# The Impact of COVID-19 on Musical Characteristics of Popular Songs on Spotify

Abdel Shehata

## Introduction

The COVID-19 pandemic profoundly disrupted daily life around the world beginning in early 2020.Social distancing measures and lockdown orders meant more time spent at home, altering work, school, social interaction and entertainment.For many, especially young people, music provided a crucial outlet during this challenging time.With 73% of Spotify's users between the ages of 18 and 34 in 2021, the platform offers a unique opportunity to examine how the listening habits of this demographic shifted during the pandemic (Cross River Therapy, 2024).

This project aims to investigate the potential impact of the COVID-19 pandemic on the music consumption patterns of young Spotify users in the United States. By analyzing data from Spotify's "Top 200" charts between 2016 and 2021, we will explore whether there were notable changes in the popularity and characteristics of the most-streamed songs before and during the pandemic period.

Our analysis will focus on various aspects of the top-charting songs, including their musical attributes, artist diversity, and thematic content. We will examine if there were any significant shifts in the types of music that resonated with young listeners during the pandemic, such as increased preferences for certain genres, emotional themes, or artist demographics.

Through this study, we seek to understand how the COVID-19 pandemic may have influenced the musical preferences and consumption habits of young people, who make up a significant portion of Spotify's user base. By exploring the relationship between societal challenges and music consumption, this project aims to contribute to the broader discussion of the cultural and psychological impacts of the pandemic on younger generations.

As music often serves as a reflection of the times and a coping mechanism during difficult periods, understanding the changes in listening habits during the pandemic can provide valuable insights into the experiences and emotional well-being of young people. This project ultimately aims to shed light on the role that music played in providing comfort, connection, and self-expression during a time of unprecedented global upheaval.

## Data

The dataset used in this project, collected by user Sunny Kakar on Kaggle, contains information on the "Top 200" and "Viral 50" charts published globally by Spotify. The data was obtained by utilizing the Spotify API to retrieve additional information for each song entry. For the purpose of this analysis, we will focus specifically on the musical charactertistics of "Top 200" charts, which represent the most popular songs on the platform. The charts are updated daily, and the dataset covers the period from January 1, 2017 until December 31st, 2021.

Each row in the dataset corresponds to a single day and contains aggregated information about the top 200 songs for that day. The musical characteristics of these songs, such as tempo, danceability, energy, and valence, are averaged across the top 200 tracks to provide a daily snapshot of the popular music landscape. For the purpose of the interrupted time series analysis, these daily values were further aggregated to calculate weekly averages due to computational limitations.The weekly averages of the musical characteristics will serve as our response variables in the analysis

Our response variables are nine musical characteristics: danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, and tempo.These variables capture various aspects of a song's sound and feel, such as its suitability for dancing, perceived intensity, presence of spoken words, acoustic nature, instrumental composition, audience presence, musical positivity, and overall tempo.For a comprehensive list and detailed descriptions of all the musical characteristics included in the dataset, please refer to the data dictionary provided with the dataset.

Our main predictor variables will be time, measured in days since the start of the dataset and a binary variable indicating the pre-COVID and during-COVID periods, with March 15, 2020, serving as the starting point of the COVID period.

We will also include an interaction term between time and the COVID indicator variable to capture any potential changes in the relationship between time and the musical characteristics during the pandemic. Additionally, a binary 'holiday' variable will be included to account for potential seasonal effects on music consumption patterns during the holiday months of November and December..

By analyzing the relationship between these predictor variables and the response variables, we aim to understand how the COVID-19 pandemic may have influenced the musical preferences and consumption habits of Spotify users over time.

This dataset provides an opportunity to explore the potential impact of a global crisis on the music industry and the listening behaviors of a young people.

Table 1: Summary Statistics and Biggest Track Name Based on Streams

| feature | PreCOVID | COVID |
|---|---|---|
| Acousticness | 0.21 | 0.25 |
| Danceability | 0.71 | 0.69 |
| Duration (Min) | 3.49 | 3.3 |
| Energy | 0.61 | 0.6 |
| Instrumentalness | 0.01 | 0.01 |
| Key | 1 | 1 |
| Liveness | 0.17 | 0.18 |
| Loudness | -6.44 | -6.77 |
| Mode | 1 | 1 |
| Speechiness | 0.14 | 0.13 |
| Tempo | 121.91 | 122.56 |
| Time Signature | 4 | 4 |
| Valence | 0.45 | 0.49 |
| Biggest Song (Streams) | Nonstop by Drake | Girls Want Girls by Drake |

## Data Exploration

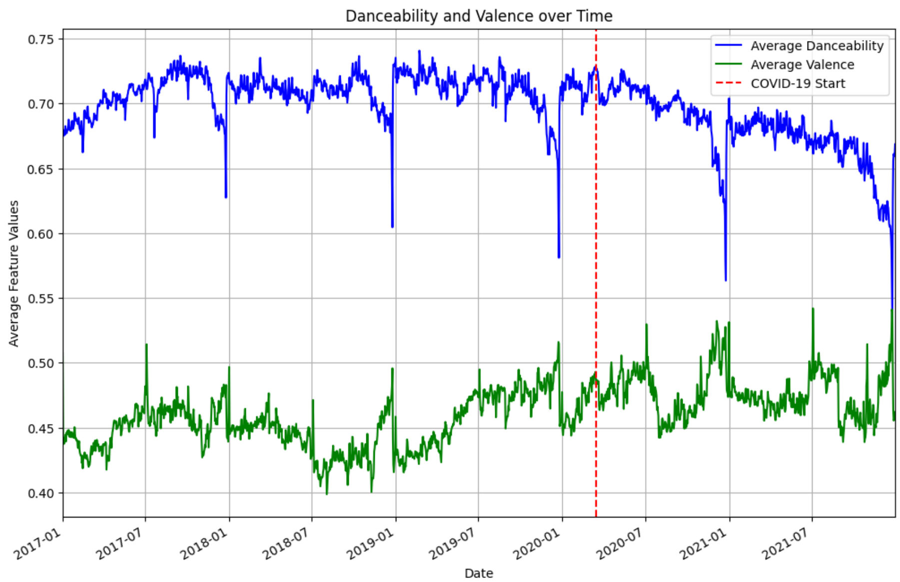
To begin our data exploration, we compared the average values and most common modes of various musical attributes between the pre-COVID and COVID periods.

