Electrical Engineering

Christmas Music Streaming on Spotify: A Machine Learning Approach to Exploratory, Temporal, and Predictive Analysis (2017-2025)

(2017-2025)

Exploratory, temporal, and predictive analysis of 7 classic Christmas songs

based on Spotify streaming data

Author: Jonas Souza
Education: Electrical Engineer
Date: December 18, 2025

1 Executive Summary

This report presents a comprehensive analysis of the streaming patterns of 7 classic Christmas songs on Spotify between 2017 and 2025. Using data science techniques and visualization, temporal trends, seasonality, and the evolution of the popularity of these songs during the holiday season were investigated.

1.1 Key Findings

- **Absolute leader:** "All I Want for Christmas Is You" (Mariah Carey) with 1.77 billion total streams
- **Annual growth:** 23.34% CAGR (2017-2024)
- **Historical peak:** All 7 songs reached their maximum peaks in Week 52 of 2024 (December 26)
- **Seasonality:** Average coefficient of variation of 64%, indicating high concentration during the holiday period
- **2025 forecast:** Projected growth of 12-13% for the most popular songs

2 Introduction

2.1 Context

Music streaming has fundamentally transformed the music industry over the past two decades. In the Christmas context, classic songs exhibit unique and highly seasonal behavior, with peaks concentrated in specific weeks of the year.

2.2 Objectives

1. Analyze descriptive streaming patterns of the top 7 Christmas songs
2. Investigate temporal trends and seasonality between 2017-2025
3. Develop predictive models for the 2025 holiday season
4. Identify unique characteristics of each song in the streaming landscape

2.3 Songs Analyzed

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 Methodology

3.1 Data Collection

Data was collected via web scraping from Kworb.net, a platform specializing in tracking Spotify streams.

3.1.1 Data Characteristics

- **Period:** 01/01/2017 – 11/28/2025
- **Granularity:** Weekly (updates on Thursdays)
- **Availability:** Only when songs are in the Global Top 200
- **Variables:** Date, song, artist, weekly streams, day of the week, week of the year

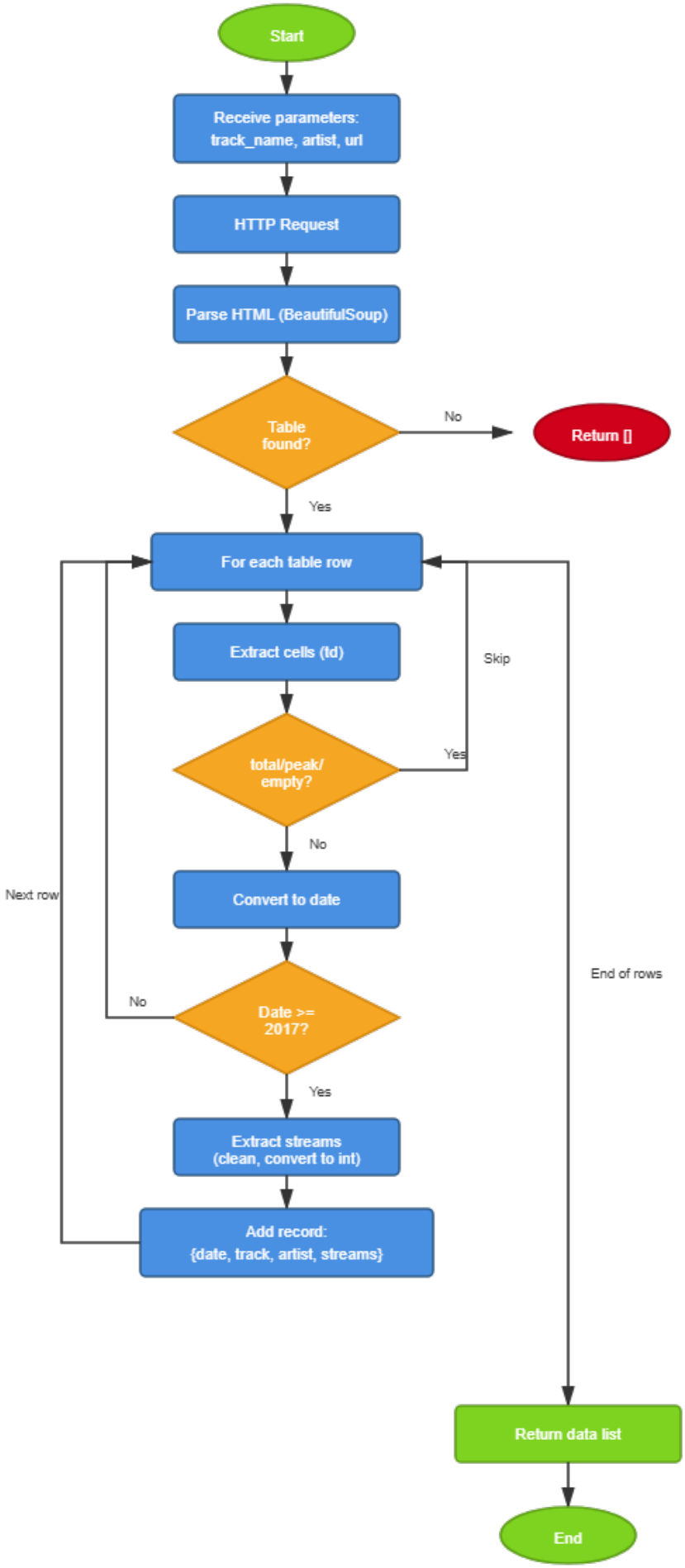

Figure 1: Data collection process flowchart

3.1.2 Data Limitations

- Data available only when songs are in the Spotify Global Top 200
- Gaps between February and October (off-season period)
- Weekly granularity (not daily)
- 2025 data partial (up to November 28)

3.2 Tools and Technologies

- **Python 3.x:** Main language
- **Pandas:** Data manipulation and analysis
- **NumPy:** Numerical computation
- **Matplotlib/Seaborn:** Data visualization
- **BeautifulSoup:** Web scraping
- **Statsmodels:** ARIMA modeling
- **Jupyter Notebook:** Development environment

4 Exploratory Data Analysis

4.1 Descriptive Statistics

Descriptive analysis was performed to understand the individual behavior of each of the 7 Christmas songs. Each record represents weekly streaming data, typically recorded on Thursdays.

4.1.1 Total Streams by Song (2017-2025)

The popularity ranking revealed:

Table 1: Total streams by song (2017-2025)

Song	Total Streams
All I Want for Christmas Is You	1.77 billion
Last Christmas	1.50 billion
Rockin' Around the Christmas Tree	1.16 billion
Jingle Bell Rock	1.03 billion
Santa Tell Me	1.01 billion
It's Beginning to Look a Lot Like Christmas	927 million
Feliz Navidad	633 million

4.1.2 Weekly Statistics

Descriptive statistics of weekly streams revealed distinct patterns:

Table 2: Descriptive statistics of weekly streams by song

Song	Count	Mean (M)	SD (M)	Min (M)	Max (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

Key observations:

- "Rockin' Around the Christmas Tree" leads the coefficient of variation (73.36%), showing great relative fluctuation
- "It's Beginning to Look a Lot Like Christmas" presents less variation, with more consistent streams
- The difference between mean and median in all songs confirms positive skewness in distributions

