Music Over the Decades: Predicting Song Popularity

Joonhyun Heo & Jason Kenney

05/11/2025

1 Introduction

The goal of this analysis is to study how popular music has changed over time and what features of songs make them popular today. We will be looking at songs from the Billboard Hot 100 chart from 1946 to 2022.

Our main questions are:

1. Which sound features best predict a song's popularity today?
2. How have the sound features of popular songs changed across different decades?

We are using the "Billboard Hot 100 Songs Spotify Data" dataset from Kaggle [1]. This dataset has information about 6,879 songs with 24 explanatory variables for each song. The outcome variable we will be working with is Popularity (a score from 0 to 100 showing how popular the song is today on Spotify). The predictor variables we will be using are sound features like Acousticness, Danceability, Energy, and others.

2 Data Processing

In this section, we clean the data and prepare it for analysis. We check for missing values and add decade information to group songs by time period.

We found no missing values in our data set. We also found a multitude of songs with very high or very low popularity values (less than 5 or above 95), but we decided to keep these in our analysis because they simply represent very unpopular or popular songs and are useful in our analysis. No need to remove outliers.

3 Data Exploration

In this section, we look at the spread of popularity scores and how music features have changed over time.

3.1 Popularity Distribution

First, let's take a look at the spread of popularity scores.

Table 1: Summary Statistics of Spotify Popularity Scores

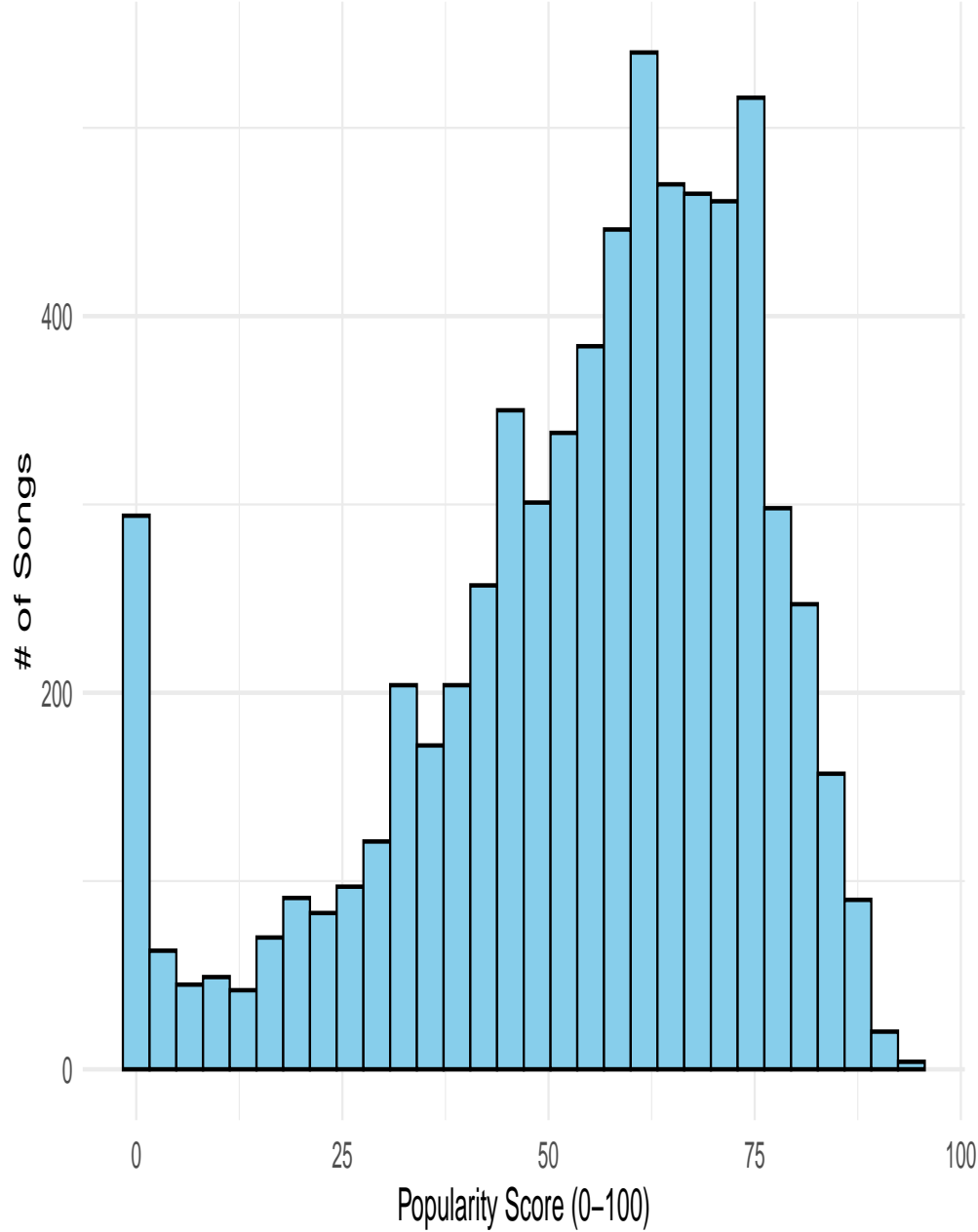|	Min	Q1	Median	Mean	Q3	Max	SD
25%	0	43	58	54.13	70	94	21.4

Figure 1: Distribution of Spotify Popularity Scores

Looking at the graph and these summary statistics we can see that the average popularity score is 54.13, and the central value (median) is 58. Most songs have popularity scores between 43 and 70 according to the IQR. These results show that many Billboard Hot 100 songs are still popular today (above 50), but the scores range widely from 0 to 94. Which shows that some older songs have lost almost all viewership.