The summary statistics table highlights some interesting patterns in the musical landscape between the pre-COVID and COVID periods. One notable observation is the slight decrease in the average danceability score, suggesting a shift towards less danceable tracks during the pandemic.This could be attributed to the closure of social gathering spaces and a preference for more introspective music.

Another change is the increase in the average acousticness score, indicating a slight preference for more acoustic-based music during the COVID period.This may be due to people seeking comfort and intimacy through music while spending more time at home. The average valence score also saw a small increase, suggesting a desire for more uplifting and positive music during challenging times.

Interestingly, the average duration of songs decreased slightly from 3.49 minutes in the pre-COVID period to 3.3 minutes during the COVID period.This could indicate a preference for shorter, more concise tracks, possibly reflecting shorter attention spans or a desire for quick musical experiences in a time of uncertainty.

Despite the changes in the musical landscape, Drake maintained his dominance, with his tracks being the biggest hits on spottily in both periods.However, the specific tracks differed, showcasing his adaptability to changing trends.




The plot illustrates the trends in average danceability and valence of songs over time, with a notable decrease in danceability after the start of the COVID-19 pandemic. In contrast, the valence of songs, representing the positivity or happiness of the music, continues its pre-existing upward trend even during the pandemic.

These observations suggest that the pandemic had a selective impact on certain musical attributes, particularly danceability, while the emotional positivity of music remained relatively unaffected. The extreme dips and upticks around December and January can be attributed to the influence of Christmas music, which often has distinct musical characteristics.

Overall, the analysis reveals that COVID-19 had an effect on the danceability of popular songs but not on their emotional tone, shedding light on the research question of whether the pandemic influenced musical characteristics.The plot also highlights the need to consider seasonal factors, such as Christmas music, when interpreting the data

## Methodology

In this study, an interrupted time series (ITS) analysis was performed to investigate the impact of the COVID-19 pandemic on the musical characteristics of songs appearing in Spotify's weekly Top 200 charts. The ITS design is particularly suitable for evaluating the effects of population-level interventions implemented at a clearly defined point in time on a specific outcome. In this case, the intervention is the onset of the COVID-19 pandemic, and the outcomes are the average musical attributes of popular songs on Spotify.

The analysis followed a step-by-step approach. First, the data was preprocessed to ensure the 'date' column was in the proper format and to create a binary 'holiday' variable indicating whether each week fell within the holiday months of November and December. The data was then aggregated to calculate weekly averages for each musical characteristic and the holiday indicator.

Next, variables for time, the COVID-19 intervention, and post-intervention time were created. The 'time' variable represented the number of weeks elapsed since the start of the study period, while 'intervention' was a binary variable indicating the pre-pandemic (0) and pandemic (1) periods. The 'post_intervention_time' variable was an interaction term between 'time' and 'intervention', capturing the time elapsed since the start of the pandemic.

To account for the autocorrelation in the time series data and to identify the best model for each musical characteristic, a function was created to fit generalized least squares (GLS) models with different combinations of autoregressive (AR) and moving average (MA) terms. The function used the corARMA correlation structure to model the autocorrelation, with p and q values ranging from 1 to 5. The best model for each musical characteristic was selected based on the Akaike Information Criterion (AIC), which balances the goodness-of-fit with the model complexity.

The GLS models included terms for the pre-pandemic trend ('time'), the level change at the start of the pandemic ('intervention'), the change in slope during the pandemic period ('post_intervention_time'), and the holiday indicator ('holiday'). By including the holiday variable, the models accounted for the potential impact of seasonal trends on the musical characteristics.

The selected best models for each musical characteristic were then used to assess the impact of the COVID-19 pandemic on the musical attributes of popular songs on Spotify. The coefficients of the 'intervention' and 'post_intervention_time' variables provided insights into the immediate and gradual changes in musical preferences following the onset of the pandemic, while controlling for pre-existing trends and seasonal effects.

In summary, the ITS analysis using GLS models with autocorrelation corrections and holiday adjustments provided a robust framework for evaluating the impact of the COVID-19 pandemic on the musical characteristics of popular songs on Spotify. The step-by-step approach, combined with the use of appropriate statistical methods and model selection techniques, ensured that the findings were reliable and interpretable.

