SPOTIFY BUSINESS CASE - WHAT MAKES A TRACK POPULAR?

Forecasting Spotify's Track Popularity

TEAM 6

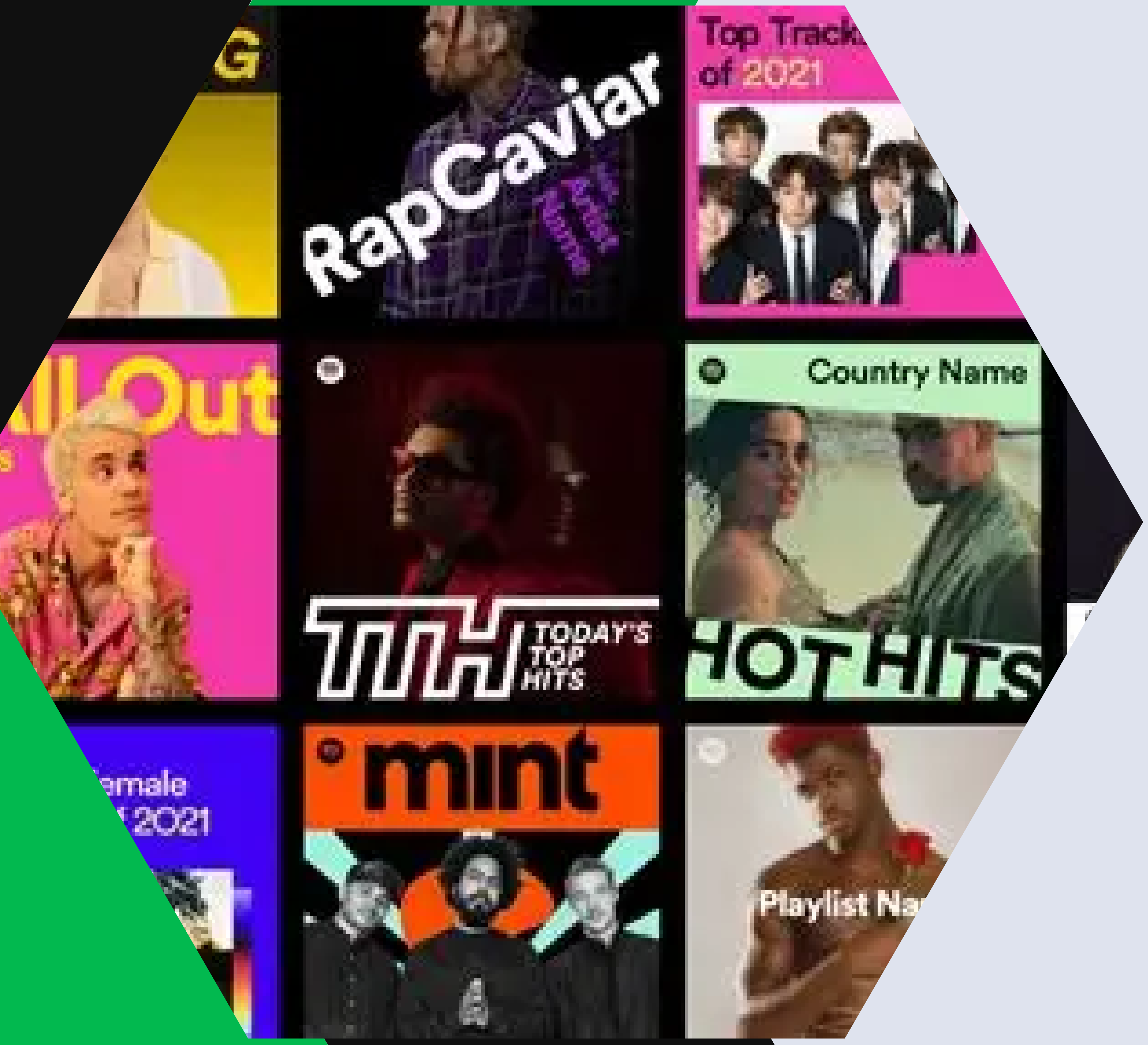
Meet the Team

Aishwarya Mathur

Abi Joshua George

Fernando Caballol

Sebastian Hernandez



Business Case & Objectives

Business Challenge

→ Spotify needs to understand what tracks are most likely to succeed to maximize marketing ROI

The Opportunity

→ We have leveraged data-driven models to

- **Predict Hits -** i.e., top-performing songs
- **Forecast Popularity** - over the next 12 months

Ojectives

→ **Optimize Spend:** Optimally allocate budget toward tracks with the highest success probability of being a "hit"
→ **Plan Release:** Align launch calendars with forecasted peaks and troughs

What is a “Hit”

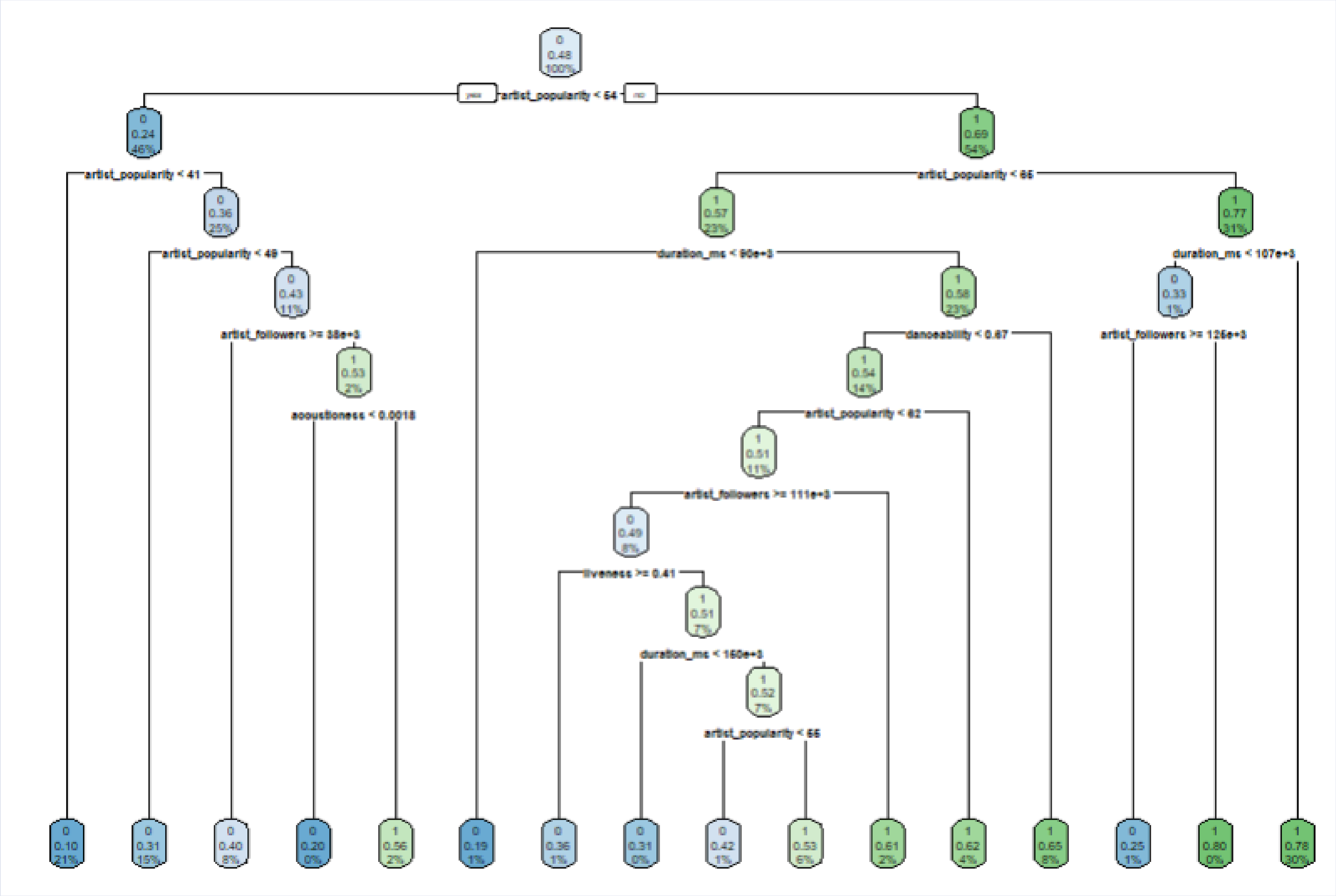
Top 25% → **1-in-4**

Any track whose popularity score ranks in the top 25%

This allows us to focus promotion campaigns on the 1-in-4 tracks most likely to generate engagement

Why top 25%? Captures the lion’s share that drives streams, shares and revenues + Balances selectivity with a managable promo pipeiline

GINI Decision Tree
[CHALLENGER MODEL]



