

Customer Analytics

Spotify Project: Audio Feature Analysis of Popular Songs Report

Names: Shashank Baluni

Explore patterns in audio features to understand trends and preferences in popular songs



Introduction

Spotify is a popular digital music streaming service that allows users to access a vast library of songs, podcasts, and other audio content. It offers both free, ad-supported plans and premium subscription plans that provide additional features such as ad-free listening, offline downloads, and higher audio quality.

With Spotify, users can create playlists, discover new music based on their preferences, and share their favorite songs with friends. The platform uses algorithms to analyze user listening habits and provide personalized recommendations.

Spotify is available on various devices, including smartphones, tablets, computers, smart speakers, and more. Users can access the service through the Spotify app or web player, making it convenient for music lovers to enjoy their favorite tunes anytime, anywhere, with an internet connection.

About the project

This is a project required to explore the patterns in audio features to understand trends and preferences in popular songs. In the dynamic realm of the music industry, understanding the intricate patterns and nuances within audio features is imperative for unraveling the secrets behind the success of songs. The dataset at our disposal provides a unique and extensive catalog of the most influential songs of 2023, curated from the renowned music streaming platform, Spotify. Beyond the conventional song attributes, this dataset delves deeper into the multifaceted dimensions of each musical composition, offering insights into factors such as artist collaboration, release details, streaming statistics, and presence on various music platforms like Apple Music, Deezer, and Shazam.

Dataset

The provided Spotify data on Kaggle seems to represent information about music tracks with various features related to their characteristics, release details, and popularity on different streaming platforms. Here's a brief overview of the features:

Song Details	Artist Details	Release Details	Streaming Platforms
<ul style="list-style-type: none">• Track Name• Streams• bpm, key, mode• danceability, valence, energy, acousticness, instrumentalness, liveness, speechiness	<ul style="list-style-type: none">• Artist Name• Artist Count	<ul style="list-style-type: none">• Released Year• Released Month• Released Day	<ul style="list-style-type: none">• Spotify Playlist and Charts• Apple Playlist and Charts• Deezer Playlist and Charts• Shazam Charts

Figure 1: A list of features available in the dataset

Description of features:

- track_name: Name of the music track.
- artist(s)_name: Name of the artist or artists associated with the track.
- artist_count: Number of artists involved in creating the track.
- released_year: Year in which the track was released.
- released_month: Month in which the track was released.
- released_day: Day of the month on which the track was released.

- in_spotify_playlists: Indicates whether the track is present in Spotify playlists (binary: 0 or 1).
- in_spotify_charts: Indicates whether the track is present in Spotify charts (binary: 0 or 1).
- streams: Number of streams the track has received.
- in_apple_playlists: Indicates whether the track is present in Apple Music playlists (binary: 0 or 1).
- in_apple_charts: Indicates whether the track is present in Apple Music charts (binary: 0 or 1).
- in_deezer_playlists: Indicates the presence of the track in Deezer playlists (data type: object).
- in_deezer_charts: Indicates whether the track is present in Deezer charts (binary: 0 or 1).
- in_shazam_charts: Indicates the presence of the track in Shazam charts (data type: object).
- bpm: Beats per minute, a measure of the tempo or speed of the track.
- key: Musical key of the track (data type: object).
- mode: Indicates the modality of the track (major or minor).
- danceability_%: Danceability percentage of the track.
- valence_%: Valence percentage, representing the musical positiveness of the track.
- energy_%: Energy percentage, representing the intensity and activity of the track.
- acousticness_%: Acousticness percentage, representing the acoustic quality of the track.
- instrumentalness_%: Instrumentalness percentage, indicating whether the track is instrumental.
- liveness_%: Liveness percentage, representing the likelihood of the track being performed live.
- speechiness_%: Speechiness percentage, indicating the presence of spoken words in the track.

This dataset appears to capture a comprehensive set of features related to music tracks, including information about the artists, release details, popularity on different platforms, and various audio characteristics.

By exploring the plethora of audio features available, we aim to decipher patterns that resonate with listeners and discern the trends that captivate the contemporary music landscape. This exploration extends beyond conventional metrics, providing a nuanced understanding of the musical ecosystem, including the impact of streaming platforms, chart rankings, and the intricacies of audio attributes.

What makes Audio Feature Patterns significant?

In the age of digital music consumption, platforms like Spotify have become key influencers in shaping musical preferences and trends. Understanding the patterns in audio features becomes crucial for artists, producers, and enthusiasts alike. It unveils the science behind what makes a song stand out in the vast sea of musical content. From the tempo and key to danceability, valence, and acousticness, each feature contributes to the overall appeal and success of a song. By unraveling these patterns, we aim to provide valuable insights into the factors that resonate with listeners, contributing to the collective understanding of musical preferences and shaping the future of the music industry.

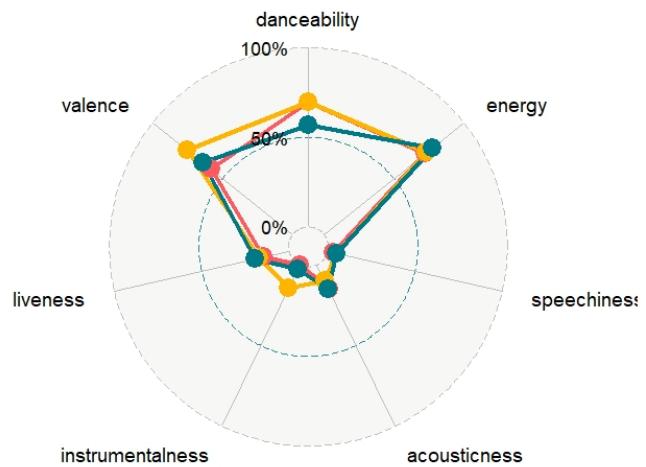


Figure 2: Spider chart to show the contribution of music features in its popularity (illustrative)

Methodology

Software Used

The exploration of patterns in audio features is conducted using Python, leveraging various libraries such as:

- Pandas for data manipulation and analysis,
- Matplotlib and Seaborn for visualization, and
- potentially Scikit-Learn for modeling if required.

