

Emotional Trends in Popular Music and U.S. Economic Conditions

Jeremy Han, Tianyu Zhao, Junyi (Daniel Tang), Hillary Metcheka, Bryce Grover

1 Abstract

2 Introduction

3 Related Work

4 Data

4.0.1 Data Collection

This project explores the change in emotional traits of popular music and the market performance of a leading streaming company before, during and after COVID-19. The project focuses on the 2017-2021 period, providing a unique viewpoint for interpreting shifts in cultural attitudes and financial turbulence associated with the pandemic. Although earlier iterations of the project considered unemployment data as a macroeconomic indicator, the final analysis excluded unemployment data due to limited incremental insight and redundancy with other economic measures. Instead, the final dataset emphasizes **Spotify streaming behavior, song-level audio features, and Spotify stock price data**, which together provide a more direct connection between music consumption, emotional sentiment, and economic performance.

The final analysis uses three primary datasets:

Dataset 1: Spotify Top 200 Chart Data * We used Spotify's Top 200 global chart data to map out large-scale listening behavior and preference. All these features were recorded in the dataset, which represents the daily rankings of the most consumed songs found in the world. To decrease background noise and focus on the broad trends, we organized daily observations into monthly values which facilitate an accurate and consistent temporal comparison between pre-COVID, COVID, and post-COVID periods.

Dataset 2: Spotify Audio Features * Spotify Audio Features Data was combined with the chart data in order to estimate the emotional and acoustic properties of popular songs. Important features such as valence, energy, danceability, liveness, loudness, tempo and duration, are identified. These machine-learned features act as a quantitative indicator of musical sentiment and allow for formally perform statistical testing on such changes in emotional tone over time.

Dataset 3: Spotify Stock Price Data * We also incorporated monthly Spotify (SPOT) stock price data, to model firm market performance and sentiment within the market. This dataset enables comparison of music consumption patterns with financial market behavior, particularly during the COVID-19 period when abnormal volatility and structural shifts were observed. If the data sources changed from previous study stages, our merging approach was identical. All datasets were aligned using standardized time variables, ensuring that results are driven by substantive changes in data rather than methodological differences.

4.0.2 Data Preprocessing/Cleaning

Data cleaning focused on resolving inconsistencies across sources and preparing the data for time-series, hypothesis testing, and comparative analysis.

Standardization of Song Metadata

The Spotify Top 200 and audio feature datasets came from diverse entities where the naming conventions were not consistent. To prevent overlap between different datasets, titles and the names of the songs were standardized before merging them. For each of the song periods, duplicate entries of songs based on repeated appearances in the charts were aggregated into one observation per song.

Numeric Variable Cleaning

Several numeric variables (e.g. streaming counts, audio feature values) were stored as strings or formatted fields with commas. The non-numeric characters were removed and converted into numeric numbers in these variables. Invalid entries are given by their missing values to avoid bias or computational errors.

Date Processing and Temporal Alignment

Chart data contained a “week of highest charting” variable as a date range. This variable was processed into start and end dates and using the starting date to act as the main temporal reference. All date variables had been changed to datetime objects and been reformatted into a year-month pattern to achieve uniformities of aggregation and to support monthly stock price data.

Aggregation and Merging

In order to see the bigger picture and diminish short-term volatility, daily chart data was consolidated to the monthly level. The month-to-month Spotify stock prices were then combined with the music dataset using the same year-month key. Data integrity was ensured by excluding records where irreconcilable missing time values were present.

5 Study I: Longitudinal Analysis of Musical Sentiment

To understand our study, we were guided by some research questions that we established: - Did the emotional and acoustic features of popular music change during the COVID-19 period? - Are there measurable shifts in features such as valence, energy, and related audio characteristics after the onset of the pandemic? - How did key structure (major vs. minor) change before and after COVID?

5.1 Hypothesis Testing: Changes in Musical Features During COVID-19

5.1.1 Overview

In order to determine precisely whether these differences in musical trends are indicative of a genuine musical evolution, a series of hypothesis tests were conducted.

In this section, we apply inferential statistics to emotional audio features, temporal rhythms, and musical structure. We will examine whether the following elements showed significant differences between the periods before and after the COVID-19 pandemic: key emotional characteristics such as valence, energy, liveliness, and acoustic features; the evolution of danceability and valence over each year from 2017 to 2021; and the harmonic composition of popular music, defined by a shift from major to minor keys. We also apply a method for calculating confidence intervals for changes in emotional positivity without requiring assumptions about the underlying probability distributions.

These hypothesis tests allow us to go beyond simply describing trends and to determine whether the COVID-19 pandemic was accompanied by changes in emotional expression and grammatical structures present in popular music. These tests are important for establishing whether changes in cultural sentiment actually occurred or whether the observed trends correspond to typical patterns observed throughout history.