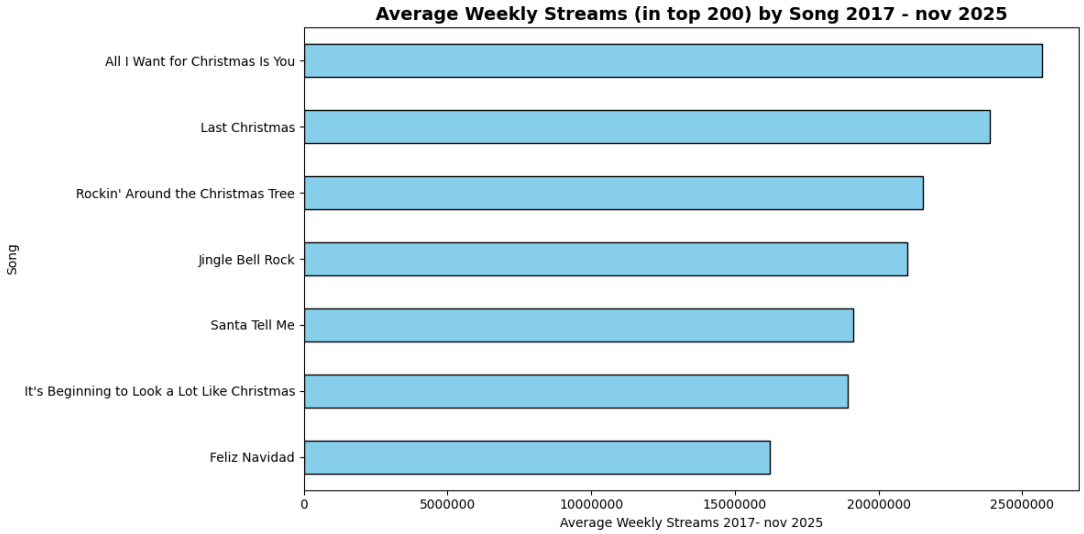

Figure 2: Average weekly streams by song (2017-2025)

4.1.3 Coefficient of Variation

The coefficient of variation (CV%) allows comparing the stability of songs, revealing which ones exhibit more consistent or volatile behavior:

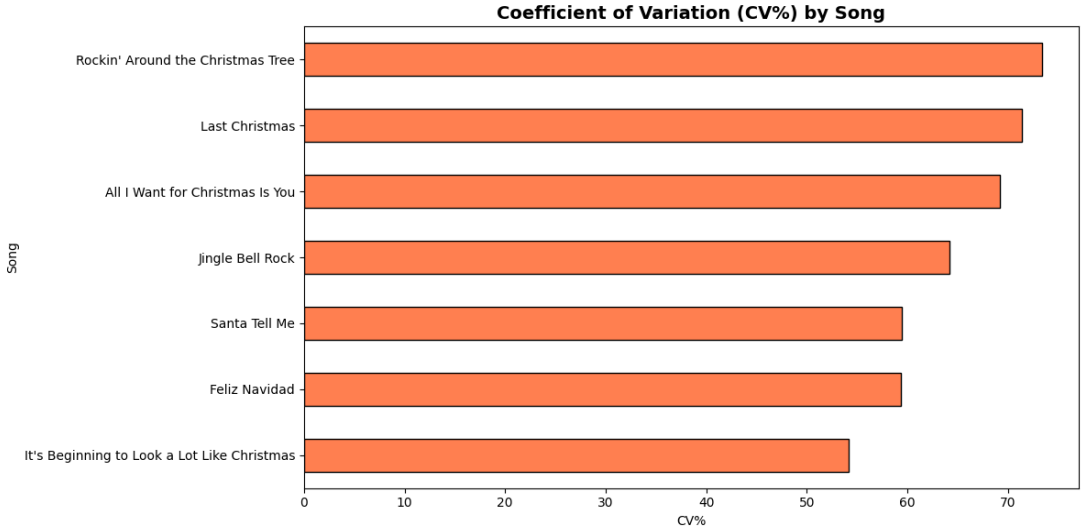

Figure 3: Coefficient of variation by song

4.2 Distribution and Frequency Analysis

4.2.1 Distribution of Weekly Streams

Histogram analysis revealed that **none of the 7 songs have a normal distribution**, showing clearly skewed patterns with concentration of values below the mean and long right tails.

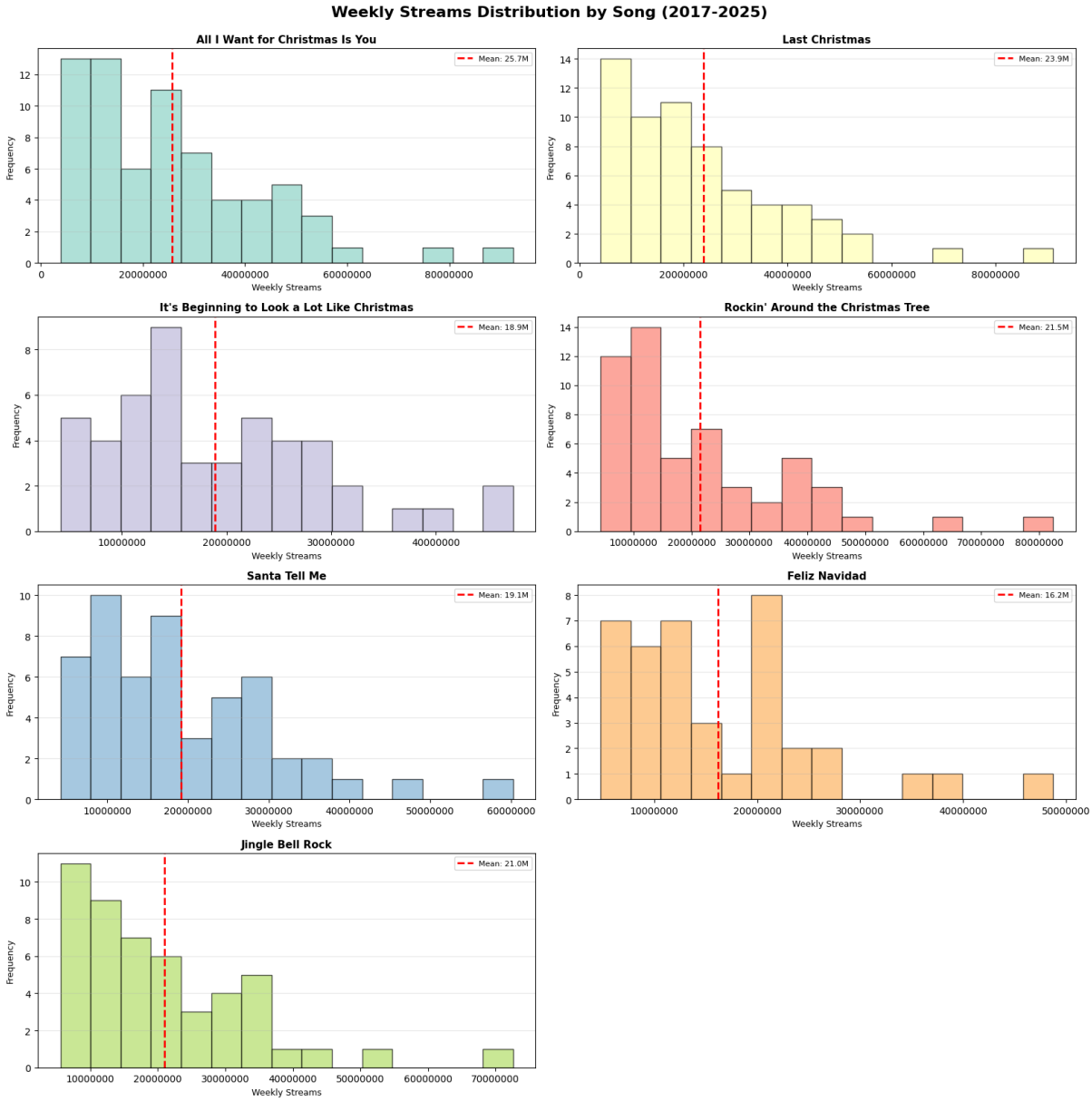

Figure 4: Distribution of weekly streams by song (histograms)