These tools offer a powerful and flexible environment for handling datasets and extracting meaningful insights.

Python, with its rich ecosystem of data science libraries, provides a robust and efficient platform for exploring and analyzing complex datasets. Pandas facilitates the manipulation and transformation of data, while Matplotlib and Seaborn enable the creation of insightful

visualizations. If necessary, Scikit-Learn offers machine learning capabilities that can be employed for predictive modeling or clustering analysis. The choice of these tools is driven by their versatility, popularity, and the extensive support they offer for data-driven tasks.

Data Exclusions

The dataset at hand is assumed to be comprehensive and inclusive of the most famous songs of 2023 on Spotify. However, certain data points may be excluded if they are deemed irrelevant to the specific analysis of audio features and their influence on song success. For instance, auxiliary information like release date and streaming statistics might not directly contribute to the exploration of audio feature patterns and could be set aside for a more targeted investigation.

Also, in “Streams” column there is a row in which wrong values are there, so we removed this row.

Data Wrangling

- Missing Data Handling: Check for and handle missing values in the dataset, ensuring the integrity of the analysis.
 - “in_shazam_charts” has missing values which we replaced with 0.
 - “Key” has missing values as well and we have removed this column from our analysis.
- Data Types and Encoding: Verify data types and encode categorical variables if needed.
- Exploratory Data Analysis (EDA):
 - Descriptive Statistics: Understand the distribution and central tendencies of the audio features.
 - Visualization: Utilize visualizations to identify patterns, correlations, and outliers within the dataset.
- Feature Engineering:
 - Derive new features if necessary to enhance the analysis.
 - Normalize or scale numerical features to ensure a consistent scale.
- Statistical Analysis:
 - Conduct statistical tests if required to validate hypotheses or relationships between variables.
- Machine Learning (if applicable):

- Employ machine learning models to predict song success or perform clustering based on audio features.
- Evaluate model performance using appropriate metrics.
- Insights and Reporting:
 - Summarize key findings, insights, and trends.
 - Provide actionable recommendations based on the analysis.

The overarching goal is to uncover patterns and trends in audio features that correlate with the success of songs on Spotify, offering valuable insights for artists, producers, and the music industry as a whole. The methodology is designed to be iterative, allowing for flexibility in adapting to emerging patterns and refining the analysis as needed.

Analysis of the data set

In the course of our project, our focus is on analyzing the top 10 most streamed songs to identify commonalities and trends among them. Additionally, we aim to delve into the characteristics of the top 10 most streamed artists and make comparisons across their songs. Before delving into these aspects, our initial exploration involves examining the relationship between the number of streams and various audio features. This preliminary analysis is crucial for gaining insights into the factors that contribute to a song or artist's success.

Notably, our observation highlights the significance of a song's release year, revealing that songs released during 2017-2018 tend to have higher stream counts compared to those released before or after this period.

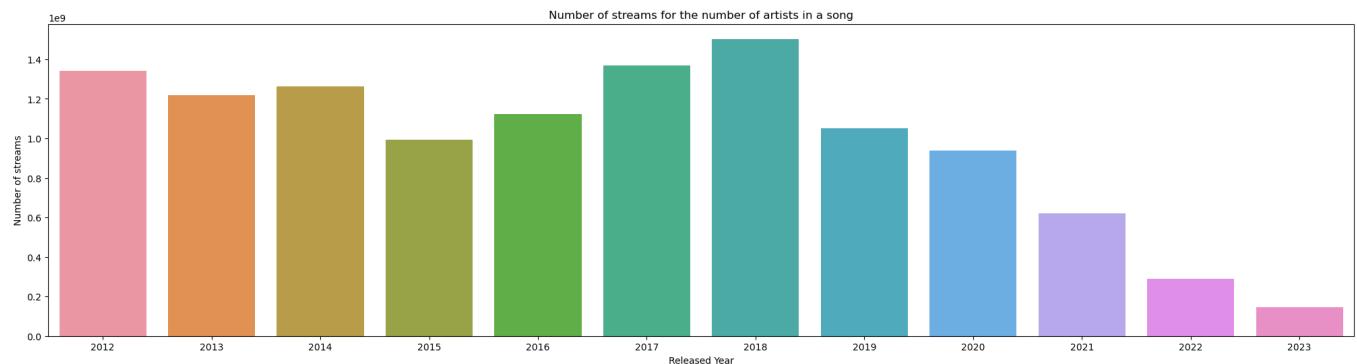


Figure 3: Average streams by release year

Initially, we examined the correlation between the number of artists in a song and its streaming count. The majority of songs feature either one or two artists, with a notable rarity of songs having five or more artists. The maximum number of artists found in a single song is eight. Our analysis indicates a general trend of decreasing streams as the number

of artists in a song increases. However, intriguingly, a song with seven artists outperformed songs with four artists in terms of stream count. To ensure fairness in our analysis, we calculated the mean, given the prevalence of songs with one or two artists.