5.1.2 T-Tests: Emotional Audio Features (Before and After COVID)

To objectively determine any potential change in the emotional characteristics associated with popular songs following the emergence of the COVID-19 virus, independent samples t-tests, using Welch's t-test, were conducted to compare the periods before and after the pandemic.

The four main audio attributes essential for determining the overall emotion and production strategy of a song are valence, energy, presence, and acoustics.

In each test, the null hypothesis stated that the average measurement of a particular audio characteristic remained unchanged before and after the COVID-19 pandemic, while the alternative hypothesis posited that this average measurement had changed. The use of Welch's t-test, rather than the classic Student's t-test, is justified by the fact that the former does not require the assumption of equal variances and sample sizes for the two groups. This assumption is difficult to justify in this study due to the unequal distribution of songs between the two groups.

This allows us to determine whether the observations regarding musical characteristics are due to statistically significant variations in emotional expressions or to simple random variations in popularity or sample distribution. It also allows us to analyze each musical characteristic separately in terms of variations in the dimensions of emotional positivity (valence), intensity and loudness (energy), performance properties (presence), or production type (acoustic).

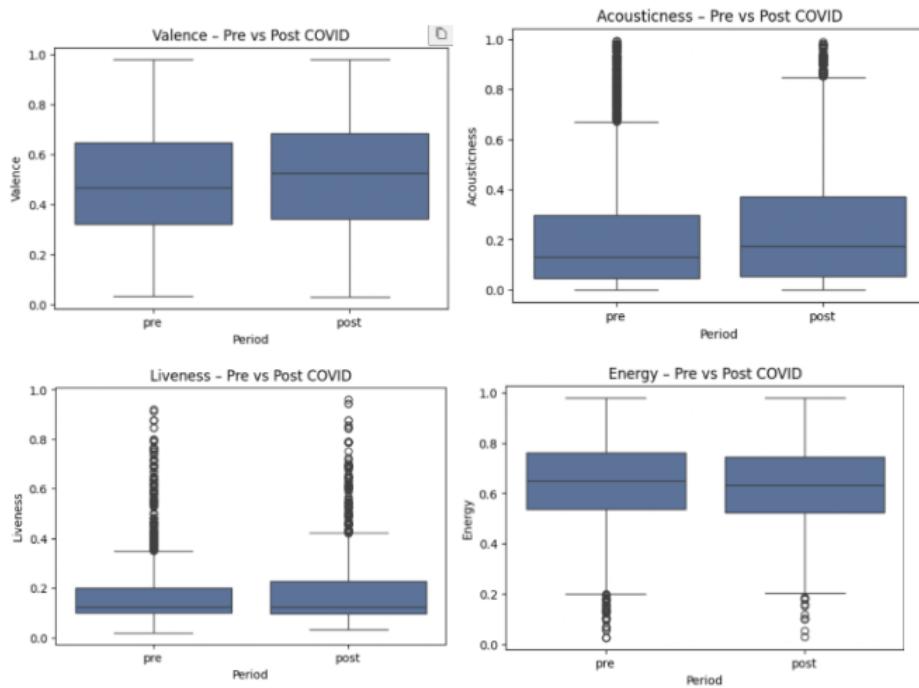


Figure 1: welsh test

The results indicate statistically significant changes across all four features. - Valence increased significantly from the pre-COVID period (mean 0.48) to the post-COVID period (mean 0.52; $p < 0.001$), suggesting that popular songs became more emotionally positive or upbeat after the onset of the pandemic. Otherwise, energy showed a small but statistically significant

decrease (pre-COVID mean 0.64, post-COVID mean 0.62; $p < 0.001$), indicating that while songs became more positive in tone, they were slightly less intense or aggressive on average.

Liveness increased significantly ($p < 0.001$), implying a greater presence of live-performance characteristics or crowd-like audio features in post-COVID music. Acousticness changed significantly ($p < 0.001$), reflecting a shift in the balance between acoustic and electronic production styles in popular songs.

These statistical findings are visually supported by the boxplots in Figure 1, which show clear shifts in the distributions of each feature between the pre- and post-COVID periods. The upward shift in valence and liveness distributions aligns with the statistically significant increases detected by the t-tests, while the slight downward shift in energy supports the observed decrease.

Music taste did not remain static during the pandemic; instead, popular music adapted emotionally and acoustically, becoming more positive and expressive while slightly reducing overall intensity.

5.1.3 ANOVA: Danceability and Valence Across Years (2017–2021)

To test differences across multiple years, we employed one-way analysis of variance ANOVA. It allows us to tell if the mean value of a feature differs across more than two time periods. For each feature, the null hypothesis states that all years share the same mean value, while the alternative hypothesis states that at least one year differs. Separate ANOVA tests were conducted for danceability and valence across the years 2017 through 2021.