## Results

Table 2: Transposed Results of GLS Models for Musical Features

| | Intercept | time | intervention | post intervention time | holiday | Log Likelihood | AIC. |
|---|---|---|---|---|---|---|---|
| danceability | 0.061*** | 0.0001*** | -0.023 | -0.001*** | -0.172*** | 291.003 | -566.007 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| energy | 0.100*** | -0.0001*** | -0.013 | 0.0004*** | -0.188*** | 276.662 | -537.323 |
| loudness | 0.083*** | -0.00002 | -0.042 | -0.0001 | -0.262*** | 202.519 | -389.039 |
| speechiness | 0.050*** | 0.00003** | -0.019 | -0.0004*** | -0.130*** | 352.766 | -689.533 |
| acousticness | -0.256*** | 0.0003*** | -0.039 | -0.0003*** | 0.236*** | 236.878 | -457.755 |
| instrumentalness | -0.060*** | 0.0001*** | -0.059*** | 0.0003*** | 0.020** | 380.319 | -744.639 |
| liveness | -0.109*** | 0.0001*** | 0.039*** | -0.0001* | 0.046*** | 388.855 | -761.709 |
| valence | -0.101*** | 0.0001*** | 0.088*** | -0.0002*** | 0.063*** | 297.246 | -578.491 |
| tempo | 0.009 | -0.00002** | 0.017 | 0.00003 | 0.017* | 403.639 | -791.278 |

*Note:* Note: *p<0.1; **p<0.05; ***p<0.01

The interrupted time series analysis using generalized least squares (GLS) models provided valuable insights into the impact of the COVID-19 pandemic on the musical characteristics of popular songs on Spotify's weekly Top 200 charts. The results suggest that the pandemic was associated with significant changes in several musical attributes, with some features exhibiting immediate level changes and others showing gradual trend changes over time.

One notable finding was the decrease in the danceability trend following the onset of the pandemic. The model showed that prior to the pandemic, the danceability of songs exhibited a slight positive trend (0.0001, $p<0.01$), indicating a gradual increase in danceable features over time. However, after the start of the pandemic, there was a significant decrease in the trend (-0.001, $p<0.01$), suggesting that songs became less danceable during the pandemic period. To put this into context, if a song had a danceability score of 0.7 at the beginning of the pandemic, the model suggests that after 10 weeks, the danceability score would have decreased by 0.01 ($0.7 - 0.001 \times 10$) to 0.69, assuming all other factors remained constant. This change could be attributed to the closure of clubs, dance venues, and the cancellation of live events, which may have shifted listeners' preferences towards less danceable music. Additionally, the holiday period was associated with a significant decrease in danceability (-0.172, $p<0.01$), which might be due to the prevalence of slower-paced, nostalgic holiday songs during November and December.

Interestingly, the energy of songs increased during the pandemic period, despite the decrease in danceability. The model indicated that prior to the pandemic, the energy of songs showed a slight negative trend (-0.0001, $p<0.01$). However, after the start of the pandemic, there was a significant positive change in the trend (0.0004, $p<0.01$), suggesting that songs became more energetic during the pandemic period. This could indicate that listeners sought out songs with higher energy levels to cope with the stress and uncertainty of the pandemic, even if the songs were not necessarily designed for dancing. The increase in energy might also reflect a desire for more uplifting and motivational music during challenging times.

The analysis also revealed a decrease in the speechiness trend during the pandemic. The model showed that before the pandemic, the speechiness of songs had a slight positive trend (0.00003, $p<0.05$). However, following the pandemic's onset, there was a significant decrease in the trend (-0.0004, $p<0.01$), indicating that songs contained less spoken word content during the pandemic period. This change could be related to a shift in listener preferences towards more melodic and instrumental music, as a means of escape or relaxation during the pandemic. The holiday period was also associated with a significant decrease in speechiness (-0.130, $p<0.01$), which might be attributed to the traditional emphasis on sung lyrics in holiday songs.

Another notable finding was the increase in the instrumentalness of songs during the pandemic period, following an initial level decrease. The model revealed that prior to the pandemic, the instrumentalness of songs exhibited a slight positive trend (0.0001, $p<0.01$). However, immediately following the pandemic's onset, there was a significant level decrease (-0.059, $p<0.01$), suggesting an initial drop in instrumental content. But as the pandemic progressed, there was a significant positive change in the trend (0.0003, $p<0.01$), indicating that songs became more focused on instrumental elements over time. This could reflect a growing appreciation for the emotional and expressive qualities of music without lyrics, or the rise of certain genres, such as ambient and electronic music, which often emphasize instrumental compositions.

The valence of songs, which represents the positivity or happiness of the music, exhibited an initial level increase but then decreased over time during the pandemic period. The model showed that prior to the pandemic, the valence of songs had a positive trend (0.0001, $p<0.01$), indicating a gradual increase in positivity over time. Immediately following the pandemic's onset, there was a significant level increase (0.088, $p<0.01$), suggesting an initial boost in the positivity of songs. However, as the pandemic progressed, there was a significant decrease in the trend (-0.0002, $p<0.01$), indicating that songs became less positive over time. This suggests that while songs initially became more positive, possibly as a response to the need for uplifting content during the early stages of the pandemic, the prolonged duration of the crisis may have led to a gradual shift towards more introspective and emotionally complex music. The holiday period was associated with an increase in valence (0.063, $p<0.01$), which is consistent with the generally positive and celebratory nature of holiday music.

Lastly, the tempo of songs did not show significant changes associated with the onset of the pandemic, although there was a slight decrease in the overall trend (-0.00002, $p<0.05$). This suggests that the pace of popular music remained relatively stable during the pandemic period, despite changes in other musical attributes. The holiday period was associated with a slight increase in tempo (0.017, $p<0.1$), which might be due to the presence of more upbeat and festive songs during the holiday season.

In conclusion, the interrupted time series analysis revealed that the COVID-19 pandemic was associated with significant changes in the musical characteristics of popular songs on Spotify. The model coefficients provided valuable insights into the immediate and gradual changes in musical attributes, such as the decrease in danceability, increase in energy, decrease in speechiness, increase in instrumentalness, and the initial increase followed by a decrease in valence. These changes likely reflect the emotional and social impact of the pandemic on listeners, as well as the ways in which music was used as a coping mechanism and a means of connection during challenging times.The holiday period also had a significant impact on musical attributes, highlighting the influence of seasonal and cultural factors on music consumption patterns.While these findings provide valuable insights into the relationship between the pandemic and music preferences, it is important to note that they are associative and do not imply a causal relationship.