distribution of number of artists in each song

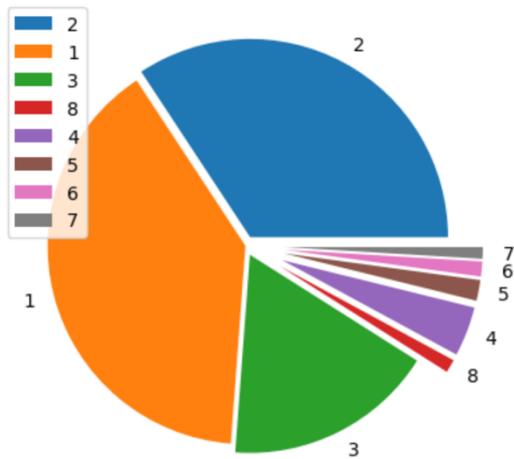


Figure 4: Distribution of songs by no. of artists

Number of streams for the number of artists in a song

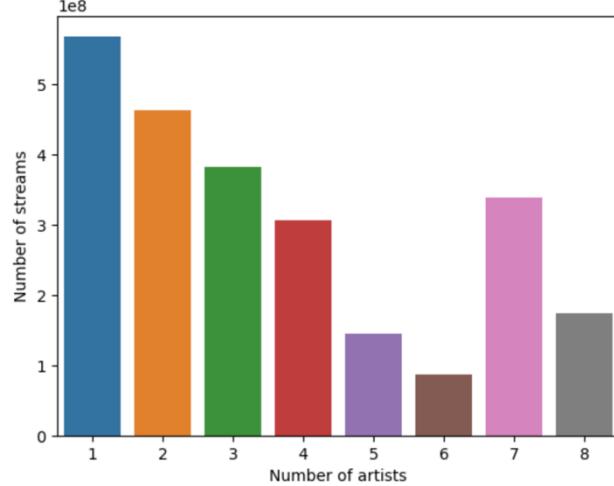


Figure 5: Average streams by no. of artists in a song

The Beats Per Minute (BPM) of a song is a crucial element that plays a significant role in shaping its musical characteristics and overall impact. It influences the rhythm, energy, danceability, and emotional impact of a song. It is a tool that artists, producers, and listeners use to understand and connect with the musical content, contributing to the diversity and richness of the musical landscape across various genres and styles. For the given data, BPM ranges from 65 to 206. The songs with bpm in the range of 166-185 have the highest number of streams. The songs with bpm in the range of 126-145 have the least number of streams.

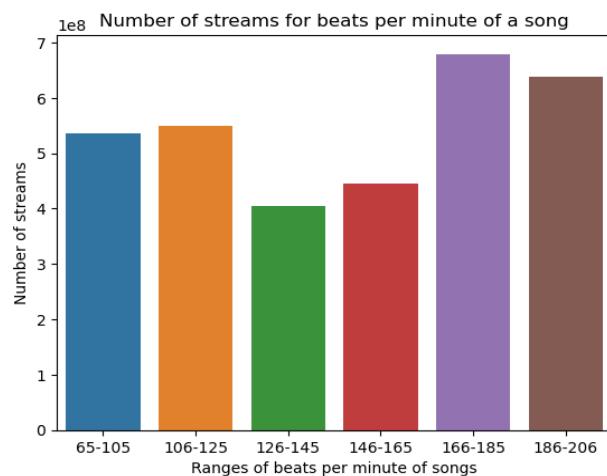


Figure 6: Average streams by BPM

It's also important to understand what effect bpm has on audio features like danceability, liveliness and instrumentalness, etc. The chart below shows that there is a steep increase in the liveliness of a song for bpm around 160-180. There is a steep decrease in the danceability of a song for bpm around 80-100. The instrumentalness of a song is equally distributed among all ranges of bpm.

