COVID-19's Impact on Music Trends: Pre- and Post-Pandemic Shifts

I. Research Question

The research question that our project aims to answer is how the COVID-19 pandemic influenced musical trends, focusing on both pre-pandemic and post-pandemic periods. We are interested in exploring whether there was a shift in the music genres that became popular, the musical characteristics of songs (such as energy, rhythm, and modality), or if, on the contrary, the music preferences of users on platforms like Spotify remained relatively constant during the lockdown and the post-pandemic phase. Additionally, we looked to answer the related question of whether emotions and content of lyrics of popular songs shifted or changed in any way as a result of COVID-19.

II. Suitability of Data

The primary dataset that we used for this study is one retrieved from *kaggle.com* titled "Spotify Top Hit Playlist (2010-2023)". It provides insights into music trends, artist popularity, and song characteristics over time. This dataset was extracted through the Spotify API directly and comprises track details from various **Spotify Top 100 playlists** for each year. The dataset was constructed by extracting track data from publicly available Spotify "Top Hits" playlists, which feature popular songs from a given year. The raw dataset consists of 23 columns, covering metadata about each song, including track details, artist information, popularity scores, and key audio features.

Below is a breakdown of the variables and their description:

- **Playlist_url:** URL of the Spotify playlist from which the track was sourced.
- **Year:** The year associated with the playlist (not necessarily the track's release year).
- **track_id:** Unique identifier for the song in Spotify's database.
- **track_name:** The title of the song.
- **track_popularity:** A Spotify popularity score (0-100), based on recent streaming activity.
- **album:** Name of the album that features the track.
- **artist_id:** Unique identifier for the artist in Spotify's database.
- **artist_name:** Name of the artist or band.
- **artist_genres**: List of genres associated with the artist.
- **artist_popularity**: Spotify's popularity rating for the artist (0-100).

Additionally, Spotify provides audio analysis features for each track, helping in understanding musical characteristics, which would be fundamental for our analysis:

- **danceability:** A score (0-1) indicating how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.
- **energy:** Numerical, Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.

- **Key:** Numerical, the estimated overall key of the track. Integers map to pitches using standard Pitch Class notation . E.g. 0 = C, 1 = C#/Db, 2 = D, and so on. If no key was detected, the value is -1 .
- **Loudness**: Numerical, Overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typically range between -60 and 0 db.
- **Mode:** Numerical, mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.
- **speechiness:** Presence of spoken words in the track (higher values for rap/spoken tracks).
- **acousticness:** Numerical, Confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
- **Instrumentalness:** Likelihood of the track having no vocals.
- **liveness**: Probability that the track was recorded live.
- **Valence:** Numerical, Measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).
- **Tempo:** Numerical, Overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
- **Duration_ms:** Length of the track in milliseconds.
- **Time_signature:** The time signature (e.g., 4/4, 3/4).

This dataset was very suitable for our research question given that its structure allowed us to view trends in pre-Covid years (2010-2019), the first year of Covid (2020), and also post-Covid years (2021-2023) and identify any potential relationships between music characteristics of popular songs and the Covid-19 pandemic. This technique is ideal for identifying patterns within the data without any prior assumptions, making it highly relevant for our project, where we aim to explore how music preferences have shifted over time in response to external factors like the pandemic. Spotify is also one of the largest music streaming platforms in the world, giving us an accurate representation of user listening trends year over year. Furthermore, the inclusion of various different musical characteristic variables in this dataset allowed us to get a detailed picture of how the characteristics of popular songs changed over time.

For some of our analysis, we also used another dataset from *kaggle.com* called **"Genius Song Lyrics"**. *Genius* is a popular online database for song lyrics and this specific database pulled information, including song lyrics from *Genius* for over 7.88 million unique songs. This dataset was merged with the Spotify dataset by 'track_name' and 'artist_name'. We then used these merged datasets to get further insight into whether lyrical content and emotions of popular songs were at all influenced by the COVID-19 pandemic. This dataset has a few additional columns like: "duration_ms", "time_signature", "track_genres", "release_year", "views", "features", "id", "language" and"lyrics" which specifically was an important column that we used in our analysis. This dataset was also perfectly suitable for our area of research. Since the dataset was so large, we were able to generate a large number of matches between songs in this dataset and those in the Spotify dataset. Further, similar to the first dataset, the "Genius

Song Lyrics" dataset also allowed us to view lyrical content for songs released before, during, and after Covid, ranging from songs released before the year 2000 up until 2022. These two datasets provided us with a great framework to perform our analysis.

III. Choice of Analytical Techniques

We decided on a few different analysis techniques to help answer our questions about the impact that Covid-19 had on music which were time series analysis, cluster analysis, and text mining. However, we started by simply performing an initial exploratory analysis, plotting the variables *acousticness, danceability, energy, key, tempo, and valence* in the years before, during and after Covid. This would give us an early insight into music characteristic trends in the years covered in our dataset. We would then move on to the more advanced analytical methods:

Time Series Analysis

Time series analysis was an appropriate analytical technique for this study, as it allowed us to observe how key musical characteristics—such as valence, danceability, energy, and tempo—evolved across the years represented in the dataset. The Spotify dataset was well-suited for time series analysis at the annual level, as it met the key assumptions: observations were equally spaced (with each year from 2000 to 2023 represented), and songs were uniformly sampled by year. The "year" column served as a natural, discrete time index, enabling consistent longitudinal analysis.

We began by performing autocorrelation analysis to examine whether the features of interest (valence, danceability, energy, and tempo) followed consistent year-over-year patterns or fluctuated randomly over time. Autocorrelation allowed us to measure the degree to which a value in one year correlated with values in previous years. To assess whether COVID-19 disrupted these temporal patterns, we conducted separate autocorrelation analyses for the pre-COVID period (2000–2019) and the post-COVID period (2020–2023).

Next, we fitted drift models for each of the four features using their average annual values from 2000 to 2019. These models established a baseline of what the post-2020 trends would have looked like if pre-COVID trajectories had continued uninterrupted. This approach enabled us to generate forecasts for the years 2020 through 2023—spanning the COVID-19 and immediate post-COVID era—and visually compare the predicted values with the actual observed ones.

To evaluate whether the differences between predicted and actual values were statistically meaningful, we conducted residual analysis for each variable. Residuals were calculated as the difference between observed and forecasted values. We then used a Wilcoxon rank-sum test to assess whether post-COVID residuals were significantly different from those in the pre-COVID period. This step was essential to assess whether observed deviations reflected meaningful structural shifts or random variation.

Overall, time series analysis enabled us to visualize potential disruptions in musical trends, quantify their magnitude, and test whether they were statistically significant in the context of COVID-19's cultural impact.

Cluster Analysis

We determined cluster analysis to be a highly suitable analytical method to answer our research questions. We selected clustering because it allows us to group music data based on similarities in key features such as *energy*, *tempo*, *loudness*, *danceability*, etc. This technique is ideal for identifying patterns within the data without any prior assumptions, making it highly relevant for our project, where we aim to explore how music preferences have shifted over time in response to external factors like the pandemic. Additionally, we used R to generate both plots and tables, helping to further illustrate and quantify the differences in the clusters based on various musical features. By clustering the data, we can segment listeners into distinct groups that share similar music tastes at different points in time. This enables us to uncover hidden patterns in music preferences across three key phases: Pre-pandemic: Establishing baseline music preferences before the pandemic; During the pandemic: Investigating how global changes affected musical tastes; Post-pandemic: Analyzing whether music preferences reverted or evolved in response to the pandemic's end. The result is an overview of trends across all years that the dataset covers in order to identify long-term shifts in music preferences.