Main findings:

- "All I Want for Christmas Is You" exhibits the greatest dispersion and highest mean (25.7M weekly streams)
- "Feliz Navidad" has the lowest mean (16.2M) and more concentrated distribution at low values
- Positive skewness in all songs indicates predominance of weeks with moderate streams and few very high-volume events (above 60-80M)
- Non-Gaussian distributions are consistent with data selection bias (only Top 200)

4.2.2 Boxplot Analysis

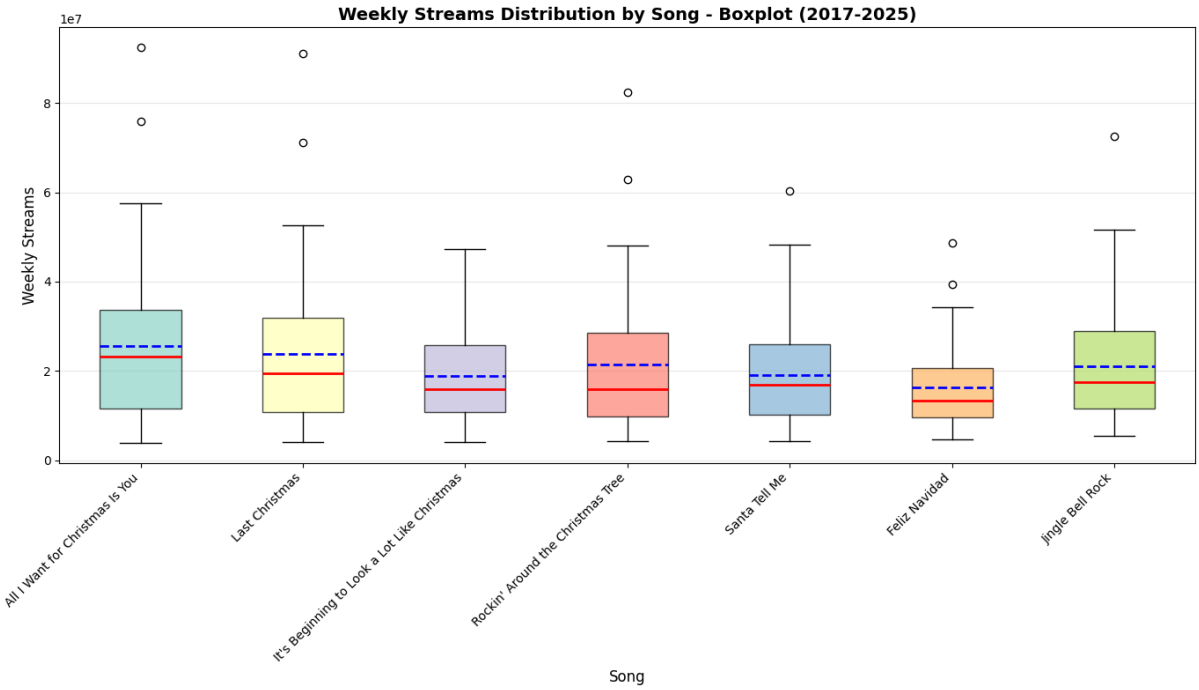

Figure 5: Distribution of weekly streams – Comparative boxplot

The comparative boxplot revealed:

- Clear popularity hierarchy, with "All I Want for Christmas Is You" and "Last Christmas" leading
- Significant upper outliers (ranging from 6M to 10M) in all songs
- "Feliz Navidad" shows lower median and more compact interquartile range
- Blue dashed lines (means) consistently above red lines (medians), confirming positive skewness

4.2.3 Total Streams Distribution

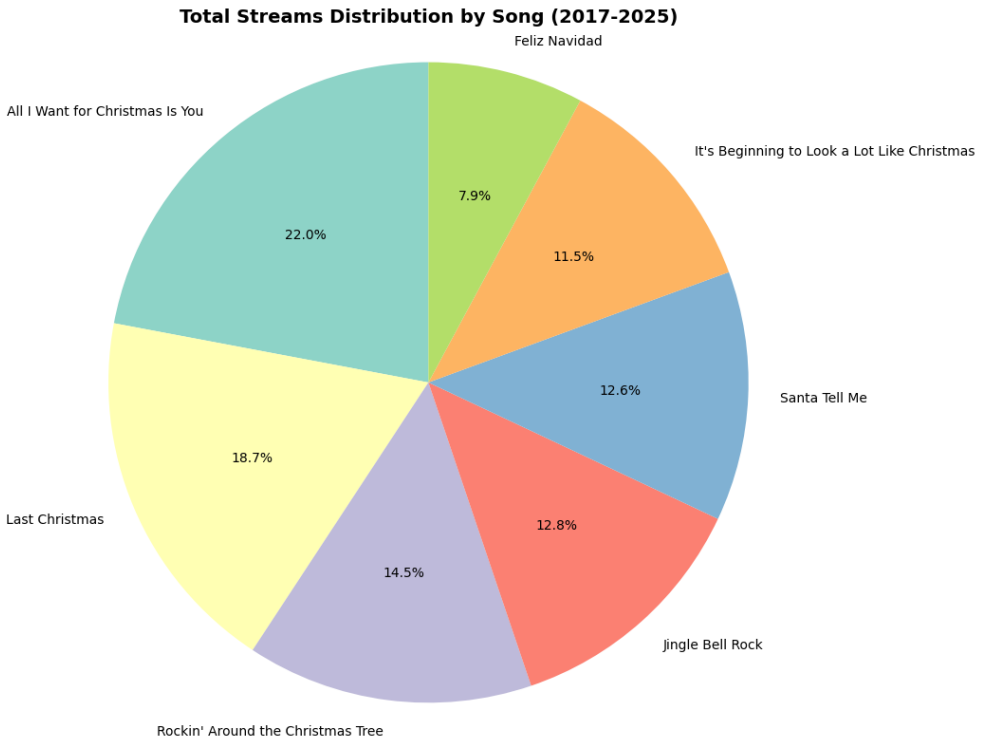

Figure 6: Percentage distribution of total streams (2017-2025)

4.3 Peak Identification

Analysis of maximum weekly stream peaks revealed a synchronized pattern:

Table 3: Maximum weekly stream peaks by song

Song	Maximum Streams (M)	Date/Week
All I Want for Christmas Is You	92.51	12/26/2024 (Week 52)
Last Christmas	91.09	12/26/2024 (Week 52)
Rockin' Around the Christmas Tree	82.41	12/26/2024 (Week 52)
Jingle Bell Rock	72.63	12/26/2024 (Week 52)
Santa Tell Me	60.26	12/26/2024 (Week 52)
Feliz Navidad	48.74	12/26/2024 (Week 52)
It's Beginning to Look a Lot Like Christmas	47.36	12/26/2024 (Week 52)

Conclusion: All 7 songs reached historical peaks in the same week (Week 52 of 2024), indicating a synchronized surge during the post-Christmas period, suggesting unprecedented growth in 2024 or that it represents the most complete/recent collection period in the dataset.

5 Temporal Analysis

5.1 Complete Time Series (2017-2025)

Complete time series analysis revealed clear and consistent seasonal patterns over the years.

