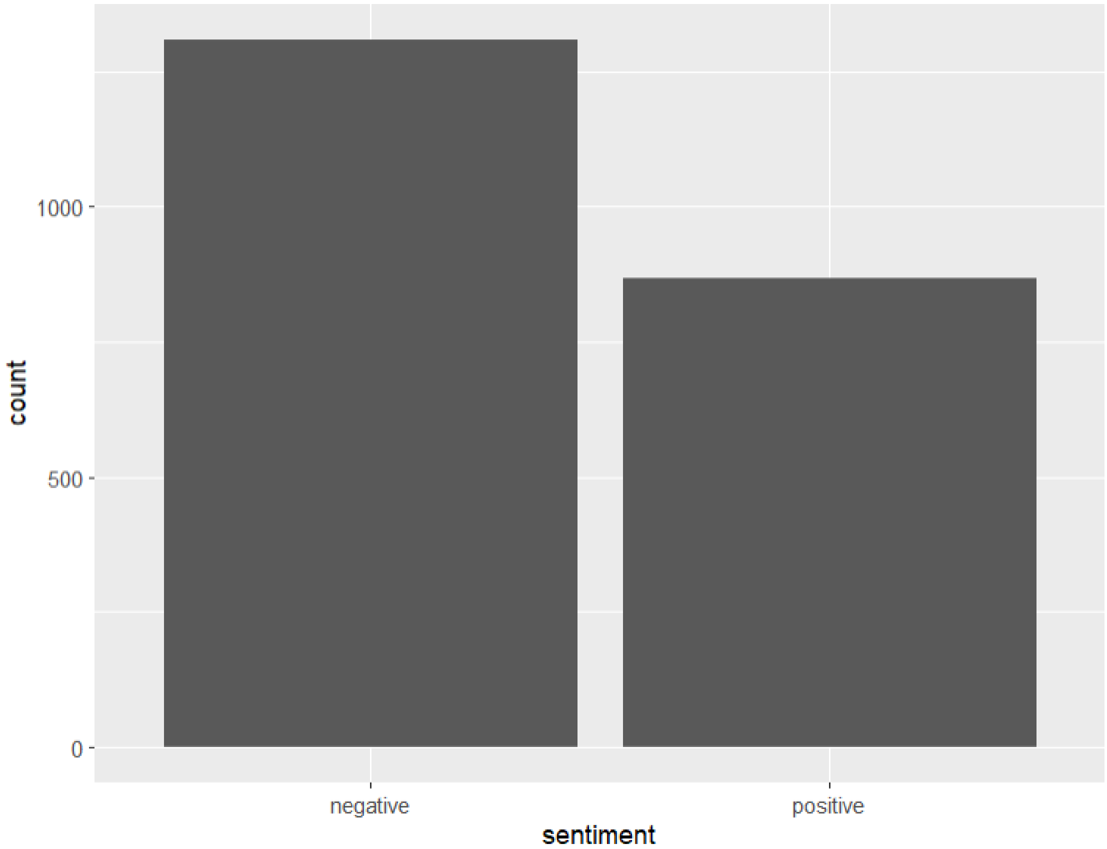


Fig : Overall Sentiment

