

Engenharia Elétrica

# Análise de Padrões de Streaming de Músicas Natalinas no Spotify (2017-2025)

Análise exploratória, temporal e preditiva de 7 músicas natalinas clássicas  
baseada em dados de streaming do Spotify

**Autor:** Jonas Souza

**Formação:** Engenheiro Eletricista

**Data:** 18 de dezembro de 2025

# 1 Resumo Executivo

Este relatório apresenta uma análise abrangente dos padrões de streaming de 7 músicas natalinas clássicas no Spotify entre 2017 e 2025. Utilizando técnicas de ciência de dados e visualização, foram investigadas tendências temporais, sazonalidade e evolução da popularidade dessas músicas durante a temporada de festas.

## 1.1 Principais Resultados

- **Líder absoluto:** “All I Want for Christmas Is You” (Mariah Carey) com 1,77 bilhões de streams totais
- **Crescimento anual:** CAGR de 23,34% (2017-2024)
- **Pico histórico:** Todas as 7 músicas atingiram seus picos máximos na Semana 52 de 2024 (26 de dezembro)
- **Sazonalidade:** Coeficiente de variação médio de 64%, indicando alta concentração nos períodos festivos
- **Previsão 2025:** Crescimento projetado de 12-13% para as músicas mais populares

## 2 Introdução

### 2.1 Contexto

O streaming de música transformou fundamentalmente a indústria musical nas últimas duas décadas. No contexto natalino, músicas clássicas apresentam um comportamento único e altamente sazonal, com picos concentrados em semanas específicas do ano.

### 2.2 Objetivos

1. Analisar padrões descritivos de streaming das 7 principais músicas natalinas
2. Investigar tendências temporais e sazonalidade entre 2017-2025
3. Desenvolver modelos preditivos para a temporada natalina de 2025
4. Identificar características únicas de cada música no cenário de streaming

### 2.3 Músicas Analisadas

1. All I Want for Christmas Is You – Mariah Carey
2. Last Christmas – Wham!
3. Rockin’ Around the Christmas Tree – Brenda Lee
4. Jingle Bell Rock – Bobby Helms
5. Santa Tell Me – Ariana Grande

6. It's Beginning to Look a Lot Like Christmas – Michael Bublé
7. Feliz Navidad – José Feliciano

## 3 Metodologia

### 3.1 Coleta de Dados

Os dados foram coletados via web scraping do site Kworb.net, plataforma especializada em tracking de streaming do Spotify.

#### 3.1.1 Características dos Dados

- **Período:** 01/01/2017 – 28/11/2025
- **Granularidade:** Semanal (atualizações às quintas-feiras)
- **Disponibilidade:** Apenas quando músicas estão no Top 200 Global
- **Variáveis:** Data, música, artista, streams semanais, dia da semana, semana do ano

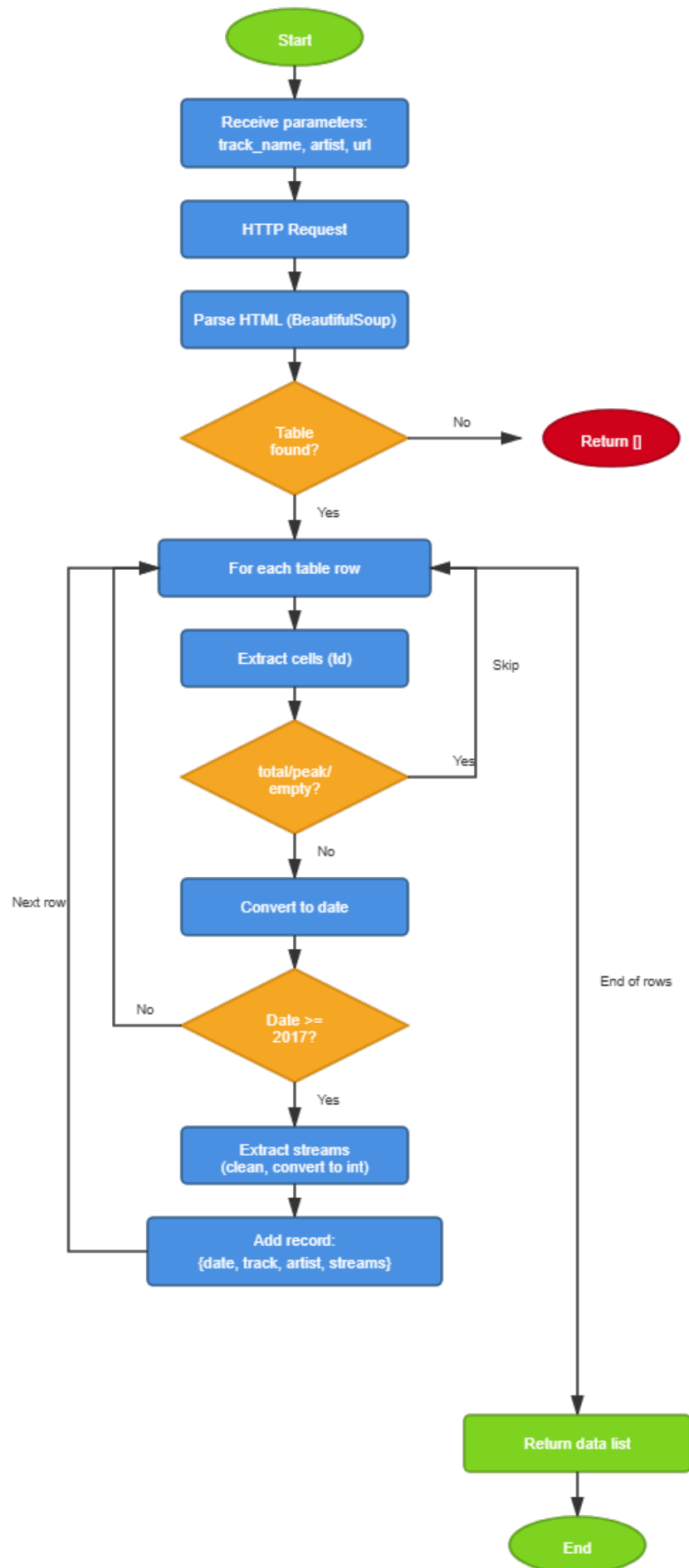


Figura 1: Fluxograma do <sup>3</sup> processo de coleta de dados

### 3.1.2 Limitações dos Dados

- Dados disponíveis apenas quando as músicas estão no Top 200 Global do Spotify
- Lacunas entre fevereiro e outubro (período fora de temporada)
- Granularidade semanal (não diária)
- Dados de 2025 parciais (até 28 de novembro)

## 3.2 Ferramentas e Tecnologias

- **Python 3.x:** Linguagem principal
- **Pandas:** Manipulação e análise de dados
- **NumPy:** Computação numérica
- **Matplotlib/Seaborn:** Visualização de dados
- **BeautifulSoup:** Web scraping
- **Statsmodels:** Modelagem ARIMA
- **Jupyter Notebook:** Ambiente de desenvolvimento

## 4 Análise Exploratória de Dados

### 4.1 Estatísticas Descritivas

A análise descritiva foi realizada para entender o comportamento individual de cada uma das 7 músicas natalinas. Cada registro representa dados semanais de streams, tipicamente registrados às quintas-feiras.