## Discussion

The interrupted time series analysis conducted in this study provides valuable insights into the impact of the COVID-19 pandemic on the musical characteristics of popular songs on Spotify's weekly Top 200 charts.The results suggest that the pandemic was associated with significant changes in several musical attributes, reflecting shifts in listener preferences and the emotional and social impact of the crisis. These findings align with the original research question and provide evidence that the pandemic indeed influenced the music consumption patterns of young Spotify users.

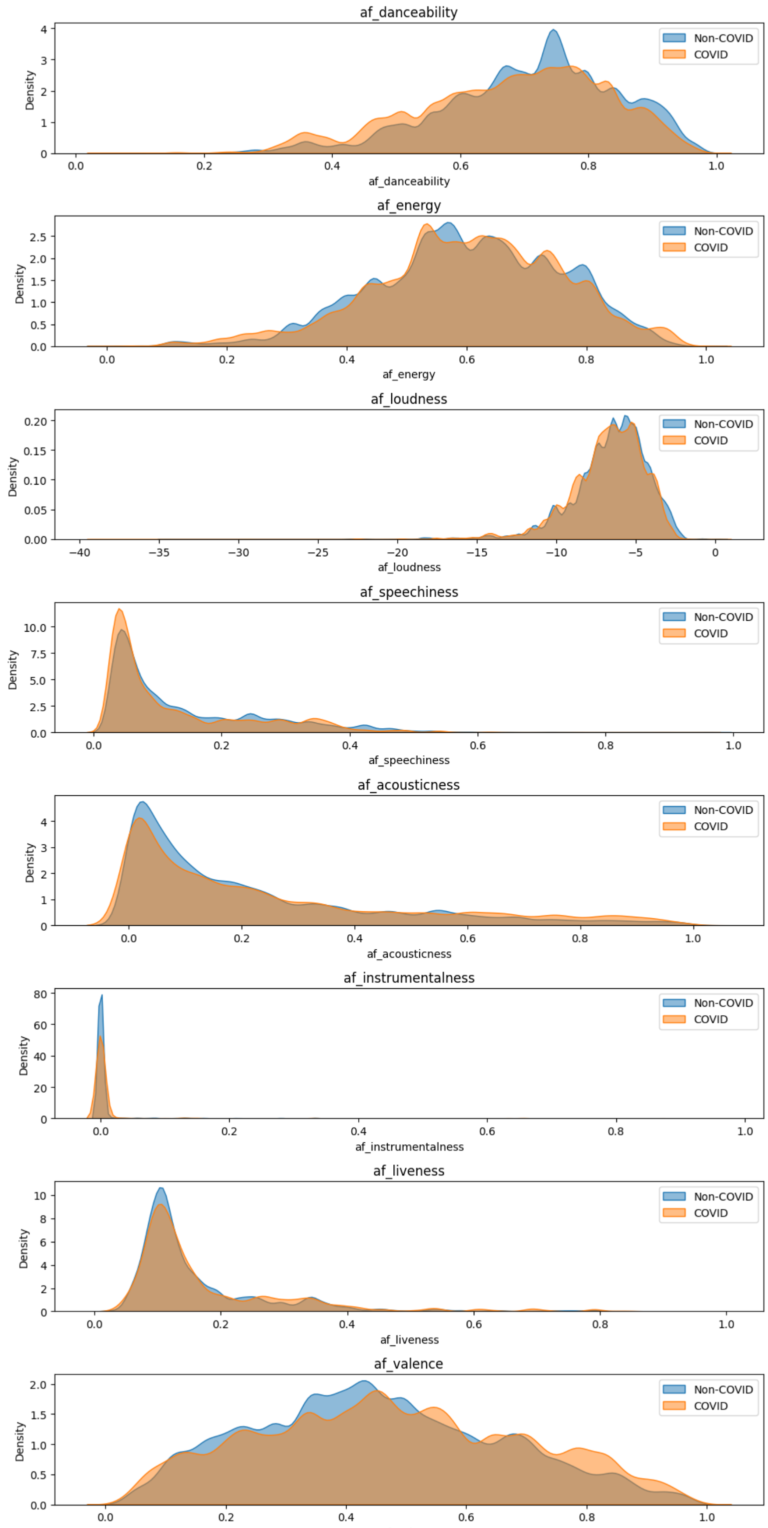
The observed changes in musical attributes, such as the decrease in danceability, increase in energy, decrease in speechiness, increase in instrumentalness, and the initial increase followed by a gradual decrease in valence, suggest that listeners gravitated towards music that reflected their emotional states and served as a coping mechanism during the pandemic.These changes highlight the potential for music to serve as a lens through which to understand collective emotional states and coping mechanisms during times of crisis.

However, it is important to acknowledge the limitations of this study. The reliance on data from a single platform (Spotify) and the focus on the Top 200 charts may not capture the full range of musical diversity or preferences. Additionally, the potential for confounding factors to influence the results cannot be ruled out. Future research could explore the impact of the pandemic on music consumption patterns across a wider range of platforms and genres, control for a broader range of potential confounding factors, and employ experimental or quasi-experimental designs to test causal hypotheses. Despite these limitations, this study makes a valuable contribution to the understanding of the cultural and psychological impacts of the COVID-19 pandemic on young people.

Future research could build on these findings by exploring the impact of the pandemic on music consumption patterns across a wider range of platforms and genres, controlling for a broader range of potential confounding factors, and employing experimental or quasi-experimental designs to test causal hypotheses.Additionally, qualitative research methods could provide valuable insights into the subjective experiences and motivations of listeners during the pandemic, complementing the quantitative findings of this study.

