2017 Spotify Trend Analysis Using SAP

Authors: Melody Gulliver, Grecia Hernandez Lugo, Jair Rocha, Diana Roman
California State University Los Angeles BUS5100-93: Introduction To Business Analytics
{mgulliv; ghernde; jrocha; droman23;}@calstatela.edu

Abstract: Our paper will analyze shifts in Spotify's daily ranking of the top 200 most popular songs within 53 countries. The data set spans from January 1st, 2017 through January 9th, 2018; and was obtained through Kaggle. It consists of information on 6,629 artists in 18,598 songs, totaling one hundred five-billion listening streams. The data itself comprises over two million rows and seven columns that include: chart position, track name, artist, number of streams, URL, date, and region. As a result, we will analyze potential correlations between the streaming popularity and geospatial variables, such as region and date. These variables mold popular culture, which in turn, informs the ways businesses interact with their consumers. Additionally, trending media influences cultural paradigms, destabilizes barriers, and drives global economies. Understanding these consumption behaviors and global influences provides valuable insight toward predicting future trends and strategizing business decisions, such as marketing campaigns.

1. Introduction

Our research focuses on the effects of consumer music behavior on mainstream culture and popular media. By examining the music-related activities of millions of users, across tens of millions of individual interactions with the popular music streaming platform Spotify, we can study how such behavior unwittingly generates trends in pop culture and develops into an extension of other consumption behaviors. To determine future plans, understanding these activities and their dynamics is vital for businesses aiming to adapt their strategies to new consumer trends and tastes. Our results will therefore point to correlations between consumer behavior and future prediction, organizational decision-making, and deployment of marketing plans and campaigns.

2. Related Work

Prior research has investigated how music consumption aligns with and predicts broader cultural patterns. Barata, Lopes, and Coelho (2021) studied how streaming platforms drive music consumption and discovery, finding that algorithm-operated user interfaces tailor the music to user-specific tastes. They also explore how music affects brand attitude and purchase intention, finding that advertising and music selection should align with the dynamic music landscape. While our analysis is motivated by insights from these studies, we differentiate ourselves by focusing on using Spotify's data and SAP database technology to (1) assess global streaming behavior and (2) predict future streaming metrics.

Our study furthers the existing research that positions music consumption as a driver of changes in cultural and business strategy. Literature has shown that music is a significant driver in influencing consumer behavior and its perceptions of brands which has become particularly salient with the advent of streaming services, such as Spotify. By utilizing matched data and the implementation of a robust finding according to analytical techniques, research has been able to discern usage patterns and signals of changes in the dynamic music landscape which can be worthwhile for businesses in any industry to recognize.

3. Specifications

The data set was retrieved from Kaggle. Kaggle is a free platform that publishes datasets from various sources. Our dataset, Spotify's Worldwide Daily Song Ranking, has rankings from January 1, 2017, to January 9, 2018. It contains the daily Spotify stream rankings of the 200 most-listened songs in 53 countries. The dataset was derived from Spotify's regional chart data and posted on Kaggle. The size of the dataset is 54 megabytes.

4. Implementation of Flowchart

The initial dataset was downloaded from Kaggle. It included chart position, track name, artist, number of streams, URL, date, and region. The flowchart below (Figure 1) demonstrates the process of data handling. The data file was uploaded to the SAP Analytics cloud and converted to a data set, which was used to create a data model. The model was transformed into a story and predictive analysis scenario. The data was examined and then transformed into several resource materials, including a lab tutorial, presentation, and term paper.
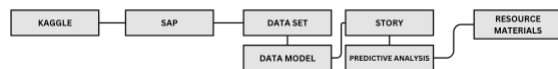

Figure 1: Flowchart

5. Data Cleaning

The CSV file was uploaded into the SAP Analytics cloud as a datasheet. From there, the first step in cleaning the data was removing the "URL" column. This was removed because it was insignificant to our dataset. Second, the column labeled "Region" was capitalized so it could be consistent. Next, the column "Region" was Geo-Enriched by Area so that each region's demographics could be added, and we were able to identify trends. Lastly, we added a Level-Based Hierarchy, ArtistHierarchy, to establish relationships between variables.

Once the dataset was cleaned it was ready to create a story.