Model Performance:

```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 14705  4427
         1  6554 15138

               Accuracy : 0.731
                 95% CI : (0.7267, 0.7353)
    No Information Rate : 0.5207
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.4634

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.6917
            Specificity : 0.7737
         Pos Pred Value : 0.7686
         Neg Pred Value : 0.6979
             Prevalence : 0.5207
         Detection Rate : 0.3602
   Detection Prevalence : 0.4686
      Balanced Accuracy : 0.7327
```

Key Predictors:

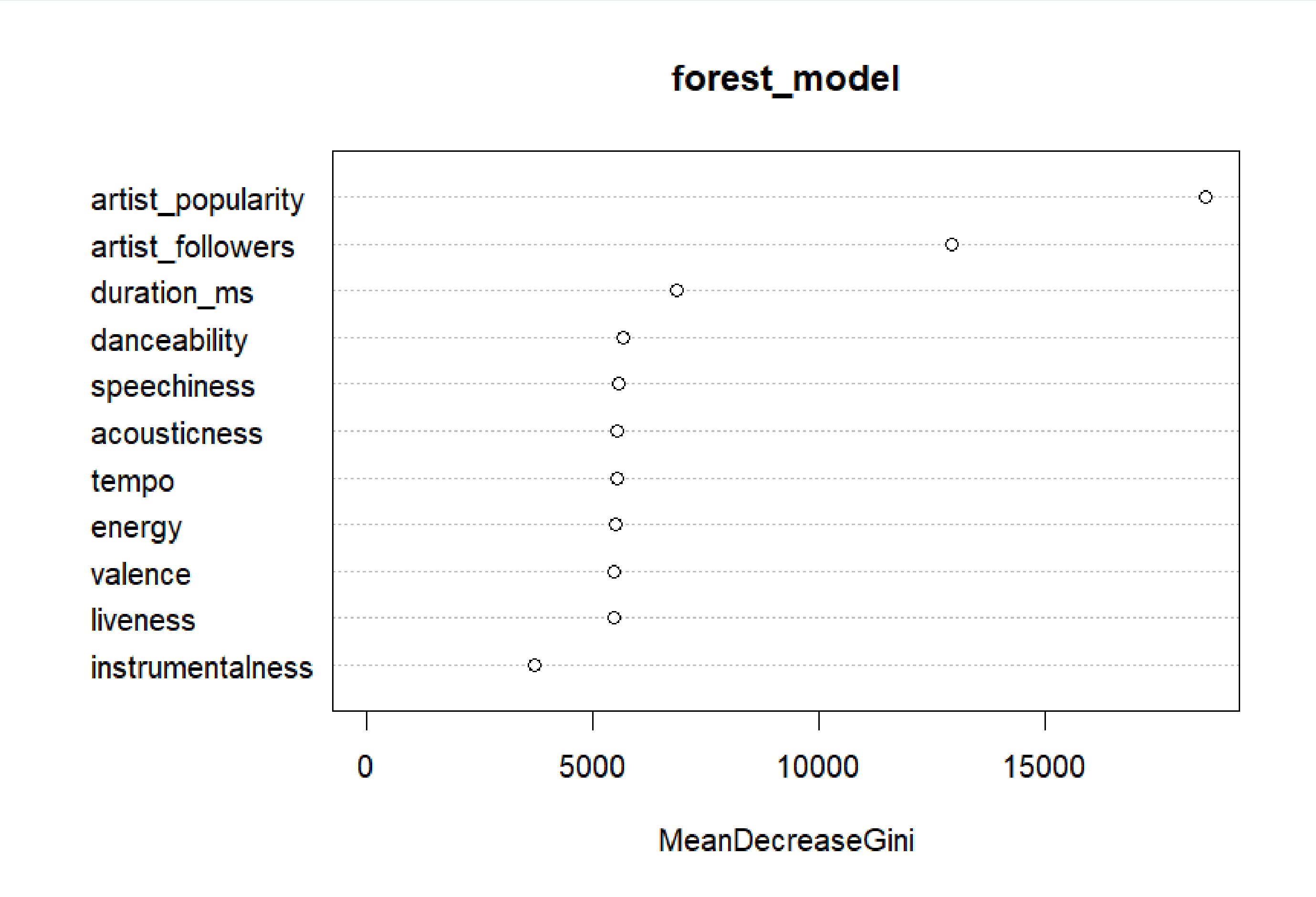
- artist_popularity was the root node → strongest predictor
- Other key splits: artist_followers, duration_ms, liveness, acousticness

Random Forest
[CHAMPION MODEL]

Model Performance

```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 16604  4770
         1  4655 14795

               Accuracy : 0.7691
                 95% CI : (0.765, 0.7732)
    No Information Rate : 0.5207
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.5374

 Mcnemar's Test P-Value : 0.2403

            Sensitivity : 0.7810
            Specificity : 0.7562
         Pos Pred Value : 0.7768
         Neg Pred Value : 0.7607
             Prevalence : 0.5207
         Detection Rate : 0.4067
   Detection Prevalence : 0.5236
      Balanced Accuracy : 0.7686
```

Variable Importance



Random Forest was the best-performing model, with **strong accuracy and balance**.

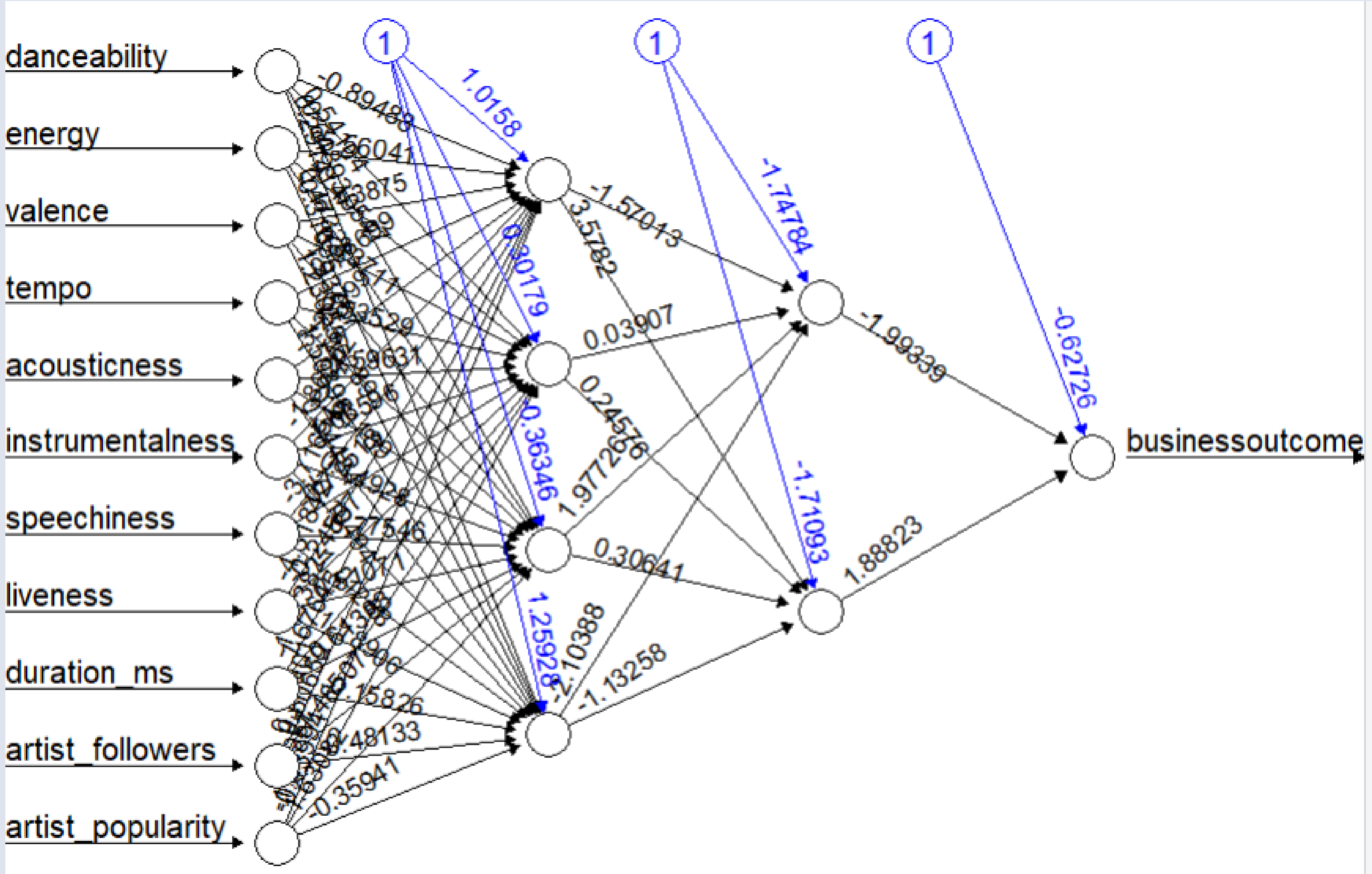
Top predictors: artist popularity, followers, and danceability.