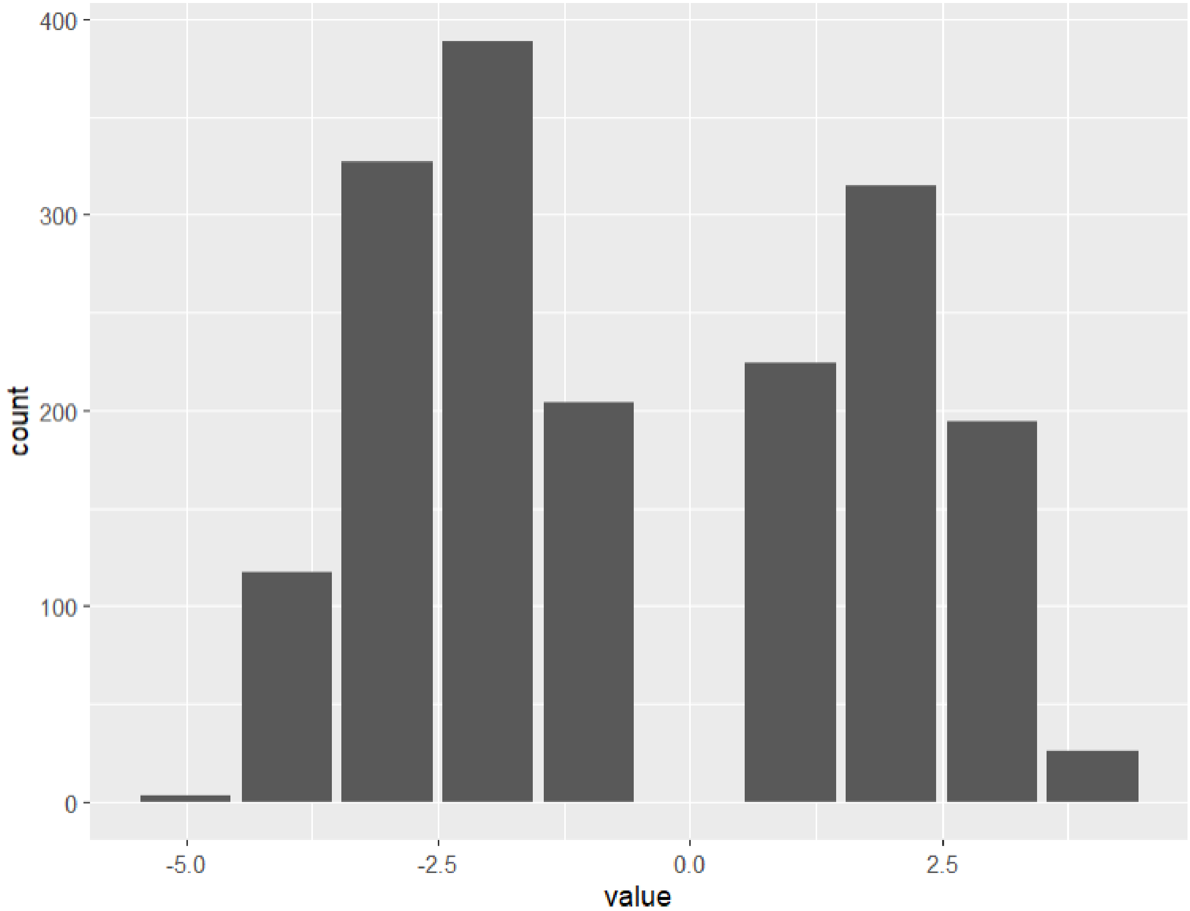
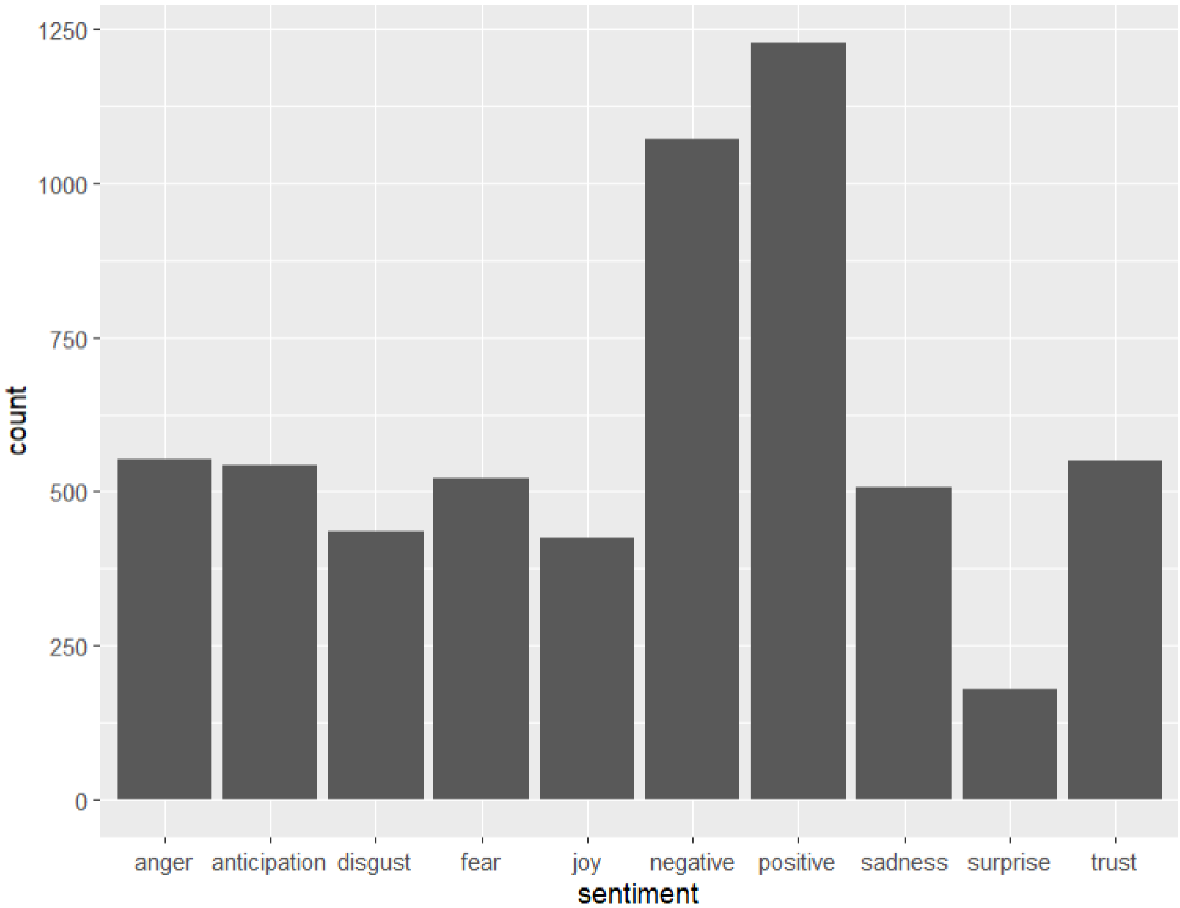