6. Analysis and Visualization

We used the Spotify Worldwide Daily Song Ranking dataset to launch a global survey into the listening habits of 53 countries. We used various data visualization techniques, such as time series and bar plots, to visually depict changes in chart positions, artist trends, and regional patterns. From our analysis and results, we found that the charts tend to reflect mainstream culture by ranking mainstream artists and genres higher. We identified strong correlations between chart popularity and current events, further emphasizing the global impact of popular music and its ability to reflect and shape culture. We also aim for our analysis to offer some insights for businesses to better understand musical trends and how they can incorporate them into their overall branding and marketing strategies.

6.1 Most Popular Artist and Song in 2017

The bar chart below (Figure 2) illustrates the leading artists of 2017. Ed Sheeran emerged as the most popular artist, amassing approximately 4.3 billion streams. Notably, this figure nearly doubles the streams of the second artist on the list, Drake, who had 2.3 billion streams. It is worth mentioning that Ed Sheeran's dominance coincided with the release of his third studio album, "Divide," in March 2017.

Total Streams by Artists in 2017
Top 25 - All Dimensions

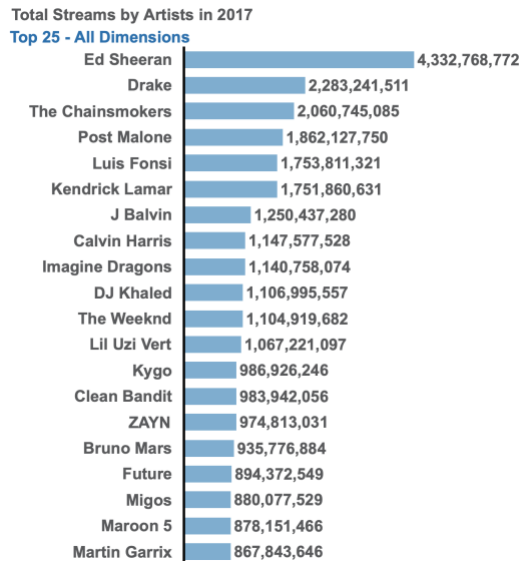| Artist	Streams
Ed Sheeran	4,332,768,772
Drake	2,283,241,511
The Chainsmokers	2,060,745,085
Post Malone	1,862,127,750
Luis Fonsi	1,753,811,321
Kendrick Lamar	1,751,860,631
J Balvin	1,250,437,280
Calvin Harris	1,147,577,528
Imagine Dragons	1,140,758,074
DJ Khaled	1,106,995,557
The Weeknd	1,104,919,682
Lil Uzi Vert	1,067,221,097
Kygo	986,926,246
Clean Bandit	983,942,056
ZAYN	974,813,031
Bruno Mars	935,776,884
Future	894,372,549
Migos	880,077,529
Maroon 5	878,151,466
Martin Garrix	867,843,646

Figure 2: Total Streams by Artists in 2017

The line graph depicted in Figure 3 showcases the five most streamed songs in 2017. Once more, Ed Sheeran takes the lead with his track "Shape of You," accumulating around 1.5 billion streams. "Shape of You" is featured in his album, "Divide," but was separately released as a single in January 2017. This graph also underscores the diversity of genres within the top 5 streamed songs.

5 Most Popular Songs Worldwide
Top 5 - All Dimensions

1,495,408,816
906,156,951
716,064,459   686,425,672   647,809,182
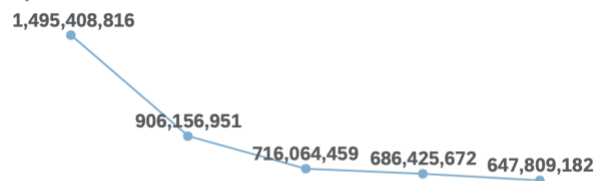
Ed Sheeran	Luis Fonsi	Luis Fonsi	The Chai...	Kendrick ...
Shape of...	Despaci...	Despaci...	Somethi...	HUMBLE.