Ideal for spotting tracks with high hit potential.

Neural Network

Model Performance:

```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 17649  9975
         1  3610  9590

               Accuracy : 0.6672
                 95% CI : (0.6626, 0.6718)
    No Information Rate : 0.5207
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3246

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.8302
            Specificity : 0.4902
         Pos Pred Value : 0.6389
         Neg Pred Value : 0.7265
             Prevalence : 0.5207
         Detection Rate : 0.4323
   Detection Prevalence : 0.6767
      Balanced Accuracy : 0.6602
```



- Model captures complex relationships between track features and hit likelihood.
- Achieved 66.7% accuracy with high sensitivity (83%).
- Weakest predictive model contender overall

Logistic Regression

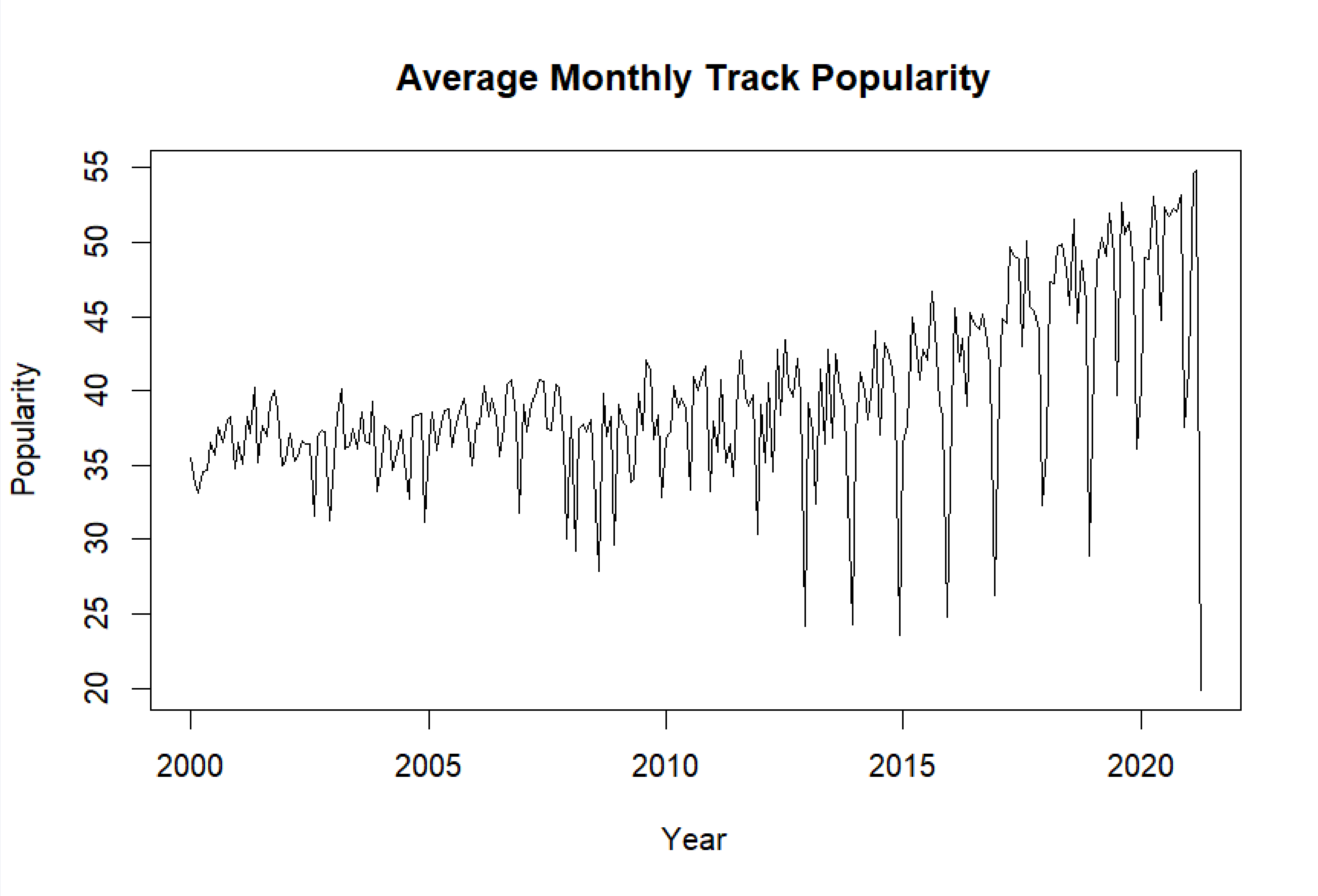
```
Coefficients:
      (Intercept)       danceability             energy            valence              tempo
       -5.021e+00          1.535e+00         -2.780e-01         -5.089e-01          1.668e-03
     acousticness   instrumentalness        speechiness           liveness        duration_ms
        3.347e-02         -7.071e-01         -9.777e-01         -7.092e-01         -1.422e-07
 artist_followers  artist_popularity
       -6.247e-08          8.357e-02

Degrees of Freedom: 163294 Total (i.e. Null);  163283 Residual
Null Deviance:      226100
Residual Deviance: 175600       AIC: 175600
```

Model Performance:

```
Confusion Matrix and Statistics

          Reference
Prediction     0     1
         0 17779  8665
         1  3480 10900

               Accuracy : 0.7025
                 95% CI : (0.698, 0.7069)
    No Information Rate : 0.5207
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3976

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.8363
            Specificity : 0.5571
         Pos Pred Value : 0.6723
         Neg Pred Value : 0.7580
             Prevalence : 0.5207
         Detection Rate : 0.4355
   Detection Prevalence : 0.6478
      Balanced Accuracy : 0.6967
```

- Baseline linear model delivers solid accuracy (70%) and strong sensitivity (84%)—making it a good classifier for detecting hits.
- Coefficients suggest danceability and artist popularity are positively associated with success, while speechiness and valence negatively impact hit likelihood.
- Lower specificity (56%) meaning more false positives compared to other models.

Average Monthly Popularity (2000–2025)



Distinct seasonal patterns:

- Sharp dips in February to March (likely post-holiday lull)
- Spikes in December, likely due to holidays, year-end playlists, and music releases

Consistent fluctuations highlight the importance of seasonal campaign timing and forecasting models like ARIMA to support release planning and marketing strategies.

Understanding key patterns: Trends, Seasonality & Stationarity

Augmented Dickey-Fuller (ADF) Test: Shows stationarity ($p < 0.05$), indicating the series is stable around a constant mean/variance, no non-seasonal differencing needed.

Last Month's Momentum: The ACF bar at lag 1 is nearly at 1.0, showing this month's popularity is highly tied to last month's performance.

Yearly Pattern: Both ACF and PACF spike strongly at lag 12, confirming a clear annual cycle in listener engagement.

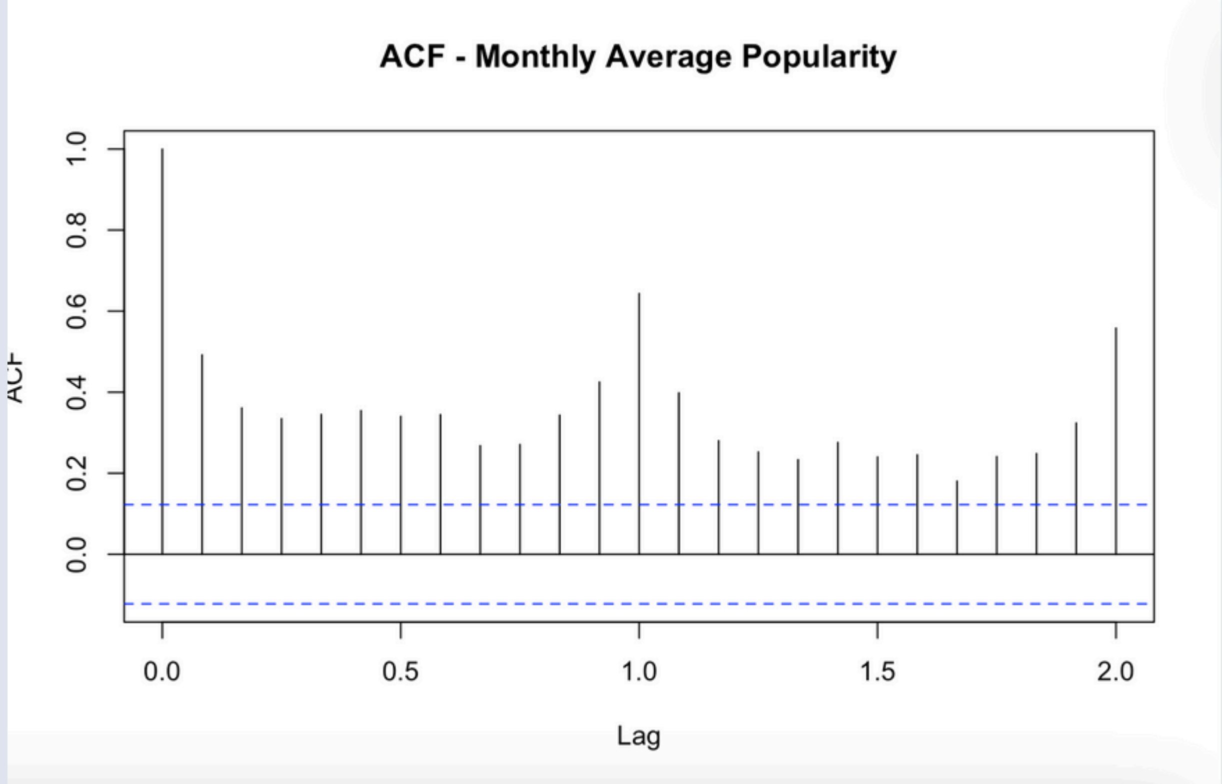
Quick Fade of Memory: After the first lag, correlations drop off quickly, so beyond one month, older data adds little new insight.