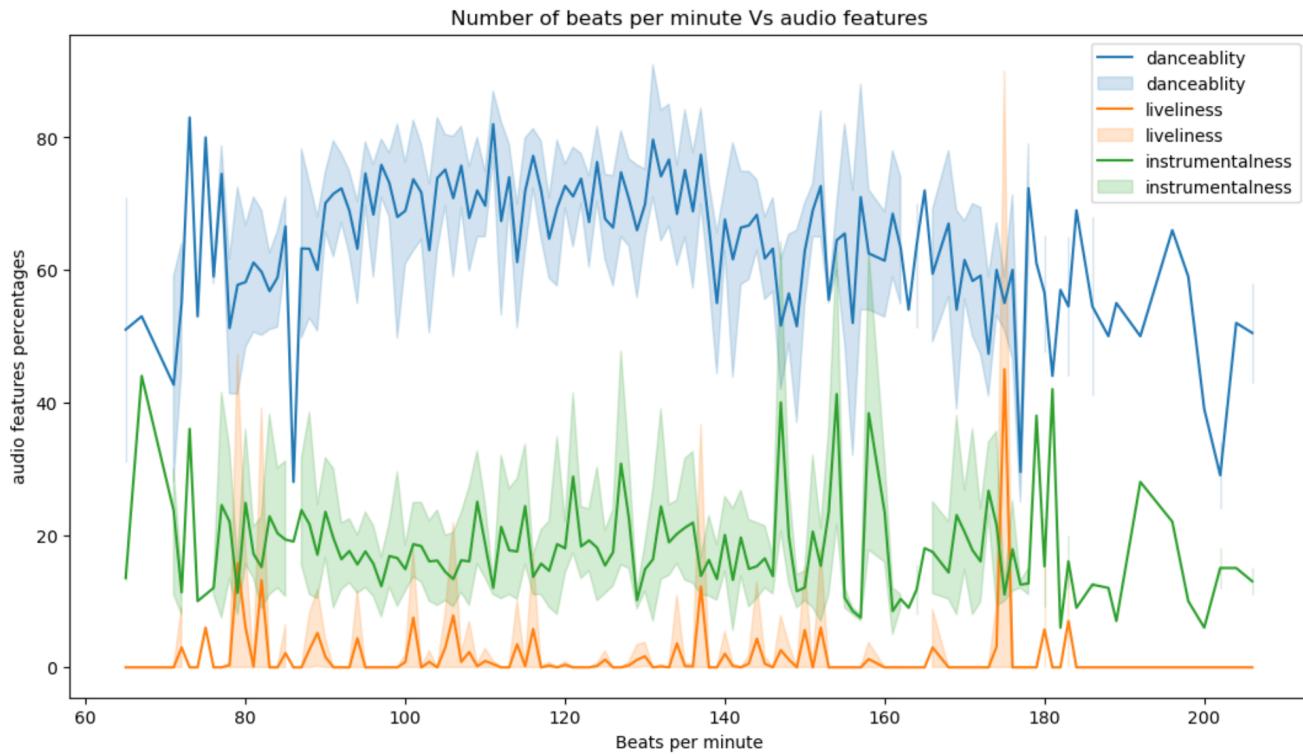


Figure 7: Line plot of BPM vs audio features

Next, let's see how different audio features are related to each other. For this we will plot a correlation matrix of audio features. From the matrix we can see that energy and acousticness don't go well together and are negatively correlated. valence and danceability have a good positive correlation with each other.

Now, let us plot a scatter plot of different audio features with the number of streams a song gets and find their correlation. Speechiness and streams are negatively correlated. Instrumentalness and streams are negatively correlated.