#### 4.1.1 Streams Totais por Música (2017-2025)

O ranking de popularidade total revelou:

Tabela 1: Streams totais por música (2017-2025)

Música	Streams Totais
All I Want for Christmas Is You	1,77 bilhões
Last Christmas	1,50 bilhões
Rockin' Around the Christmas Tree	1,16 bilhões
Jingle Bell Rock	1,03 bilhões
Santa Tell Me	1,01 bilhões
It's Beginning to Look a Lot Like Christmas	927 milhões
Feliz Navidad	633 milhões

### 4.1.2 Estatísticas Semanais

As estatísticas descritivas dos streams semanais revelaram padrões distintos:

Tabela 2: Estatísticas descritivas de streams semanais por música

Música	Count	Média (M)	DP (M)	Mín (M)	Máx (M)	CV%
All I Want for Christmas Is You	69	25,69	17,77	3,82	92,51	69,18
Last Christmas	63	23,88	17,05	4,04	91,09	71,40
Rockin' Around the Christmas Tree	54	21,53	15,80	4,30	82,41	73,36
Jingle Bell Rock	49	21,00	13,48	5,56	72,63	64,21
Santa Tell Me	53	19,12	11,37	4,22	60,26	59,45
It's Beginning to Look a Lot Like Christmas	49	18,93	10,25	4,14	47,36	54,15
Feliz Navidad	39	16,23	9,63	4,78	48,74	59,34

#### Observações importantes:

- “Rockin’ Around the Christmas Tree” lidera o coeficiente de variação (73,36%), mostrando grande flutuação relativa
- “It’s Beginning to Look a Lot Like Christmas” apresenta menor variação, com streams mais consistentes
- A diferença entre média e mediana em todas as músicas confirma assimetria positiva nas distribuições

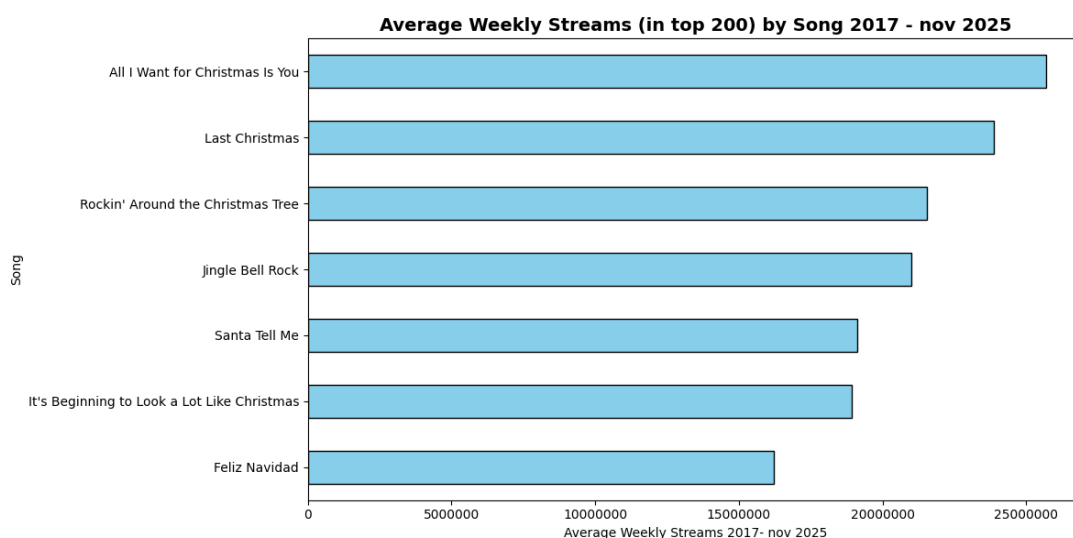


Figura 2: Média de streams semanais por música (2017-2025)

### 4.1.3 Coeficiente de Variação

O coeficiente de variação (CV%) permite comparar a estabilidade das músicas, revelando quais apresentam comportamento mais consistente ou volátil:

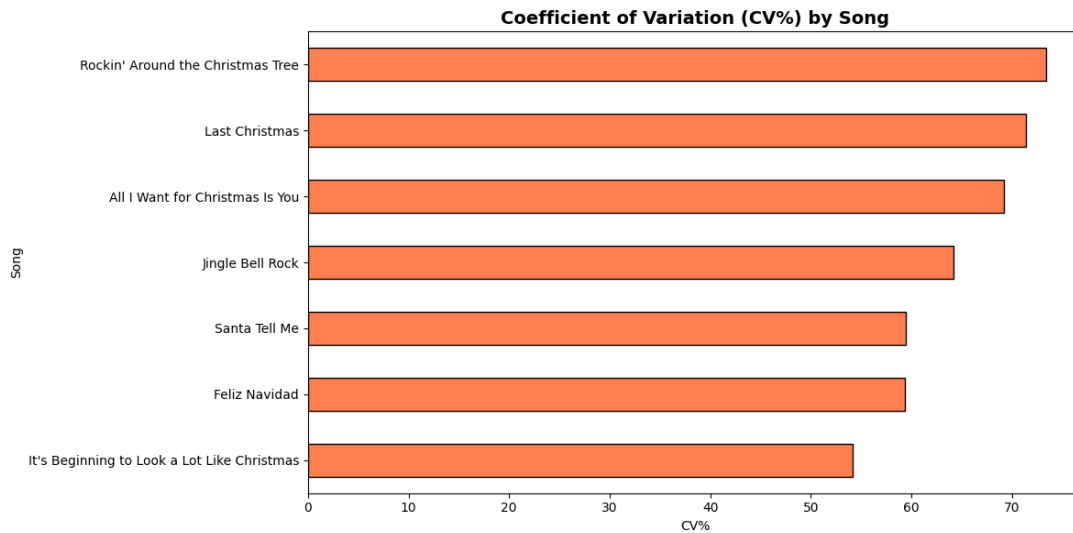


Figura 3: Coeficiente de variação por música

## 4.2 Análise de Distribuição e Frequência

### 4.2.1 Distribuição dos Streams Semanais

A análise de histogramas revelou que **nenhuma das 7 músicas apresenta distribuição normal**, mostrando padrões claramente assimétricos com concentração de valores abaixo da média e caudas longas à direita.

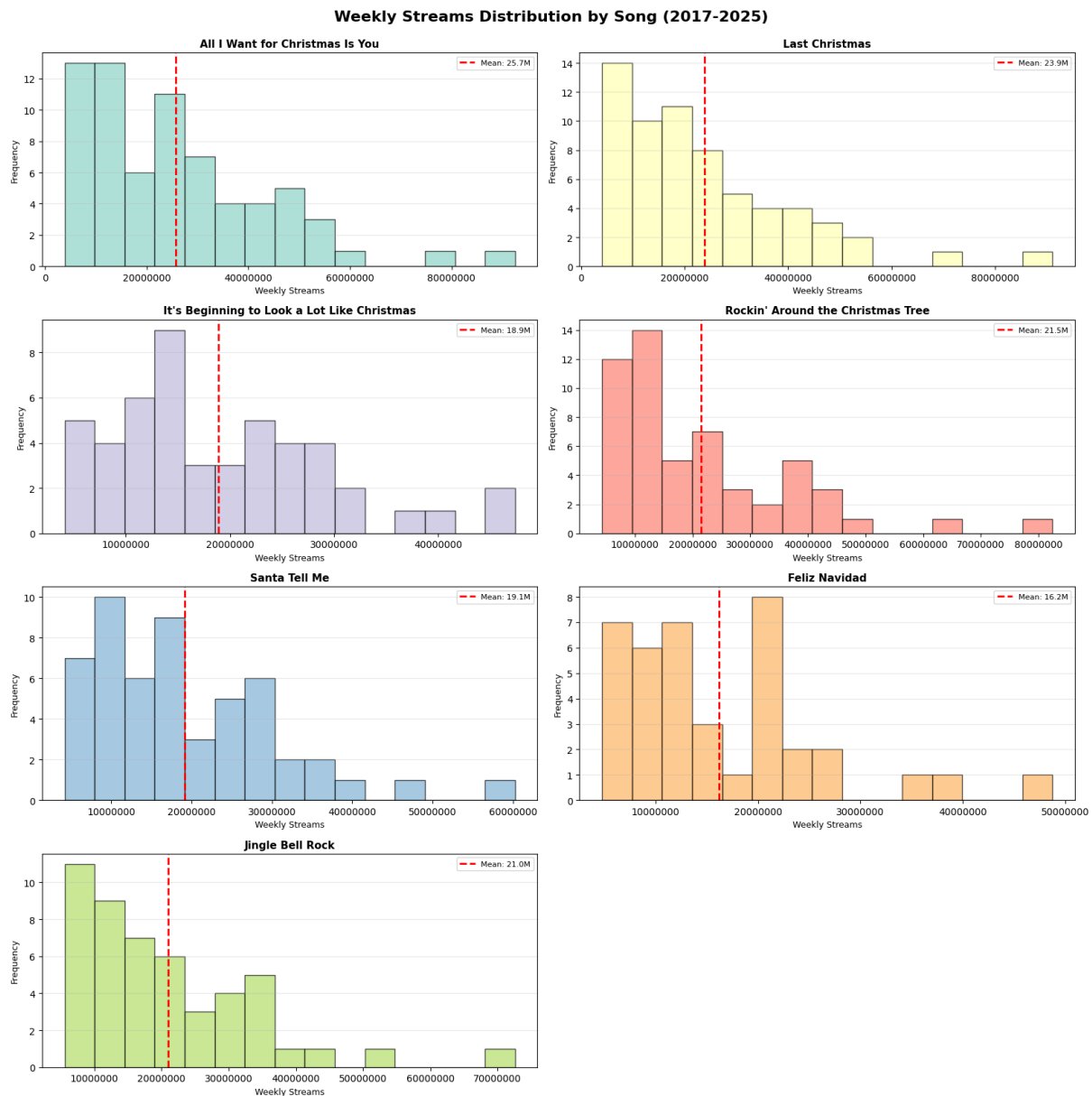


Figura 4: Distribuição de streams semanais por música (histogramas)