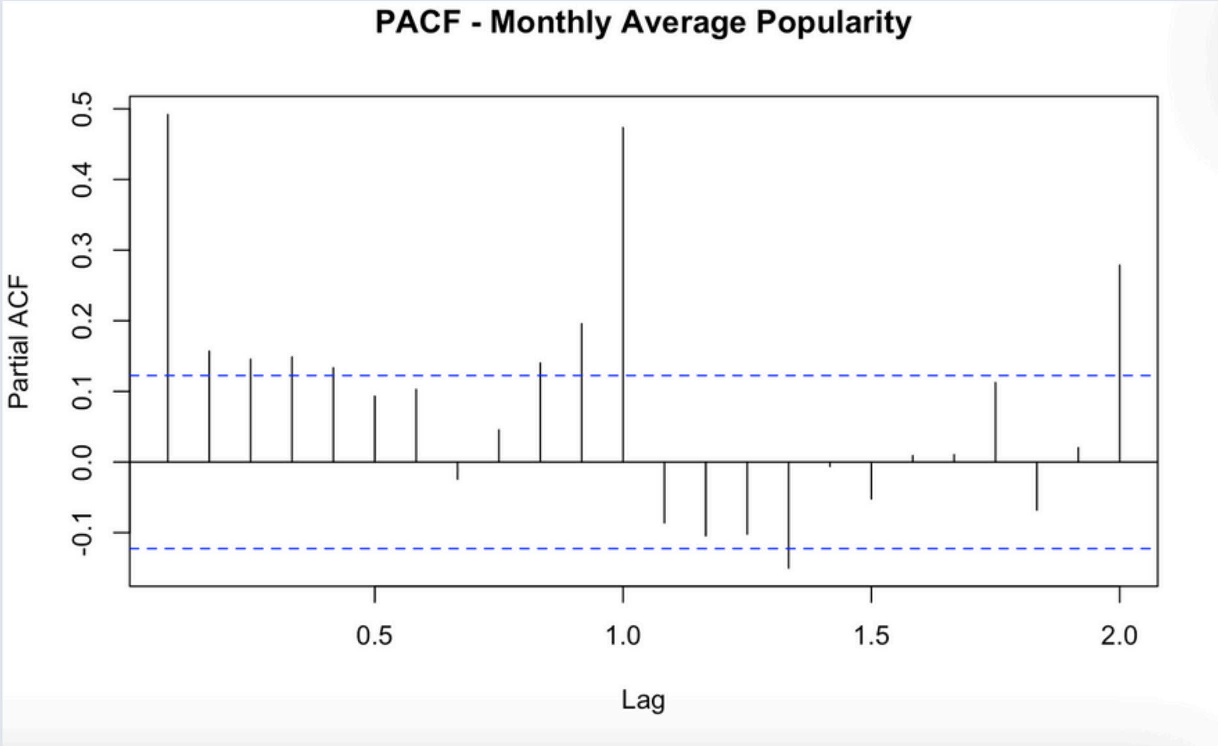
Seasonal and Short-Term Mix: These patterns tell us to include:

- One-month AR term (to capture immediate momentum)
- One-month MA term (to smooth random swings)
- One-year seasonal AR term (to lock in the annual cycle)