Figure 3: Five Most Popular Songs Worldwide

6.2 Streams by Nation and Region in 2017
The map in Figure 4 illustrates the number of streams by nation in 2017. Leading global statistics is the United States (US), depicted in lime green on the map, with over 26.3 billion streams. Surprisingly, non-English speaking countries such as Mexico, Brazil, etc. (highlighted in orange) follow with the next largest number of streams.
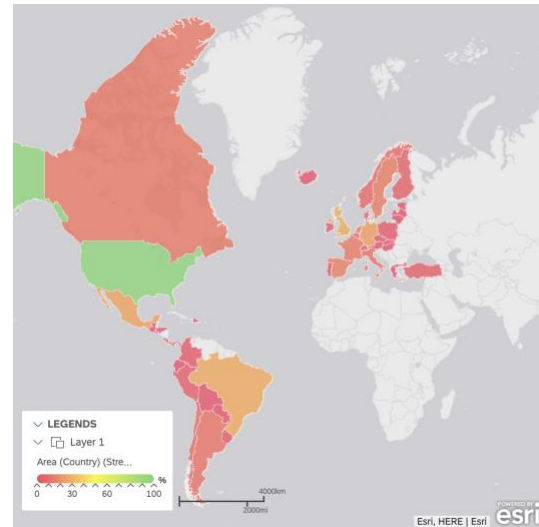

Figure 4: Streams by Nation 2017

The pie chart displayed in Figure 5 further demonstrates the significant influence the US has on mainstream music, representing 37.61% of total streams. While population density does contribute to the number of streams per region, it's worth noting countries like Sweden (SE) and the Philippines (PH) are included in the figure despite their comparatively smaller populations.

Streams by Region in 2017
Top 10 - All Dimensions

US GB BR MX DE ES NL AU
SE PH

37.61%
10.24%
9.82%
9.04%
8.66%
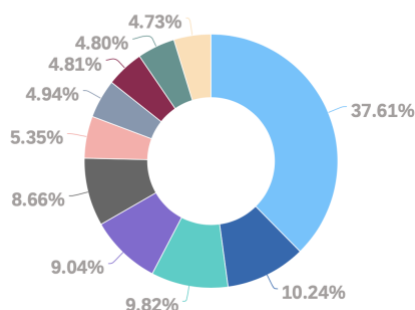5.35%
4.94%
4.81%
4.80%
4.73%

Figure 5: Streams by Region 2017

6.3 Consumption Patterns

The line graph in Figure 6 shows how consumption patterns are affected by different periods, specifically, highlighting the fluctuation of music streams over months. Globally, it is evident that the highest music consumption occurs in December, January, and March, while February shows the least consumption. The increased streaming activity during December and January can be attributed to holiday gatherings and festivities.
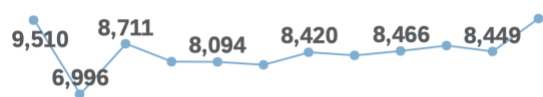
Streams per Month

in Million

Figure 6: Global Streams per Month

6.4 Time Series Analysis and Forecasts

Our research utilized SAP's machine learning tools to predict stream metrics by date. We utilized a time series graph to analyze historical streaming patterns and applied an automatic forecast to measure future streaming behaviors. The graph below (Figure 7) shows a time series analysis of streams per day. This helped visualize how specific days influence the number of Spotify streams. Time series, as both an analytical and predictive measure, is a valuable data-driven tool. For instance, in Figure 7 there's a peak in December and January. It would be strategic for a business to align its marketing strategy during the months of peak streams.

To strengthen our predictive analytics, we implemented an additional predictive model. We utilized SAP's machine learning service, Smart Predict, to build a predictive scenario. Like the time series, our regression-based scenario sought to make predictions about streaming trends (our target) based on historical streaming data (Figure 8). To ground our analysis, we excluded all influencers except region, date, and artist. We selected a regression model over other predictive scenarios, like classification, due to our non-binary data set. After the model was trained, it had a prediction confidence of over 99.4%. The root mean square error was over 114k, which is acceptable considering the data size and streaming counts. The model also demonstrated high accuracy levels in its predicted vs. actual line assessments. These assessments indicate our model is a viable tool for predictive analysis.
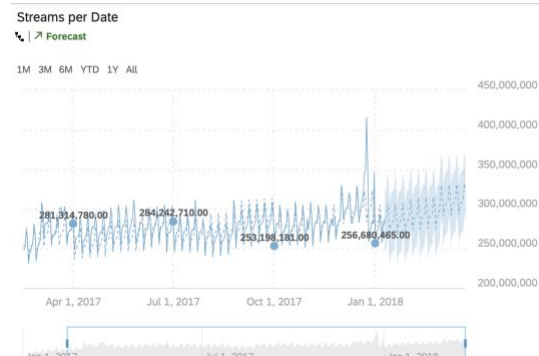
Streams Per Date Forecast