In conclusion, this study provides evidence that the COVID-19 pandemic was associated with significant changes in the musical characteristics of popular songs on Spotify, reflecting the emotional and social impact of the crisis on young listeners. The findings contribute to the broader discussion of the cultural and psychological impacts of the pandemic and underscore the importance of considering the role of music in people's lives during challenging times.

# Appendix


af_danceability
Density
af_danceability
Non-COVID
COVID
af_energy
Density
af_energy
Non-COVID
COVID
af_loudness
Density
af_loudness
Non-COVID
COVID
af_speechiness
Density
af_speechiness
Non-COVID
COVID
af_acousticness
Density
af_acousticness
Non-COVID
COVID
af_instrumentalness
Density
af_instrumentalness
Non-COVID
COVID
af_liveness
Density
af_liveness
Non-COVID
COVID
af_valence
Density
af_valence
Non-COVID
COVID

Results of GLS Models for Musical Features

| | Dependent variable: |
|---|---|
| | af_tempo |
| time | -0.001** |
| | (0.0004) |
| intervention | 0.486 |
| | (0.398) |
| post_intervention_time | 0.001 |
| | (0.001) |
| holiday | 0.506* |
| | (0.265) |
| Constant | 122.526*** |
| | (0.236) |
| Observations | 261 |
| Log Likelihood | -477.986 |
| Akaike Inf. Crit. | 971.972 |
| Bayesian Inf. Crit. | 1,000.488 |
| Note: | *p<0.1; **p<0.05; ***p<0.01 |

Results of GLS Models for Musical Features

| | Dependent variable: | | | |
|---|---|---|---|---|
| | af_acousticness | af_instrumentalness | af_liveness | af_valence |
| | (1) | (2) | (3) | (4) |
| time | 0.0001*** | 0.00000*** | 0.00001*** | 0.00002*** |
| | (0.00001) | (0.00000) | (0.00000) | (0.00000) |
| intervention | -0.009 | -0.004*** | 0.005*** | 0.019*** |

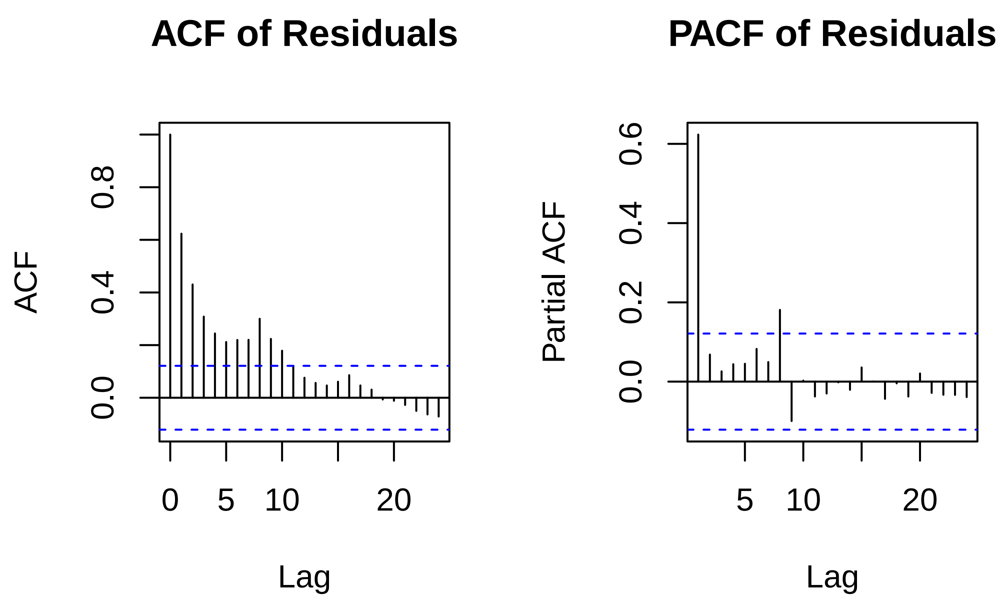
| | | | | |
|---|---|---|---|---|
| | (0.006) | (0.001) | (0.002) | (0.004) |
| post_intervention_time | -0.0001*** | 0.00002*** | -0.00001* | -0.00003*** |
| | (0.00001) | (0.00000) | (0.00000) | (0.00001) |
| holiday | 0.057*** | 0.001** | 0.006*** | 0.013*** |
| | (0.004) | (0.001) | (0.001) | (0.003) |
| Constant | 0.166*** | 0.005*** | 0.156*** | 0.439*** |
| | (0.004) | (0.001) | (0.001) | (0.003) |
| Observations | 261 | 261 | 261 | 261 |
| Log Likelihood | 608.521 | 1,116.912 | 938.128 | 701.711 |
| Akaike Inf. Crit. | -1,201.042 | -2,217.824 | -1,860.256 | -1,387.422 |
| Bayesian Inf. Crit. | -1,172.526 | -2,189.308 | -1,831.740 | -1,358.906 |
| Note: | | | | *p<0.1; **p<0.05; ***p<0.01 |