These diagnostics guided our choice of ARIMA(1,0,1)(1,1,0)[12], balancing trend, seasonality, and noise.

```
        Augmented Dickey-Fuller Test

data:  popularity_ts
Dickey-Fuller = -4.5274, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```





Next 12-Month Popularity Forecast

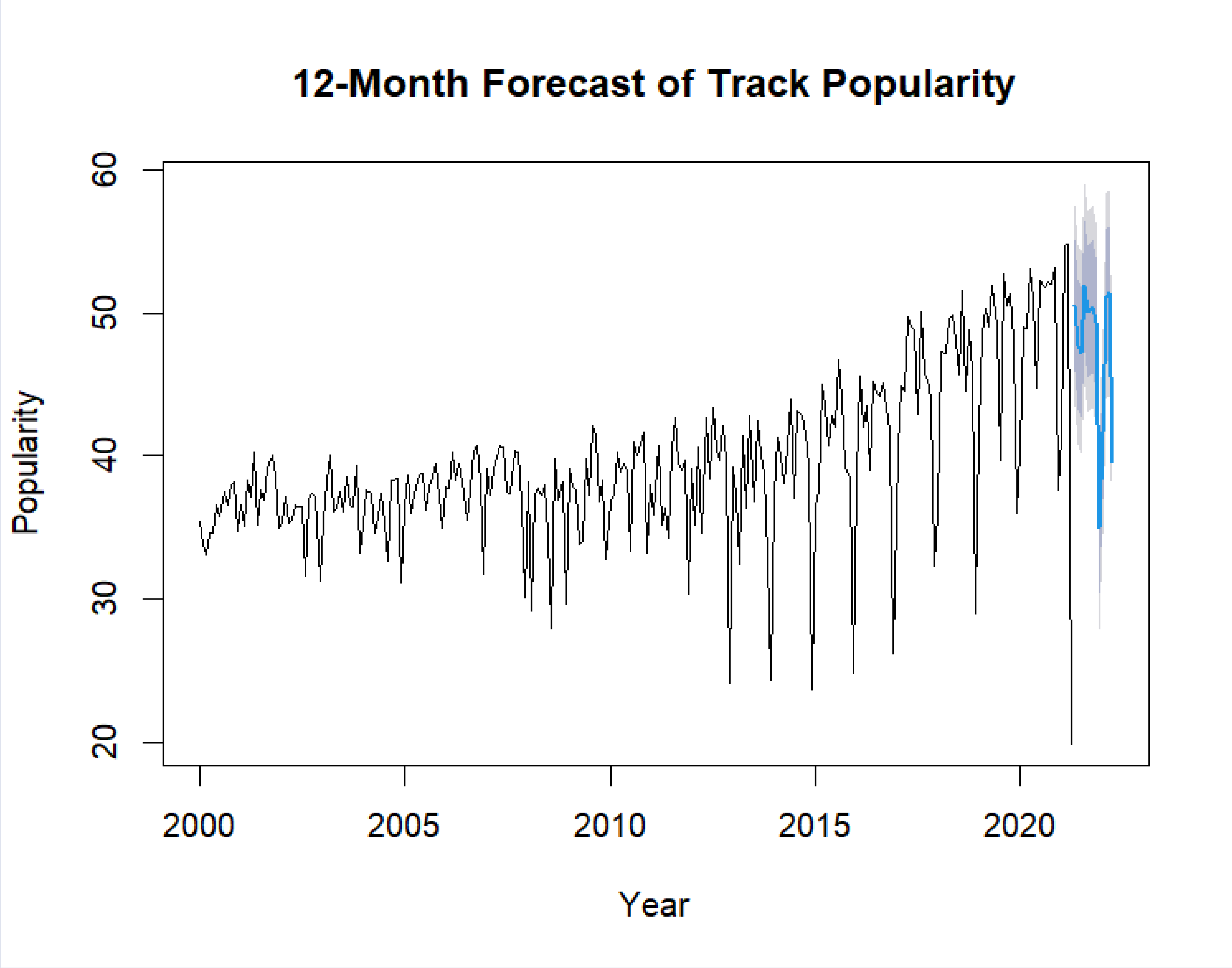
Steady Growth Expected: Forecast line sits above historical average, projecting a rise to around 48–52 popularity points over the coming year.

Seasonal Variability Persists: Confidence bands widen around typical holiday months and troughs, indicating larger swings—plan marketing pushes in Dec and buffer for slow Feb/Mar.

Confidence in Forecast: Narrow bands in mid-year suggest more predictable performance during Q2–Q3—optimize promotional budgets then for maximum ROI.

Actionable Insights:

- Q4 Releases: Leverage elevated December demand by scheduling major track launches in Q4.
- Risk Management: Set aside contingency ad spend for months with higher forecast uncertainty.
- Quarterly Reviews: Update model with new data each quarter to refine forecasts and adjust strategy.



Key Insights & Recommendations

1

Random Forest Classifier

Achieved 77% accuracy in identifying hit tracks (top 25% by popularity). Its superiority over baseline models, and its value for precise campaign targeting.

2

ARIMA Forecasting

The ARIMA(1,0,1)(1,1,0) captured trend and seasonality in track popularity, enabling reliable 12-month forecasts for planning release calendars.