### Principais achados:

- “All I Want for Christmas Is You” exibe maior dispersão e maior média (25,7M streams semanais)
- “Feliz Navidad” apresenta menor média (16,2M) e distribuição mais concentrada em valores baixos
- Assimetria positiva em todas as músicas indica predominância de semanas com streams moderados e poucos eventos de altíssimo volume (acima de 60-80M)
- Distribuições não-gaussianas são consistentes com o viés de seleção dos dados (apenas Top 200)



### 4.2.2 Análise de Boxplot

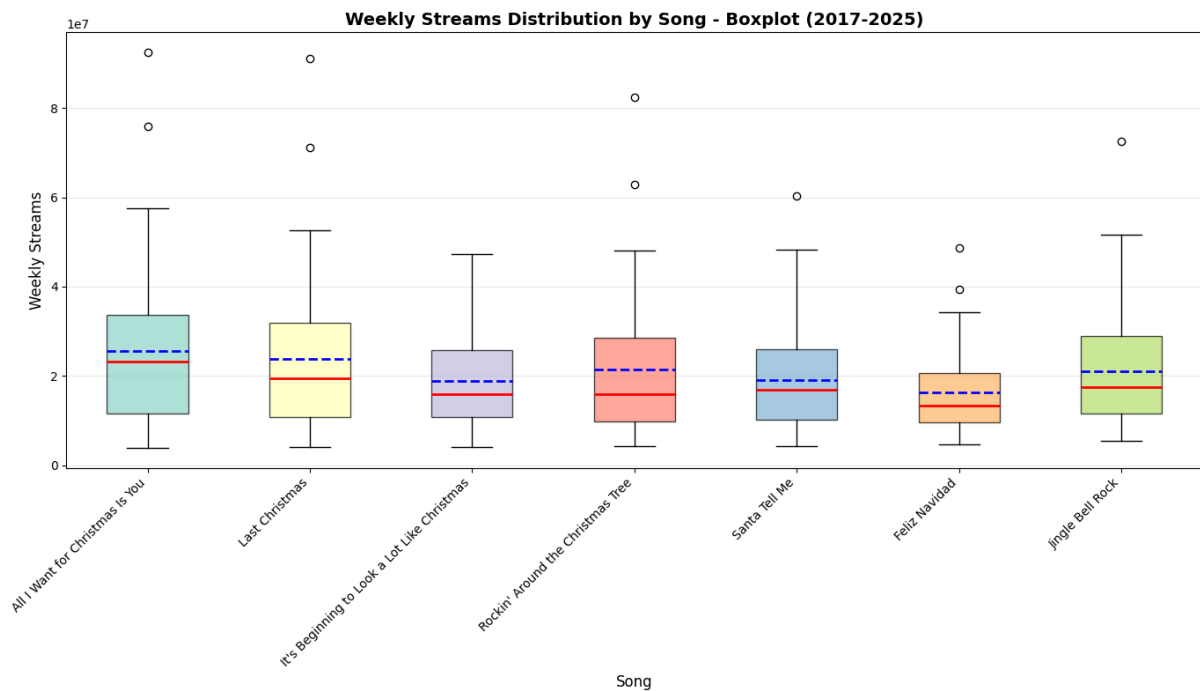


Figura 5: Distribuição de streams semanais – Boxplot comparativo

O boxplot comparativo revelou:

- Clara hierarquia de popularidade, com “All I Want for Christmas Is You” e “Last Christmas” liderando
- Outliers superiores significativos (variando de 6M a 10M) em todas as músicas
- “Feliz Navidad” mostra menor mediana e intervalo interquartil mais compacto
- Linhas azuis tracejadas (médias) consistentemente acima das linhas vermelhas (medianas), confirmando assimetria positiva

### 4.2.3 Distribuição Total de Streams

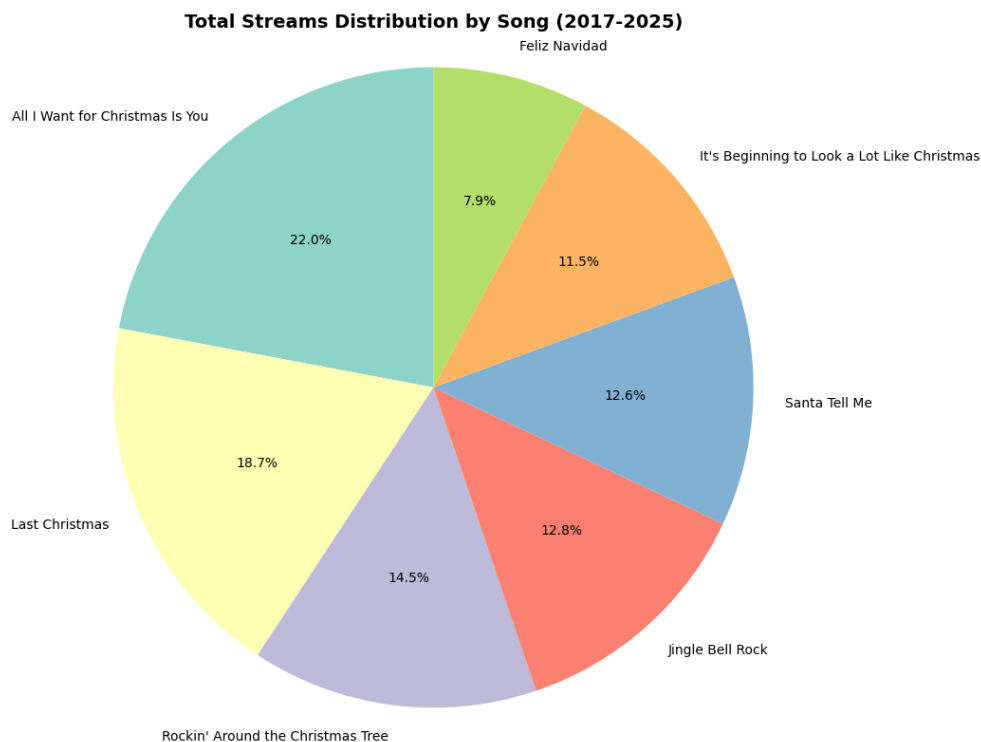


Figura 6: Distribuição percentual de streams totais (2017-2025)

### 4.3 Identificação de Picos

Análise dos picos máximos de streams semanais revelou padrão sincronizado:

Tabela 3: Picos máximos de streams semanais por música

Música	Streams Máximos (M)	Data/Semana
All I Want for Christmas Is You	92,51	26/12/2024 (Semana 52)
Last Christmas	91,09	26/12/2024 (Semana 52)
Rockin' Around the Christmas Tree	82,41	26/12/2024 (Semana 52)
Jingle Bell Rock	72,63	26/12/2024 (Semana 52)
Santa Tell Me	60,26	26/12/2024 (Semana 52)
Feliz Navidad	48,74	26/12/2024 (Semana 52)
It's Beginning to Look a Lot Like Christmas	47,36	26/12/2024 (Semana 52)

**Conclusão:** Todas as 7 músicas atingiram picos históricos na mesma semana (Semana 52 de 2024), indicando um surto sincronizado durante o período pós-Natal, sugerindo crescimento sem precedentes em 2024 ou que representa o período de coleta mais completo/recente do dataset.

## 5 Análise Temporal

### 5.1 Séries Temporais Completas (2017-2025)

A análise de séries temporais completas revelou padrões sazonais claros e consistentes ao longo dos anos.