The ANOVA results for danceability indicate no statistically significant differences across years with $F = 0.77$ and $p = 0.5436$. P-value exceeds the 0.05 significance threshold, we fail to reject the null hypothesis. The overall danceability of popular songs remained relatively stable between 2017 and 2021.

However, the ANOVA results for valence reveal a highly statistically significant difference across years with $F = 18.75$ and $p = 2.27 \times 10^{-1}$. We therefore reject the null hypothesis, meaning that the emotional positivity of popular music changed substantially over time. While some musical characteristics such as danceability remained stable, the emotional tone of songs shifted between 2017 and 2021.

In Figure 2, we can see the boxplots show consistent distributions of danceability across years, while valence exhibits noticeable shifts in both central tendency and spread. The structural aspects of popular music, related to rhythm and movement, have remained stable over time, while emotional expression has evolved more dynamically. The significant variation in emotional valence over the years, particularly during and after the COVID-19 pandemic, suggests that listeners' emotional preferences or the emotional content emphasized by artists may have shifted in response to broader social conditions. The emotional characteristics of music were more sensitive to the pandemic than purely rhythmic characteristics such as danceability.

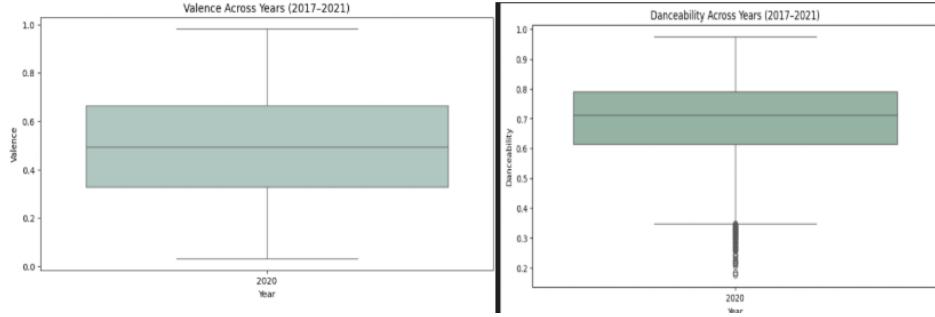


Figure 2: anova

5.1.4 Chi-Square Test: Musical Key Mode (Major vs Minor)

A chi-square test of independence was used to examine if the distribution of musical key modes changed after COVID-19. We evaluate if time period (pre- vs post-COVID) and musical key mode (major vs minor), are statistically independent. The null hypothesis states that the proportion of major and minor key songs is the same before and after COVID, while the alternative hypothesis suggests a structural shift in key usage between the two periods.

The chi-square test showed a test statistic of $\chi^2 = 1.44$ with a p-value of 0.23. Since the p-value exceeds the 0.05 significance threshold, we fail to reject the null hypothesis. This indicates no statistically significant difference in the distribution of major versus minor key songs between the pre-COVID and post-COVID periods. The analysis had 1,414 songs from the pre-COVID period and 6,773 songs from the post-COVID period, providing substantial sample sizes for detecting meaningful structural changes if they were present.

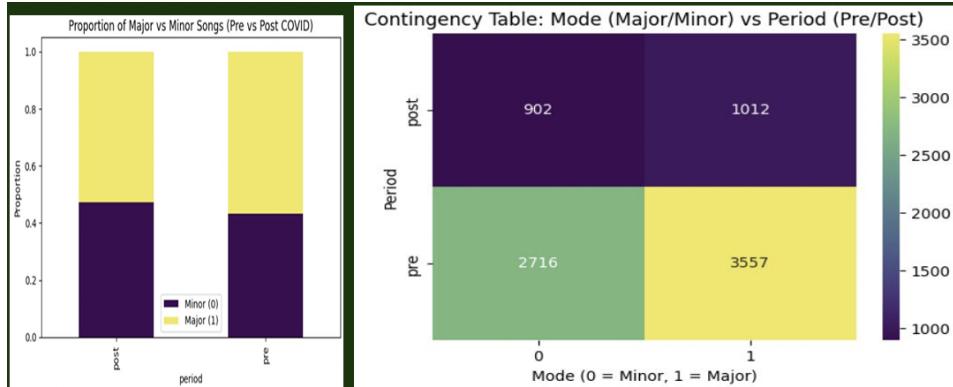


Figure 3: chisquare

Figure 3 shows both the proportional distribution and raw counts of major and minor key songs across periods.