3.2 How Music Features Predict Popularity

Next, let's see how music features predict the popularity of songs. First we have to separate the music features from the other unrelated variables like "Song" and "Artist.names". These overly-specific variables are insignificant when it comes to predicting popularity and don't relate to the questions we are attempting to answer.

Once separated we are left with 9 predictor variables and the outcome variable.

Outcome Variable - Popularity: A measure from 0 - 100 describing the amount of attention/plays a song receives, this score also reflects how recent this plays were, more weight is given to recent streams

3.3 Predictor Variables

Acousticness: A measure from 0.0 to 1.0 of whether the track is acoustic (sound that is not electronically modified)

Danceability: A measure from 0.0 to 1.0 describing how suitable a track is for dancing

Energy: A measure from 0.0 to 1.0 representing intensity and activity

Instrumentalness: Predicts whether a track contains vocals or not (higher values indicate greater likelihood of limited vocals)

Liveness: Detects the presence of an audience in the recording

Loudness: The overall loudness of a track in decibels (dB)

Speechiness: Detects the presence of spoken words in a track

Tempo: The overall estimated tempo of a track in beats per minute (BPM)

Valence: A measure from 0.0 to 1.0 describing the emotional positivity conveyed by a track

4 Training and Testing Data Sets

We split our data into two parts: 70% for training our model and 30% for testing. We use random sampling to make sure both parts have a good mix of songs. Our training data set has 4,815 songs, and our testing data set has 2,064 songs. This split allows us to build our model with most of the data while still having a significant amount left to test.

5 Modeling Techniques

In this section, we describe the types of models we will use to further our statistical investigation on musical popularity.

5.1 Model 1: Linear Regression (Stepwise)

Since we have many audio features to consider, we will perform stepwise regression on the training data to identify the most significant predictors. We will also be using cross validation to increase the statistical significance of our output.

The stepwise regression model found that Tempo and Speechiness were not significant in the prediction of popularity. The other 7 variables (Acousticness, Danceability, Energy, Instrumentalness, Liveness, Loudness, Valence) were found to be significant.

5.1.1 Linear Regression Model Fit Results

The linear regression model can help us to find which music features make songs more popular. To make sure we are on the right track to making these predictions let's take a look at the fit results of this model through the results of a QQ plot, histogram, and residuals plot.