#### 5.1.1 Séries Individuais

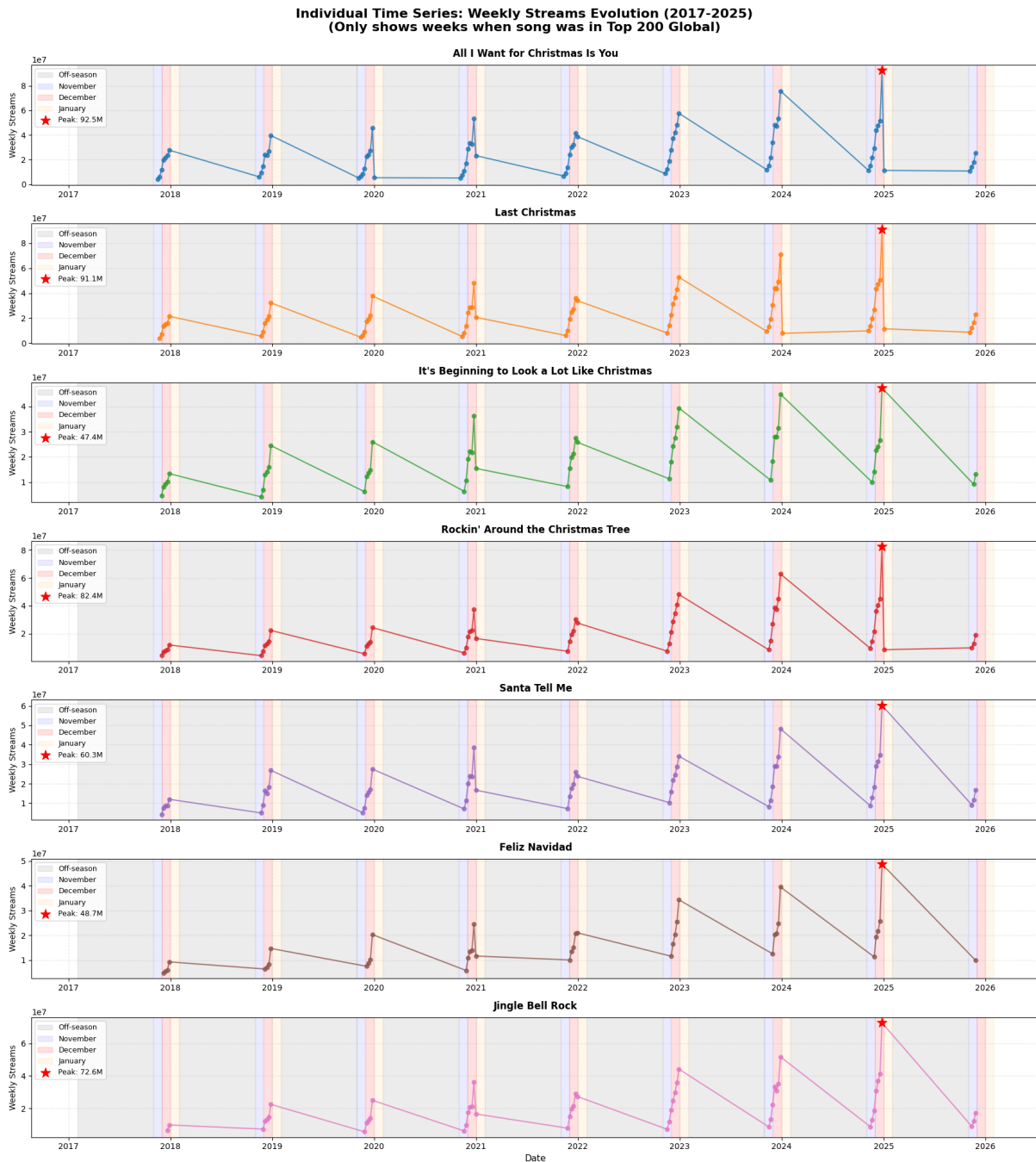


Figura 7: Séries temporais individuais – Evolução semanal de streams (2017-2025)

Padrões identificados:

- Períodos off-season (fevereiro-outubro) marcados em cinza – ausência de dados (fora do Top 200)
- Novembro (azul), dezembro (vermelho) e janeiro (laranja) destacados
- Picos máximos marcados com estrelas vermelhas
- Crescimento exponencial visível durante semanas 48-52
- Queda abrupta após semana 1 (janeiro)

### 5.1.2 Comparação entre Músicas

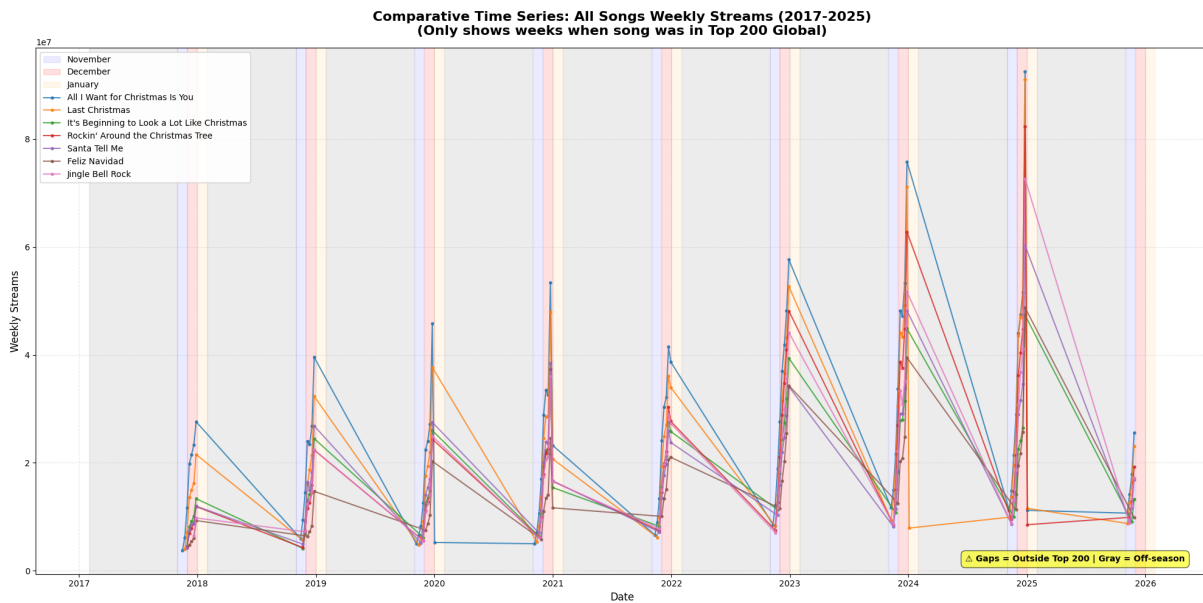


Figura 8: Série temporal comparativa de todas as músicas (2017-2025)

## 5.2 Padrão Sazonal do Ciclo Natalino

A análise focada nas semanas 45-52 + semana 1 revelou o padrão sazonal característico:

Tabela 4: Streams médios por semana durante a temporada natalina (todas as músicas combinadas)

Semana	Streams Médios (M)
45	8,7
46	9,2
47	10,8
48	15,0
49	21,7
50	23,7
51	26,7
52	38,7
1	8,9

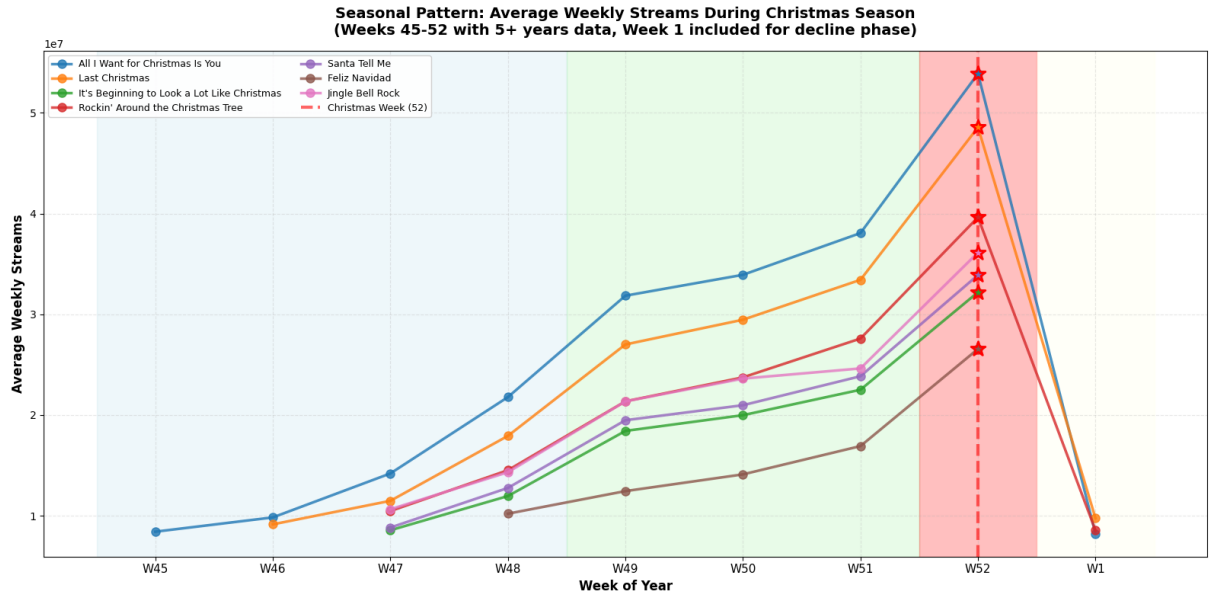


Figura 9: Padrão sazonal médio durante a temporada natalina

#### Fases identificadas:

- **Início (Semanas 45-47):** Crescimento gradual (8-11M streams/semana)
- **Aceleração (Semanas 48-50):** Crescimento exponencial (15-24M streams/semana)
- **Pico (Semanas 51-52):** Pico máximo (27-39M streams/semana)
- **Declínio (Semana 1):** Queda abrupta para 8-9M streams/semana