3

Key Predictors and Feature Importance

- Artist popularity and follower count are the strongest predictors of hit tracks.
- Social and artist metrics outperform audio features for forecasting success.

4

Seasonality and Strategic Implications

- Popularity peaks in December, with additional surges in Q2–Q3.
- Guide optimal timing for flagship releases and budget allocation

Future Directions & Implementation Plan

- **Expand Data Signals:** Integrate lyrics sentiment analysis and real-time social media metrics (e.g., TikTok/Instagram mentions, playlist adds) to boost prediction accuracy and trend detection.

- **Granular Segmentation:** Develop genre- and region-specific models to tailor recommendations and forecasts for different listener segments.

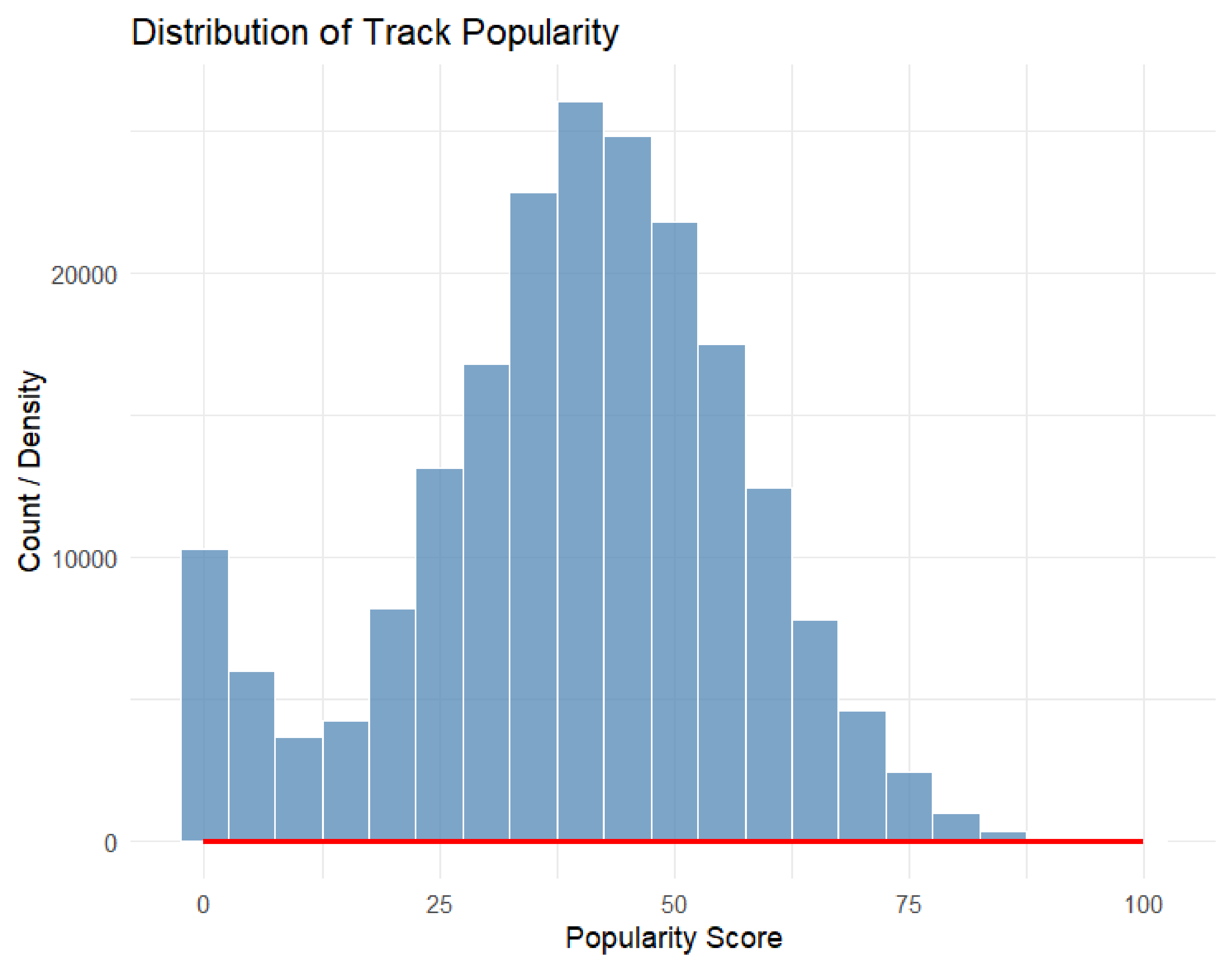
- **Process Automation**: Build a live dashboard for real-time monitoring and automate quarterly model retraining to keep insights current.

- **Pilot, Testing & Scalability**: Run Q4 campaign pilots and A/B tests to validate impact, while establishing a scalable analytics framework for continuous improvement and future growth.

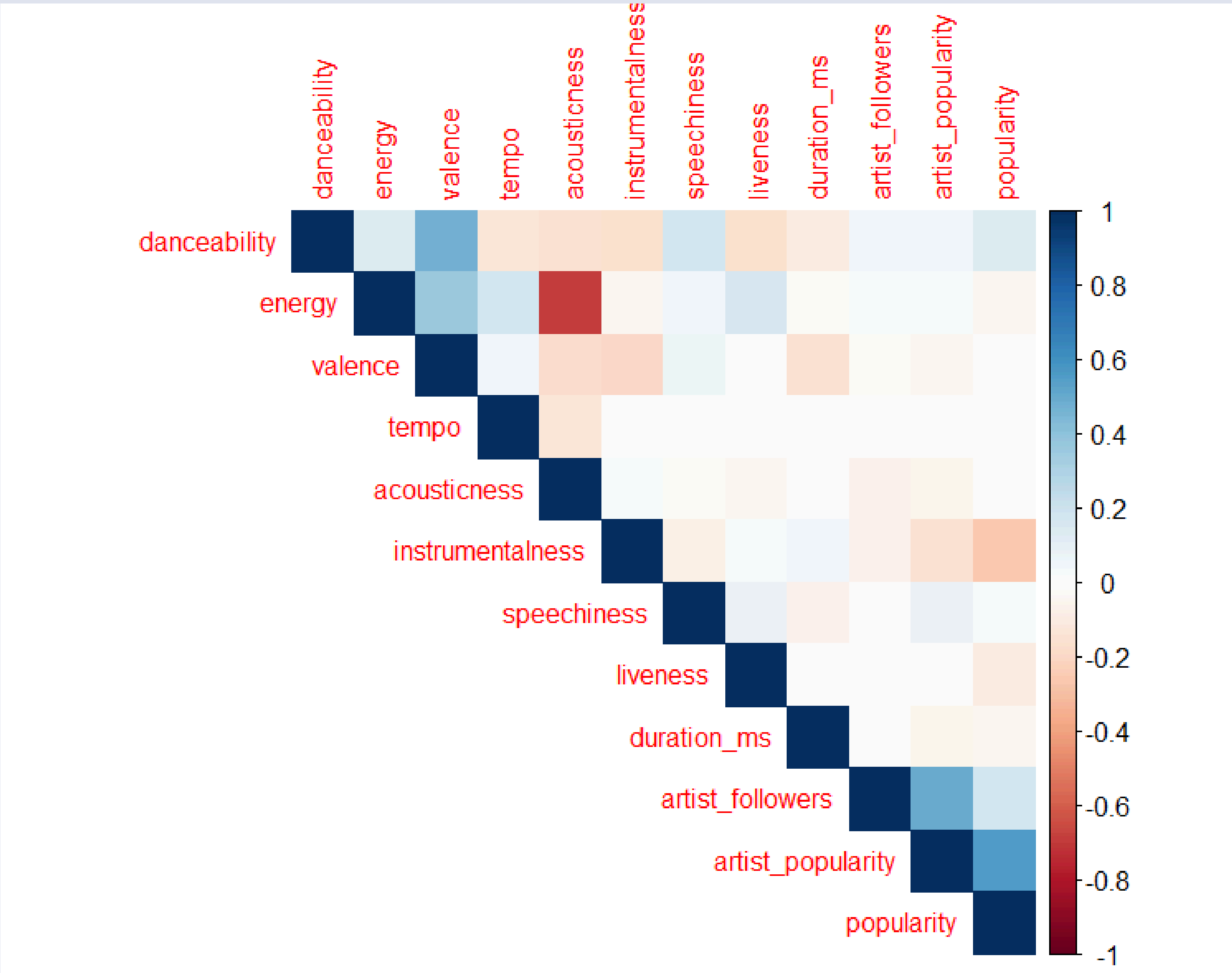
Thank you!

Feel free to approach us if you have any questions.

Appendix

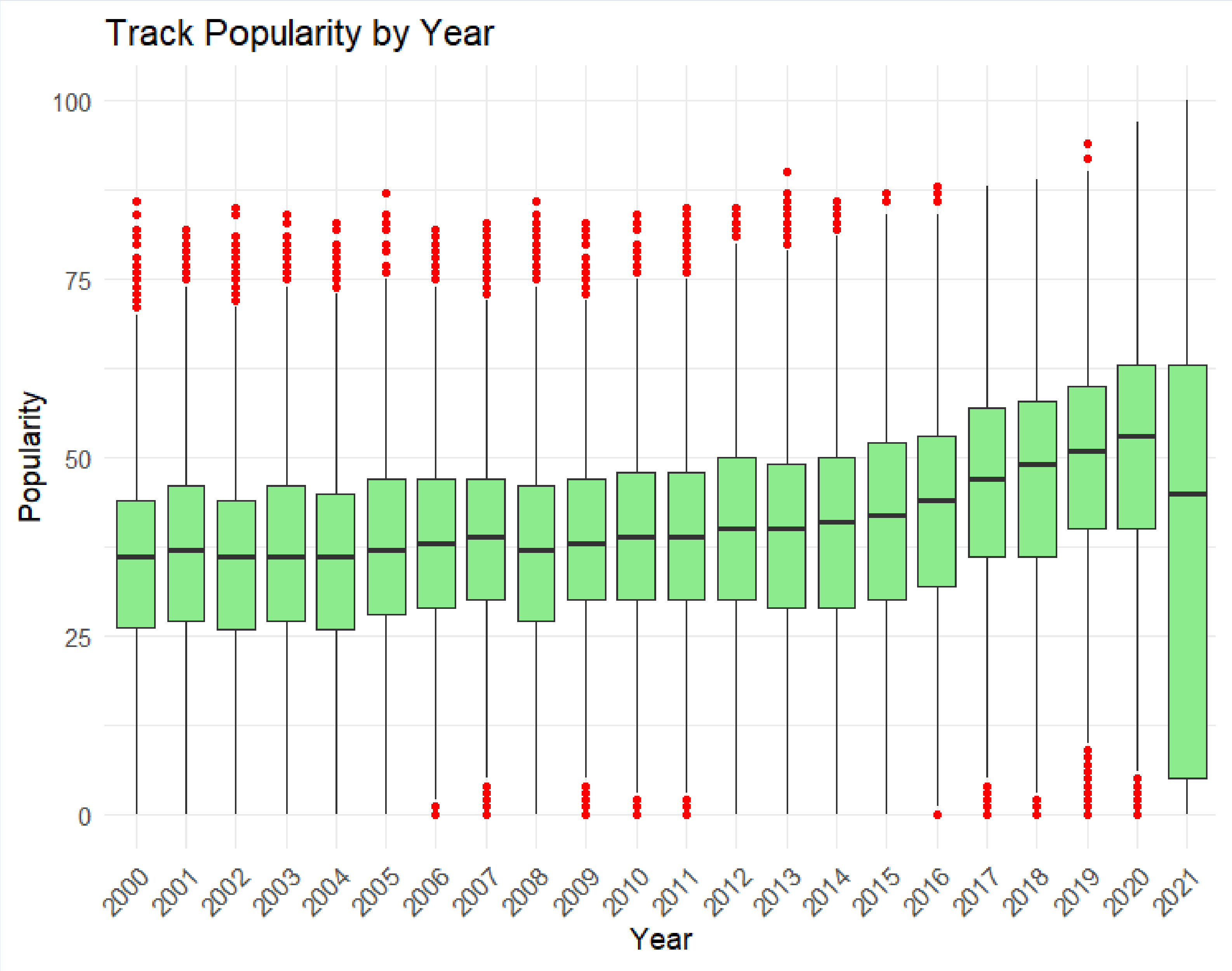


Distribution of Track Popularity
Count / Density
20000
10000
0
0
25
50
75
100
Popularity Score

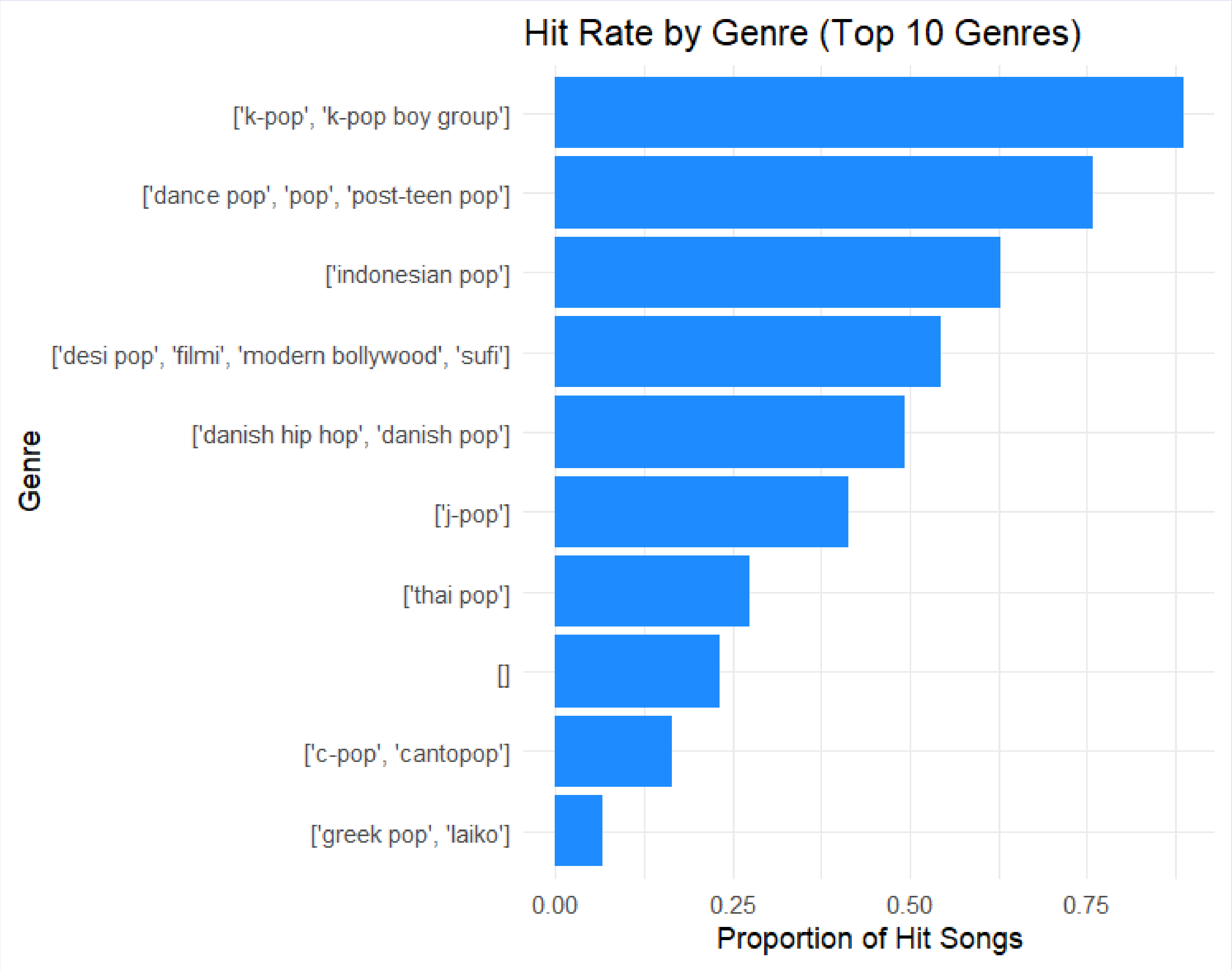