5.1.1 Individual Series

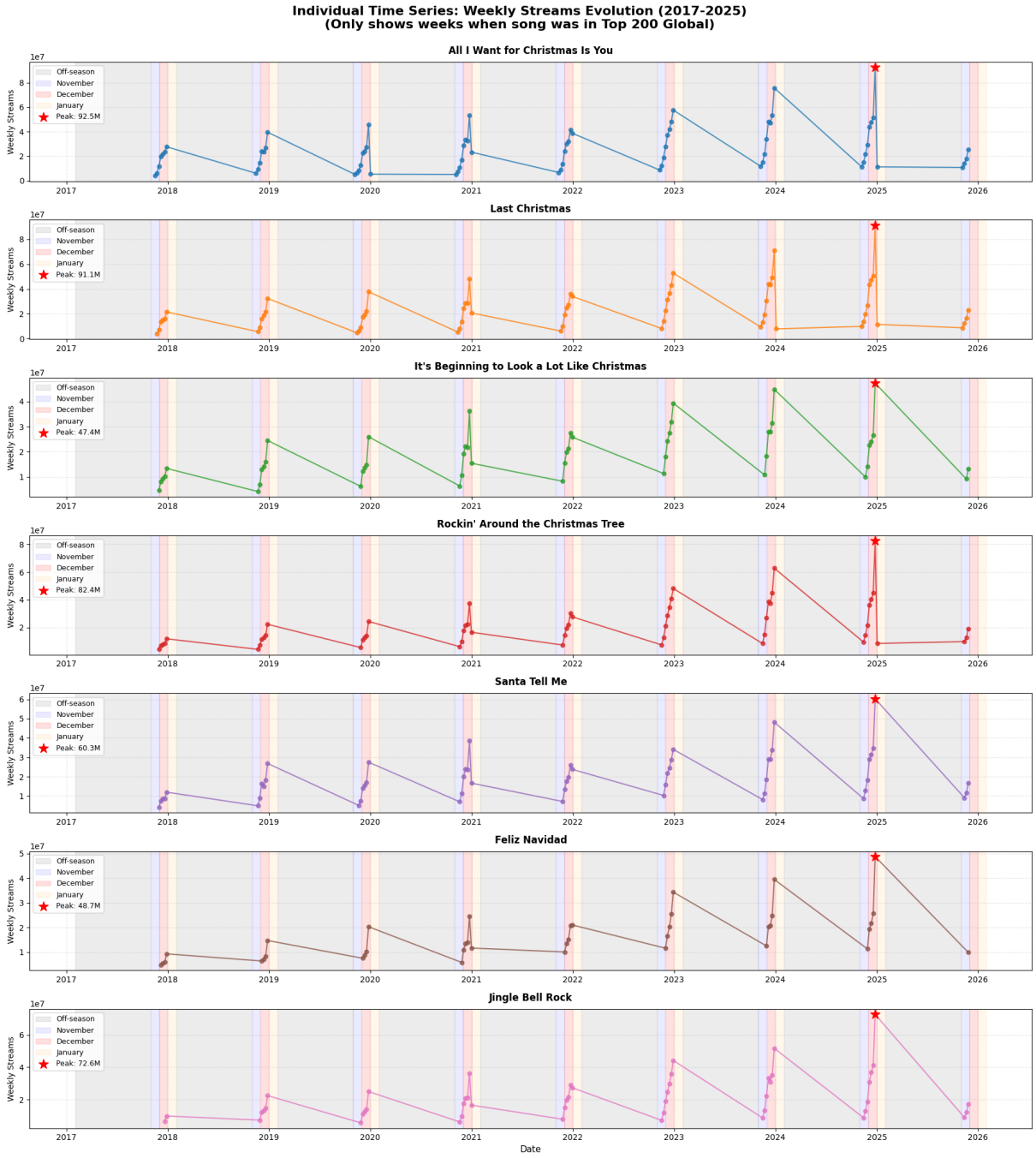

Figure 7: Individual time series – Weekly streams evolution (2017-2025)

Identified patterns:

- Off-season periods (February-October) marked in gray – absence of data (outside Top 200)
- November (blue), December (red), and January (orange) highlighted
- Maximum peaks marked with red stars
- Exponential growth visible during weeks 48-52
- Sharp drop after week 1 (January)

5.1.2 Comparison Between Songs

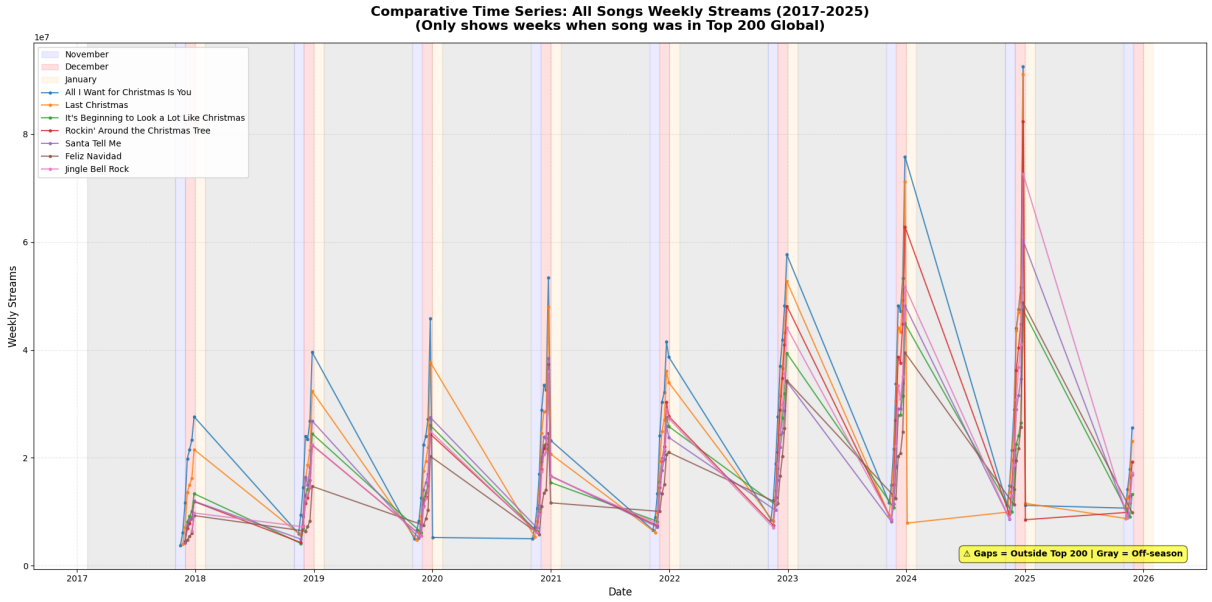

Figure 8: Comparative time series of all songs (2017-2025)

5.2 Seasonal Pattern of the Christmas Cycle

Analysis focused on weeks 45-52 + week 1 revealed the characteristic seasonal pattern:

Table 4: Average streams per week during the Christmas season (all songs combined)