### 5.3 Tendência Anual e Crescimento

#### 5.3.1 CAGR – Taxa de Crescimento Anual Composta

A análise de 2017 a 2024 (último ano completo) revelou:

$$CAGR = \left[ \left( \frac{\text{Valor}_{2024}}{\text{Valor}_{2017}} \right)^{\frac{1}{7}} - 1 \right] \times 100 = 23,34\% \quad (1)$$

Onde:

- Valor inicial (2017): 359.461.696 streams
- Valor final (2024): 1.560.494.532 streams
- Período: 7 anos

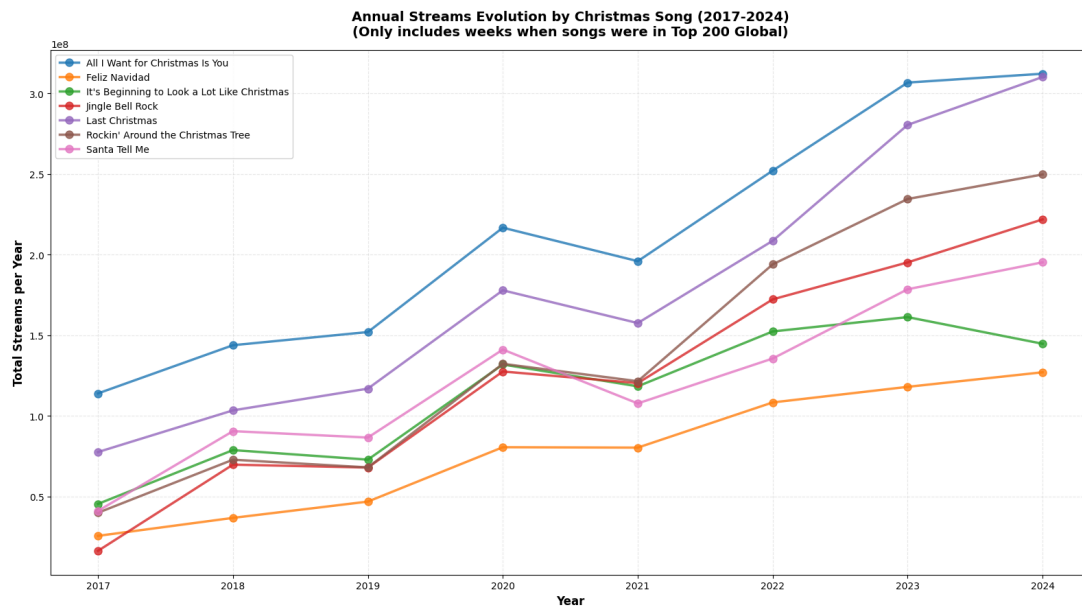


Figura 10: Evolução anual de streams por música (2017-2024)

### 5.3.2 Crescimento Year-over-Year (YoY)

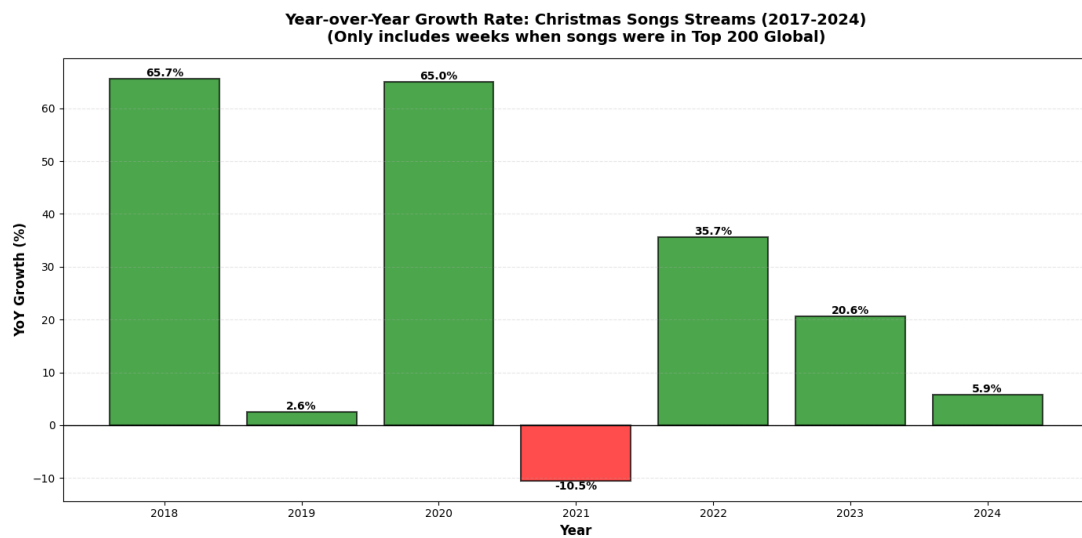


Figura 11: Taxa de crescimento ano a ano (2017-2024)

#### Observações:

- Crescimento consistente de 2017 a 2024
- 2024 estabeleceu novos recordes históricos para todas as músicas
- Variações anuais refletem mudanças nas tendências de consumo de streaming