Results of GLS Models for Musical Features

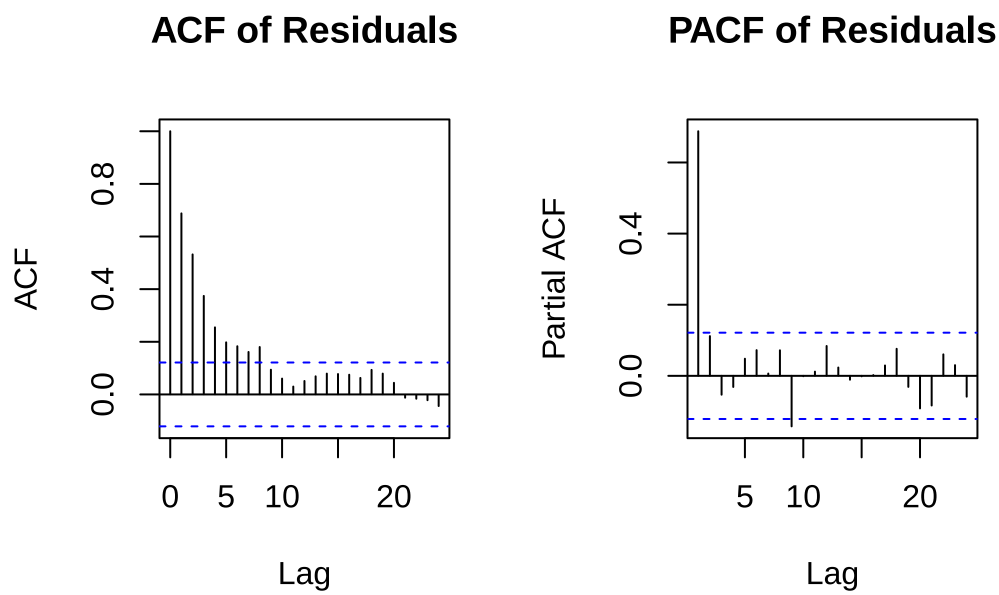
| | Dependent variable: | | | |
|---|---|---|---|---|
| | af_danceability | af_energy | af_loudness | af_speechiness |
| | (1) | (2) | (3) | (4) |
| time | 0.00001*** | -0.00002*** | -0.00004 | 0.00000** |
| | (0.00000) | (0.00000) | (0.0001) | (0.00000) |
| intervention | -0.003 | -0.002 | -0.099 | -0.002 |
| | (0.003) | (0.003) | (0.070) | (0.002) |
| post_intervention_time | -0.0001*** | 0.0001*** | -0.0002 | -0.0001*** |
| | (0.00001) | (0.00001) | (0.0002) | (0.00000) |
| holiday | -0.024*** | -0.029*** | -0.624*** | -0.016*** |
| | (0.002) | (0.002) | (0.047) | (0.001) |
| Constant | 0.707*** | 0.618*** | -6.458*** | 0.143*** |
| | (0.002) | (0.002) | (0.041) | (0.001) |

| | | | | |
|---|---|---|---|---|
| Observations | 261 | 261 | 261 | 261 |
| Log Likelihood | 799.872 | 762.116 | -24.010 | 905.889 |
| Akaike Inf. Crit. | -1,583.744 | -1,508.233 | 64.019 | -1,795.778 |
| Bayesian Inf. Crit. | -1,555.228 | -1,479.717 | 92.536 | -1,767.261 |
| Note: | | | | *p<0.1; **p<0.05; ***p<0.01 |