Week	Average Streams (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

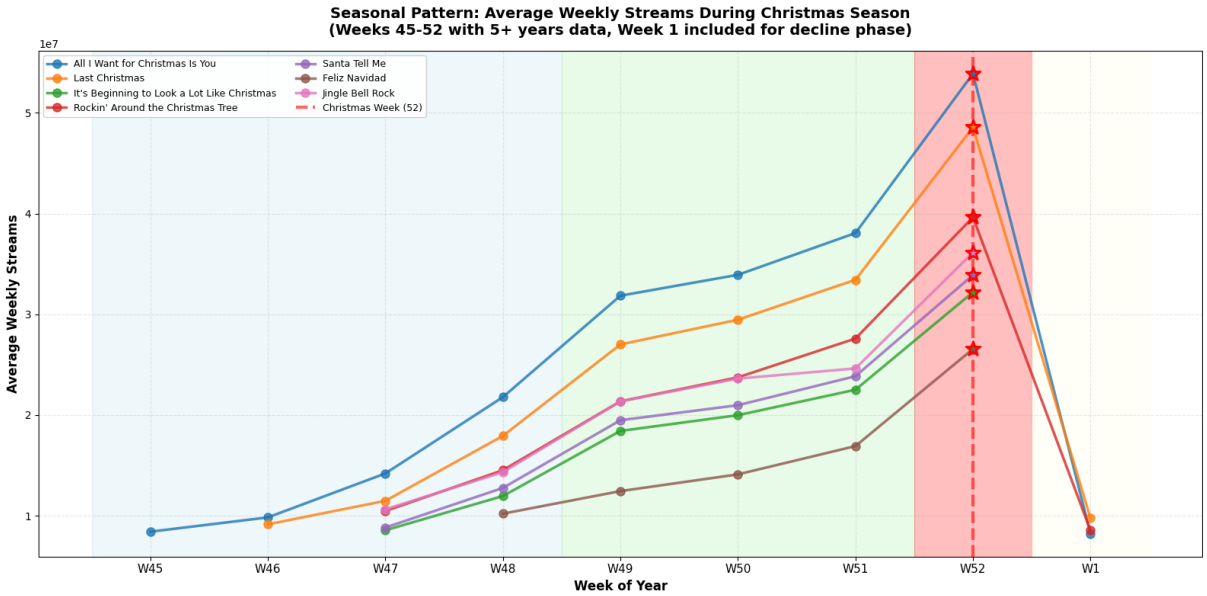

Figure 9: Average seasonal pattern during the Christmas season

Identified phases:

- **Start (Weeks 45-47):** Gradual growth (8-11M streams/week)
- **Acceleration (Weeks 48-50):** Exponential growth (15-24M streams/week)
- **Peak (Weeks 51-52):** Maximum peak (27-39M streams/week)
- **Decline (Week 1):** Sharp drop to 8-9M streams/week

5.3 Annual Trend and Growth

5.3.1 CAGR – Compound Annual Growth Rate

Analysis from 2017 to 2024 (last complete year) revealed:

$$CAGR = \left[\left(\frac{\text{Value}_{2024}}{\text{Value}_{2017}} \right)^{\frac{1}{7}} - 1 \right] \times 100 = 23.34\% \tag{1}$$

Where:

- Initial value (2017): 359,461,696 streams
- Final value (2024): 1,560,494,532 streams
- Period: 7 years

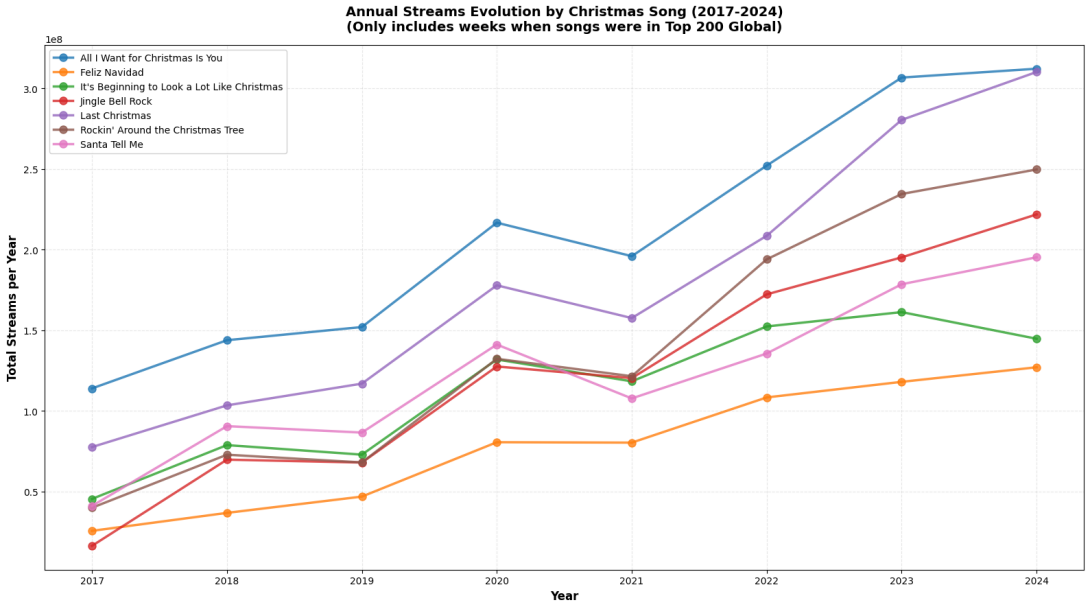

Figure 10: Annual evolution of streams by song (2017-2024)

5.3.2 Year-over-Year (YoY) Growth

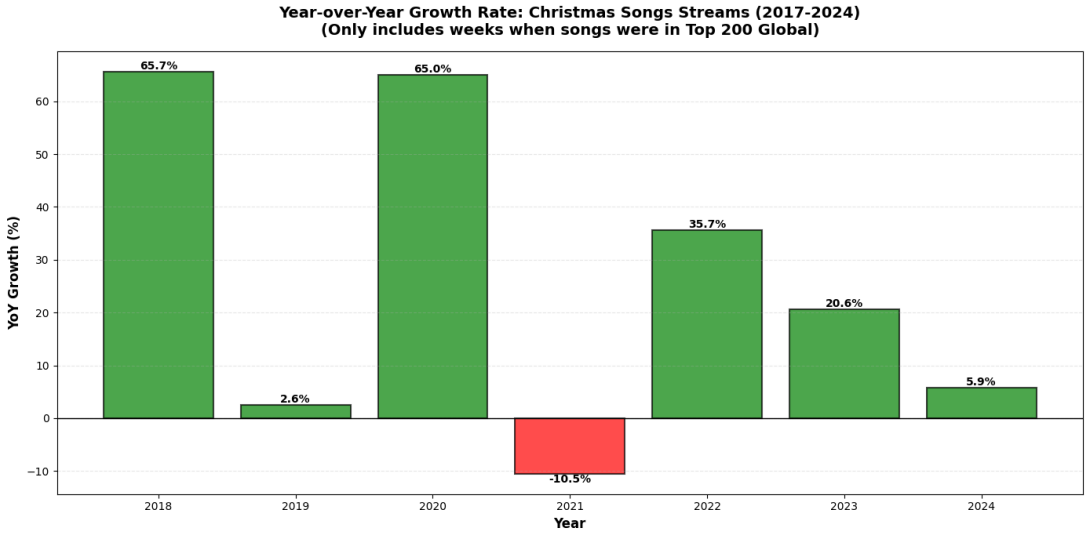

Figure 11: Year-over-year growth rate of Christmas songs streams (2017-2024)

Observations:

- Consistent growth from 2017 to 2024
- 2024 set new historical records for all songs
- Annual variations reflect changes in streaming consumption trends