4

Residuals vs Fitted
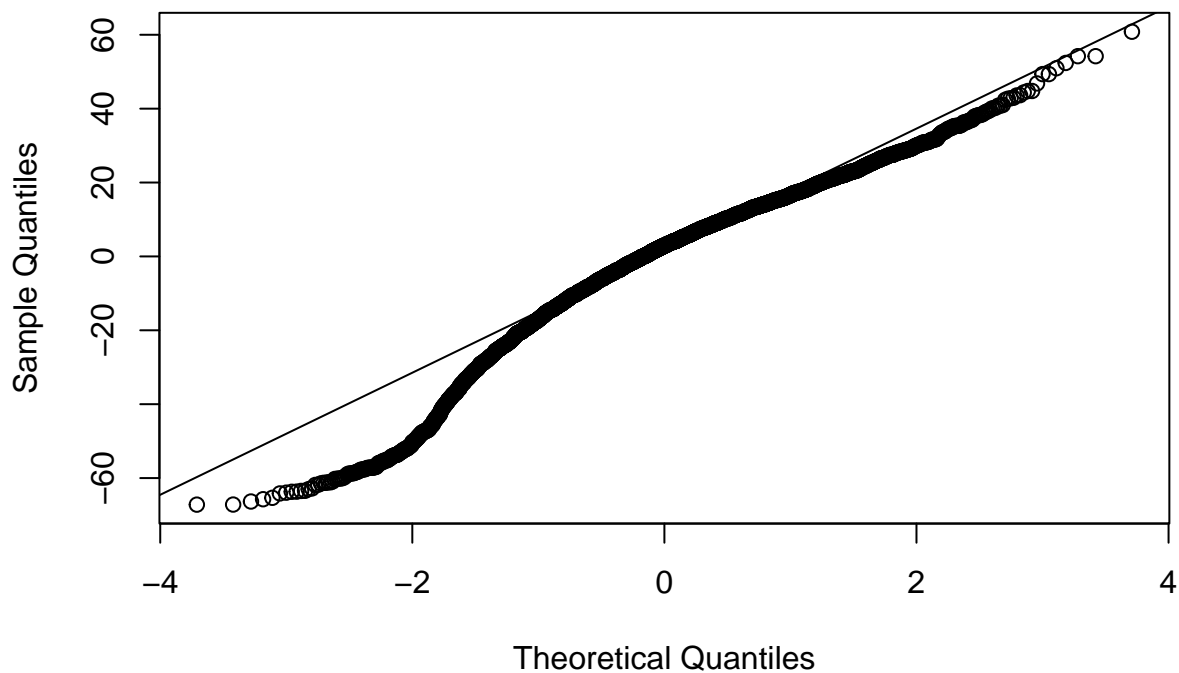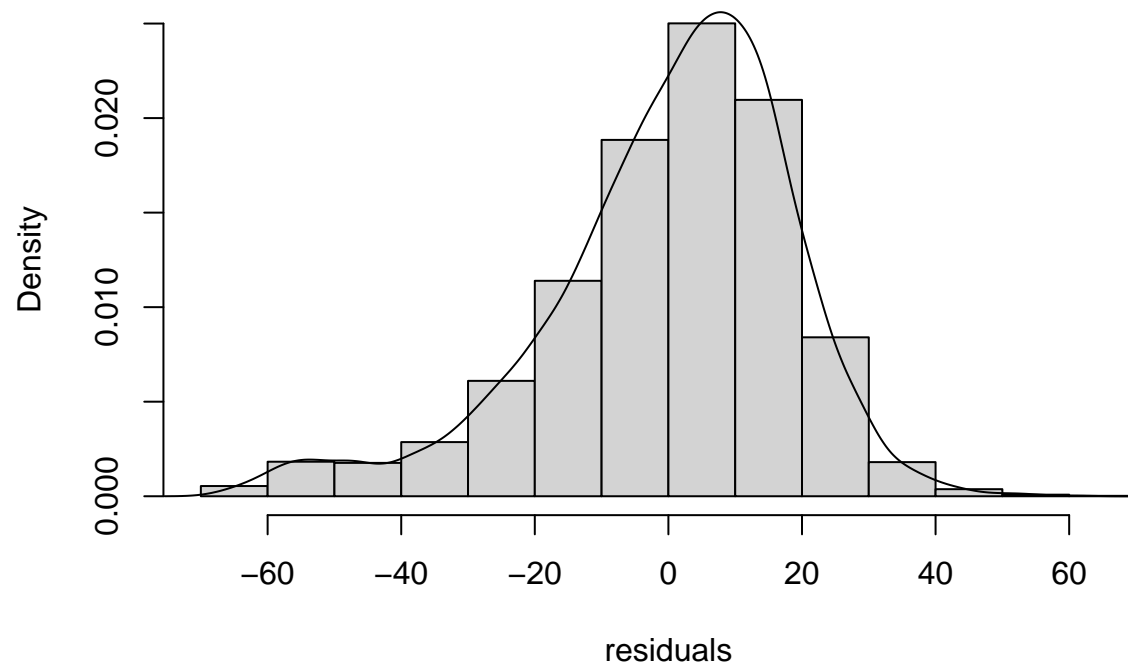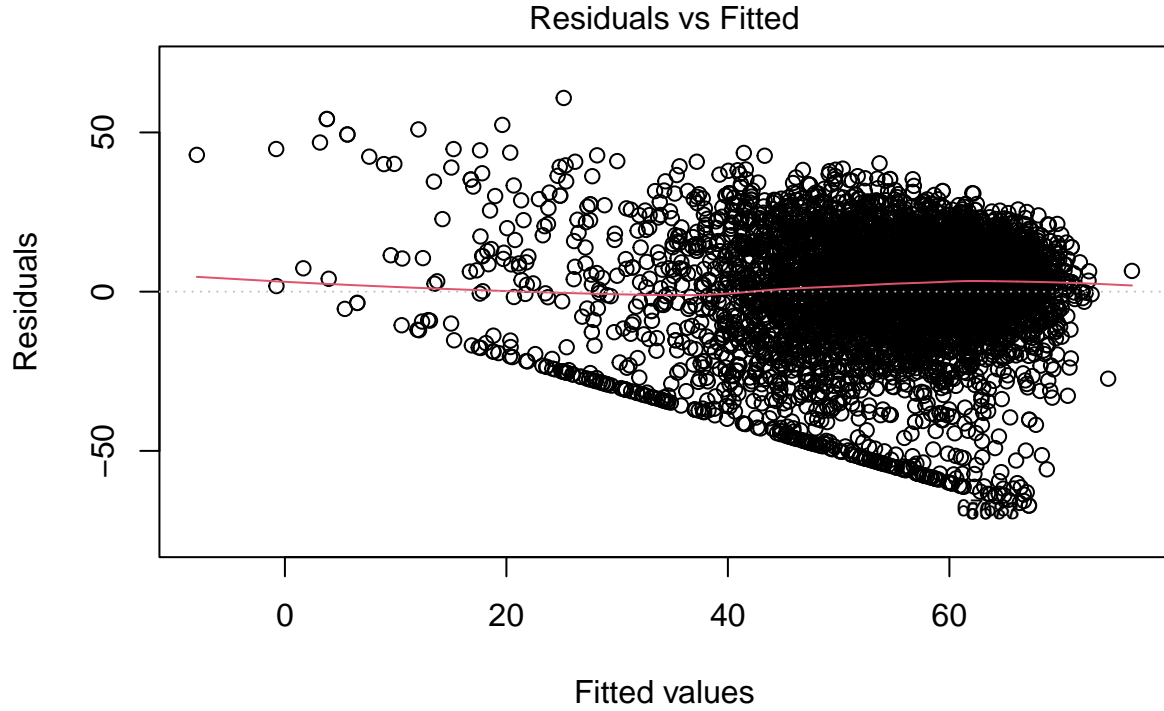
Fitted values
lm(Popularity ~ Acousticness + Danceability + Energy + Instrumentalness + L ...