We used K-Means clustering to group songs into clusters based on musical characteristics. We again segmented our data into three time periods: pre-Covid, during Covid, and post-Covid and clustering was performed separately for each period. We then sought to determine the optimal number of clusters for each of the three periods which was determined using the elbow method and replicated for all three periods. We then repeated the process for each of the three periods, this time using a ratio plot, which compared teh between-cluster sum of squares (BSS) to the within-cluster sum of squares (WSS) for different numbers of clusters. This helped to decide what number of clusters would yield meaningful separation. Silhouette analysis followed for the purpose of determining how well each song fit in its assigned cluster. Once again, this was performed for all three time periods. This provided validation of the clustering structure and allowed us to see what number of clusters provided us with the best balance between cluster separation and cohesion.

We also explored cluster analysis even further, using Gaussian Mixture Models (GMM) and Hierarchical Clustering. GMM was suitable for our analysis because unlike k-means clustering, it allows for overlapping of clusters. For each time period, GMM was run with varying numbers of clusters and the BIC value was used to determine which number of clusters best fit our data. Additionally, hierarchical clustering was also used for all three time periods, giving us a visual representation of different cluster combinations using dendrograms.

The results of k-means, GMM, and hierarchical clustering were noted and compared, giving us an in-depth look at all of the different possible cluster solutions from multiple analytical perspectives.

Text Mining

Text mining was the final analytical approach that we used in our research. Text mining was an important addition to our analysis because it allowed us to analyze the lyrics of the songs in our dataset and make observations about music trends from them. We joined the Spotify dataset with the Genius dataset by ‘track_name’ and ‘artist_name’ and began by performing sentiment analysis on the lyrics, removing stop words and one and two letter words. We then created plots of the top 25 most popular songs pre-Covid (2016-2019) and post-Covid (2020-2023), and used this to compare if lyrical content had changed over the two periods. We then performed an analysis using binary sentiment lexicon *bing*. This was suitable because it allowed us to analyze if the lyrical content had gotten more or less positive from before Covid to after Covid. We dove further into this analysis by using the *nrc* emotion lexicon, which categorized words into eight different emotional categories, which helped to generate a more descriptive idea of how the emotions represented in songs changed over our periods of interest. Furthermore, we used the sentiment score lexicon *afinn*, to give us information about how the intensity of emotions displayed in song lyrics shifted from pre-Covid to post-Covid which we thought was a very important consideration.

The Bag of Words approach gave us another method to analyze lyrics in songs pre-Covid and post-Covid by converting text into numerical features. We pre-processed the data by creating a corpus and then performing text cleaning which included: converting all text to lowercase, removing URLs, removing punctuation, removing stopwords, stripping extra whitespace, and stemming words and a dictionary was created from unstemmed words for stem completion. A document term matrix was then created from the cleaned lyrics corpus where rows are songs, columns are unique words, and values are the frequency of each word in each song. Sparse terms were removed using a threshold of 0.95 and terms were restored using stem completion.

A new DTM was created using TF-IDF weighting. This was suitable for our research because it allowed us to view top terms in songs for both TF and TF-IDF weighting methods when comparing the pre-Covid and post-Covid years. Linear regression modeling was used to evaluate the summary of models and determine which terms in lyrics were most significant in influencing popularity before and after Covid. It gave us a way to visualize how words in lyrics influence popularity.

Text mining was not only suitable, but crucial, for our analysis since it gave us the ability to make predictions and compare pre-Covid and post-Covid trends for lyrical content in popular songs- something our other analytical methods could not capture.

IV. Results from Analyses

Time Series & Statistical Modeling Results:

Between 2010 and 2023, the emotional and structural characteristics of popular music on Spotify shifted in clear and meaningful ways. The onset of the COVID-19 pandemic in 2020 marked a distinct turning point. Using time series analysis, we examined how four key musical features—valence, danceability, energy, and tempo—evolved across this period. The data suggest that the pandemic disrupted not only the global social landscape, but also the emotional tone and rhythmic trajectory of popular music.

Danceability, which reflects how likely a song is to inspire physical movement, followed a consistent upward trend in the years leading up to 2020. However, this trend reversed sharply with the onset of the pandemic and remained suppressed through 2023. This decline may reflect the long-term effects of reduced social interaction and the temporary closure of clubs, concerts, and dance venues. The broader shift away from high-tempo, rhythm-forward genres like EDM and dance-pop further reinforces this interpretation

Energy, a measure of the intensity or drive of a track, also dropped noticeably in 2020. There was a slight recovery in 2021 and 2022—potentially mirroring renewed optimism as vaccines became available—but by 2023, energy levels declined again. This pattern may reflect ongoing cultural fatigue or a shift in listener preferences toward more subdued or emotionally grounded styles of music.

Valence, which measures the overall emotional positivity of a song, also fell in 2020. However, unlike energy or danceability, valence began rising again by 2023. This partial rebound may indicate a slow emotional recovery in both listeners and artists, as well as a renewed interest in more joyful or hopeful music after several years of hardship.

Tempo, which refers to the speed or pace of a track, remained relatively stable over the full time period. A modest increase after 2021 suggests that artists may be reintroducing rhythmic energy into their work, though this shift was less pronounced than those observed in other features.

To assess the regularity and predictability of these musical trends, we conducted autocorrelation analysis (ACF). Before 2020, features like valence and danceability exhibited strong autocorrelation across multiple lags, suggesting that musical characteristics evolved in a relatively stable and consistent way. After 2020, these patterns broke down entirely—autocorrelation values dropped and lost significance, indicating that the usual year-to-year momentum in music was interrupted (see Appendix A, Figure A1-A2).

To further examine this disruption, we built drift models using data from 2000 to 2019 to forecast values for 2020–2023 (see Appendix A, Figures A3 - A6). These models represent what each musical feature might have looked like if historical trends had continued without interruption. When we compared actual post-COVID values to these forecasts, clear deviations emerged. For valence, the observed values were consistently higher than predicted, suggesting an emotional rebound not accounted for by the pre-pandemic trend. In contrast, danceability was consistently lower than forecasted, reinforcing the idea that socially driven musical qualities have not yet returned to pre-pandemic levels.

To determine whether these differences were statistically significant, we performed residual analysis using Wilcoxon rank-sum tests. The results confirmed that:

- Valence was significantly higher than forecasted ($p = 0.0219$)
- Danceability was significantly lower than expected ($p = 0.0014$)
- Energy showed a marginal increase ($p = 0.09$)
- Tempo showed no significant change ($p = 0.31$)

Taken together, these findings suggest that the pandemic meaningfully disrupted the emotional and rhythmic continuity of popular music. Some features—like valence—have begun to rebound, while others, such as danceability and energy, remain altered. These shifts likely reflect a combination of changing listener behaviors, artist responses to the cultural moment, and evolving emotional needs. COVID-19 didn't just influence what people listened to—it reshaped how music feels, and how it's made.

Clustering Analysis Results

By looking at music trends across three distinct time frames, before the pandemic (2010–2019), during the pandemic (2020), and after the pandemic (2021–2023), we can clearly see how people's musical preferences evolved in response to the changing social landscape and shifting emotional experiences during these periods. Clustering techniques such as K-means, Gaussian Mixture Models (GMM), and Hierarchical Clustering were applied, and across all models, the optimal solution consistently pointed to two main clusters. However, the makeup of these clusters and the degree of separation between them evolved noticeably over time.