5.4 Temporal Heatmap Analysis

5.4.1 Year × Week Heatmap

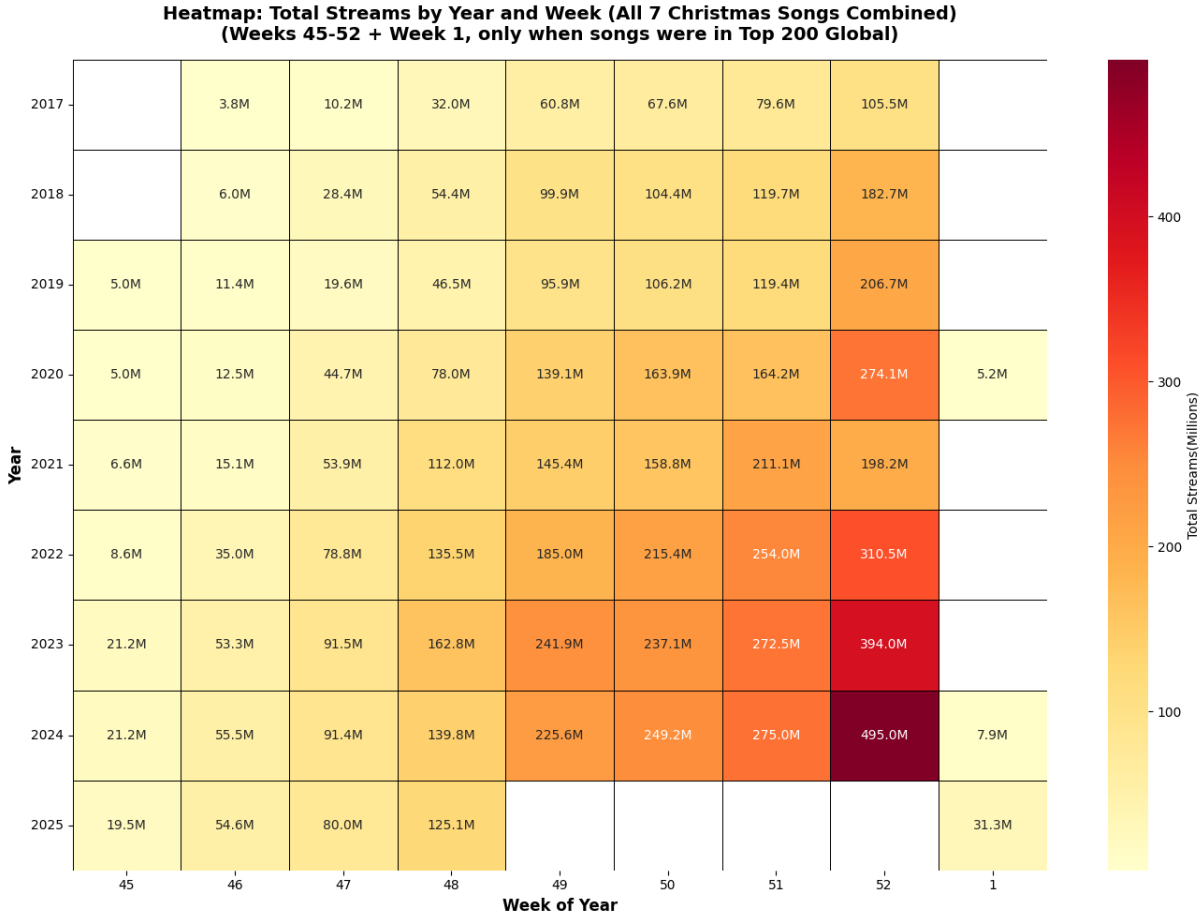

Figure 12: Heatmap of total streams by year and week (all 7 songs combined)

Main insights:

- Clear intensification of streams over the years (warmer colors)
- Week 52 consistently the strongest in all years
- More pronounced growth from 2020 onwards
- 2024 presents the highest values in all weeks

5.4.2 Comparison of the Same Week in Different Years

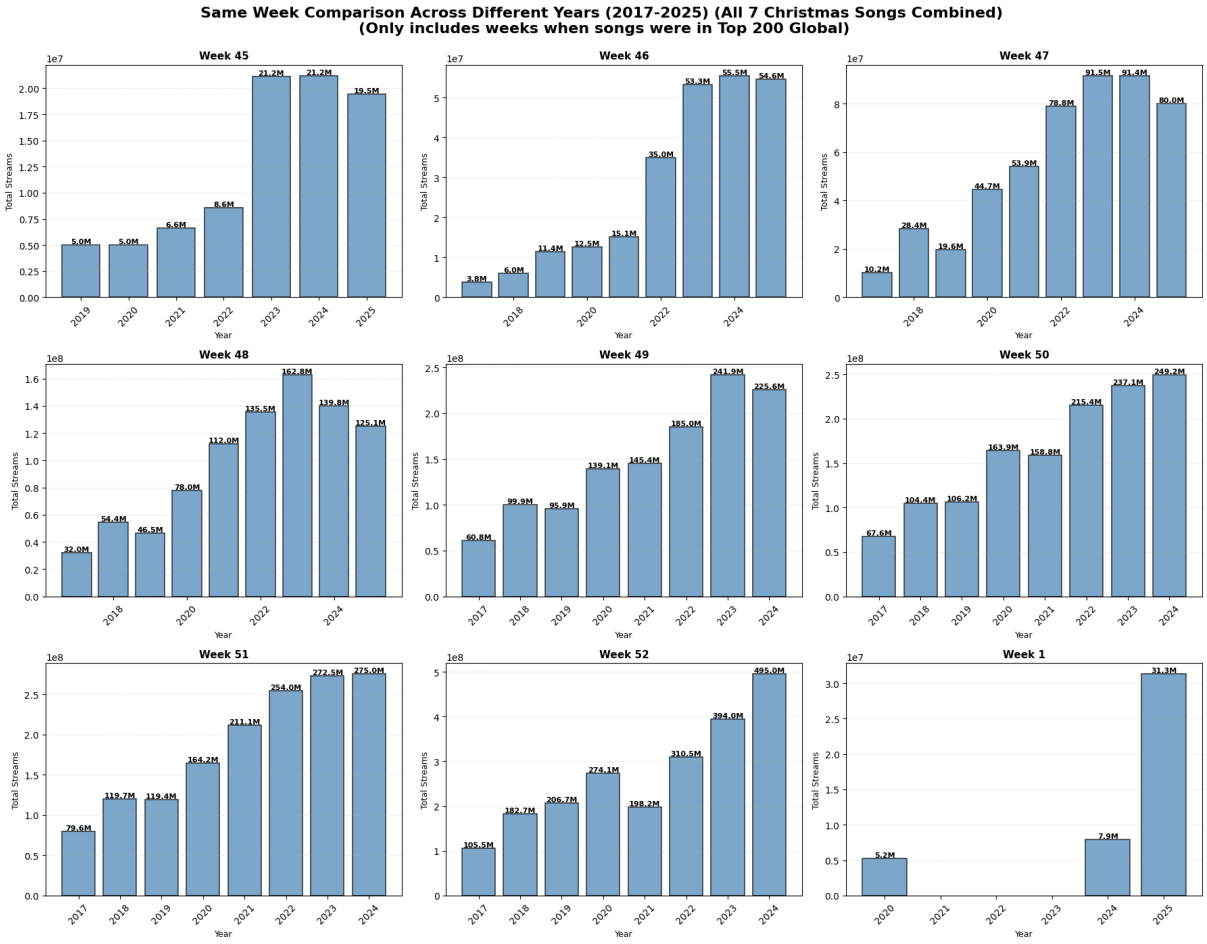

Figure 13: Comparison of the same week in different years (2017-2025)

6 Predictive Models

6.1 ARIMA Model Selection

6.1.1 Data Structure and Modeling Approach

The central methodology transformed the data into multiple independent time series. For each song, a separate time series was created for each week of the Christmas season (e.g., one series for streams in Week 45 over the years, another for Week 46, etc.).