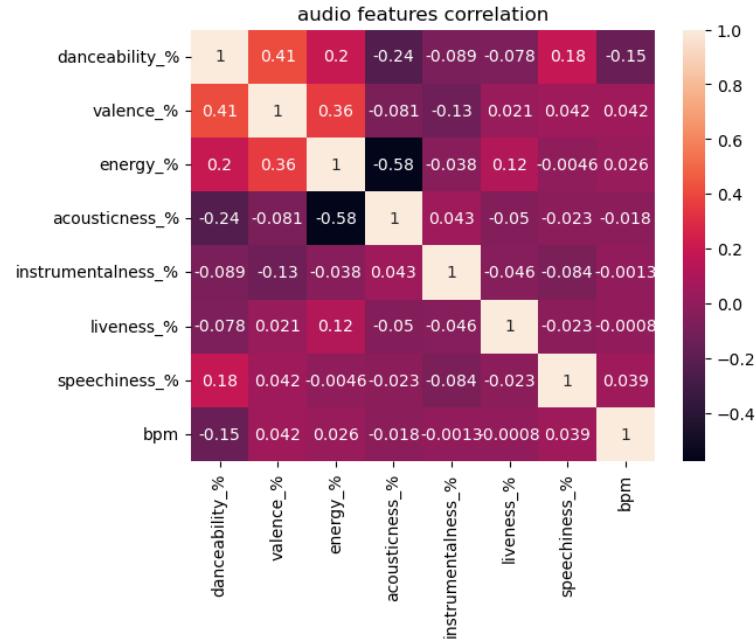


Figure 8: Correlation matrix of audio features

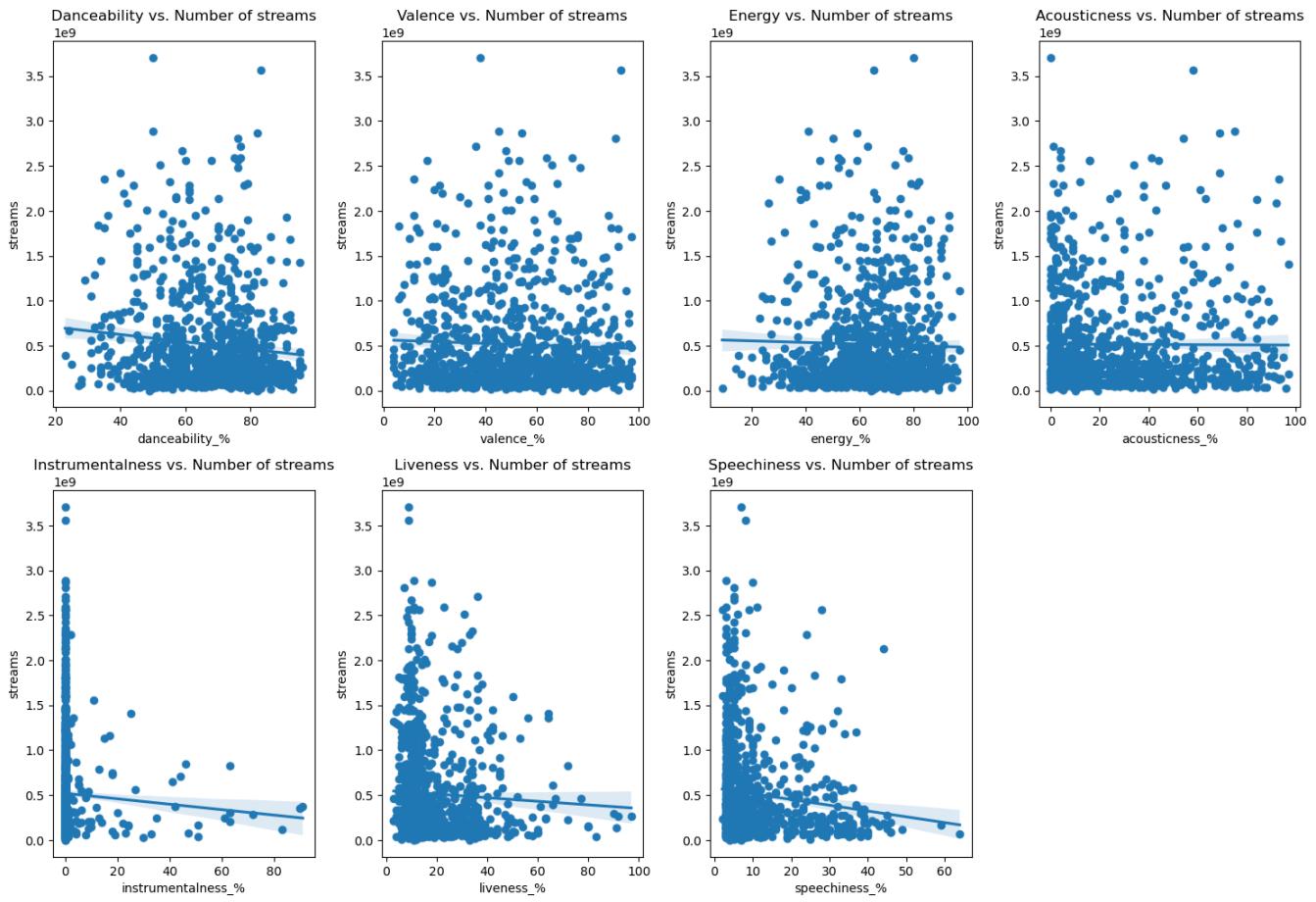


Figure 9: Scatter plot of audio features by streams

To summarize, a song is more likely to have success if it has a bpm in the range of 166-185. The songs with only one artist have the highest number of streams. There is a steep decline in the danceability if a song has around 85-90 bpm. There is a huge incline in the liveliness of a song if it has around 170-175 bpm. The more the speechiness and instrumentalness of a song the less likely it is to get more streams. Keeping this in mind now let's look at top 10 songs and artists and what makes them successful.

Top 10 Most Streamed Songs

We used “Sorting” function in descending order based on the “streams” column to display the top 10 most streamed songs in the dataset. Each song in the list has garnered a significant number of streams, making them popular among listeners.

Here are the top 10 most streamed songs and some of their features which make them special:

Track name	streams	Artists name	released_year	bpm	danceability_%	valence_%	energy_%	acousticness_%	liveness_%	speechiness_%
Blinding Lights	3,703,895,074	The Weeknd	2019	171	50	38	80	0	9	7
Shape of You	3,562,543,890		2017	96	83	93	65	58	9	8
Someone You Loved	2,887,241,814		2018	110	50	45	41	75	11	3
Dance Monkey	2,864,791,672		2019	98	82	54	59	69	18	10
Sunflower - Spider-Man: Into the Spider-Verse	2,808,096,550		2018	90	76	91	50	54	7	5
One Dance	2,713,922,350		2016	104	77	36	63	1	36	5
STAY (with Justin Bieber)	2,665,343,922		2021	170	59	48	76	4	10	5
Believer	2,594,040,133		2017	125	77	74	78	4	23	11
Closer	2,591,224,264		2016	95	75	64	52	41	11	3
Starboy	2,565,529,693	The Weeknd, Daft Punk	2016	186	68	49	59	16	13	28

Figure 10: List of top 10 songs and their audio features (sorted on streams)

**Color gradient – high to low; green for high values and yellow for low values.

1. "Blinding Lights" by The Weeknd

Solo artist with a high energy level (80%) and low acousticness (0%).

2. "Shape of You" by Ed Sheeran

Solo artist with very high danceability (83%) and moderate energy and acousticness.

3. "Someone You Loved" by Lewis Capaldi

Solo artist with a mix of danceability (50%) and relatively high acousticness (75%).

4. "Dance Monkey" by Tones and I

Solo artist with high danceability (82%) and moderate energy and acousticness.

5. "Sunflower - Spider-Man: Into the Spider-Verse" by Post Malone, Swae Lee

Collaboration with two artists, high danceability (76%), and moderate energy and acousticness.

6. "One Dance" by Drake, WizKid, Kyla

Collaboration with three artists, high danceability (77%), and low acousticness (1%).

7. "STAY (with Justin Bieber)" by Justin Bieber, The Kid Laro

Collaboration with two artists, moderate danceability (59%), and low acousticness (4%).

8. "Believer" by Imagine Dragons

Solo artist with high danceability (77%) and energy (78%).

9. "Closer" by The Chainsmokers, Halsey

Collaboration with two artists, high danceability (75%), and moderate energy and acousticness.

10. "Starboy" by The Weeknd, Daft Punk

Collaboration with two artists, moderate danceability (68%), and moderate energy and acousticness.

These songs have not only captivated audiences but have also become global phenomena, contributing to their high streaming counts. The popularity of the top 10 most streamed songs can be attributed to a combination of factors. Firstly, the majority of these songs were released in earlier years, between 2016 and 2019, allowing them ample time to accumulate streams. Additionally, the presence of well-known and globally recognized artists like The Weeknd, Ed Sheeran, Drake, and Justin Bieber contributes significantly to the songs' popularity. A notable characteristic is the variation in Beats Per Minute (BPM), with some songs featuring a higher tempo, creating an engaging and energetic listening experience. The high danceability percentages suggest a strong beat, appealing to listeners who enjoy dancing to music. Moreover, the songs generally exhibit moderate to high energy and valence percentages, contributing to an overall lively and positive musical atmosphere. Low acousticness indicates a preference for electronically produced sounds over acoustic elements. Some songs incorporate a higher liveness percentage, suggesting elements recorded in a live setting, enhancing the overall appeal. The variability in speechiness percentages reflects the presence of spoken words or lyrics in certain songs, adding a distinctive vocal element. Collectively, these factors make the top 10 most streamed songs favorites among listeners, appealing to a broad and diverse audience.