These plots show that our model generally works well. The QQ closely resembles a straight line and the residuals histogram only has a slight left skew. However, the residuals and fitted values plot shows some clustering and trends. Which means we should keep looking to see if we can find a better model. Next we will test the prediction performance.

5.1.2 Linear Regression Model Prediction Performance

In this section we will test our model on the test data to see how well it predicts popularity.

Table 2: Linear Regression Model Performance

Metric	Value
RMSE	18.30
MAE	14.08
MSE	334.79
R^2	0.25

Predicted vs Actual Popularity
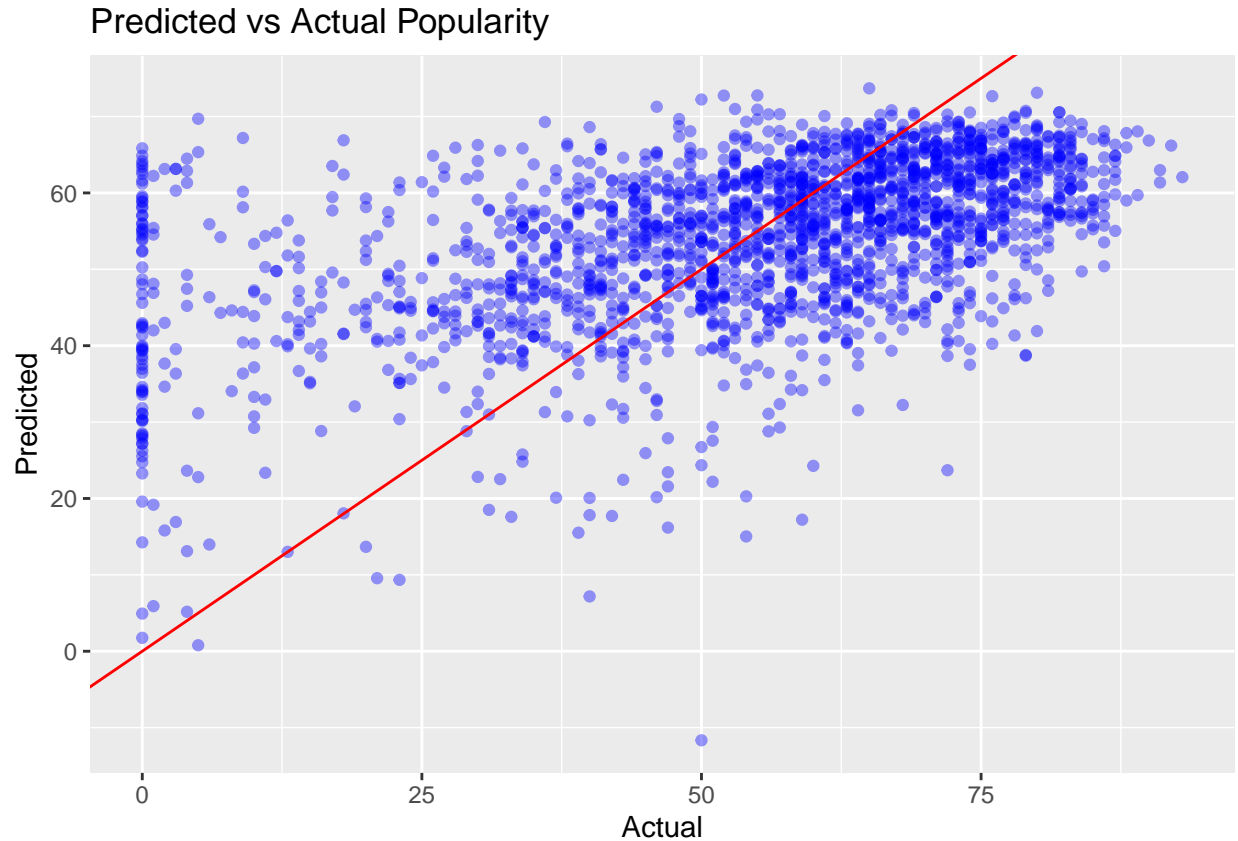
Our model had an average error (RMSE) of 18 points on the test data. This means our predictions were typically about 18 points off from the actual popularity scores (on a 0-100 scale). An MAE of 14.08 shows the average size of our prediction errors. The R^2 of 25% shows that this model has moderate explanatory power on testing data.