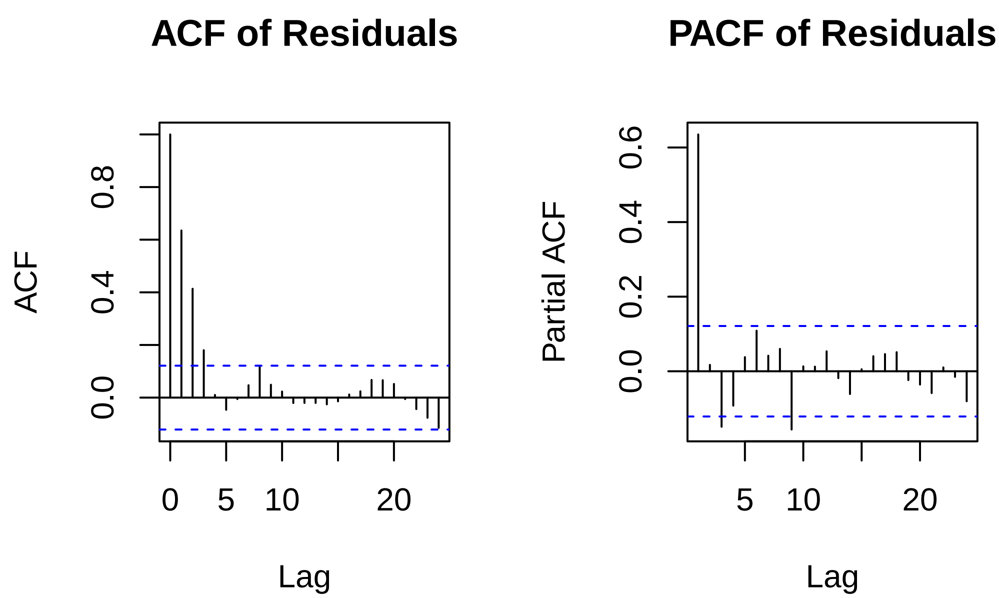
Feature: af_danceability



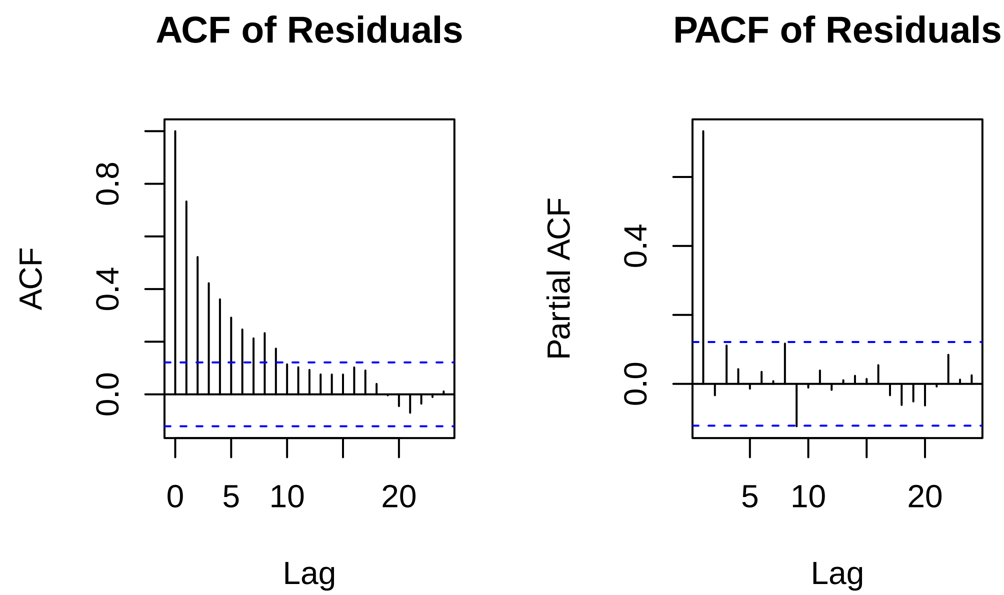

Feature: af_energy



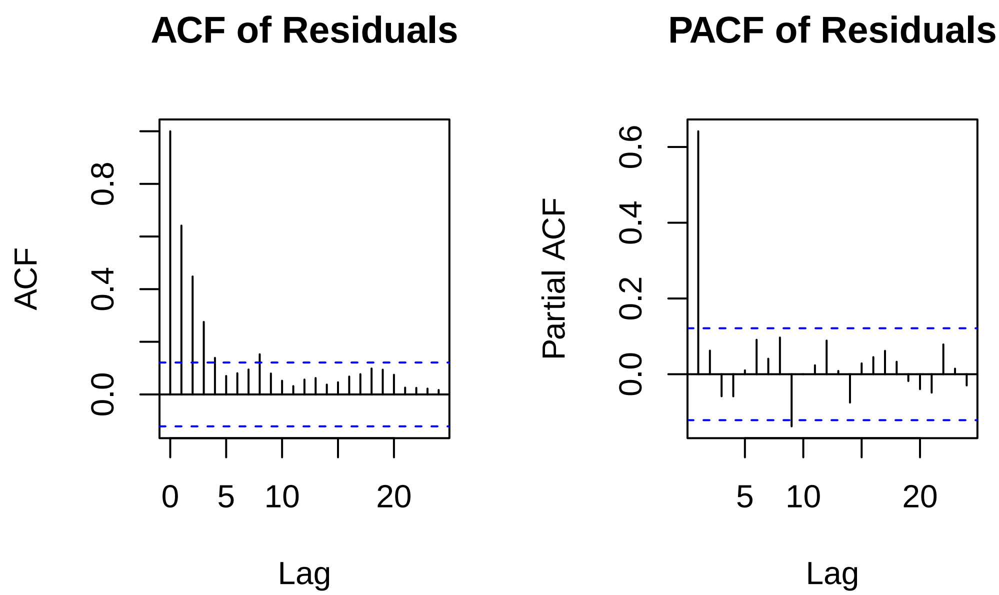

`Feature: af_loudness`



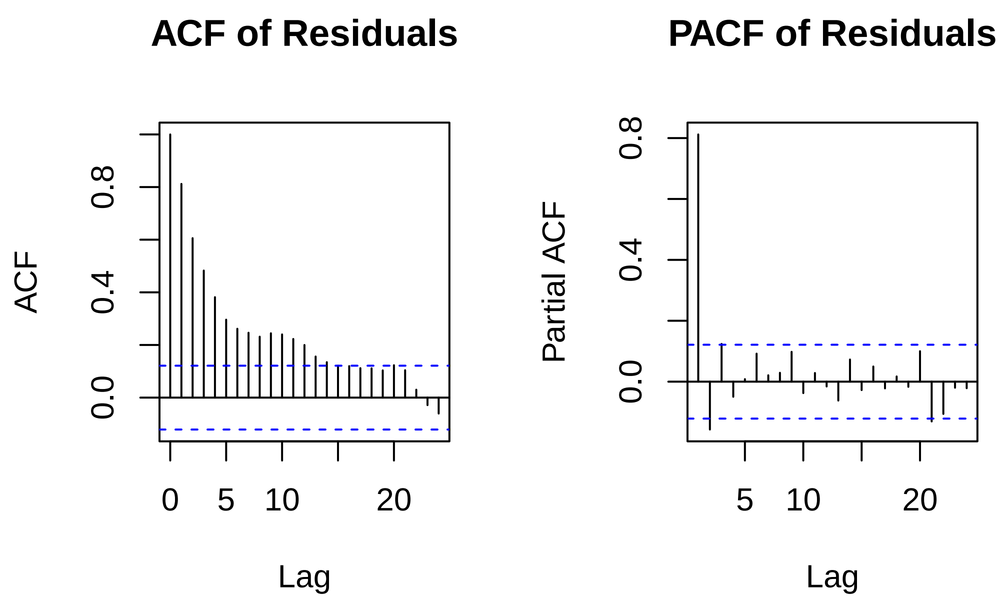
```
Feature: af_speechiness
```



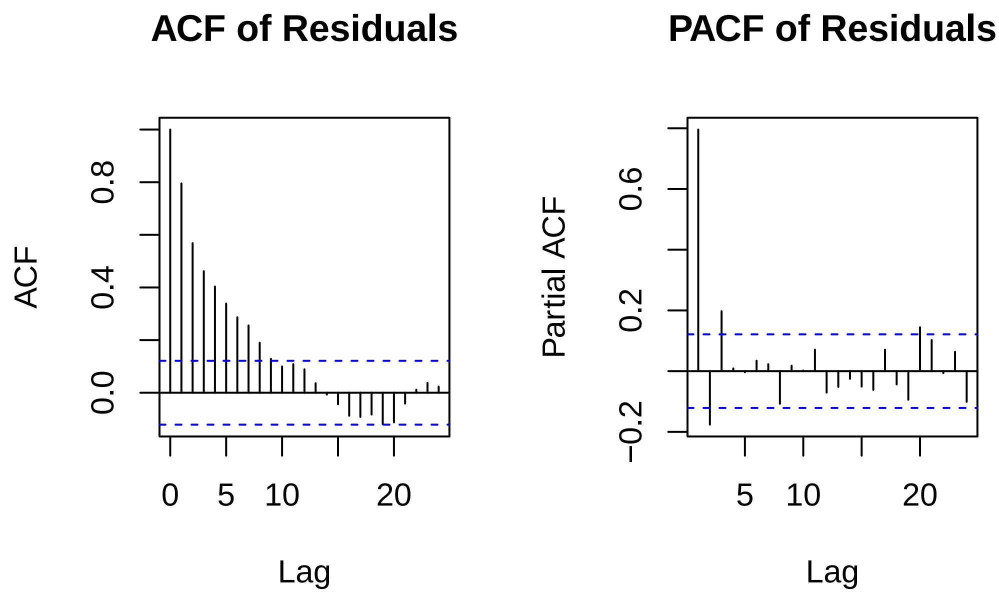

```
Feature: af_acousticness
```



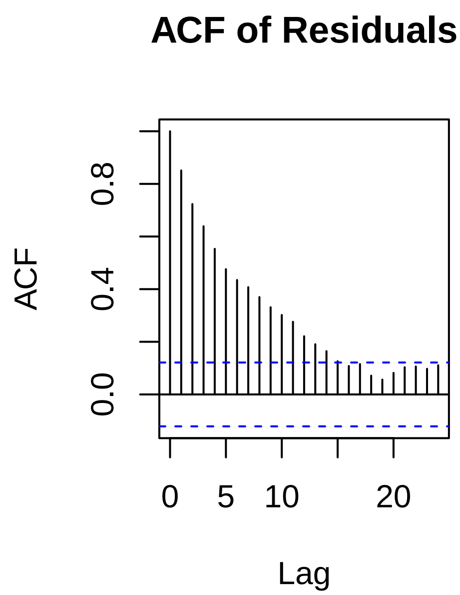

Feature: af_instrumentalness

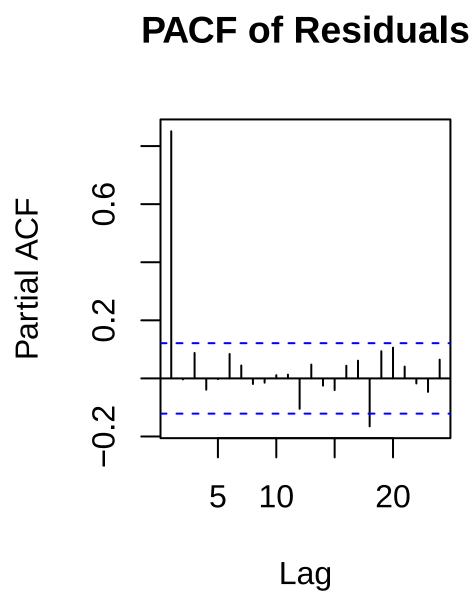