When can also look at the least popular songs and measure how do their audio features contrast to the top songs.

Track name	streams	Artists name	released_year	bpm	danceability_%	valence_%	energy_%	acousticness_%	liveness_%	speechiness_%
Que Vuelvas	2,762	Carin Leon, Grupo Frontera	2022	162	49	78	64	19	11	4
Jhoome Jo Pathaan	1,365,184	Arijit Singh, Vishal Dadlani, Sukriti Kakar, Vishal-Shekhar, Shekhar Ravjiani, Kumaar	2022	105	82	62	74	10	33	7
QUEMA	11,599,388	Sog, Ryan Castro, Peso Pluma	2023	97	79	92	89	5	6	5
Gol Bolinha, Gol Quadrado 2	11,956,641	Mc Pedrinho, DJ 900	2023	133	93	68	65	42	12	25
Overdrive	14,780,425	Post Malone	2023	140	56	48	73	0	35	4
S91	16,011,326	Karol G	2023	128	86	42	72	59	9	19
Rush	22,581,161	Troye Sivan	2023	126	74	35	84	0	11	6
Danger (Spider) (Offset & JID)	24,975,653	Offset, JID	2023	143	83	25	69	4	23	12
New Jeans	29,562,220	NewJeans	2023	134	81	53	72	51	12	5
Better Than Revenge (Taylor's Version)	30,343,206	Taylor Swift	2023	146	50	67	89	0	19	8

Figure 11: List of bottom 10 songs and their audio features (sorted on streams)

Factors contributing to the lower popularity of the top 10 least streamed songs include several elements. Firstly, these songs predominantly hail from recent years (2022-2023), implying that they may not have had sufficient time to accumulate streams compared to their more established counterparts. Additionally, some of the less streamed songs feature artists who may not enjoy widespread recognition on a global scale. The Beats Per Minute (BPM) in these songs varies, with some exhibiting a lower tempo that may impact their overall appeal. Danceability percentages also fluctuate, with some of the less streamed songs having a less engaging rhythm for listeners. Furthermore, varying levels of energy and valence are observed, but certain songs register lower values compared to the more streamed counterparts. Acousticness percentages differ, and some of the less streamed songs lean towards higher acousticness, indicating a preference for acoustic elements. The liveness percentage varies, and certain songs have lower liveness, suggesting a potentially more studio-recorded or less dynamic nature. Additionally, speechiness percentages fluctuate, with some songs featuring fewer spoken words or lyrics. Collectively, these factors contribute to the comparatively lower popularity of the top 10 least streamed songs, reflecting a range of elements influencing listener preferences.

Overall Trends:

- The popularity of songs is influenced by a combination of artist recognition, musical elements (e.g., BPM, danceability, energy), and the time since release.
- Global recognition and established artists contribute to the success of most streamed songs.
- The less streamed songs may face challenges related to recency, artist recognition, and musical characteristics that may not align with current listener preferences.

In the next section let's look at the top 10 most streamed artists.

Top 10 Most Streamed Artists

Several data pre-processing steps were implemented to enhance the analysis, including the utilization of functions such as "explode()," "split()," and "lstrip()." This was particularly relevant for the "artist(s)_name" column, which contained multiple artists in a single entry. To address this, we used the "explode()" function to separate artists into distinct rows and renamed the column to "artist_name." Subsequently, the "groupby()" function was applied to aggregate all songs of each artist, calculating the total number of streams gained by summing the streams of their respective songs. These pre-processing measures were crucial

for obtaining a more refined and insightful analysis of the streaming data. This process allowed us to identify the top 10 most streamed artists on Spotify in 2023.