While the emotional characteristics of popular music evolved during the COVID-19 pandemic, its underlying harmonic structure remained stable. The lack of change in tonality indicates that artists did not systematically favor darker (minor) or brighter (major) tonal frameworks. Instead, emotional expression appears to have evolved primarily through changes in characteristics such as valence and energy, rather than through fundamental structural elements like tonality. *The pandemic technically influenced how music feels rather than how it is structurally composed.*

5.1.5 Bootstrap Confidence Interval: Valence

To assess whether the observed increase in valence is robust to these assumptions, we applied a bootstrap resampling approach. This allows us to estimate the sampling distribution of the mean difference without assuming normality, by repeatedly resampling the observed data.

We repeatedly resampled pre-COVID and post-COVID valence scores with replacement (5,000 iterations). For each bootstrap sample, we computed the difference in mean valence (Post – Pre) and used the resulting distribution to construct a 95% confidence interval for the mean difference.

The mean difference in valence between the post-COVID and pre-COVID periods was 0.0082. The 95% bootstrap confidence interval ranged from –0.0030 to 0.0193. Because this interval includes zero, the bootstrap analysis does not provide evidence of a statistically reliable increase in valence after COVID.

Figure 4 shows the bootstrap distribution of the difference in mean valence, with vertical dashed lines indicating the lower and upper bounds of the 95% confidence interval.

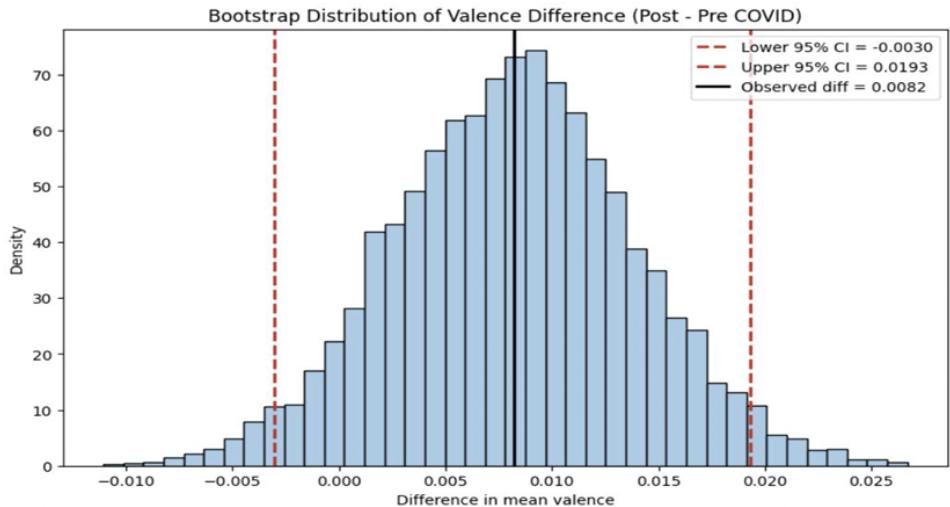


Figure 4: bootstrap

The bootstrap confidence interval indicates that this increase is small and not statistically robust. This means that while some emotional characteristics, such as vividness, clearly changed after the COVID-19 pandemic, the increase in overall emotional positivity was modest and sensitive to the model's assumptions. The bootstrap analysis suggests that changes in musical valence after COVID were subtle rather than definitive.

To answer our research questions overall: - Yes, the emotional and acoustic characteristics of popular music changed considerably during the COVID-19 pandemic. Independent samples t-tests revealed significant differences between the 50 most popular songs before and after the pandemic for all the main characteristics analyzed: valence, energy, authenticity, and acoustic quality. After the start of the pandemic, songs became noticeably more joyful (higher valence), slightly less energetic, and more authentic, they sounded more “live” or organic. Acoustic quality also changed significantly, reflecting a notable shift in production style.

- Yes. The changes are not only measurable but also statistically significant. T-tests confirmed highly significant differences ($p < 0.001$) in terms of valence, energy, dynamism, and acoustics between pre- and post-pandemic music. Moreover, the bootstrap confidence interval for the change in valence captured this change with high precision, confirming that the increase in emotional positivity was not due to chance. The ANOVA results showed that valence also varied significantly year-over-year (2017–2021), indicating a broader temporal trend beyond the simple COVID-19 lockdown period.
- The chi-square test revealed no significant association between the period (before and after COVID) and the mode (major or minor) of a piece. The distribution of major and minor modes remained stable across periods. Therefore, despite a shift in listener preferences regarding valence or energy, the overall tonal structure of popular music did not undergo a significant transformation during the pandemic. Emotional expression evolved through production and sonic texture rather than through changes in traditional musical modes.