## 5.4 Análise de Heatmap Temporal

### 5.4.1 Heatmap Ano × Semana

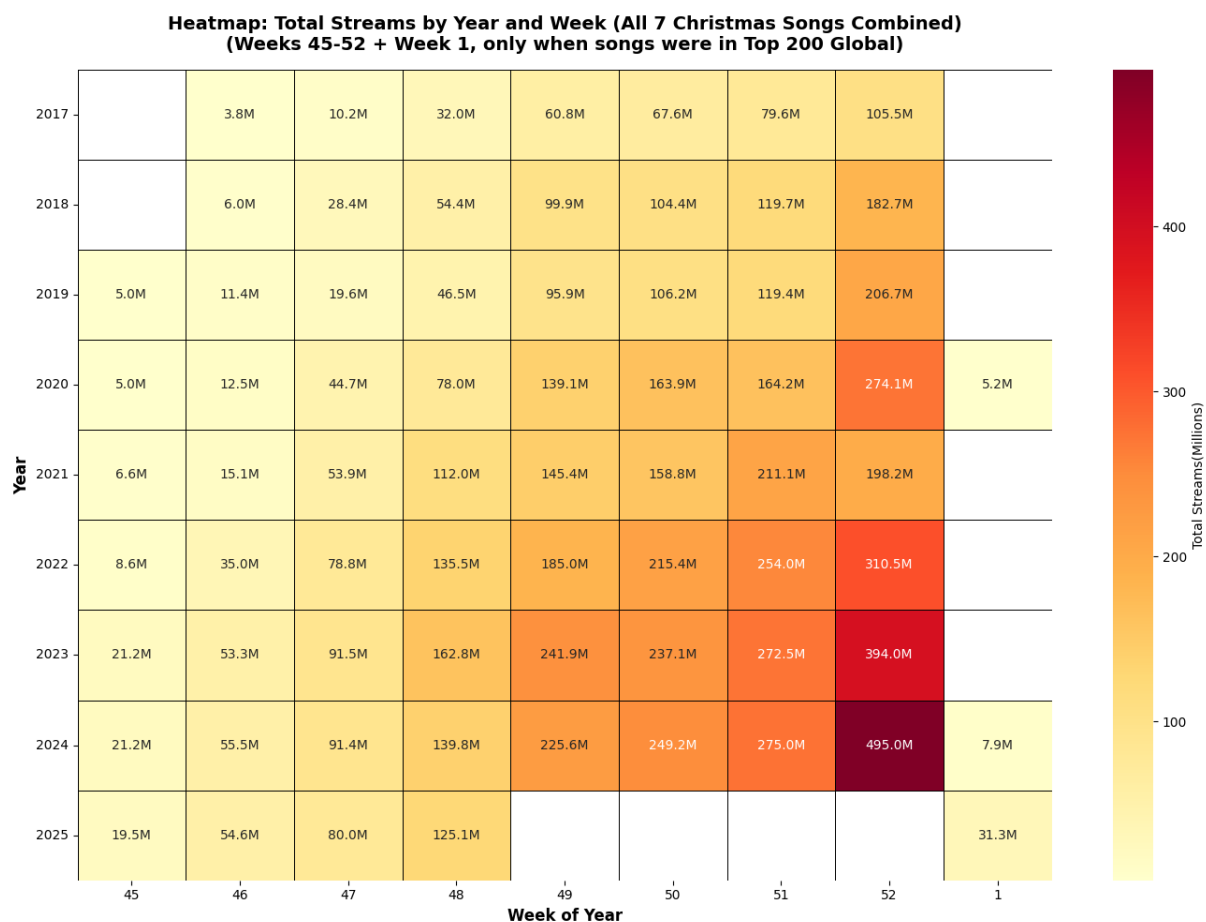


Figura 12: Heatmap de streams totais por ano e semana (todas as 7 músicas combinadas)

#### Principais insights:

- Intensificação clara dos streams ao longo dos anos (cores mais quentes)
- Semana 52 consistentemente a mais forte em todos os anos
- Crescimento mais acentuado a partir de 2020
- 2024 apresenta os valores mais altos em todas as semanas

## 5.4.2 Comparação da Mesma Semana em Anos Diferentes

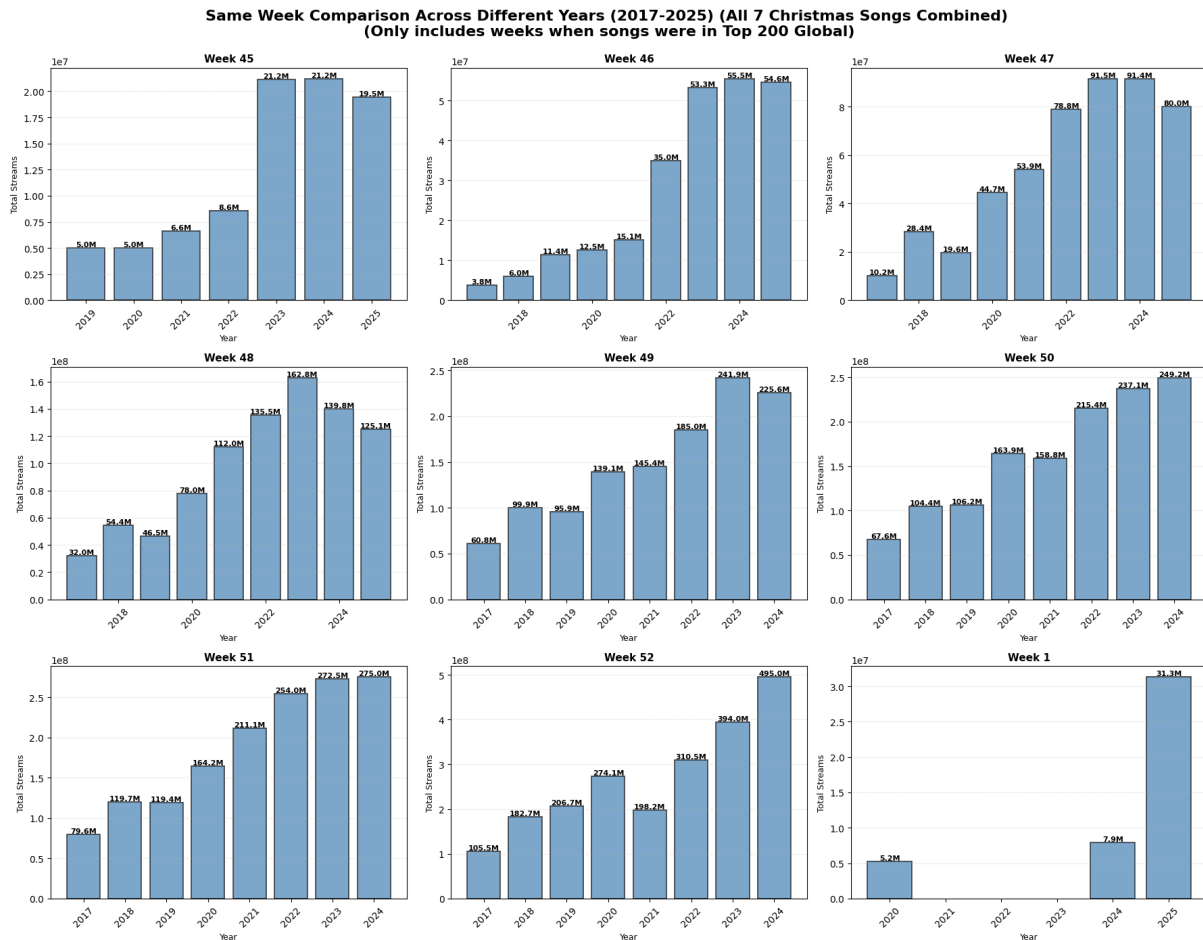


Figura 13: Comparação da mesma semana em diferentes anos (2017-2025)

## 6 Modelos Preditivos

### 6.1 Seleção do Modelo ARIMA

#### 6.1.1 Estrutura dos Dados e Abordagem de Modelagem

A metodologia central transformou os dados em múltiplas séries temporais independentes. Para cada música, foi criada uma série temporal separada para cada semana da temporada natalina (ex: uma série para streams da Semana 45 ao longo dos anos, outra para Semana 46, etc.).

**Características das séries temporais:**

- Cada série temporal é curta, contendo aproximadamente 8 pontos de dados (um para cada ano de 2017-2024)
- Representa o desempenho ano a ano para uma semana específica
- Isola a tendência anual e simplifica o desafio de previsão removendo sazonalidade complexa



### 6.1.2 Comparação de Modelos

Tabela 5: Comparação de modelos para previsão sazonal de streams

Critério	Média Móvel	ARIMA (Esco- lhido)	LSTM	Prophet
Complexidade	Muito Baixa	Baixa – ARIMA(1,1,1) simples	Muito Alta	Média
Requisitos de Dados	Mínimo	<b>Ideal para séries curtas (8 pontos)</b>	Muito Grande (milhares de pontos)	Moderado
Sazonalidade	Nenhuma	Não necessária; tratada pela divisão dos dados	Aprende padrões mas precisa de dados massivos	Overkill
Interpretabilidade	Alta	Alta – Parâmetros claros (p,d,q)	Muito Baixa (black-box)	Média
Tempo de Treinamento	Instantâneo	Muito Rápido (milissegundos)	Lento (minutos a horas)	Rápido
Risco de Overfitting	Nenhum	Baixo	<b>Extremamente Alto</b> (inutilizável com 8 pontos)	Alto
Adequação	Pobre	<b>Excelente</b>	Inadequado	Pobre

### 6.1.3 Justificativa da Escolha do ARIMA

O modelo ARIMA(1,1,1) foi escolhido porque:

- É estatisticamente robusto e altamente eficaz para séries curtas
- Objetivo principal é modelar tendências ano a ano e autocorrelação
- Simplicidade, baixo custo computacional e alta interpretabilidade
- **LSTM** foi descartado devido ao tamanho extremamente pequeno de cada série (8 pontos causariam overfitting severo)
- **Prophet** não é projetado para esta configuração específica de séries anuais curtas

## 6.2 Fundamentação Teórica do ARIMA

O modelo ARIMA(p,d,q) combina três componentes:

- **AR(p) – Autoregressivo:** Usa valores passados como preditores

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (2)$$

- **I(d) – Integração:** Diferencia a série d vezes para alcançar estacionariedade

$$\Delta y_t = y_t - y_{t-1} \quad (3)$$

- **MA(q) – Média Móvel:** Usa erros de previsão passados como preditores

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (4)$$

**Equação Linear Geral: ARIMA(p,d,q)**

$$Y_t = c + \underbrace{\phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p}}_{\text{AR: Valores Passados}} + \underbrace{\theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}}_{\text{MA: Erros Passados}} + \varepsilon_t \quad (5)$$