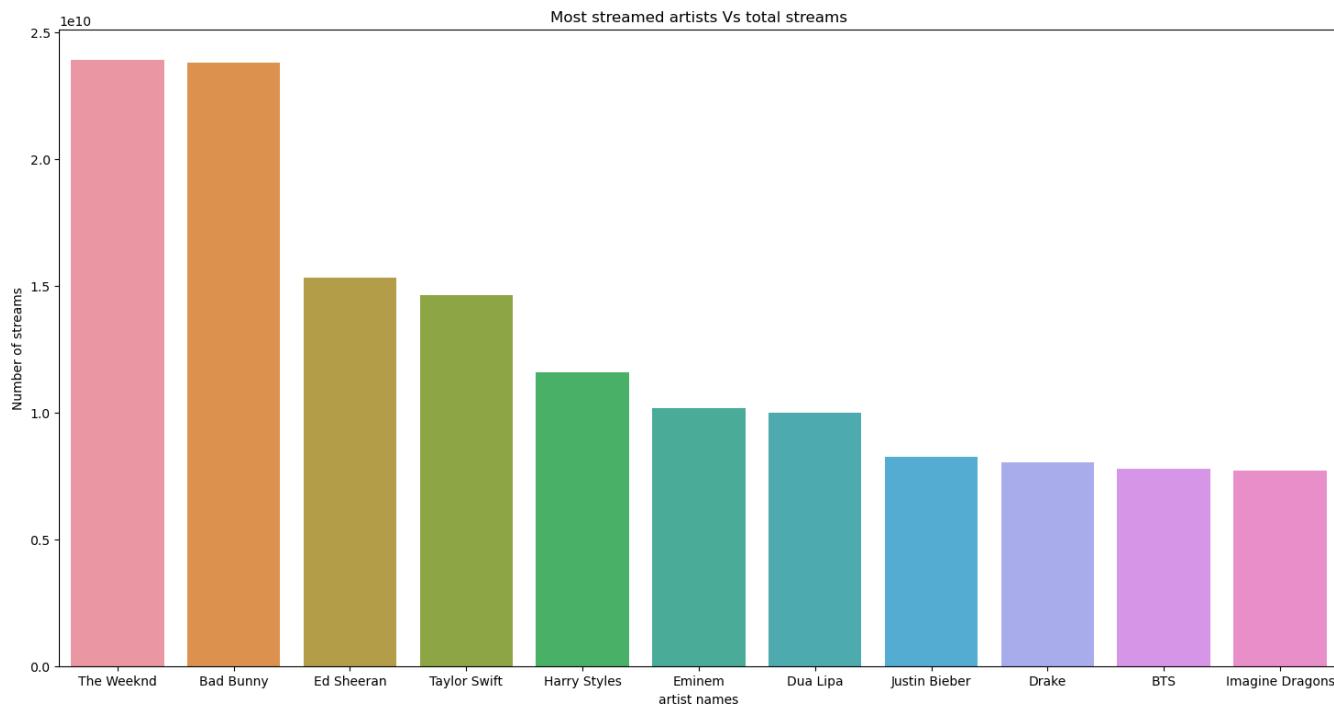


Figure 12: Most streamed artists vs total streams

We looked at the audio features of most streamed artists as it could indicate patterns in musical preferences that resonate with a large audience. These patterns might influence what is considered popular or trendy. For example, if energetic and danceable songs consistently dominate the charts, it suggests a trend towards more upbeat music.

The Weeknd and Bad Bunny dominate in total streams, affirming their widespread popularity. A prevailing trend of collaborative projects involving three artists is observed, seen in works by The Weeknd, Bad Bunny, and Dua Lipa. Ed Sheeran's 1996 debut contrasts with newer artists like Bad Bunny and BTS in the late 2010s, highlighting a temporal mix. BPM variations, from Justin Bieber's 138 to Harry Styles' slightly lower pace, underscore diverse musical expressions. Eminem and Dua Lipa favor high danceability, while Taylor Swift and The Weeknd showcase varied danceability. Dua Lipa's uplifting valence contrasts with Taylor Swift's nuanced emotional tones. Eminem and Dua Lipa consistently produce high-energy tracks, with Taylor Swift's slightly lower levels reflecting varied compositions. Ed Sheeran leans towards acoustic sounds, while Eminem prefers electronically produced ones. The likelihood of live elements is indicated by higher liveness percentages for The Weeknd, Bad Bunny, and Justin Bieber. Eminem's high speechiness contrasts with Dua Lipa's balanced mix of singing and spoken words. Overall,

the analysis unveils a diverse musical landscape, providing insights into artists' popularity and showcasing rich diversity in musical characteristics.

Track name	streams	Maximum Artists in a song	First song released in	bpm	danceability_%	valence_%	energy_%	acousticness_%	liveness_%	speechiness_%
The Weeknd	23,929,760,757	3	2015	118	60	44	64	20	21	8
Bad Bunny	23,813,527,270	3	2019	124	74	51	69	24	20	11
Ed Sheeran	15,316,587,718	2	1996	114	71	56	63	33	18	5
Taylor Swift	14,630,378,183	2	2010	124	60	34	55	31	17	7
Harry Styles	11,608,645,649	1	2019	129	61	54	59	43	14	5
Eminem	10,193,727,260	2	1999	123	80	47	74	6	23	16
Dua Lipa	9,980,020,481	3	2016	112	76	74	80	6	16	9
Justin Bieber	8,243,081,039	3	2011	138	68	58	63	31	22	13
Drake	8,043,031,261	3	2016	134	74	31	55	6	23	20
BTS	7,780,428,159	3	2019	117	69	63	72	11	22	9

Figure 13: Summarized audio features of top 10 most streamed artists

The Weeknd and Bad Bunny Dominance

"The Weeknd" and "Bad Bunny" dominate Spotify streams, showcasing broad appeal across diverse demographics. Their consistent chart presence suggests sustained popularity, potentially due to hit releases and playlist inclusions. The substantial stream difference indicates a competitive edge, reflecting their ability to engage listeners. This dominance culturally influences contemporary music and significantly impacts Spotify's success, potentially driven by effective marketing and promotional strategies.

The Weeknd's top 10 songs showcase a diverse range of musical styles and emotions. "Blinding Lights," released in 2019, leads with 3,703,895,074 streams and a lively BPM of 171, coupled with high danceability and energy. "Starboy," featuring two artists, follows with 2,565,529,693 streams, an upbeat BPM of 186, and a unique blend of danceability and speechiness. "The Hills," a 2015 release, stands out with its distinctive low BPM of 136 and a moody ambiance. Other notable tracks include "Save Your Tears" (2020), "Die For You" (2016), and "Call Out My Name" (2018), each contributing to The Weeknd's widespread acclaim. The remix of "Save Your Tears" with Ariana Grande adds a collaborative touch, while newer releases like "Creepin'" (2022) and "After Hours" (2020)

showcase The Weeknd's evolving musical journey with varying BPMs and emotional tones.