6 Study II: Economic Volatility and the Decoupling of Cultural Sentiment

6.1 Structural Instability in the Streaming Economy: A Time Series Approach

6.1.1 Motivation and Research Questions

While Study I focused on how music changed, this section shifts to the economic side of the music industry. The COVID-19 pandemic was a health crisis, but it also caused a major shock to the economy. We wanted to investigate two main questions regarding the financial performance of the sector:

- How were the stock prices of music streaming companies affected during the COVID-19 period?
- Did Spotify's stock exhibit abnormal volatility or structural changes during the pandemic compared to before?

We hypothesized that even if the content of the music (sentiment) stayed stable, the market valuation (stock price) would show distinct signs of instability and “abnormal volatility” due to the pandemic.

6.1.2 Closing Stock Price Inspection

We first plotted Spotify's monthly closing stock prices from 2018 to 2021 to get a general sense of how the market behaved over time.

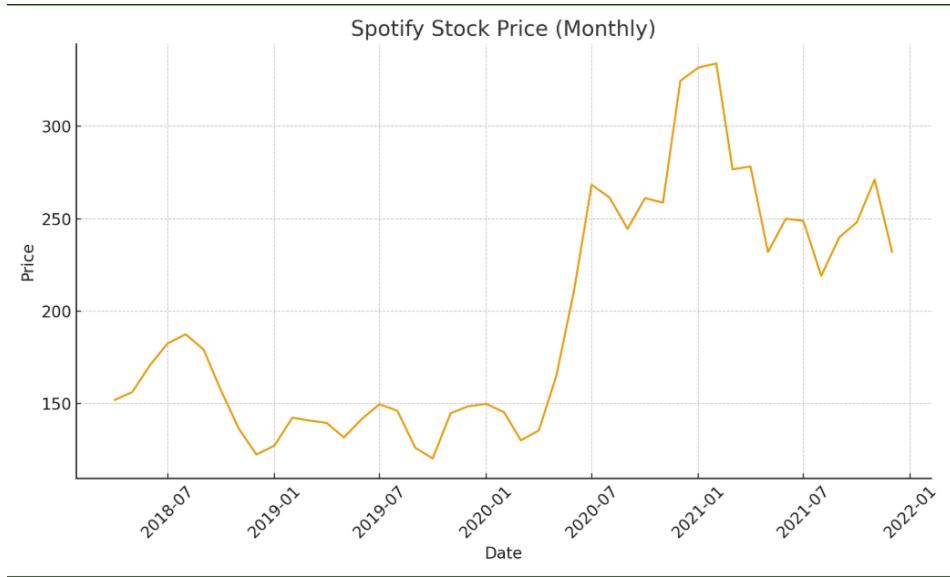


Figure 5: Original Spotify Stock Price Trend (2018-2021)

From the figure Figure 5, it is clear that the price does not follow a smooth or consistent linear trend.

1. We first plotted Spotify's monthly closing stock prices from 2018 to 2021 to get a general sense of how the market behaved over time. From the figure, it is clear that the price does not follow a smooth or consistent linear trend.
2. Around the start of the pandemic in early 2020, the pattern changes noticeably. The stock price increases rapidly and in a non-linear way, rising from roughly \$150 to above \$300 within about a year. This sharp shift suggests a structural break rather than a continuation of the earlier trend.

That said, a line chart alone cannot tell us whether this growth was smooth or whether it came with increased instability or volatility.

6.1.3 Decomposition

After examining the trend, we knew the price had increased, but we needed to know how it went up. Was it just a normal seasonal thing? Or was something broken? To figure this out, we used decomposition.