Figure 7: Time Series Analysis

6.5 Listening Behavior

Global Performance Indicators

Root Mean Square Error (RMSE)	Prediction Confidence
114,621.23	99.84%

Target Statistics

Data Partition	Minimum	Maximum	Mean	Standard Deviation
Training	1,001	11,381,520	51,888.47	201,485.78
Validation	1,001	9,849,173	51,901.64	202,754.83

Influencer Contributions

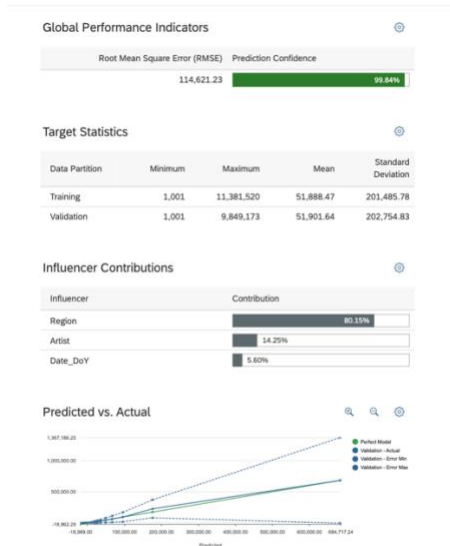| Influencer	Contribution
Region	80.15%
Artist	14.25%
Date_DoY	5.60%

Predicted vs. Actual

Figure 8: Predictive Analysis

7. Challenges with Data Set

Although our dataset contained extensive information, we faced several challenges during our analysis. Despite its size, the data lacked key information such as the genre of each song. Understanding the genre could have been valuable in identifying trends and shifts in popularity across different music genres throughout the year. Additionally, our dataset only covered the year 2017, limiting our ability to track trends over multiple years and potentially hindering our ability to identify long-term patterns.

8. Conclusion

To summarize, our analysis showcases how music consumption data can inform our understanding of cultural shifts and consumer behavior. By mining data from Spotify's daily chart positions, we gain a comprehensive view of popular music and its influence on the mainstream, which is invaluable for businesses and marketers. This study highlights the importance of region, artist, and time when assessing streaming metrics. Brands should align their business models with data-driven solutions that assess and leverage all three variables. This will help brands drive product engagement and capitalize on existing consumer behavior trends. This study functions as a case study of the confluence of music, culture, and business in today's digital world. Additional resources regarding this study may be accessed via our Github: https://github.com/BUS5100group5/2017spotifytrends.

References

1. Barata, Mariana Lopes, and Pedro Simões Coelho. "Music Streaming Services: Understanding the Drivers of Customer Purchase and Intention to Recommend." *Heliyon*, vol. 7, no. 8, Aug. 2021, https://doi.org/10.1016/j.heliyon.2021.e07783.
2. C, Vickie. "Spotify Is a Pop Culture Icon." *Medium*, 2 July 2023, medium.com/@vickiec630/spotify-is-a-pop-culture-icon-ed25e6497c8e.
3. Eduardo. "Spotify's Worldwide Daily Song Ranking." *Www.kaggle.com*, Kaggle, 2019, www.kaggle.com/datasets/edumucelli/spotifys-worldwide-daily-song-ranking. Accessed 2 Mar. 2024.
4. Janak, Heather. "How Background Music in Ads Affects Consumers." *Texas A&M Today*, 16 Aug. 2021, today.tamu.edu/2021/08/16/how-background-music-in-ads-affects-consumers/.
5. Gulliver, M, Rocha, J. Hernandez Lugo, G. Roman, D. "BUS5100GROUP5/2017spotifytrends." *GitHub*, 30 Mar. 2024, github.com/BUS5100group5/2017spotifytrends.