Time series characteristics:

- Each time series is short, containing approximately 8 data points (one for each year from 2017-2024)
- Represents year-over-year performance for a specific week
- Isolates the annual trend and simplifies the forecasting challenge by removing complex seasonality

6.1.2 Model Comparison

Table 5: Comparison of models for seasonal stream forecasting

Criteria	Moving Average	ARIMA (Chosen)	LSTM	Prophet
Complexity	Very Low	Low – Simple ARIMA(1,1,1)	Very High	Medium
Data Requirements	Minimum	**Ideal for short series (8 points)**	Very Large (thousands of points)	Moderate
Seasonality	None	Not needed; handled by data splitting	Learns patterns but requires massive data	Overkill
Interpretability	High	High – Clear parameters (p,d,q)	Very Low (black-box)	Medium
Training Time	Instantaneous	Very Fast (milliseconds)	Slow (minutes to hours)	Fast
Risk of Overfitting	None	Low	**Extremely High** (unusable with 8 points)	High
Suitability	Poor	**Excellent**	Inadequate	Poor

6.1.3 Justification for ARIMA Choice

The ARIMA(1,1,1) model was chosen because:

- It is statistically robust and highly effective for short series
- The main objective is to model year-over-year trends and autocorrelation
- Simplicity, low computational cost, and high interpretability
- **LSTM** was discarded due to the extremely small size of each series (8 points would cause severe overfitting)
- **Prophet** is not designed for this specific configuration of short annual series

6.2 Theoretical Foundation of ARIMA

The **ARIMA(p,d,q)** model combines three components:

- **AR(p) – Autoregressive:** Uses past values as predictors

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{2}$$

- **I(d) – Integration:** Differences the series d times to achieve stationarity

$$\Delta y_t = y_t - y_{t-1} \tag{3}$$

- **MA(q) – Moving Average:** Uses past forecast errors as predictors

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{4}$$

General Linear Equation: ARIMA(p,d,q)

$$Y_t = c + \underbrace{\phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p}}_{\text{AR: Past Values}} + \underbrace{\theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}}_{\text{MA: Past Errors}} + \varepsilon_t \tag{5}$$

Where:

- Y_t: Series after d times differencing (stationary)
- ϕ_p: Weights of autoregressive terms
- θ_q: Weights of moving average terms
- c: Constant (intercept or drift)
- ε_t: Current white noise (random error)

6.3 Model Training and Validation

6.3.1 Data Splitting

- **Training:** Data from 2017 to 2023
- **Validation:** Data from 2024
- **Forecast:** 2025 season

6.3.2 Validation Results (2024)

Validation with 2024 data revealed satisfactory accuracy:

Table 6: MAPE (Mean Absolute Percentage Error) by song – 2024 Validation

Song	**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
Overall Average	**15.5**

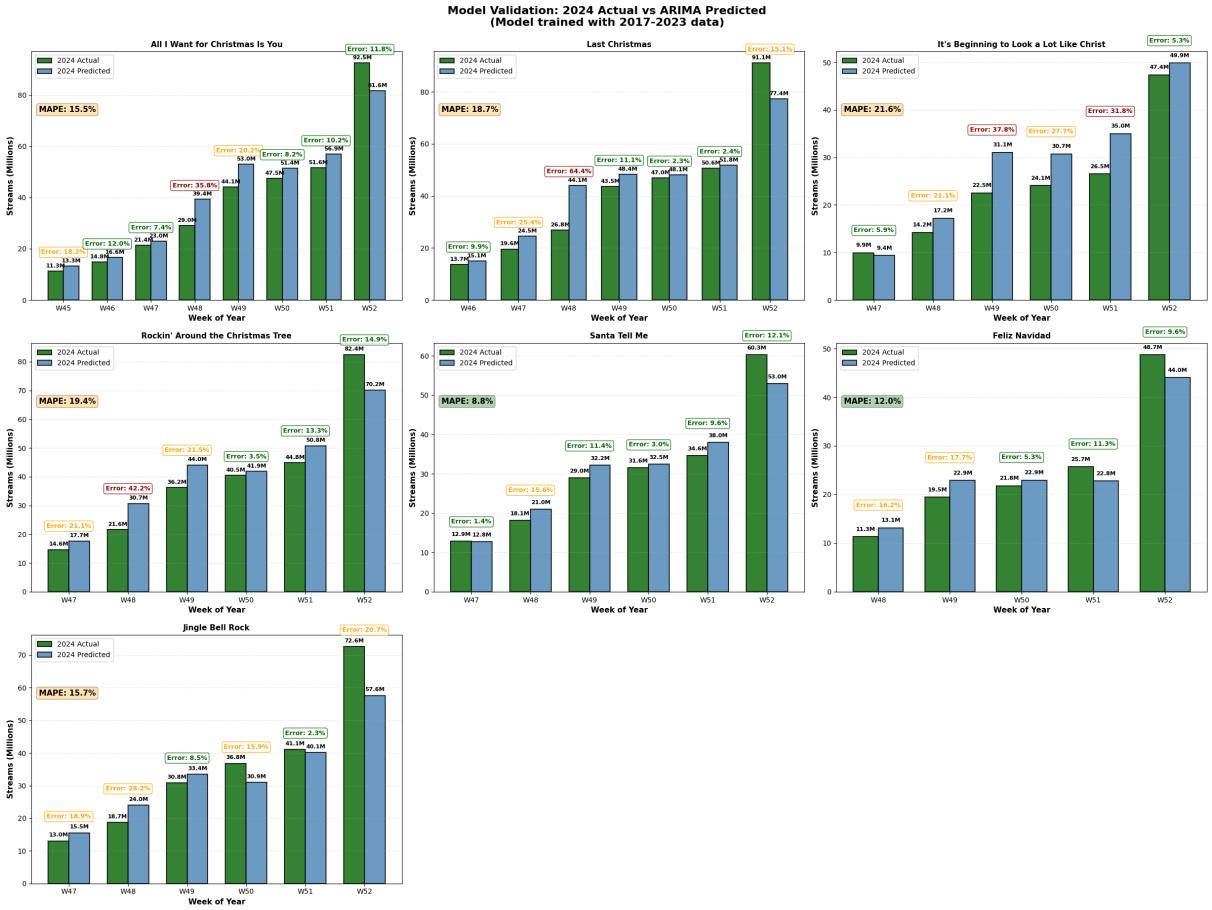

Figure 14: Model validation – 2024 Actual vs ARIMA Predicted

Error analysis by week:

- Early weeks (45-47): Generally low errors (5-20%)
- Week 48: Higher errors in several songs (up to 64% in "Last Christmas")
- Final weeks (49-52): Accuracy recovery (most below 20%)
- "Santa Tell Me" had the lowest MAPE (8.8%), indicating a more predictable pattern

6.4 2025 Forecasts

6.4.1 Forecast Results

Table 7: ARIMA forecast for the 2025 Christmas season – Comparison with 2024

Song	2024 Actual (M)	2025 Forecast (M)	Difference (M)	Variation (%)
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