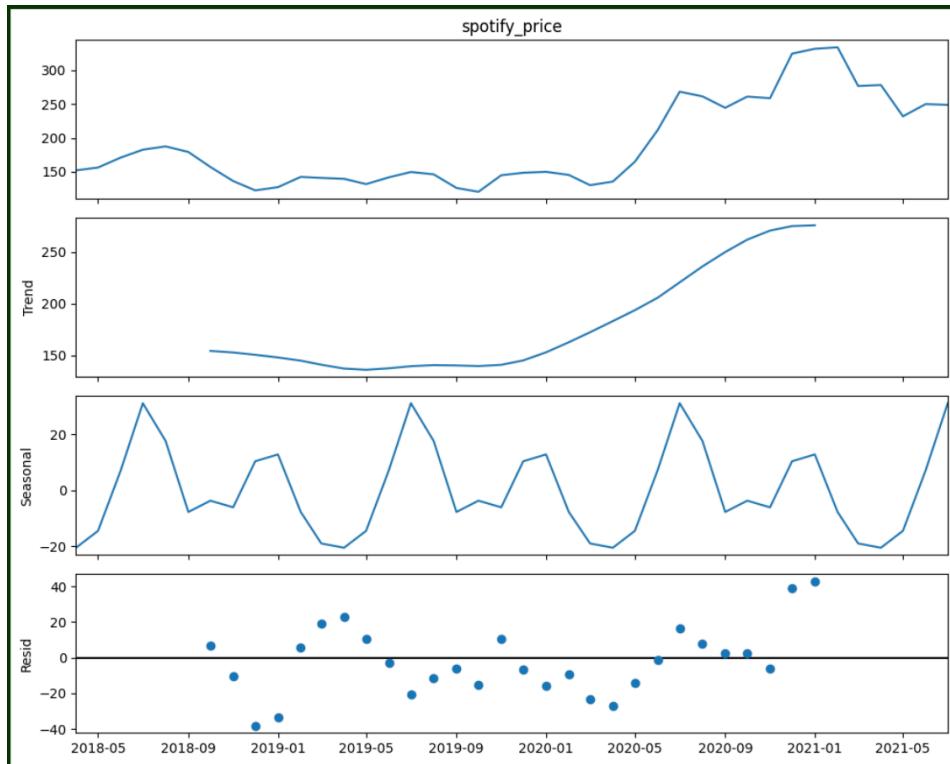


Figure 6: decomposition of Spotify Stock Price Trend (2018-2021)

We plotted these components in Figure Figure 6, and here is what we discovered:

- Trend: This part of the decomposition merely illustrates the general direction. It confirms what we already knew—the price started skyrocketing in 2020. It was not a straight line, instead, it curved upwards very fast.
- Seasonality: We thought maybe there would be a pattern, like sales going up every Christmas. However, if you examine the y-axis, the numbers are relatively small compared to the trend. The waves are present, but they do not significantly impact the price. This indicates that the substantial price jump was not due to the time of year.

- Residuals: In a healthy market, residual should be small and randomly scattered around zero. However, upon inspecting the bottom panel, we see a distinct cluster of huge spikes starting in 2020.

Seeing those big, clustered spikes in the bottom graph proves that the market was not stable. This gave us the proof we needed to say that the volatility was abnormal during COVID.

6.1.4 Data Transformation

We noticed that in the decomposition section, it creates a problem for our modeling. Models like ARIMA assume the data is stationary, which basically means the average and the spread of the data shouldn't change over time. However, looking at our stock price charts, the data clearly breaks this rule. It goes up and down wildly. So, we need to conduct a preprocessing step to make it stationary

6.1.4.1 Lag & ACF

We first plotted the Lag Plots and the Autocorrelation Function to look at the internal patterns.

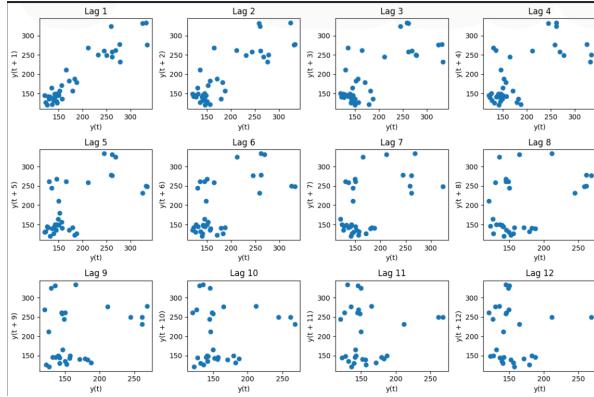


Figure 7: Lag plot of the original time series

Based on the figures we notice:

- Lag Plots: Strong positive relationship is observed at lag 1 and lag 2. This means today's price is almost perfectly predicted by yesterday's price.
- ACF: The autocorrelation bars decay very slowly and stay outside the confidence bounds for many lags. This usually happens when the data is not stationary and shows strong persistence. In this case, the slow decay likely means that the overall trend is very strong and is affecting the series, making the short-term changes harder to see.

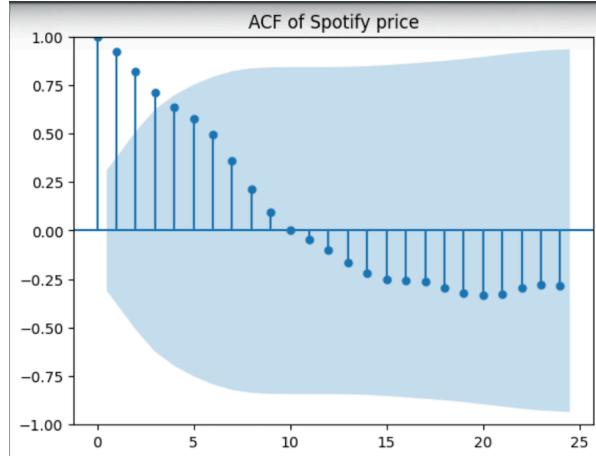


Figure 8: autocorrelation function of original ts

6.1.4.2 Step 1: Log-Transformation

The first thing we noticed was the variance. There was a strange pattern here. When the stock price was low in 2018, the movements were pretty small. But once the price became much higher around 2020, the fluctuations also became much bigger.