Fig: Sentiment Using AFINN model

Are there any specific sentiments or emotions expressed in the text? If so, what are they and how prevalent are they?

To find specific sentiments in the text we can use the NRC model lexicon where the texts are categorized in the 10 elementary categories that are Anger, Anticipation, Disgust, Fear, Joy, Negative, Positive, Sadness, Surprise, Trust

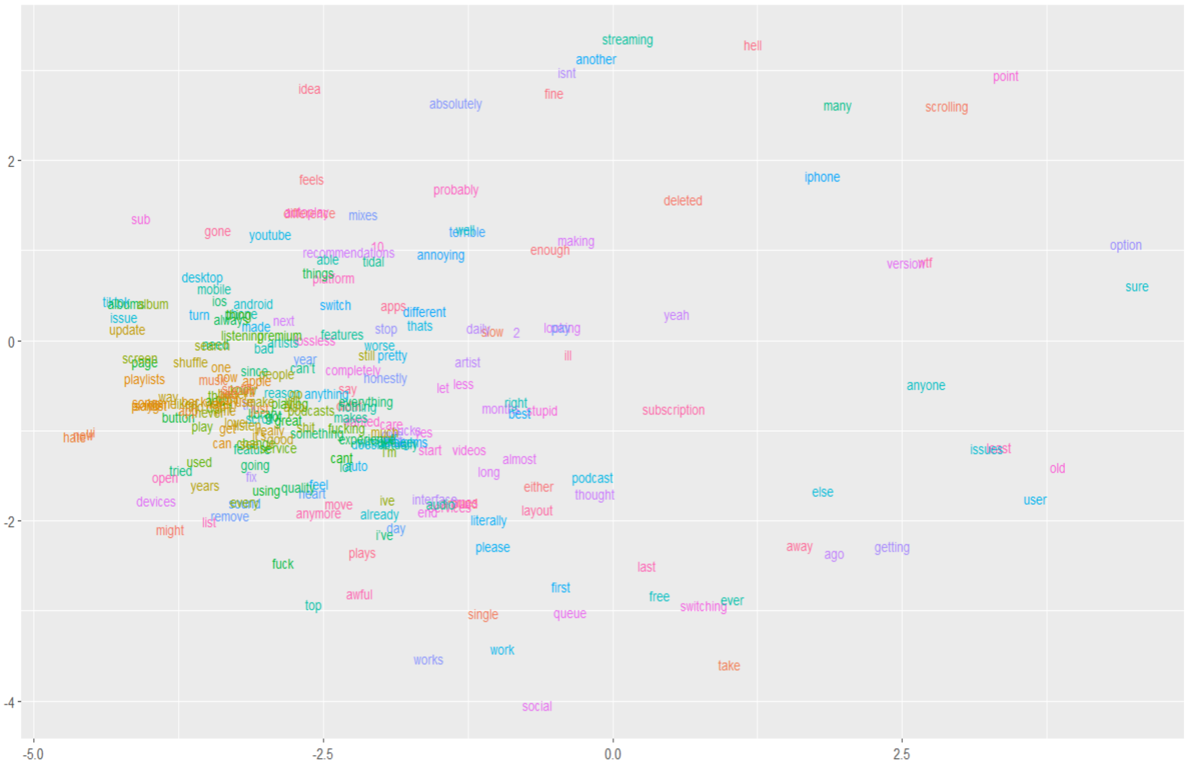


How to better understand the reviews?

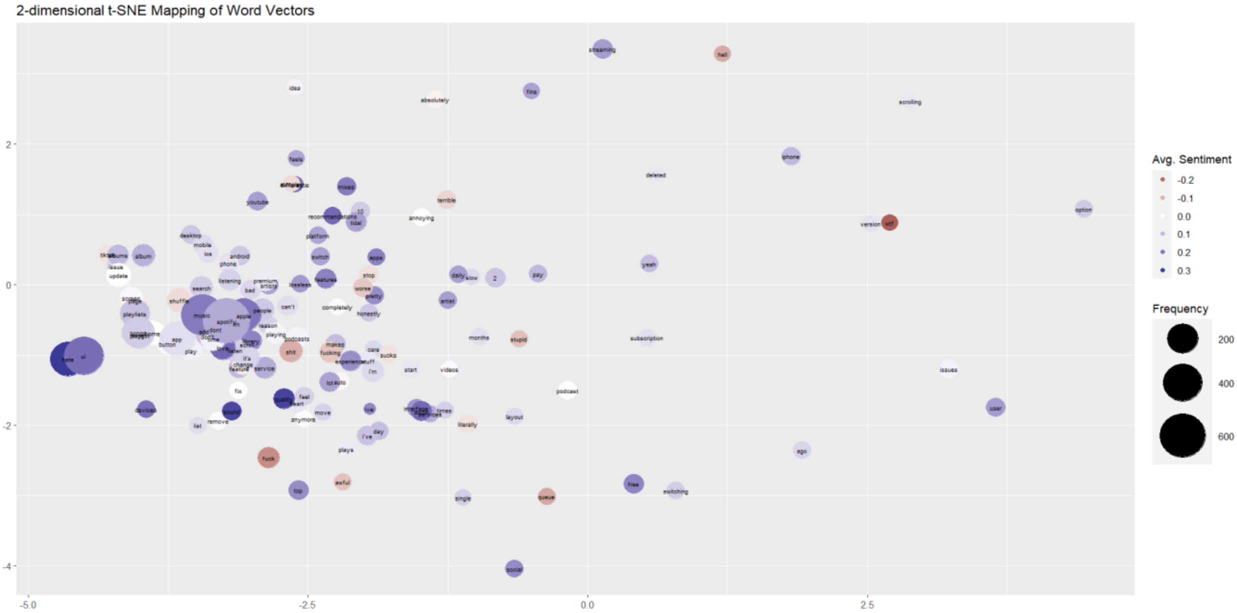
To get a better access of what the user is saying we can use word embedding to relate the words with the words closer to it and getting a collective sentiment rather than sentiments of individual words.

The t-SNE chart below (it represents data points in a lower-dimensional space in a way that preserves their similarities or dissimilarities, as measured by their pairwise distances in the original high-dimensional space) shows how the words are related for example the words Hate and UI are right next to each other which indicates that the users are not liking the new User interface from the updates, similarly the words update and Tiktok are close by an Tiktok has a really negative sentiment (as you can see in the t-SNE bubble chart) which could mean that the new update looks like Tiktok and is not performing great for the users, also the word Music with –ve sentiment Shuffle refers to the new update where you have to listen to music in shuffle if you don't have a premium account which is a problem I faced as well , also there is

Podcast with the word Shit which might mean that the new podcasts are not good and is getting strongly critiqued , and finally the word Experience and the word Sucks comes together which might meant the overall user experience is bad. And these might be some of the problems that have to be tackled by the company.