danceability
energy
valence
tempo
acousticness
instrumentalness
speechiness
liveness
duration_ms
artist_followers
artist_popularity
popularity
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Appendix

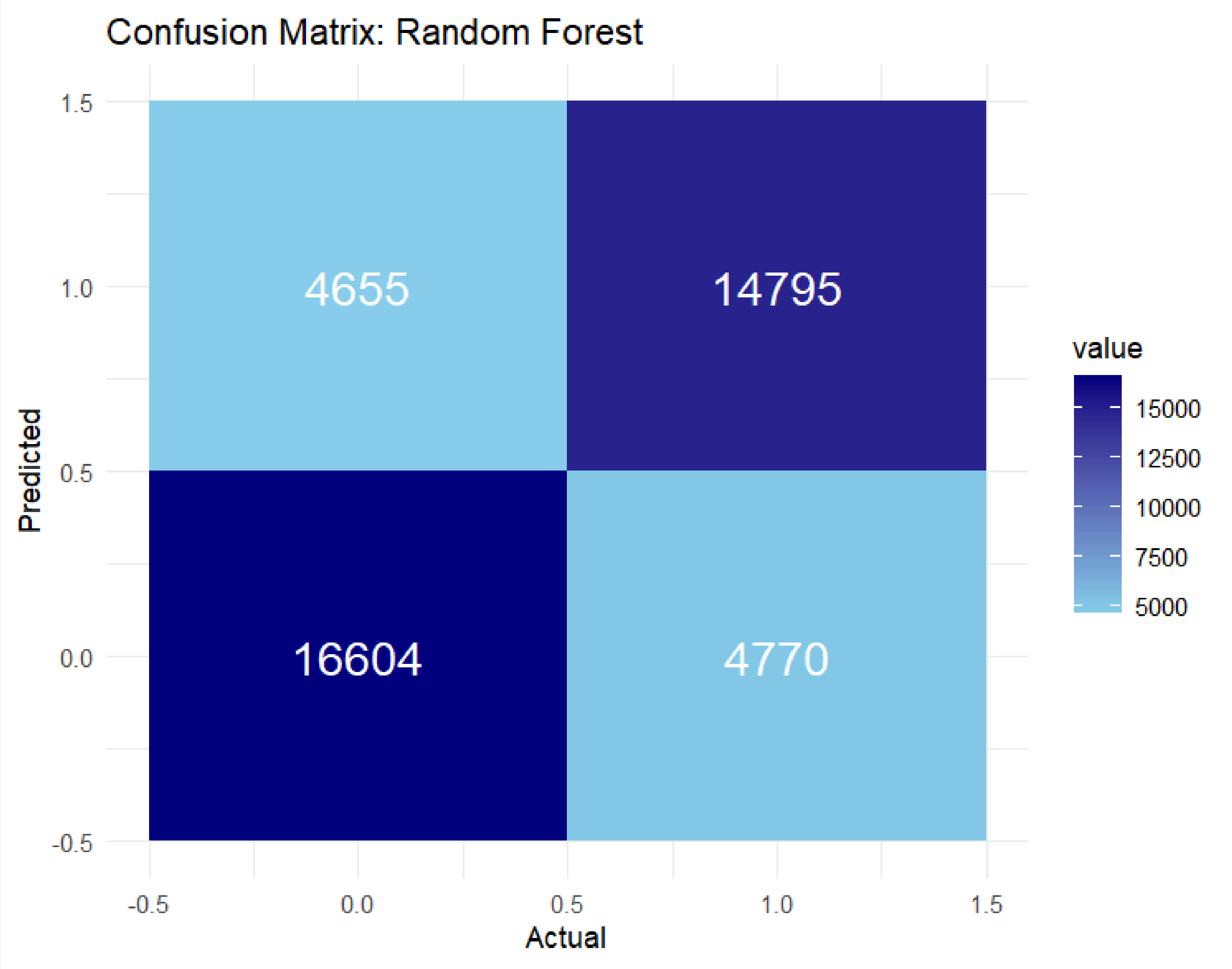


Track Popularity by Year
Popularity
100
75
50
25
0
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
Year

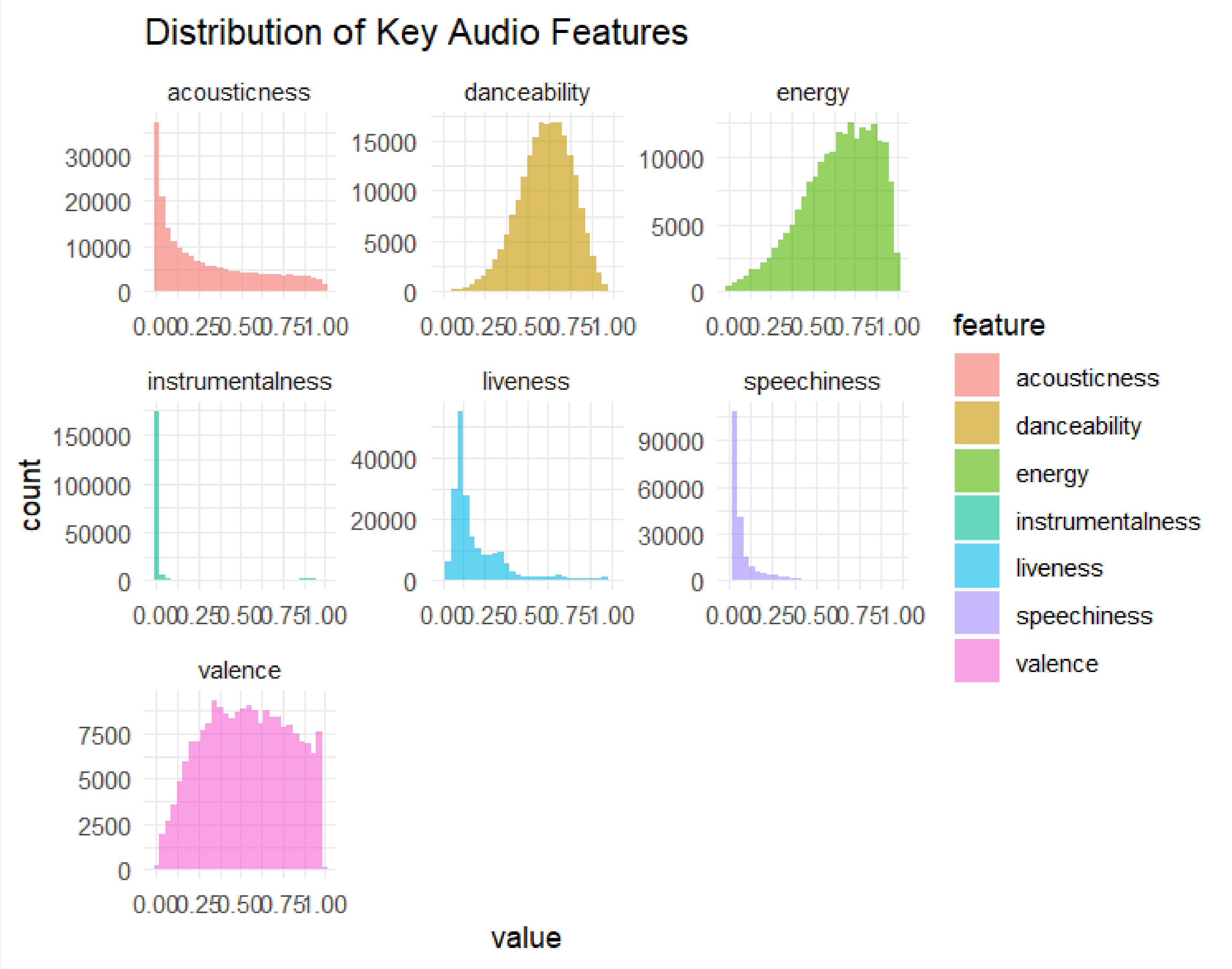


Hit Rate by Genre (Top 10 Genres)
Genre
['k-pop', 'k-pop boy group']
['dance pop', 'pop', 'post-teen pop']
['indonesian pop']
['desi pop', 'filmi', 'modern bollywood', 'sufi']
['danish hip hop', 'danish pop']
['j-pop']
['thai pop']
[]
['c-pop', 'cantopop']
['greek pop', 'laiko']
0.00
0.25
0.50
0.75
Proportion of Hit Songs

Appendix

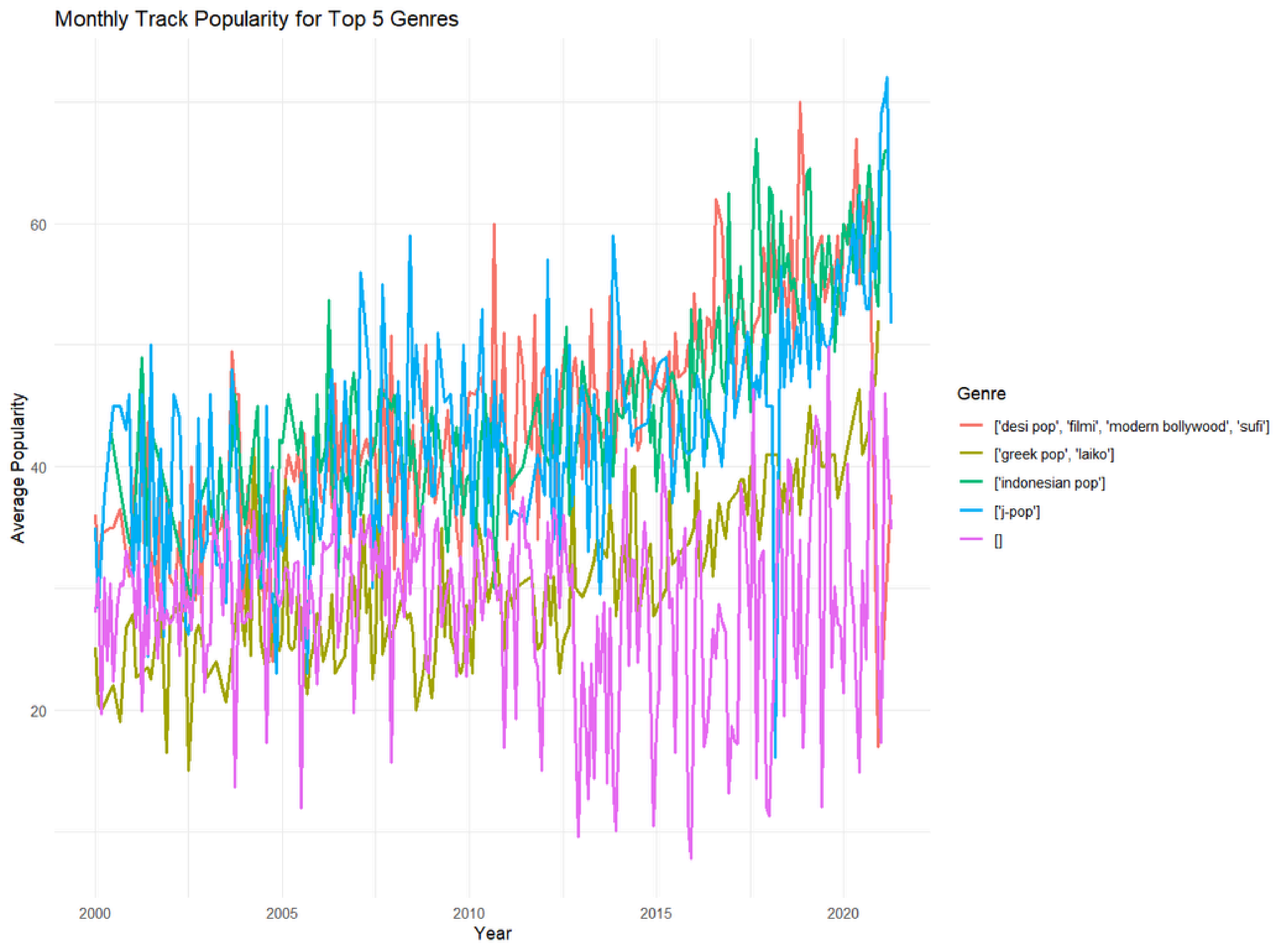


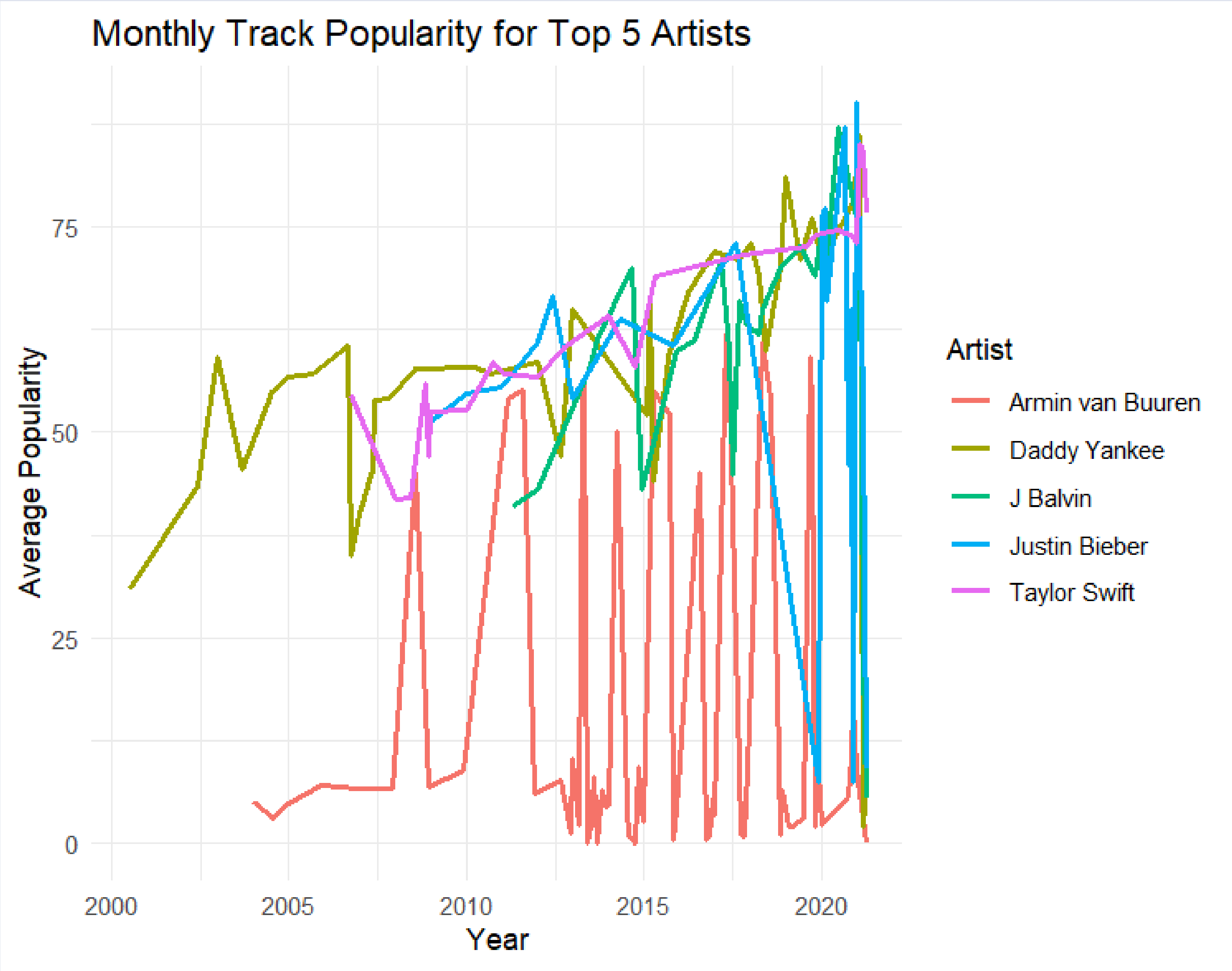
Confusion Matrix: Random Forest
4655
14795
16604
4770
Predicted
Actual
-0.5
0.0
0.5
1.0
1.5
value
15000
12500
10000
7500
5000



Distribution of Key Audio Features
acousticness
danceability
energy
instrumentalness
liveness
speechiness
valence
count
value
feature
acousticness
danceability
energy
instrumentalness
liveness
speechiness
valence

Appendix


Monthly Track Popularity for Top 5 Genres
Average Popularity
Year
20
40
60
2000
2005
2010
2015
2020
Genre
['desi pop', 'filmi', 'modern bollywood', 'sufi']
['greek pop', 'laiko']
['indonesian pop']
['j-pop']
[]


Monthly Track Popularity for Top 5 Artists
Average Popularity
Year
0
25
50
75
2000
2005
2010