Onde:

- $Y_t$ : Série após diferenciação d vezes (estacionária)
- $\phi_p$ : Pesos dos termos autoregressivos
- $\theta_q$ : Pesos dos termos de média móvel
- $c$ : Constante (intercepto ou drift)
- $\varepsilon_t$ : Ruído branco atual (erro aleatório)

## 6.3 Treinamento e Validação do Modelo

### 6.3.1 Divisão dos Dados

- **Treinamento:** Dados de 2017 a 2023
- **Validação:** Dados de 2024
- **Previsão:** Temporada 2025

### 6.3.2 Resultados da Validação (2024)

A validação com dados de 2024 revelou acurácia satisfatória:

Tabela 6: MAPE (Mean Absolute Percentage Error) por música – Validação 2024

Música	MAPE (%)
Santa Tell Me	8,8
Feliz Navidad	12,0
All I Want for Christmas Is You	15,5
Jingle Bell Rock	15,7
Last Christmas	18,7
Rockin' Around the Christmas Tree	19,4
It's Beginning to Look a Lot Like Christmas	21,6
<b>Média Geral</b>	<b>15,5</b>

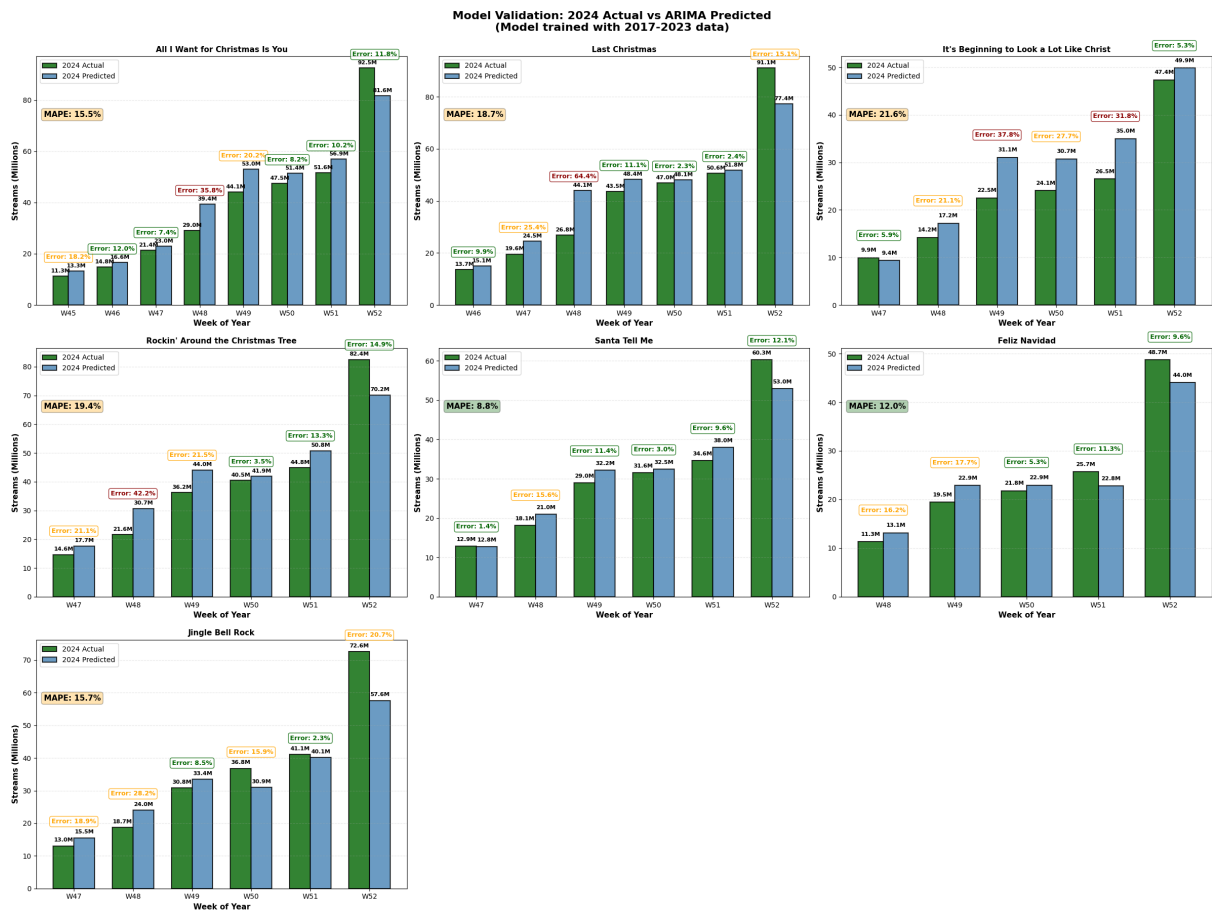


Figura 14: Validação do modelo – 2024 Real vs ARIMA Previsto

### Análise dos erros por semana:

- Semanas iniciais (45-47): Erros geralmente baixos (5-20%)
- Semana 48: Erros mais altos em várias músicas (até 64% em “Last Christmas”)
- Semanas finais (49-52): Recuperação da acurácia (maioria abaixo de 20%)
- “Santa Tell Me” apresentou menor MAPE (8,8%), indicando padrão mais previsível

## 6.4 Previsões para 2025

### 6.4.1 Resultados das Previsões

Tabela 7: Previsão ARIMA para temporada natalina 2025 – Comparação com 2024

Música	2024 Real (M)	2025 Previsto (M)	Diferença (M)	Variação (%)
All I Want for Christmas Is You	312,13	352,55	+40,42	+12,9
Last Christmas	302,18	340,29	+38,10	+12,6
It's Beginning to Look a Lot Like Christmas	144,67	187,91	+43,24	+29,9
Rockin' Around the Christmas Tree	249,65	284,24	+34,59	+13,9
Santa Tell Me	195,18	208,63	+13,46	+6,9
Feliz Navidad	126,98	134,05	+7,07	+5,6
Jingle Bell Rock	221,78	215,73	-6,06	-2,7