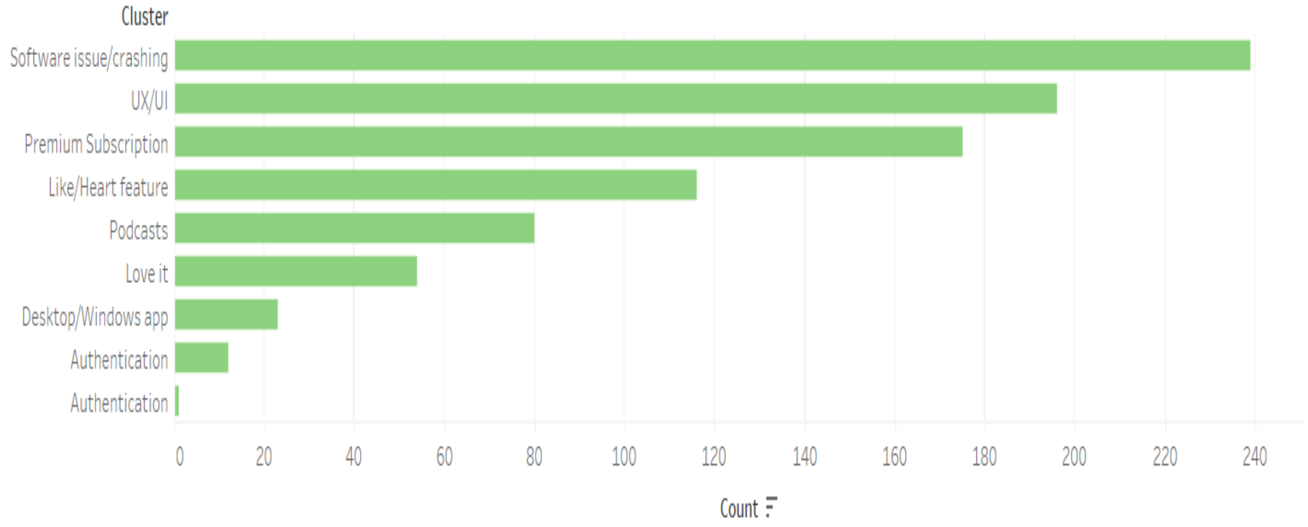


2-dimensional t-SNE Mapping of Word Vectors



How can the company pinpoint exactly where the problem is and how to prioritize the problems?

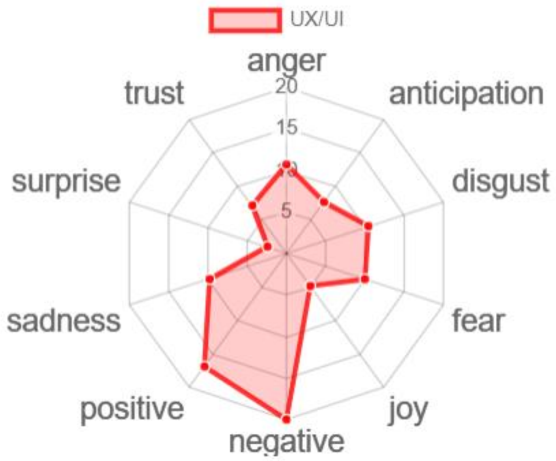
Here using topic modelling method, I have analysed the texts and created clusters to segregate the reviews and get an overall count for each cluster as we can see below. Now here we can see that software crashing, UX/UI are few of the top most discussed topics and should be the ones to be tackled initially as it covers a larger set of users.



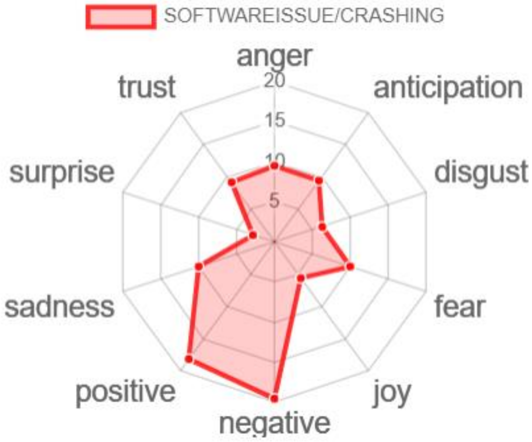
How to better understand the sentiment of these clusters individually.