Before the pandemic, the clustering result was quite clear. The songs could be separated into two stable groups. The first group mainly included songs that had low energy, slow tempo, more acoustic feeling, and not very strong emotion, these songs were often from acoustic, indie, or soft pop styles. The second group was just the opposite, with songs full of energy, easy to dance to, louder sound, and more happy feelings, these songs matched pop, electronic, and dance music. This kind of clear separation might have reflected the music choices of people in normal social life at that time, when going to parties, dancing in clubs, or watching live shows was very common.

During the pandemic, although the two-cluster structure remained the most statistically reasonable choice (as verified by the Elbow and Silhouette methods), the differences between the two clusters became less obvious. Cluster 1 still mainly contained soft, reflective, and emotionally profound tracks, often using acoustic instruments and having slower tempos. This shift can be seen as a reflection of people's collective move toward introspection. During a time of isolation, uncertainty, and limited movement, listeners were seeking comfort and peace. In contrast, Cluster 2 was composed of more upbeat and energetic tracks, which likely offered a form of psychological balance, injecting energy and vitality into life during a period of restrictions. However, the greater overlap between the two clusters indicated a more diverse and fragmented emotional state of people. At that time, music was not only a way for people to cope with difficulties and escape from reality but also a means to build resilience and distract themselves. Sometimes, its emotional functions even seemed contradictory.

After the pandemic, the musical clusters began to show more of their previous distinct characteristics. Cluster 1 increasingly concentrated on high-energy, fast-tempo, and rhythm-driven songs. This trend seemed to reflect the growing demand for lively and expressive music, which was in line with the resumption of social activities like music festivals, concerts, and gatherings. Meanwhile, Cluster 2 retained its association with calmer, acoustic, and emotionally intimate tracks, suggesting that even as social activity resumed, a portion of listeners still valued introspective and emotionally grounding music. This dual-track recovery matches behavioral theories that propose parallel emotional needs: the desire to celebrate renewed freedom and the lingering need for emotional healing.***(Figure B1 (see Appendix B) showing cluster separation(post-COVID)).***

Looking at the data from 2010 to 2023 as a whole, this analysis clearly shows that two main musical clusters have persisted. Cluster 1 generally represents songs that are high-energy, positive, and very danceable, with louder, more processed sounds and an emphasis on upbeat or euphoric moods. However,

Cluster 2 usually includes tracks with lower energy levels, softer emotional tones, slower tempos, and more acoustic elements. These tracks often convey a more intimate or emotionally subdued atmosphere. Over the past years, these clusters have remained consistent. However, their specific compositions and the ways listeners interact with them change according to evolving cultural and emotional trends.
The bigger implication of this analysis is that music plays a dual emotional role, serving both as a catalyst for social connection and physical energy, and as a reflective outlet for processing personal and collective emotions. This insight can be useful for music producers, streaming platforms, and cultural researchers to better personalize recommendations, understand mood trends, and create playlists that reflect changing listener emotions. As music consumption is increasingly influenced by social context, the clustering approach provides a useful framework to track and respond to emotional shifts on a larger scale.

Text Mining & Sentiment Analysis Results

Text mining and sentiment analysis showed that the lyrics of popular songs had an obvious change before and after the COVID-19 pandemic. Before the pandemic, the lyrics often praised bold, carefree, and party-centered themes. Slang and bold expressions such as "money," "bitch," and "woo" were often used, which reflected the high-energy, social music that was dominant in the charts at that time. These lyrics usually focused on materialism, nightlife, and self-assertion, reflecting a cultural period that was obsessed with escapism and seeking external excitement.

However, after the pandemic broke out, the lyrical themes clearly began to shift towards introspection, emotional depth, and personal contemplation. Words like "feel," "life," "mind," and "time" appeared more frequently, indicating a greater focus on the inner self. Both musicians and listeners seemed to engage more with emotions such as uncertainty, vulnerability, and resilience. Visual representations of word frequency, including bar charts and word clouds, clearly demonstrated this transformation, revealing a more diverse and emotionally evocative vocabulary in the songs of the post-pandemic period.***(Figure C1 (see Appendix C) Bar plot of top 25 words (pre- vs. post-COVID)).***

Sentiment polarity analysis based on the Bing lexicon further confirmed this trend. After the pandemic, the proportion of positive words in song lyrics increased, while the use of negative words slightly decreased. This indicates that songwriters, in response to the collective emotional challenges, began to convey hope and optimism through their music. Emotion analysis using the NRC lexicon also backed up this change, revealing a rise in words tied to joy, trust, and anticipation, while words linked to fear, sadness, anger, and disgust became less frequent. This shift mirrors a broader societal move toward emotional healing, where music evolved beyond mere entertainment, becoming a vital source of comfort and motivation.***(Figure C2 (see Appendix C) Proportion of Sentiment in Song Lyrics(pre- vs. post-COVID)), (Figure C3 (see Appendix C) Proportion of various Emotions in Lyrics (pre- vs. post-COVID))***

The AFINN sentiment score analysis added more detail, showing that while pre-COVID lyrics displayed a broad range of emotional intensity—with more extreme positive and negative sentiments—post-COVID lyrics became more neutral to mildly positive. The sentiment distribution after the pandemic was higher and narrower, indicating fewer emotional extremes and a more consistent mood. This indicates that there is a cultural shift towards emotional stability, most likely as a coping mechanism in response to the long standing uncertainty. ***(Figure C4: Sentiment Score Distribution in Lyrics (pre- vs. post-COVID))***

When analyzing song popularity using the regression tree model, the alteration in what attracted listeners became distinctly noticeable. The linear regression model identified the top 20 predictors influencing song popularity before and after the COVID-19 pandemic, using term frequency (TF) as input features. Pre-COVID, popularity was associated with a diverse set of moderately weighted words such as *fight*, *buy*, and *fine*, suggesting that lyrical themes like materialism resonated with listeners. Negative predictors like *broke* and *front* pointed to a lower preference for themes related to struggle or confrontation. In contrast, the post-COVID model showed that the words like *lotta*, *whole*, and *street* are strongly linked to higher popularity scores, while terms such as *gone*, *house*, and *cold* were negatively associated. These results suggest that after the pandemic, listeners gravitated toward lyrics that were more emotionally immediate, authentic, and easy to relate to—highlighting a shift in audience preferences toward sincerity and simplicity in music. ***(Figure C5: Top Predictors of Popularity Score from Linear Regression Model)***

V. Conclusion & Recommendations

Our evidence strongly suggests that the COVID-19 pandemic caused a measurable and persistent disruption to musical trends on Spotify. The time series analysis showed that several musical features, most significantly acousticness, danceability, energy, and valence experienced extreme changes beginning in 2020. Acousticness increased, reflecting a turn toward stripped-down, emotionally focused music. Danceability and energy fell sharply, reflecting a decrease in social interaction and a mass shift away from high-energy music. Although valence fell at the pandemic's height, it reflected emotional recovery during 2023, predicting a slow recovery to positive values in music narratives ***(Figure A1, A2 of Appendix A)***

Statistical testing of deviations from forecast confirmed that such changes were not random. Valence levels arrived significantly higher than anticipated, and danceability arrived lower than anticipated, highlighting a durable impact on how people relate to movement and mood in music.In short, emotional tone in popular music began to recover, but rhythmic and social features like danceability are slow to recover.