Feature: af_liveness

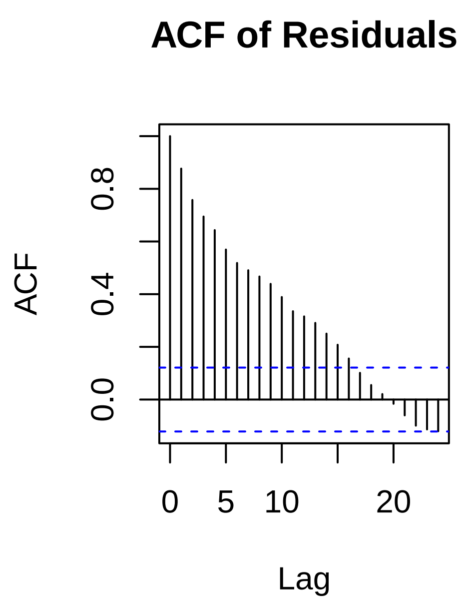


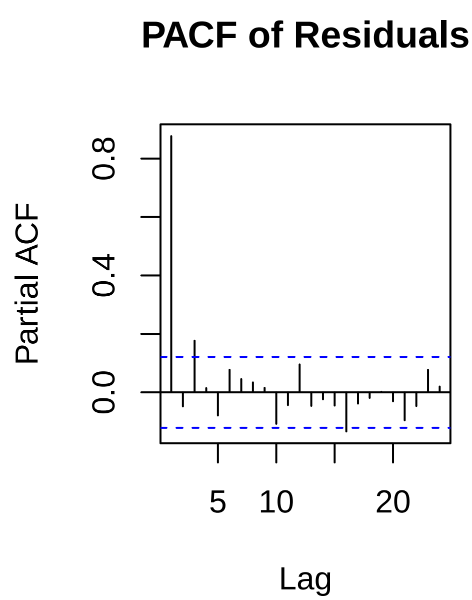

Feature: af_valence




Feature: af_tempo



## Work Cited

"Spotify Statistics: 79+ Intriguing Consumption Statistics in Music." Cross River Therapy, 2024, www.crossrivertherapy.com/research/spotify-statistics. Accessed 30 Apr. 2024 Sunny Kakar. Spotify Charts: All Audio Data. Kaggle, 2024. www.kaggle.com/datasets/sunnykakar/spotify-charts-all-audio-data. Accessed 30 Apr. 2024.