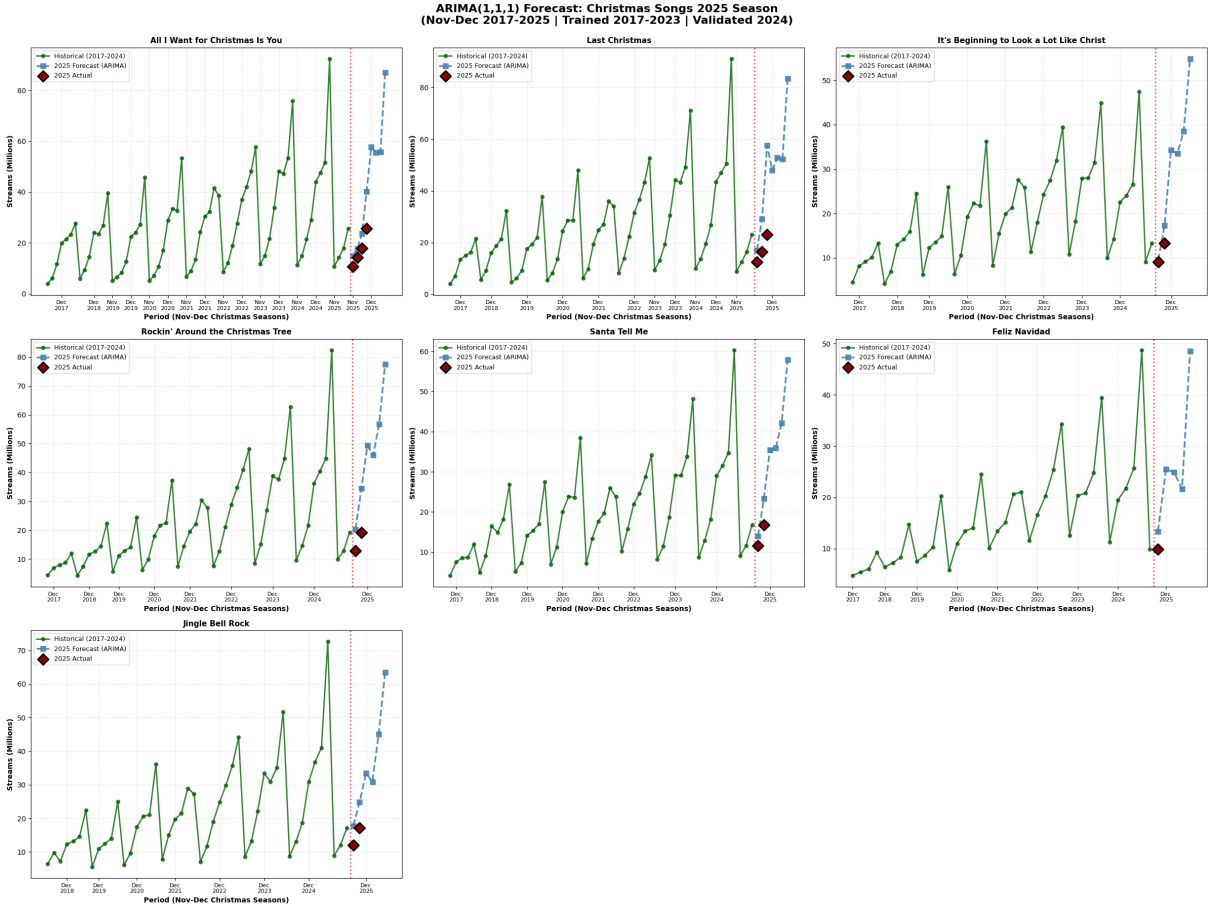

Figure 15: ARIMA(1,1,1) forecast for the 2025 Christmas season

6.4.2 Forecast Analysis

Analyzing the 2025 forecast results:

- **General growth:** 6 out of 7 songs show growth forecast
- **Biggest jump:** "It's Beginning to Look a Lot Like Christmas" (+29.9%), indicating accelerated popularization trend
- **Established leaders:** "All I Want for Christmas Is You" and "Last Christmas" maintain dominance with forecasts above 340 million streams each (13% growth)
- **Modest growth:** "Feliz Navidad" and "Santa Tell Me" (5-7%), already established songs
- **Only decline:** "Jingle Bell Rock" (-2.7%), suggesting possible saturation

This dynamic reveals a market where classics remain strong, but with different growth trajectories as audience engagement evolves.

7 Animated Visualizations

In addition to static analyses, animated visualizations (GIF format) were created for better understanding of temporal evolution. The animations are available in the repository:

- **Accumulated Streams Race:** Visualizes the competition between songs as they accumulate total streams from 2017 to 2025 (race_accumulated_total.gif)
- **Weekly Growth Pattern:** Shows how streaming patterns evolve week by week during the Christmas season (growth_weekly_season.gif)

Visualization of how streaming patterns evolve week by week during the Christmas season.

8 Acknowledgment

This dataset was featured in **Kaggle's Trending Datasets**, demonstrating the community's interest in music streaming analysis and seasonal patterns.

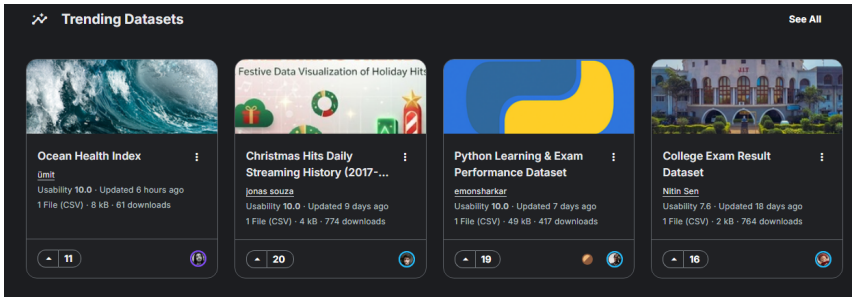

Figure 16: Recognition as Trending Dataset on Kaggle

9 Conclusions

9.1 Main Findings

1. **Mariah Carey's Dominance:** "All I Want for Christmas Is You" is unquestionably the most popular Christmas song on Spotify, with 1.77 billion total streams and an average of 25.7 million weekly streams.
2. **Exponential Growth:** 23.34% CAGR (2017-2024) indicates that Christmas songs are becoming increasingly popular on streaming platforms, not only maintaining relevance but expanding significantly.
3. **Highly Concentrated Seasonal Pattern:** The average coefficient of variation of 64% confirms that these songs are extremely seasonal, with 95%+ of streams concentrated in 9 weeks of the year (weeks 45-52 + week 1).
4. **Synchronized Peaks:** All 7 songs reached historical peaks in the same week (Week 52 of 2024), demonstrating collective audience behavior.
5. **Effective ARIMA Model:** With an average MAPE of 15.5% in 2024 validation, the ARIMA(1,1,1) model proved suitable for short-term forecasts, especially considering the limited data structure (short series).
6. **Positive 2025 Forecast:** Continued growth is expected in 6 out of 7 songs, with "It's Beginning to Look a Lot Like Christmas" (+29.9%) and established leaders maintaining 13% growth.

9.2 Study Limitations

- Data available only when songs are in the Global Top 200
- Significant gaps in off-season periods (February-October)
- Weekly granularity prevents analysis of daily patterns
- 2025 data partial (up to November 28)
- ARIMA model with very short time series (8 annual points)
- External factors not considered (marketing campaigns, TikTok trends, etc.)

9.3 Future Work

1. Incorporate data from other streaming platforms (Apple Music, YouTube Music)
2. Sentiment analysis on social media correlated with streaming peaks
3. Study the impact of TikTok viralization on consumption patterns
4. Geographic analysis (countries/regions) of consumption patterns
5. More complex models when more years of data are available
6. Investigation of emerging Christmas songs vs. traditional classics

10 References

- **Data Source:** Kworb.net – Spotify Streaming Charts (https://kworb.net/spotify/)
- **Method:** Web scraping with BeautifulSoup and Requests
- **Analyzed Period:** 01/01/2017 – 11/28/2025
- **GitHub Repository:** https://github.com/your-username/Christmas_Songs_Spotify_Analysis
- **Kaggle Dataset:** Featured in Trending Datasets

A Repository Structure

```
Christmas_Songs_Spotify_Analysis/
  data/
    raw/
      spotify_christmas_streams_kworb_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 Detailed Statistics by Song and Week

Table 8: Average streams (in millions) per week for each song during the Christmas season

Song	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

Note: Missing values (–) indicate that the song did not have sufficient data (5+ years) for that week.