The clustering analysis further confirmed this bifurcation in listener preferences. Across all periods, two stable musical clusters emerged: one marked by high energy, danceability, and positive emotion; the other by calm, acoustic, and introspective characteristics. However, the degree of separation between these clusters decreased during the pandemic, indicating emotional fragmentation and diverse listener needs. Post-pandemic data showed evident separation returning, which means that both lively and contemplative listening experiences remain relevant.***(Figure B17 in Appendix B)***

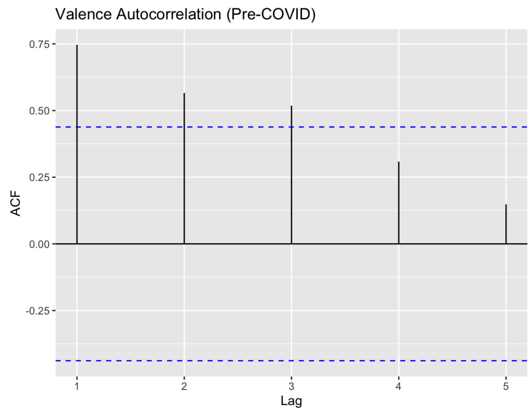
From a text mining and sentiment analysis perspective, the shift was equally profound. Pre-COVID lyrics were heavily dependent on brash, slang-infused, party-themed or materialistic language. Post-COVID lyrics were more introspective in terms of emotions, words related to emotions and self-awareness, and resilience. Sentiment score analysis and emotion lexicon analysis supported higher positive emotion scores (happiness, trust, expectation) and lower negative emotion scores (fear, anger, sadness) in the post pandemic era, indicating a consistent emotional healing. In particular, post-COVID lyrics were simpler and more emotionally authentic, and popularity in songs was more influenced by empathetic, emotionally direct language ***(Figures C1–C5 in Appendix C)***

Recommendations for Decision Makers

1. Music Streaming Platforms: To improve user experience and engagement, platforms like Spotify could leverage clustering and sentiment analysis tools to segment songs based on emotional content and mood. Creating playlists centered around specific moods (e.g., upbeat, introspective) could cater to the diverse emotional needs of users, especially post-pandemic.
2. Data Analysts & Product Teams: It is crucial for analysts to track how music features, such as energy, tempo, and lyrical content, evolve over time. Tools like sentiment analysis (e.g., Bing, NRC, AFINN) can help identify changes in listeners' emotional engagement with music. This will assist in predicting and responding to shifts in listener preferences and improving music recommendations.
3. Music Producers & Artists: Based on observed trends, artists might consider emphasizing authentic, emotionally sincere lyrics that resonate with listeners' post-pandemic emotional states. Additionally, incorporating acoustic elements and calm rhythms could align with the increasing demand for introspective, peaceful music that has gained popularity during the pandemic.
4. Our analysis of COVID-19 shows that recommender systems, using cluster analysis and time-series trends, can adapt to changing listener preferences during future disruptions, ensuring relevant content for users.

Appendix A: Time Series & Statistical Modeling Results

Figure A1: ACF plots for valence (pre vs. post-COVID)



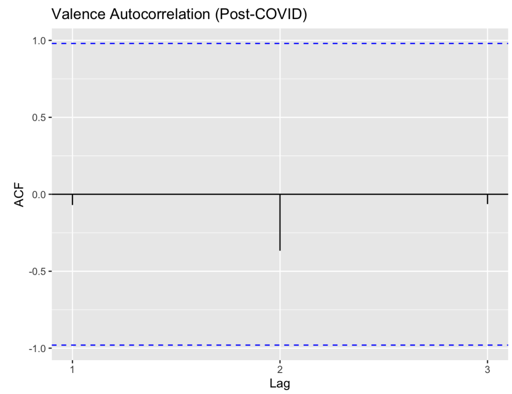
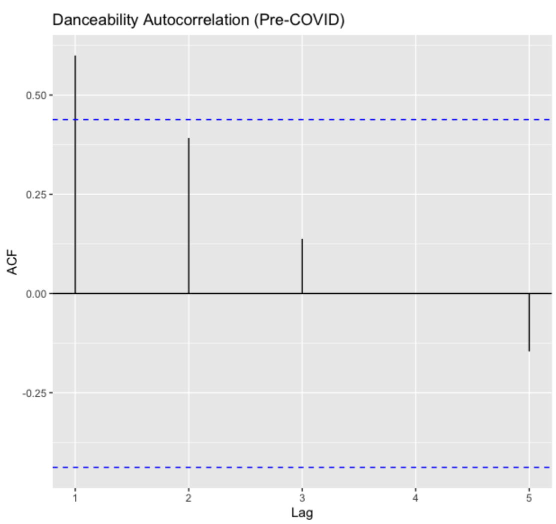


Figure A2: ACF plots for danceability (pre vs. post-COVID)