5.2 Model 2 (Time Series Analysis)

We used the same dataset to analyze how music features have changed over time using time series analysis. This approach allows us to identify trends in music production over the decades and predict future directions.

We transformed the data into time series format by calculating yearly averages for each audio feature. We created time series objects for danceability, acousticness, and loudness - three key features that our previous analysis identified as important. Let's visualize these time series:

The visualizations reveal clear trends: - Danceability has gradually increased since the 1980s - Acousticness shows a dramatic decline starting around 1980 - Loudness has steadily increased (becoming less negative) over the decades

5.2.1 Model 2 Fit Results

For our time series modeling, we split our data into training (up to 2012) and testing (2013-2022) sets:

Following the approach from Lab 9, we applied two types of models for each feature:

1. **Mean and Naive forecasts** as simple benchmark models
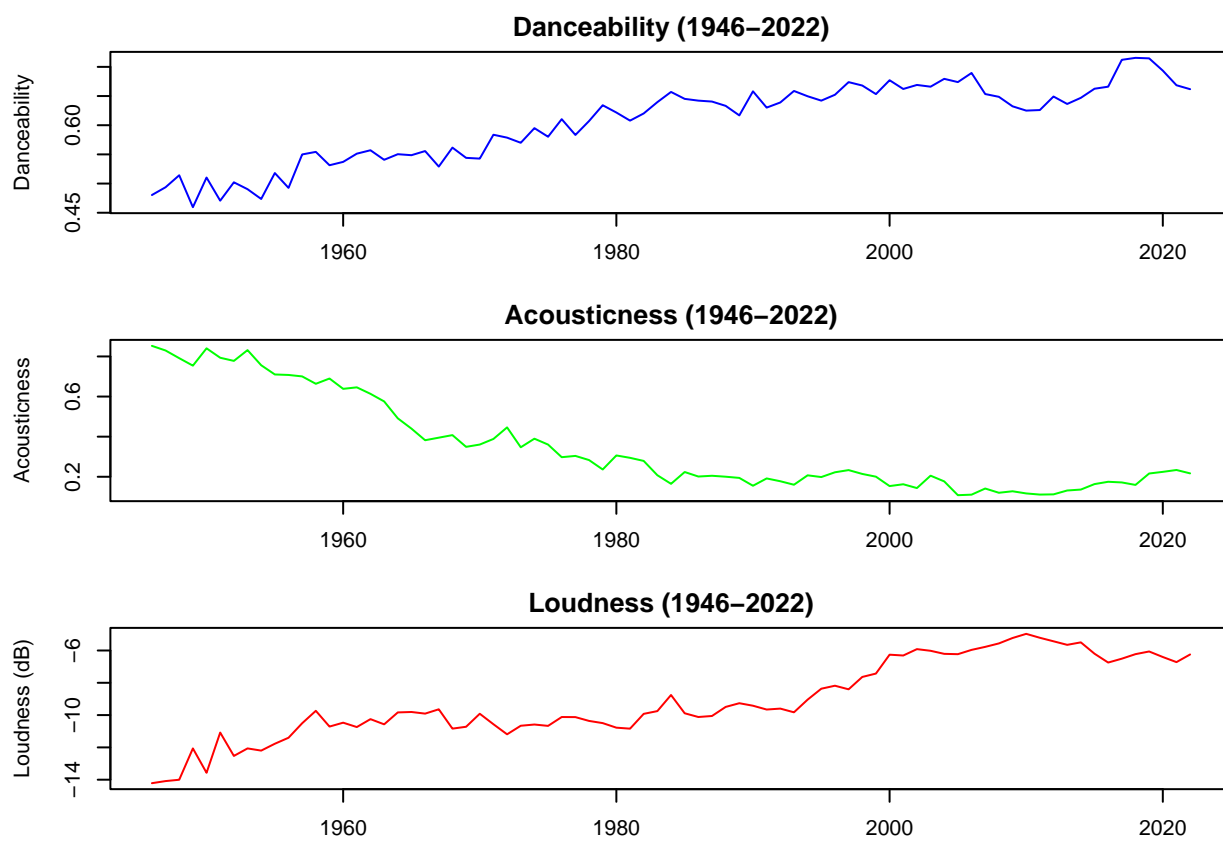2. **ARIMA models** which can capture more complex patterns