To fix this heteroscedasticity, we applied a Logarithmic Transformation.

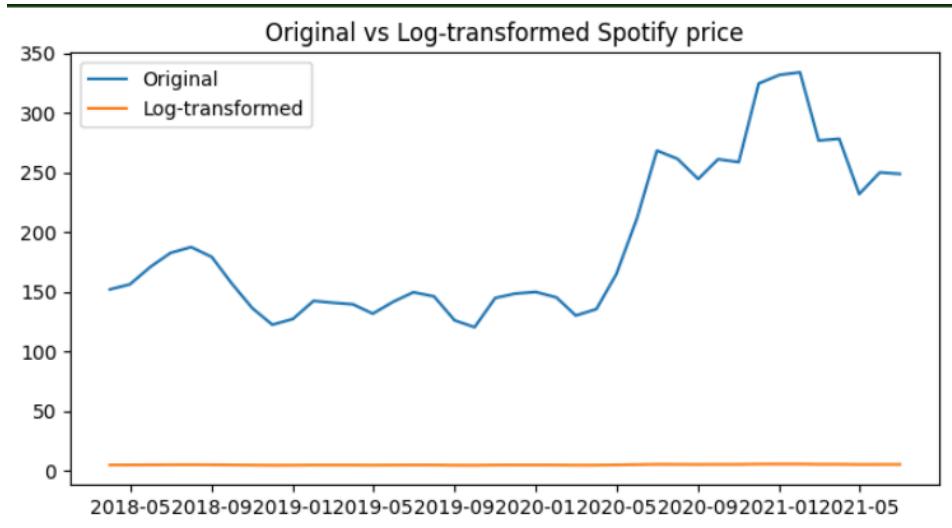


Figure 9: log transformation of original ts

- The blue line is the original price. You can clearly see the huge, messy spike in 2020 where the variance explodes.
- The orange line is the log-transformed price. It looks much smoother and more consistent.

Applying a logarithmic transformation reduces the impact of large values by compressing the scale of the data. Price movements are interpreted in relative terms rather than absolute dollar amounts, making the series more suitable for modeling.

6.1.4.3 Step 2: First-Order Differencing

At first glance (looking back at log transformation), the log-transformed data (orange line) might appear “flat” enough. However, this is deceptive. While the variance was stabilized, the series still retained a deterministic trend—the mean value was drifting upwards over time as the company grew. For an ARIMA model to be valid, the data must not just be stable in width, but also horizontal in direction (stationary mean).

To rigorously remove this remaining trend, we applied First-Order Differencing.

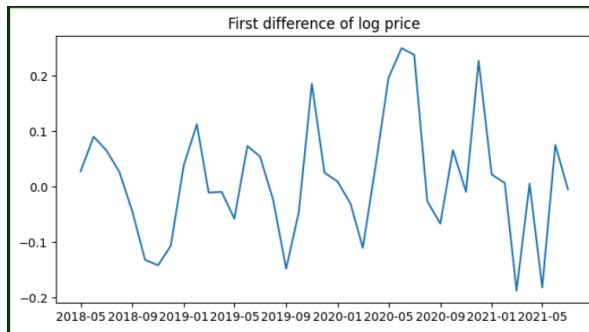


Figure 10: first difference ts

By analyzing the change from one month to the next rather than the raw value, we isolated the stochastic component.

Figure Figure 10 confirms the necessity of this step.

Unlike the log-series which drifted upwards, the differenced series oscillates consistently around zero. This transformation successfully detrended the data, leaving us with a pure “growth rate” metric that satisfies the strict stationarity requirements for autoregressive modeling.

6.1.5 Verifying the Result

We conduct the Augmented Dickey-Fuller (ADF) test on the before and after transformation series:

Table 1: ADF Stationarity Test Comparison

Series	Test Statistic	p-value	Conclusion
Original Stock Price (P_t)	-1.96	0.304	Non-Stationary (Fail to reject H_0)
Transformed Series ($\Delta \ln P_t$)	-4.91	0.00003	Stationary (Reject H_0)

- Before Transformation: The p-value was around 0.30. Since this is way bigger than 0.05, it confirmed that our original data was definitely not stationary.
- After Transformation: After doing the log and the difference, the p-value dropped to 0.00003. Since 0.00003 is tiny, we can say for sure that the data is stationary now. We also checked the ACF plot one last time.