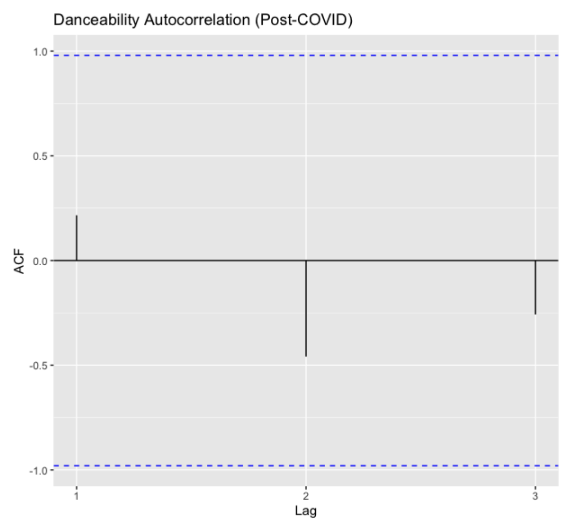


Figure A3: Drift Model, Valence

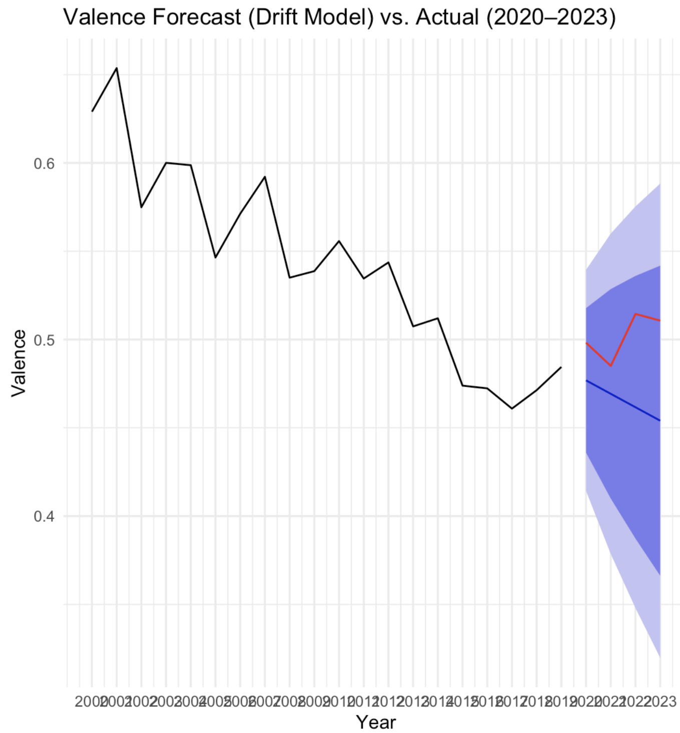


Figure A4: Drift Model, Danceability

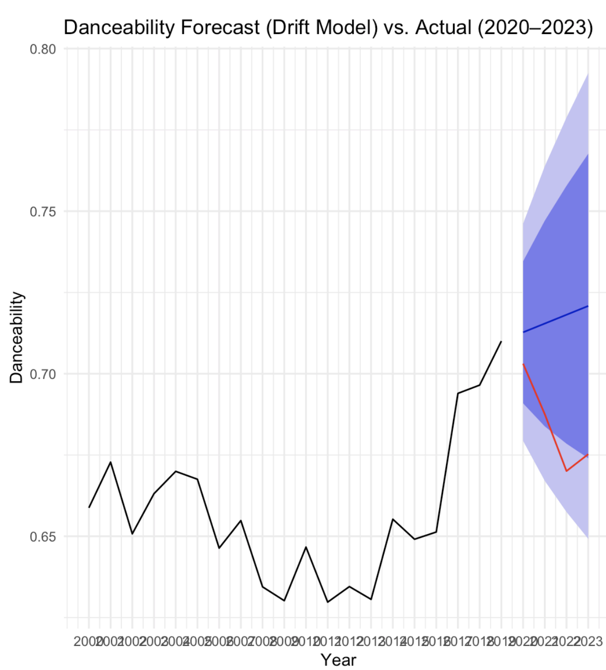


Figure A5: Drift Model, Energy

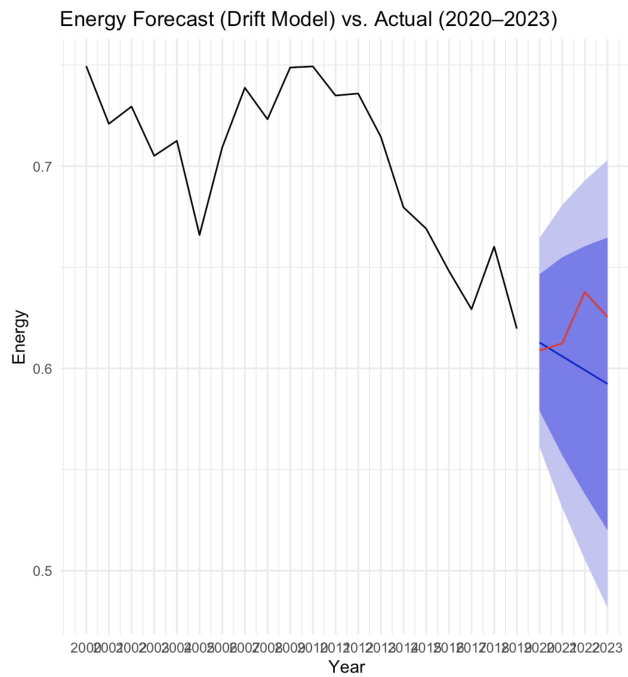
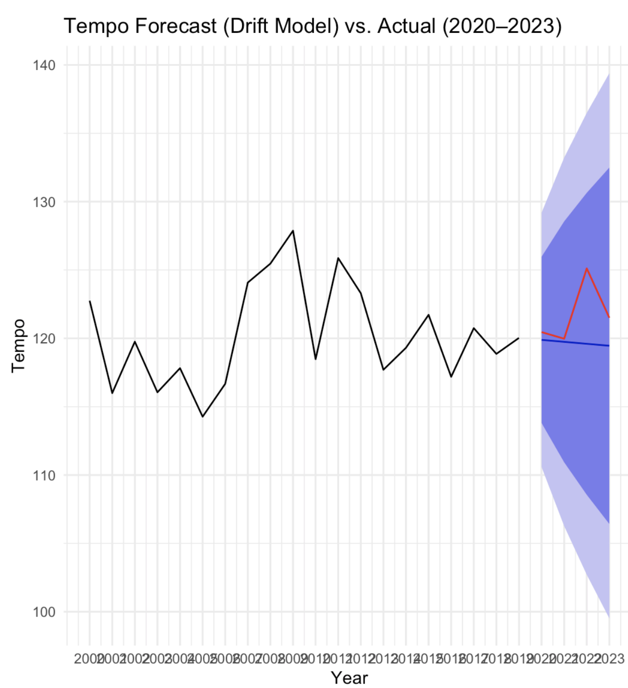


Figure A6: Drift Model, Tempo



Appendix B: Clustering Analysis Results

Figure B1: Elbow method for determining optimal number of clusters (Pre-COVID)