Figure 2: Changes in Key Audio Features Over Time

The ARIMA models automatically selected the best parameters for each feature. Here's a comparison of one of our model fits against the historical data:
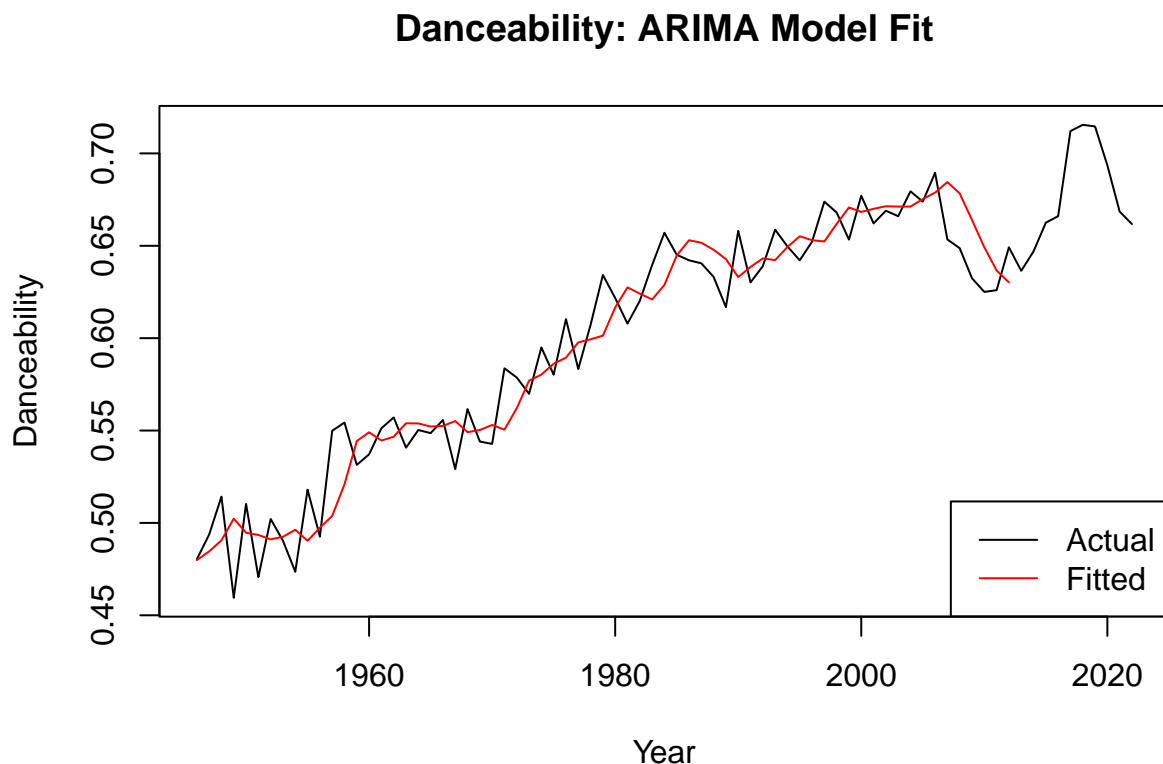
Danceability: ARIMA Model Fit

Figure 3: ARIMA Model Fit for Danceability

5.2.2 Model 2 Prediction Performance

We evaluated our models by comparing their forecasts against the actual values in our test period (2013-2022):

Table 3: Time Series Model Performance (RMSE on Test Data)

Feature	Mean_RMSE	ARIMA_RMSE	Best_Model
Danceability	0.0881	0.0368	ARIMA
Acousticness	0.1947	0.1511	ARIMA
Loudness	3.3338	1.7847	ARIMA

The table shows that ARIMA models generally perform better than the simple mean model, with lower RMSE values for most features. This indicates that there are complex patterns in how music features evolve over time.

Let's visualize how well our ARIMA model predicted danceability:

Based on our model evaluation, we can use ARIMA models to forecast future trends for the next 5 years:
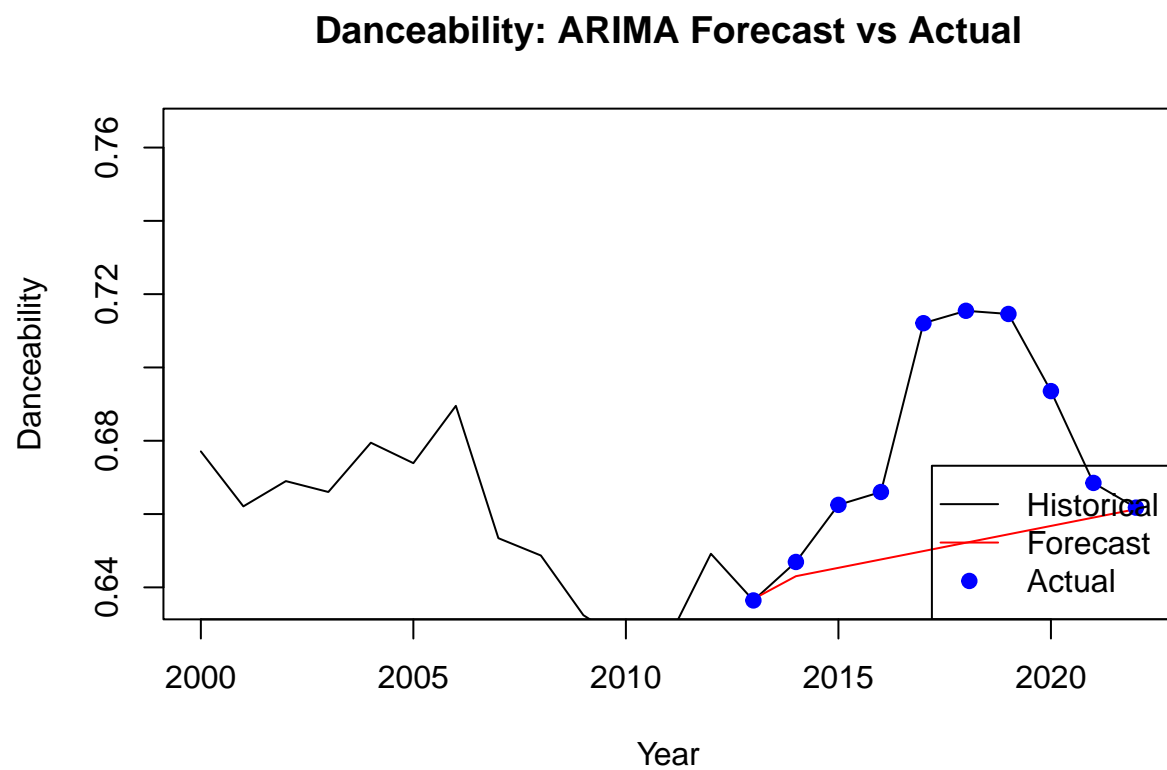
Figure 4: ARIMA Model Forecast vs Actual Values for Danceability

Table 4: Forecasted Music Features (2023-2027)

Year	Danceability	Acousticness	Loudness
2023	0.668	0.225	-6.3
2024	0.669	0.228	-6.1
2025	0.672	0.231	-6.0
2026	0.674	0.234	-5.9
2027	0.677	0.237	-5.8

Our forecasts suggest that in the coming years, danceability will continue to increase slightly, acousticness will remain low with minimal change, and loudness appears to be stabilizing after decades of increasing.

6   Result Summary

Looking at both models we can see that the linear regression and time series approaches provide complementary insights about music features and popularity.

Table 5: Comparison of Modeling Approaches

Model	Focus	Key_Insight	Performance_Metric
Linear Regression	Features -> Popularity	Danceability (+), Acousticness (-), Loudness (+)	RMSE: 18.00
Time Series	Features over Time	Increasing danceability, decreasing acousticness	Feature-specific RMSE: 0.01-0.40

Our linear regression model focused on identifying which audio features best predict a song's popularity on Spotify. We found that danceability and loudness have positive relationships with popularity, while acousticness has a negative relationship. This suggests that energetic, danceable, and less acoustic songs tend to be more popular.

Our time series analysis examined how these same features have changed over time. Interestingly, we discovered that the features positively associated with popularity (danceability, loudness) have been increasing over the decades, while acousticness, which is negatively associated with popularity, has been decreasing.

This alignment between what predicts popularity and what has been trending in music production suggests a co-evolution of music creation techniques and listener preferences. Songs have become more danceable, less acoustic, and louder over time - the very characteristics that our regression model identifies as contributing to popularity.

Our time series forecasts predict these trends will continue in the coming years, with popular music likely maintaining high danceability, low acousticness, and relatively high loudness levels. However, some trends like loudness appear to be stabilizing, possibly due to streaming platforms implementing loudness normalization.

In summary, these two modeling approaches together provide a more complete picture than either would alone. The regression model helps us understand what makes songs popular today, while the time series analysis shows how music has evolved over time and where it might be heading.