2015
2020
Artist
Armin van Buuren
Daddy Yankee
J Balvin
Justin Bieber
Taylor Swift

Why ARIMA(1,0,1)(1,1,0)[12] Is Our Forecast Engine

Stationarity Check – ADF Test

- Purpose: Verifies if our series has a constant mean and variance over time
- Result: p-value < 0.05 → The series is already stationary (no overall drift)
- Business Implication: We skip non-seasonal differencing and focus only on seasonal adjustments

Non-Seasonal Components – ARIMA(1,0,1)

- AR(1): Captures momentum from last month's popularity—today's value leans on one-month lag
- I(0): No non-seasonal differencing needed since the data is stationary after ADF confirmation
- MA(1): Smooths out random shocks by incorporating last month's forecast error

Seasonal Components – (1,1,0)[12]

- Seasonal AR(1): Models the 12-month echo (e.g., this December reflects last December's pattern)
- Seasonal I(1): One seasonal difference at lag 12 removes the repeating annual cycle before modeling
- Seasonal MA(0): No additional seasonal error term—seasonal AR suffices to capture yearly effects

Putting It All Together

1. Seasonally difference the data to strip out the annual cycle (guided by ADF and ACF/PACF)
2. Apply AR(1) to capture immediate momentum
3. Use MA(1) to smooth one-month shocks
4. Incorporate SAR(1) at lag 12 to honor the yearly rhythm

This combination—ARIMA(1,0,1)(1,1,0)[12]—balances momentum, noise smoothing, and seasonality, delivering a robust 12-month popularity forecast.