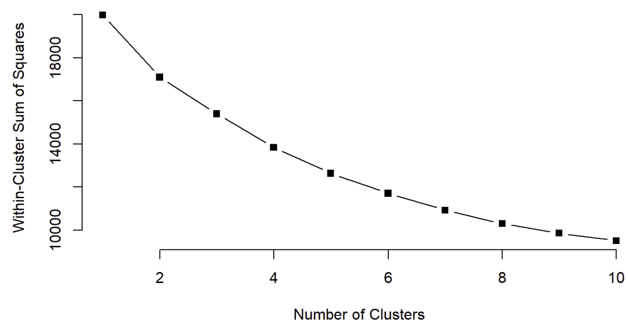


Figure B3: Ratio plot for Pre-COVID k-means clustering

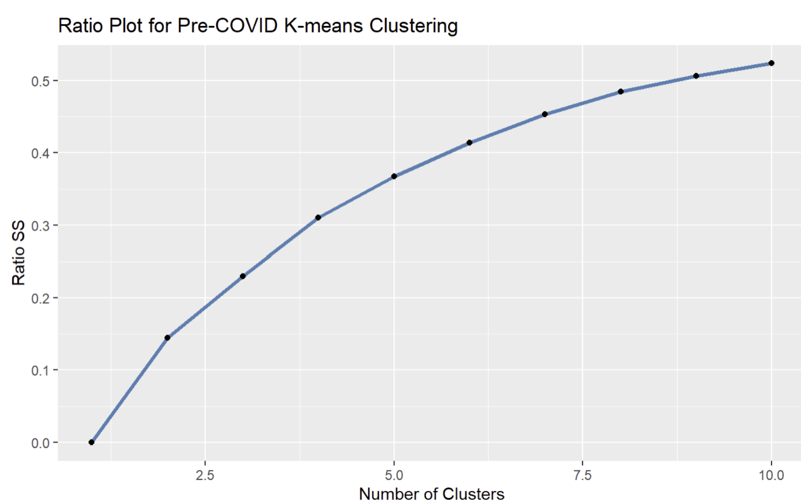


Figure B4: Silhouette analysis for Pre-COVID k-means clustering

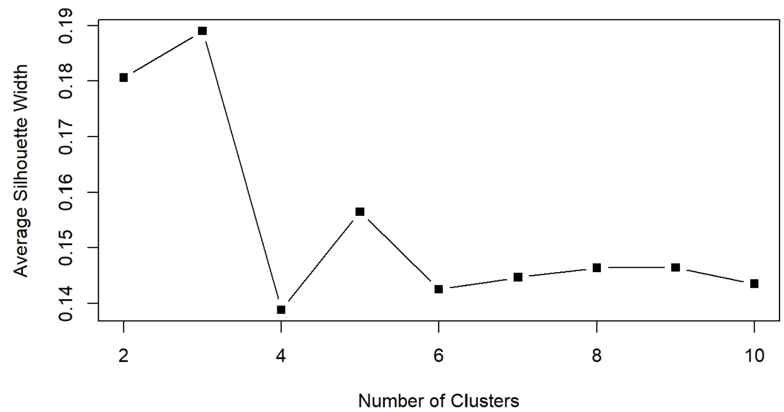


Figure B5: K-means clustering plots for Pre-COVID period

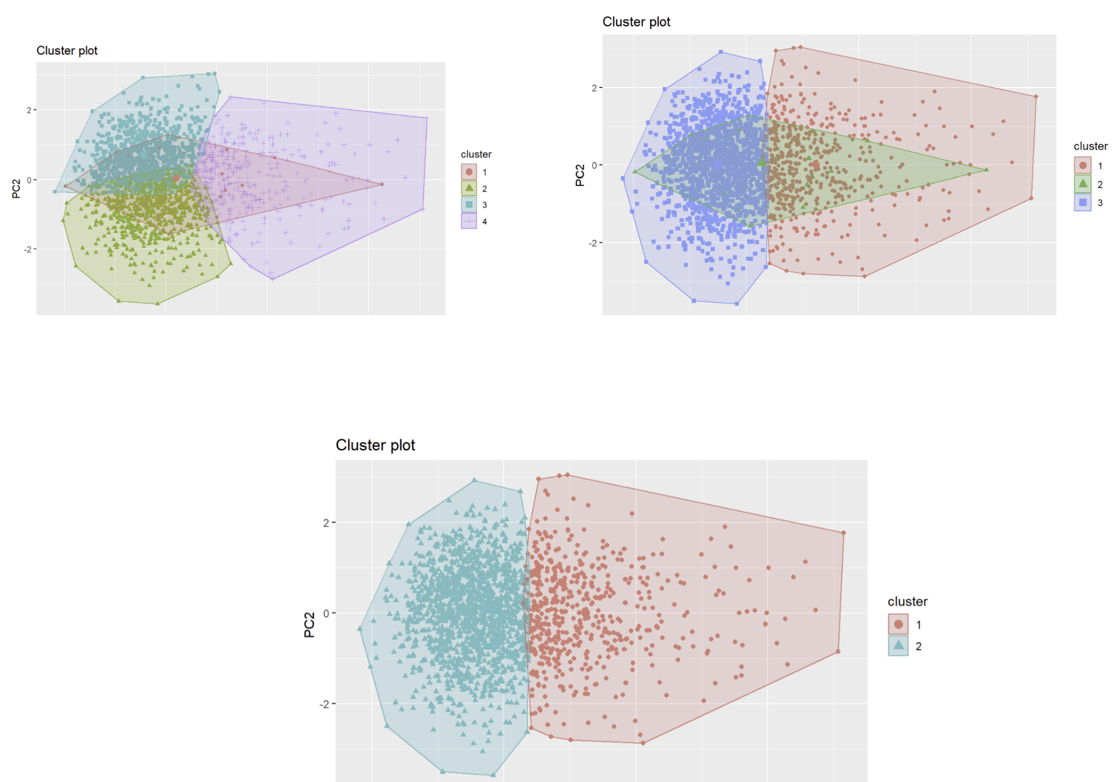


Figure B6: Pre-COVID dendrogram

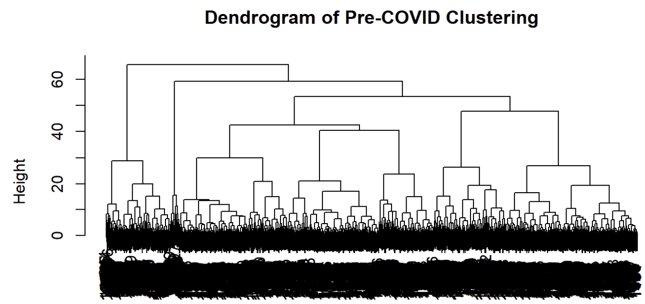


Figure B7: Pre-COVID GMM model selection (Mclust)

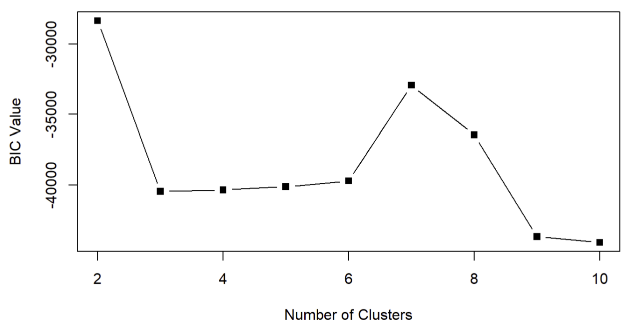


Figure B8: Elbow method for determining optimal number of clusters (pandemic)