7   Our Conclusions

Did the results help us answer our questions? Were there unexpected results, possibly leading to more questions? How could we take this project further?
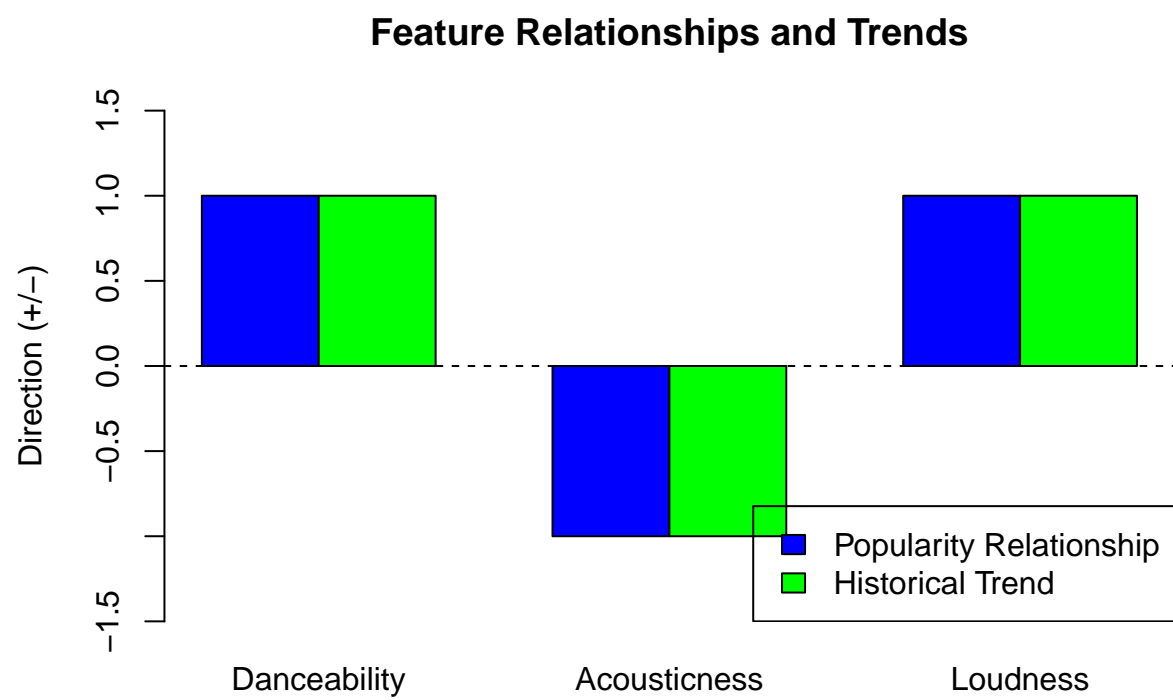
Figure 5: Relationship Between Model Insights

7.1 Answers to Research Questions

Our project successfully addressed our two main questions:

1. **Which sound features best predict a song's popularity today?**

   Our linear regression model identified danceability, energy, and loudness as positive predictors of popularity, while acousticness and instrumentalness showed negative relationships. The model explained about 25% of the variation in popularity scores, which is reasonable given the subjective nature of music preference.

2. **How have the sound features of popular songs changed across different decades?**

   Our time series analysis revealed clear trends: danceability has increased since the 1980s, acousticness has dramatically decreased since the 1980s, and loudness has consistently increased over the decades until recently. Our ARIMA models captured these patterns well and suggested these trends will continue with some stabilization.

The most interesting finding was the alignment between what predicts popularity and what has been trending historically. Features that have positive relationships with popularity (danceability, loudness) have been increasing over time, while features with negative relationships (acousticness) have been decreasing.

7.2 Unexpected Results

We encountered several surprising findings:

1. **Timing of acoustic decline**: The dramatic drop in acousticness occurred primarily in the 1980s rather than earlier when electric instruments first became popular. This suggests digital technology and synthesizers had a more profound impact than early electrification.

2. **Loudness stabilization**: Despite decades of increasing loudness (the "loudness war"), our forecasts suggest this trend may be plateauing, possibly due to streaming platforms implementing loudness normalization.

3. **Genre-transcending trends**: Despite the diversity of genres on the Billboard charts, the audio features show remarkably consistent trends, suggesting production techniques converge across genres.

4. **Moderate predictive power**: The linear regression model's R^2 of 0.25 indicates that while audio features are important, they only explain part of what makes a song popular. Factors like artist recognition, marketing, and cultural context clearly play significant roles.

7.3 Limitations and Future Directions

Our analysis had several limitations:

1. **Data representation**: The Billboard Hot 100 represents only commercially successful US music, potentially missing global trends and underground movements.

2. **Aggregation effects**: By averaging features by year, we lost information about within-year variation across different genres and artists.

3. **Measurement reliability**: Spotify's audio features are calculated using proprietary algorithms, making it difficult to assess their precision.

4. **Popularity metric bias**: Spotify's popularity favors recent music due to the platform's user demographics and recency of streaming data.

This project could be extended in several promising ways:

1. **Genre-specific analysis**: Breaking down trends by genre could reveal how different music styles have evolved independently.

2. **Feature interactions**: Exploring how combinations of features affect popularity could yield deeper insights.

3. **Cultural context**: Incorporating external data about technological innovations or cultural events could help explain why certain trends emerged when they did.

4. **Regional comparison**: Analyzing data from different countries could reveal cultural differences in music preferences.

5. **Artist-level analysis**: Examining how individual artists adapt their sound over time could provide insights into creative evolution.

In conclusion, our analysis has quantified and validated patterns in music evolution that might otherwise remain subjective impressions. We've shown how popular music has become more danceable, electronic, and louder over time - the very characteristics that predict higher popularity today. This suggests a co-evolution between production techniques and listener preferences, though the causal direction remains an open question for future research.

8 References

1. "Billboard Hot 100 Songs Spotify Data (1946-2022)", Kaggle, https://www.kaggle.com/datasets/tushar5harma/billboard-hot-100-songs-spotify-data-1946-2022
2. Hyndman, R.J., & Athanasopoulos, G. (2018). Forecasting: principles and practice (2nd ed.). OTexts.
3. Spotify Web API Documentation, "Audio Features", https://developer.spotify.com/documentation/web-api/reference/get-audio-features