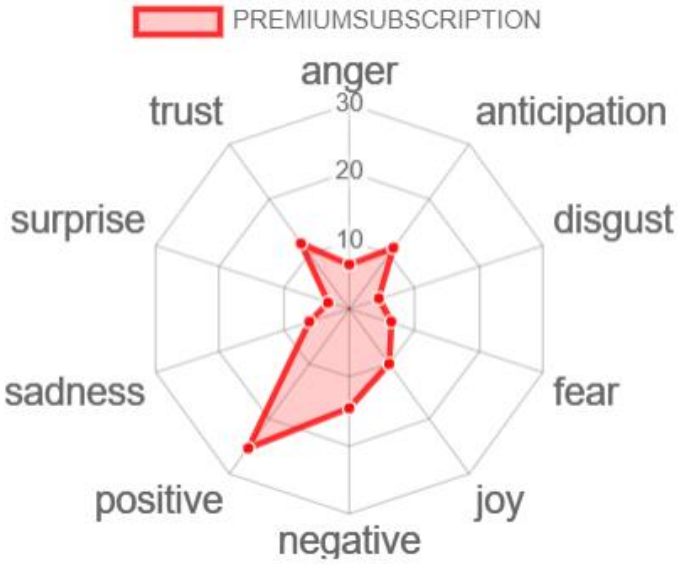
To have a clearer image of the issue we can look at the radar chart of individual clusters and can act upon the ones with highest negative reviews in order to increase the survival rate of our customers.

Our previous decision of acting on software crashing, UX/UI issues to start with still stands as these 2 clusters have the highest negative rating compared to the others and also the fact that these 2 clusters also have a good positive marker means reducing the causes for negative comments would help retain more customers.



The update might be working



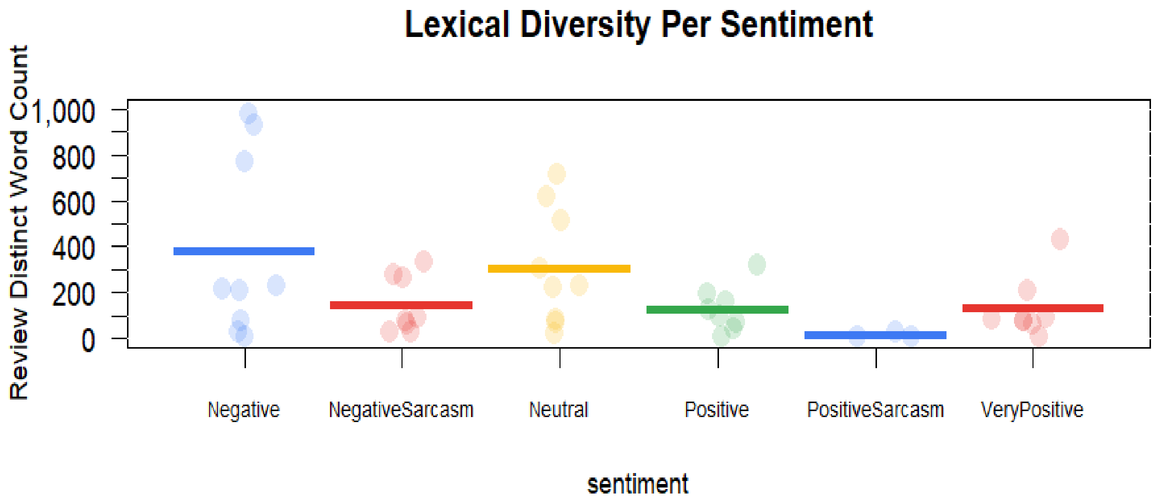


Furthermore, to understand how different complex tones of comments have a relation with the word count, I have created a new set of sentiments using the new Open AI's API's which uses multiple NLP deep learning models to analyse the reviews and it even identifies positive and negative sarcasms in the reviews which is the best for reviews extracted from platforms like Reddit.

Here in the figure we can see that the pure negative comments have the highest average word count, which is probably because the customer might be too angry and was venting out on the review sections.

The neutral comments seem to have a moderate amount of criticism as it might be a genuine review and to the point one

And the positive ones are small quick ones as they are contempt with the service and don't have much to critique about.



With all this information extracted from the comments I believe we can start our work on debugging our software for crashes and redesigning the UX/UI according to the user preferences and then advertise these changes to increase the retention rate of our customers that we might lose to our competitors.