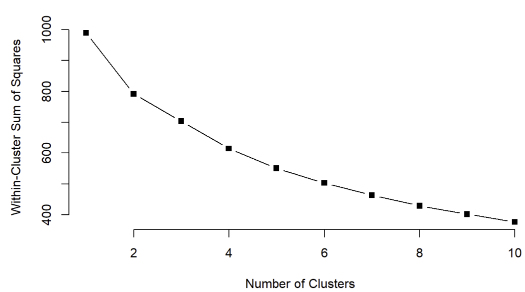


Figure B9: Ratio plot for pandemic period k-means clustering

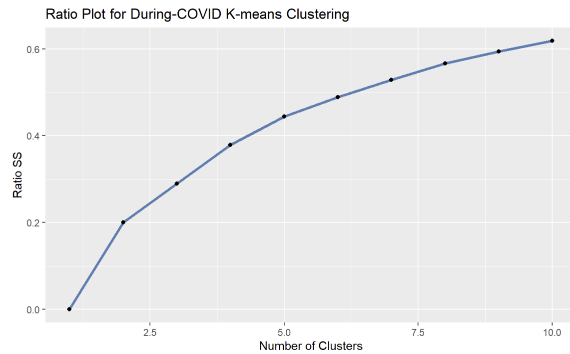


Figure B10: Silhouette analysis during pandemic k-means clustering

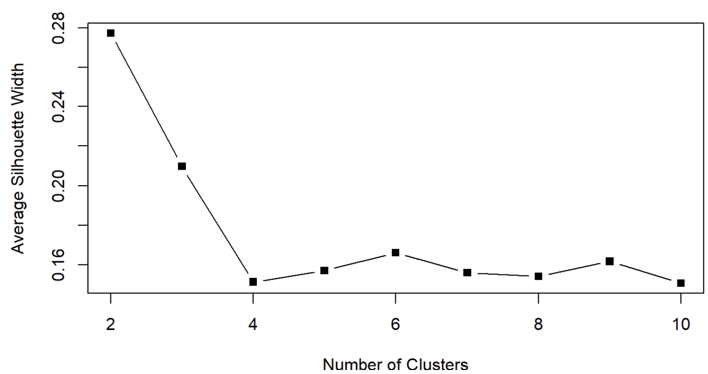


Figure B11: K-means clustering plots during COVID period

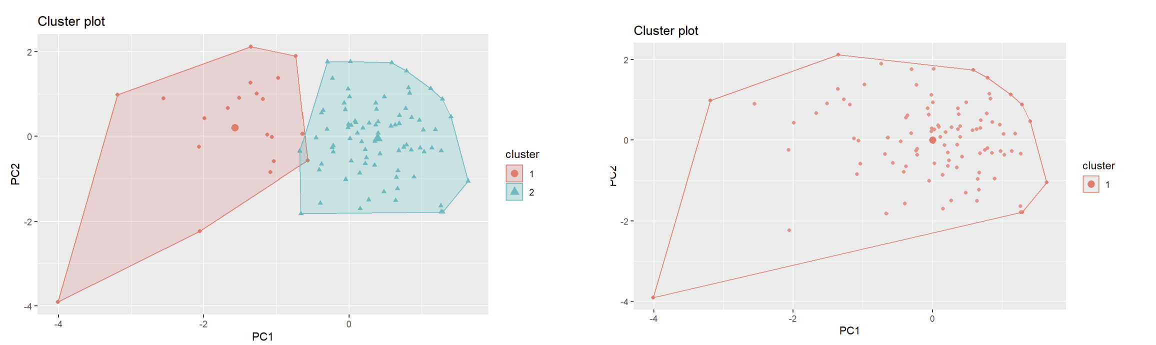


Figure B12: COVID dendrogram

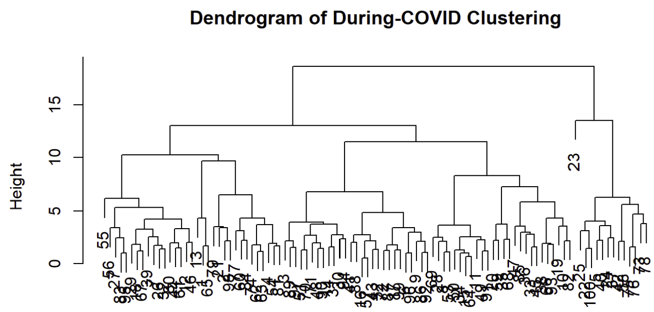


Figure B13: COVID GMM model selection (Mclust)

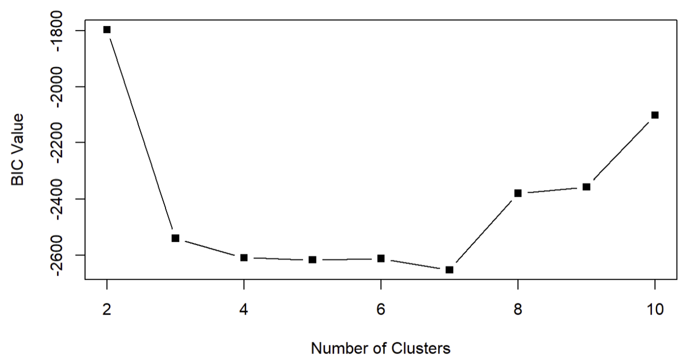


Figure B14: Elbow method for determining optimal number of clusters (Post-COVID)

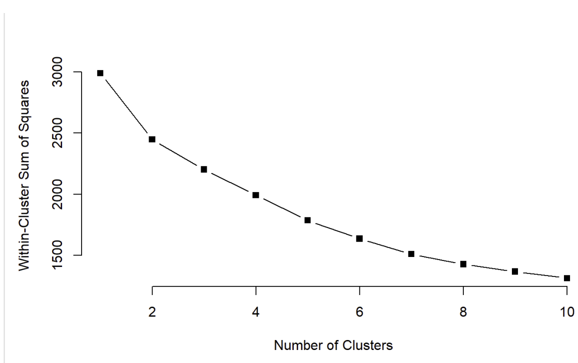


Figure B15: Ratio plot for Post-Covid period k-means clustering

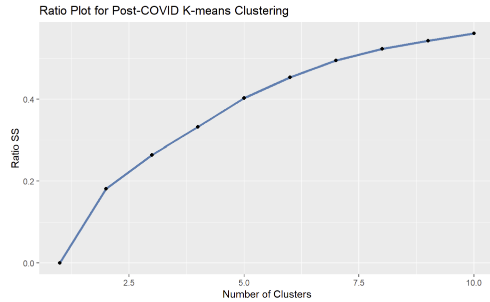


Figure B16: Silhouette analysis Post-COVID k-means clustering

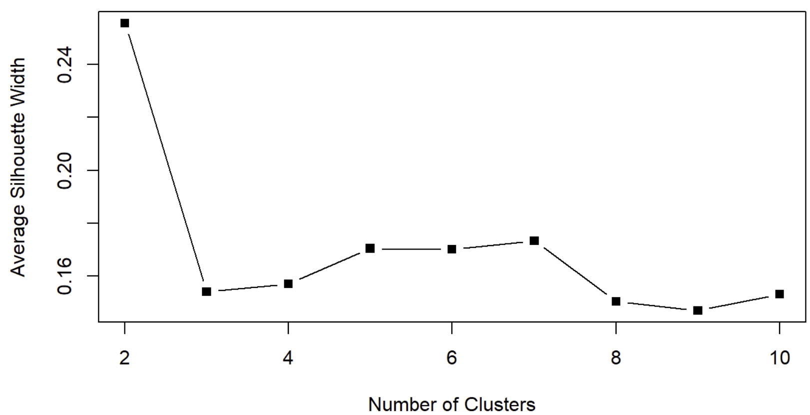


Figure B17: K-means clustering plots Post-COVID period

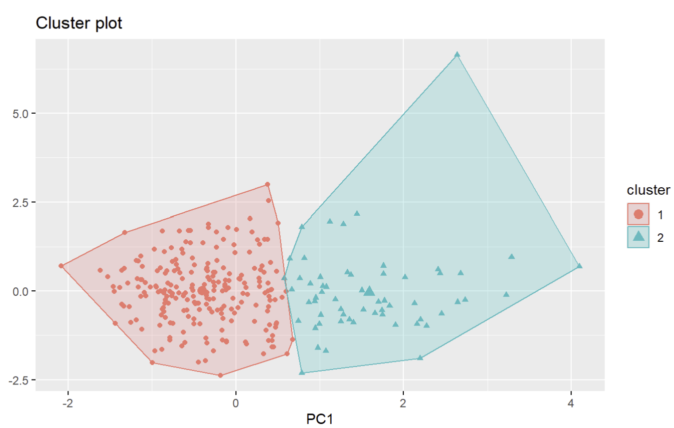


Figure B18: Post-COVID dendrogram

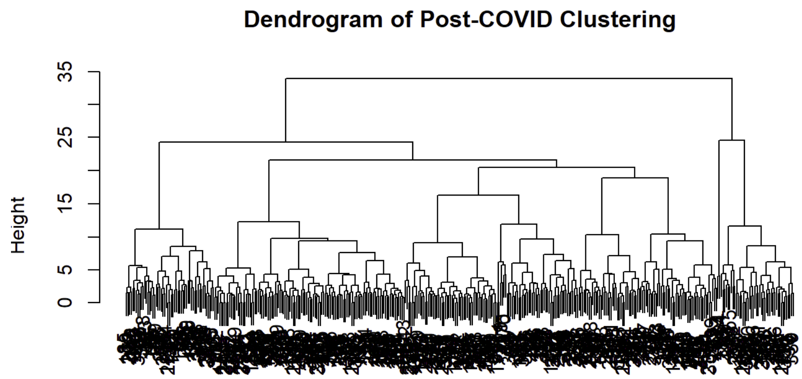
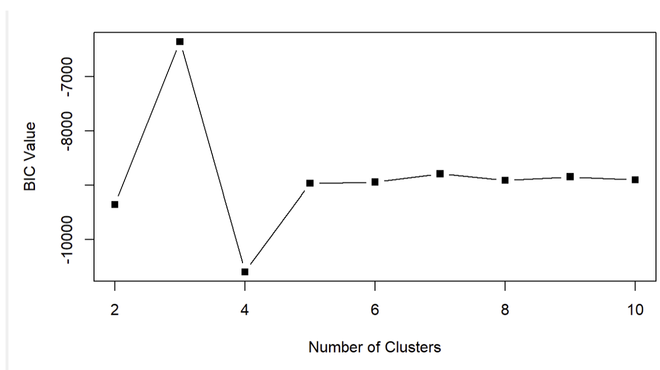