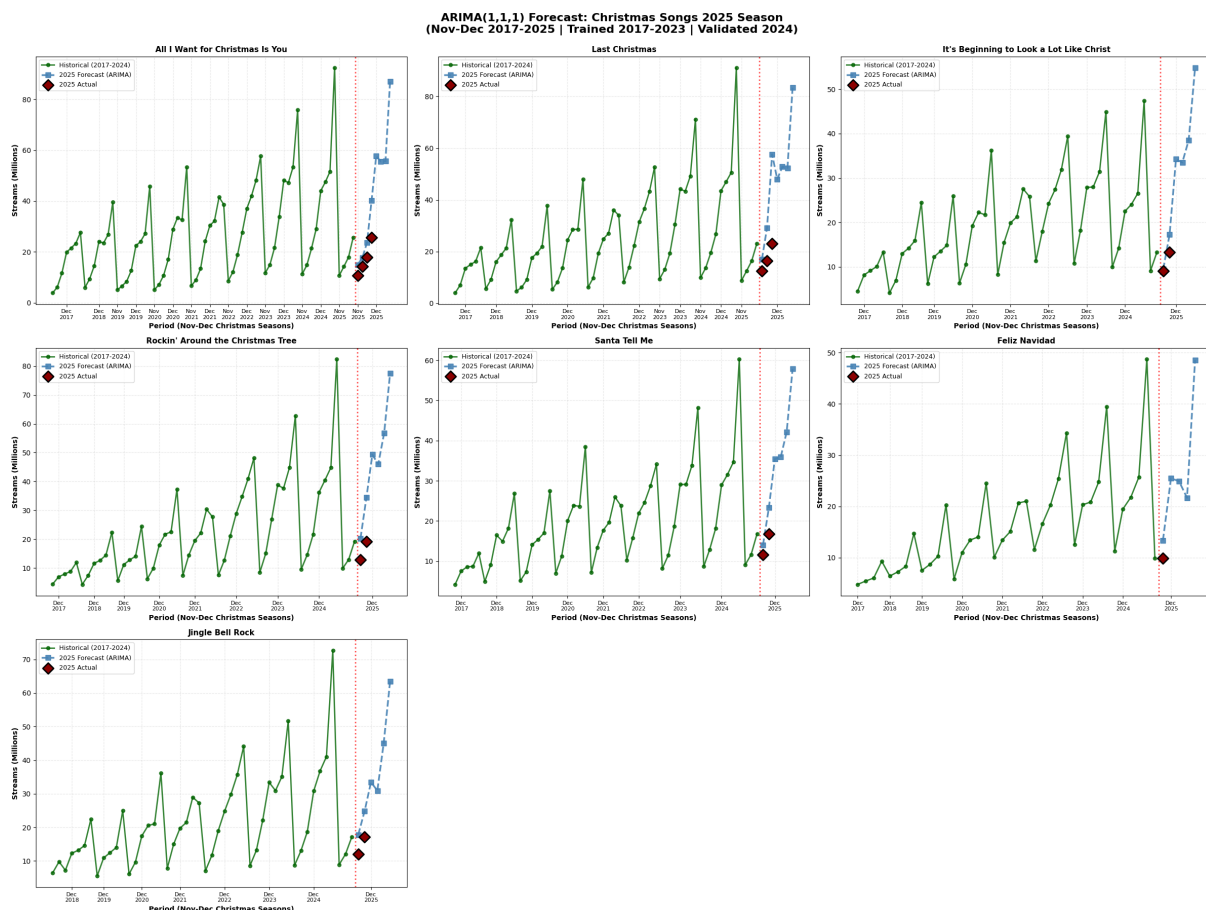


Figura 15: Previsão ARIMA(1,1,1) para temporada natalina 2025

### 6.4.2 Análise das Previsões

Analisando os resultados da previsão para 2025:

- **Crescimento generalizado:** 6 das 7 músicas apresentam previsão de crescimento
- **Maior salto:** “It’s Beginning to Look a Lot Like Christmas” (+29,9%), indicando tendência de popularização acelerada
- **Líderes consolidados:** “All I Want for Christmas Is You” e “Last Christmas” mantêm domínio com previsões acima de 340 milhões de streams cada ( 13% de crescimento)
- **Crescimento modesto:** “Feliz Navidad” e “Santa Tell Me” (5-7%), músicas já estabelecidas
- **Único declínio:** “Jingle Bell Rock” (-2,7%), sugerindo possível saturação

Esta dinâmica revela um mercado onde clássicos permanecem fortes, mas com trajetórias de crescimento diferentes conforme o engajamento do público evolui.

## 7 Visualizações Animadas

Além das análises estáticas, foram criadas visualizações animadas (formato GIF) para melhor compreensão da evolução temporal. As animações estão disponíveis no repositório:

- **Corrida de Streams Acumulados:** Visualiza a competição entre as músicas à medida que acumulam streams totais de 2017 a 2025 (race\_accumulated\_total.gif)
- **Padrão de Crescimento Semanal:** Mostra como os padrões de streaming evoluem semana a semana durante a temporada natalina (growth\_weekly\_season.gif)

Visualização de como os padrões de streaming evoluem semana a semana durante a temporada natalina.

## 8 Reconhecimento

Este dataset foi destaque nos **Trending Datasets do Kaggle**, demonstrando o interesse da comunidade em análises de streaming de música e padrões sazonais.

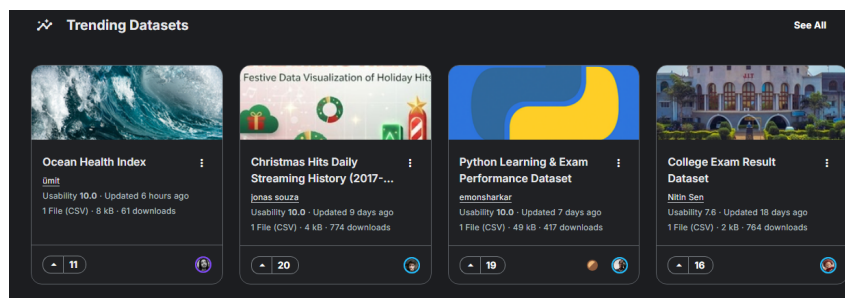


Figura 16: Reconhecimento como Trending Dataset no Kaggle

## 9 Conclusões

### 9.1 Principais Descobertas

1. **Domínio de Mariah Carey:** “All I Want for Christmas Is You” é inquestionavelmente a música natalina mais popular no Spotify, com 1,77 bilhões de streams totais e média semanal de 25,7 milhões.
2. **Crescimento Exponencial:** CAGR de 23,34% (2017-2024) indica que músicas natalinas estão cada vez mais populares no streaming, não apenas mantendo relevância, mas expandindo significativamente.
3. **Padrão Sazonal Altamente Concentrado:** O coeficiente de variação médio de 64% confirma que essas músicas são extremamente sazonais, com 95%+ dos streams concentrados em 9 semanas do ano (semanas 45-52 + semana 1).
4. **Sincronização de Picos:** Todas as 7 músicas atingiram picos históricos na mesma semana (Semana 52 de 2024), demonstrando comportamento coletivo do público.
5. **Modelo ARIMA Eficaz:** Com MAPE médio de 15,5% na validação 2024, o modelo ARIMA(1,1,1) mostrou-se adequado para previsões de curto prazo, especialmente considerando a estrutura limitada dos dados (séries curtas).
6. **Previsão 2025 Positiva:** Espera-se crescimento contínuo em 6 das 7 músicas, com destaque para “It’s Beginning to Look a Lot Like Christmas” (+29,9%) e os líderes consolidados mantendo crescimento de 13%.

## 9.2 Limitações do Estudo

- Dados disponíveis apenas quando músicas estão no Top 200 Global
- Lacunas significativas nos períodos off-season (fevereiro-outubro)
- Granularidade semanal impede análise de padrões diários
- Dados de 2025 parciais (até 28 de novembro)
- Modelo ARIMA com séries temporais muito curtas (8 pontos anuais)
- Não foram considerados fatores externos (campanhas de marketing, TikTok trends, etc.)

## 9.3 Trabalhos Futuros

1. Incorporar dados de outras plataformas de streaming (Apple Music, YouTube Music)
2. Análise de sentimento em redes sociais correlacionada com picos de streaming
3. Estudo de impacto de viralizações no TikTok nos padrões de consumo
4. Análise geográfica (países/regiões) dos padrões de consumo
5. Modelos mais complexos quando houver mais anos de dados disponíveis
6. Investigação de músicas natalinas emergentes vs. clássicos tradicionais

## 10 Referências

- **Fonte de Dados:** Kwordb.net – Spotify Streaming Charts (<https://kwordb.net/spotify/>)
- **Método:** Web scraping com BeautifulSoup e Requests
- **Período Analisado:** 01/01/2017 – 28/11/2025
- **Repositório GitHub:** [https://github.com/your-username/Christmas\\_Songs\\_Spotify\\_Analysis](https://github.com/your-username/Christmas_Songs_Spotify_Analysis)
- **Dataset Kaggle:** Featured in Trending Datasets

## A Estrutura do Repositório

```
Christmas_Songs_Spotify_Analysis/  
  data/  
    raw/  
      spotify_christmas_streams_kwordb_2017_2025.csv  
  notebooks/  
    01_data_collection.ipynb  
    02_exploratory_data_analysis.ipynb  
    03_temporal_analysis.ipynb
```

```

04_prediction_models.ipynb
images/
  animations/
  data_collection/
  exploratory_analysis/
  temporal_analysis/
  prediction_models/
scripts/
  create_animations.py
README.md
LICENSE

```

## B Estatísticas Detalhadas por Música e Semana

Tabela 8: Streams médios (em milhões) por semana para cada música durante a temporada natalina

Música	W45	W46	W47	W48	W49	W50	W51	W52
All I Want for Christmas Is You	8,4	9,8	14,2	21,8	31,8	33,9	38,1	53,9
Last Christmas	9,4	9,1	11,5	17,9	27,0	29,5	33,4	48,6
Rockin' Around the Christmas Tree	–	8,9	10,5	14,5	21,4	23,7	27,6	39,6
Jingle Bell Rock	–	8,3	10,6	14,3	21,4	23,6	24,6	36,1
Santa Tell Me	–	8,6	8,8	12,8	19,5	21,0	23,8	33,9
It's Beginning to Look a Lot Like Christmas	–	–	8,6	11,9	18,4	20,0	22,5	32,2
Feliz Navidad	–	–	–	10,2	12,4	14,1	16,9	26,6

*Nota: Valores ausentes (–) indicam que a música não teve dados suficientes (5+ anos) para aquela semana.*