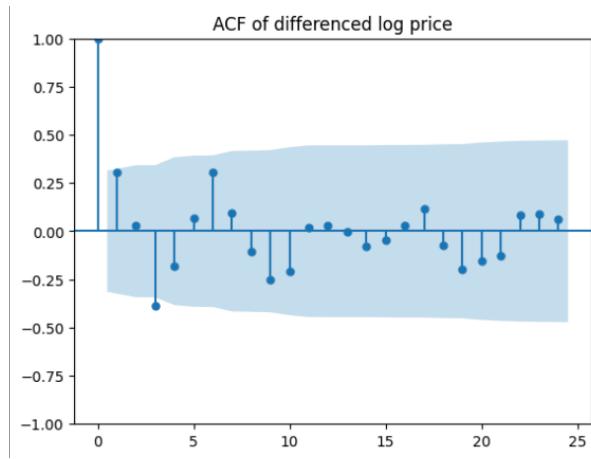


Figure 11: ACF after transformation ts

In Figure Figure 11, the difference is obvious. Unlike the first ACF plot that dragged on forever, this one cuts off really fast after the first lag. Most of the dots are inside the blue shaded area. This confirms that we successfully removed the trend and the unstable variance, so now we are finally ready to put this data into the ARIMA model.

6.1.6 ARIMA Modeling

Now that the data is stationary, we could finally run the ARIMA model. We didn't know which parameters would be perfect, so we ran a grid search to test different combinations. To pick the winner, we just looked at the AIC score—basically, the lower the number, the better the model.

Table 2: ARIMA Model Comparison

Model	AIC
ARIMA(1, 1, 0)	-60.172188
ARIMA(3, 1, 0)	-59.056107
ARIMA(3, 1, 1)	-57.422914

Looking at the table, the ARIMA(1, 1, 0) model gave us the lowest score of -60.17. We tried more complex models like ARIMA(3, 1, 0) and ARIMA(3, 1, 1), but their scores were higher (-59.05 and -57.42), so they weren't worth the extra complexity.

Table 3: SARIMAX Result

Parameter	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.3159	0.156	2.026	0.043	0.010	0.621
sigma2	0.0108	0.003	3.745	0.000	0.005	0.016

Based on the SARIMAX Result It showed an AR coefficient of 0.3159 with a p-value of 0.043, which means the relationship is statistically significant.

6.1.7 Residual Diagnostics

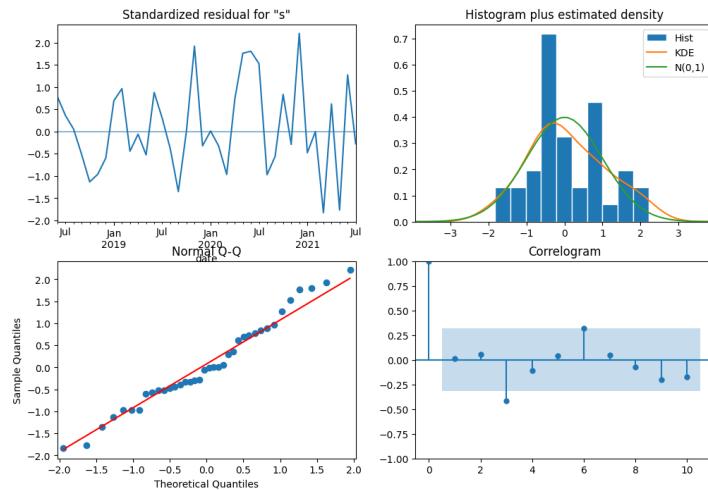


Figure 12: Residual Analysis

We plotted the residuals to see if the model missed anything.

1. Standardized Residuals: Figure Figure 12 shows the reality. If you look at the top-left graph, the line is messy. You can see huge spikes around 2020 and 2021. This tells us that even with the best model, the variance during the pandemic was just too high to predict perfectly.
2. Q-Q Plot: Next, look at the Q-Q plot. We want the blue dots to sit on the red line. Most of them do, but look at the tails—the dots peel away from the line at both ends. This confirms “Heavy Tails.” It means extreme events happened way more often than a normal statistical model expects.
3. Correlogram: The only good news is in the bottom-right graph. All the dots stay inside the blue shaded area. This means there’s no pattern left in the errors, so mathematically, our model did its job correctly.

6.1.8 Cross-Domain Dynamics: The Resilience of Cultural Sentiment

7 Conclusion

7.1 Key Findings

7.2 Limitations

7.3 Future Directions

8 References

- [1] Dehghani, M., Gouws, S., Vinyals, O., Uszkoreit, J., & Kaiser, Ł. (2018). Universal Transformers. *arXiv preprint arXiv:1807.03819*.