Figure B19: Post-COVID GMM model selection (Mclust)



Appendix C:Text Mining & Sentiment Analysis

Figure C1:Bar plot of top 25 words (pre- vs. post-COVID)

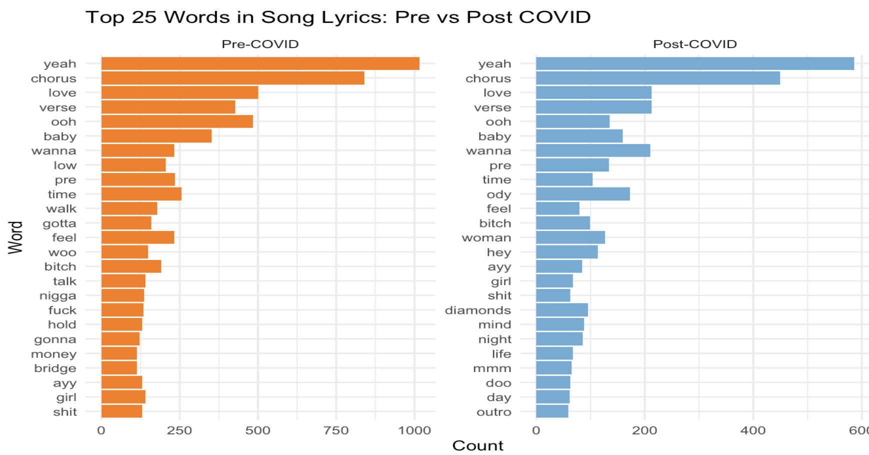


Figure C2: Proportion of Sentiment in Song Lyrics(pre- vs. post-COVID)

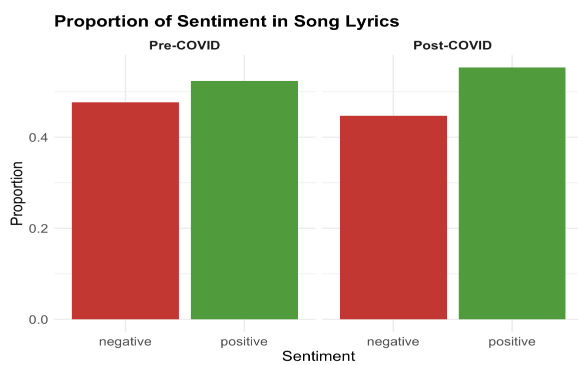


Figure C3: Proportion of various Emotions in Lyrics (pre- vs. post-COVID)

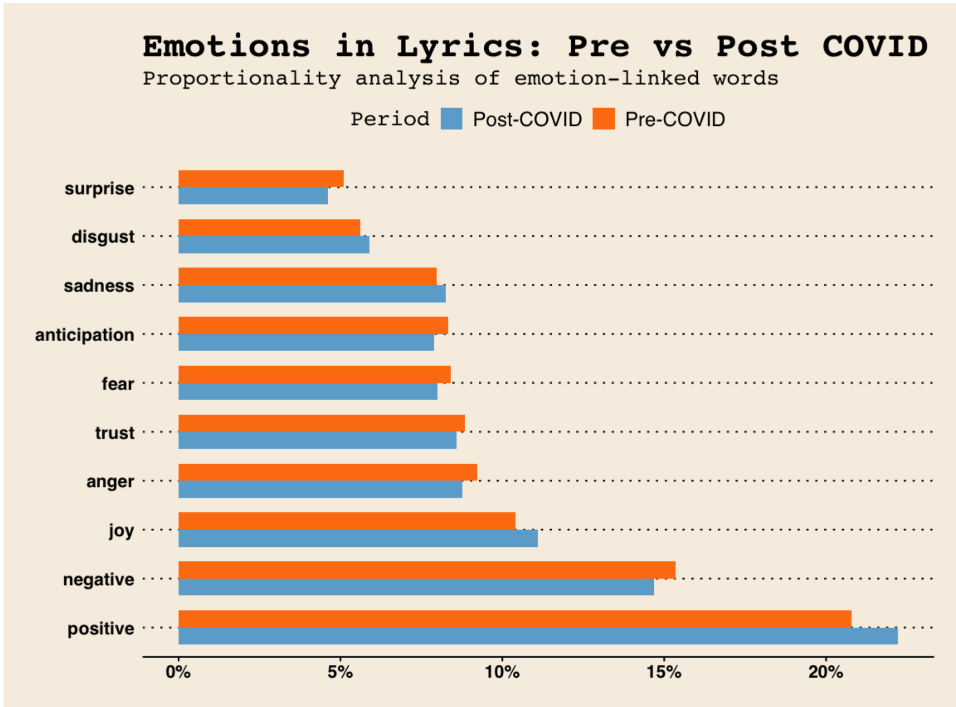


Figure C4: Sentiment Score Distribution in Lyrics (pre- vs. post-COVID)

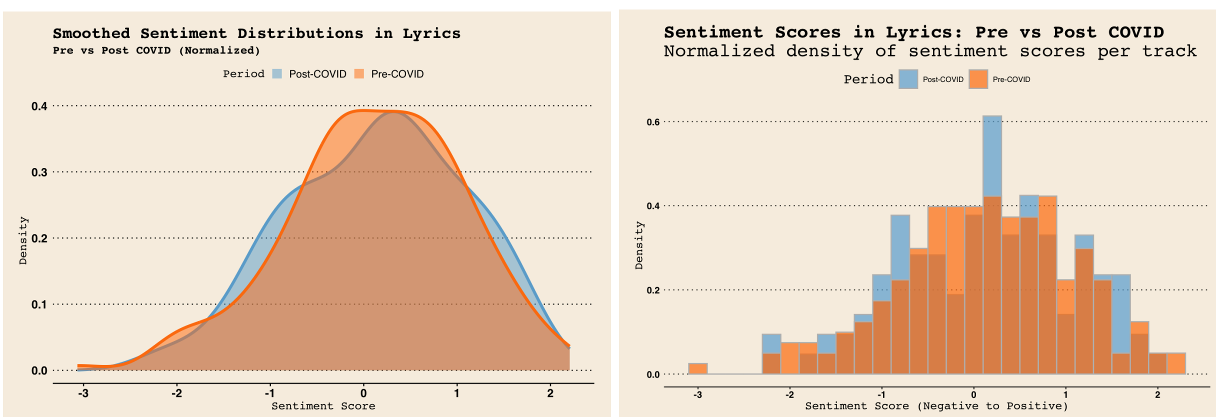


Figure C5: Top Predictors of Popularity Score from Linear Regression Model