Track name	streams	Artists count	released_year	bpm	danceability_%	valence_%	energy_%	acousticness_%	liveness_%	speechiness_%
Blinding Lights	3,703,895,074	1	2019	171	50	38	80	0	9	7
Starboy	2,565,529,693	2	2016	186	68	49	59	16	13	28
The Hills	1,947,371,785	1	2015	136	36	12	57	9	14	8
Die For You	1,647,990,401	1	2016	134	59	51	52	9	15	7
Save Your Tears	1,591,223,784	1	2020	118	68	61	82	2	50	3
Call Out My Name	1,449,799,467	1	2018	134	45	17	60	21	33	4
Save Your Tears (with Ariana Grande) (Remix)	1,221,813,483	2	2020	118	65	63	79	3	10	3
Creepin'	843,957,510	3	2022	98	71	17	61	36	8	5
I Was Never There	705,469,769	2	2018	114	32	17	74	14	17	3
After Hours	698,086,140	1	2020	109	66	16	57	10	12	3

Figure 14: Top 10 The Weeknd songs

Bad Bunny's top 10 songs present a vibrant mix of musical flavors and emotions. "Dakiti" leads with 1,763,363,713 streams, featuring a collaboration with another artist, a moderate BPM of 110, and a danceable rhythm.

Track name	streams	Artists count	released_year	bpm	danceability_%	valence_%	energy_%	acousticness_%	liveness_%	speechiness_%
Dakiti	1,763,363,713	2	2020	110	73	14	57	40	11	5
Me Porto Bonito	1,440,757,818	2	2022	92	91	43	71	9	9	8
LA CANCIÓN	1,435,127,549	2	2019	176	75	43	65	15	11	32
Callaita	1,304,313,953	2	2019	176	61	24	62	60	24	31
Titi Me Pregunté	1,264,310,836	1	2022	107	65	19	72	10	13	25
Yonaguni	1,260,594,497	1	2021	180	64	44	65	28	14	12
Efecto	1,047,480,053	1	2022	98	80	23	48	14	6	5
Moscow Mule	909,001,996	1	2022	100	80	29	67	29	12	3
Lo Siento BB:/ (with Bad Bunny & Julieta Venegas)	775,542,072	3	2021	170	64	14	70	9	9	8
La Santa	759,208,783	2	2020	93	74	59	87	3	8	5

Figure 15: Top 10 Bad Bunny songs

"Me Porto Bonito" (2022) follows closely with 1,440,757,818 streams, showcasing a slower tempo, high danceability, and positive valence. The energetic "LA CANCIÓN" (2019) with a BPM of 176 stands out for its danceable and speech-heavy composition. Other notable tracks include "Callaita" (2019) and "Yonaguni" (2021), each offering a unique blend of BPM, danceability, and acoustic elements. The collaborative effort "Lo Siento BB:/" featuring Julieta Venegas, adds diversity, while tracks like "Efecto" (2022) and "Moscow Mule" (2022) exhibit Bad Bunny's versatility with varying BPMs and danceability percentages. "La Santa" (2020) closes the list with a slower tempo, high danceability, and a positive valence, showcasing Bad Bunny's range in musical expression.

"The Weeknd" and "Bad Bunny" exhibit a consistent presence in the realm of top-streamed songs, showcasing their ability to consistently release popular and engaging music over the specified period. Both artists embrace diverse collaborations, as evidenced by the "artist_count" column, reflecting a strategic approach to attract a broader audience through joint projects. Notably, the release years of their top-streamed songs span various periods, indicating a timeless appeal that resonates with listeners over time. "Bad Bunny" further demonstrates a global approach through collaborations with artists like Dua Lipa and Julieta Venegas, contributing to a diverse and international fan base. The musical diversity of both artists is evident in their discography, encompassing a wide range of styles and genres that cater to different listener preferences, from energetic tracks to more melodic and introspective compositions. The engaging features such as "bpm," "danceability_%" , "valence_%" , and "energy_%" reveal a well-balanced musical approach, with "Bad Bunny" particularly standing out for high danceability and valence percentages, signaling music suitable for dancing and positive vibes.

Conclusion

In conclusion, our exploration of the patterns in audio features has provided valuable insights into the trends and preferences in popular songs on Spotify. The dataset, curated from the renowned music streaming platform, offered a comprehensive catalog of the most influential songs of 2023, allowing us to delve into multifaceted dimensions of each musical composition. Our analysis covered various aspects, including the top 10 most streamed songs, the least streamed songs, and the top 10 most streamed artists, shedding light on factors that contribute to the success or lesser popularity of songs in the dynamic music industry.

The analysis of the top 10 most streamed songs unveiled a combination of factors that contribute to their popularity. Time since release, artist recognition, and diverse musical elements such as Beats Per Minute (BPM), danceability, energy, and valence played crucial

roles. The presence of globally recognized artists like The Weeknd, Ed Sheeran, and Drake, alongside engaging musical features, contributed to the broad appeal and global success of these songs. The exploration of the top 10 least streamed songs highlighted the challenges faced by more recent releases or songs featuring less recognized artists. The varying musical characteristics reflected a range of elements influencing listener preferences, contributing to the lower streaming counts.

Our focus on the top 10 most streamed artists, "The Weeknd" and "Bad Bunny," uncovered consistent patterns in their music that resonated with a wide audience. Both artists demonstrated a diverse range of collaborations, spanning different release years and genres, showcasing their versatility. The global approach of "Bad Bunny" through collaborations with international artists contributed to a broader fan base. The engagement with various musical styles, coupled with balanced audio features like BPM, danceability, and valence, indicated a well-crafted and appealing musical approach. The dominance of "The Weeknd" and "Bad Bunny" in Spotify streams not only showcases their individual success but also culturally influences contemporary music trends.

The analysis of audio features like BPM, danceability, valence, and energy provided nuanced insights into their relationships and impact on the popularity of songs. For instance, a song's likelihood of success increased with a BPM in the range of 166-185, and danceability played a significant role in engaging listeners. The correlation matrix revealed interesting relationships, such as the negative correlation between energy and acousticness. The scatter plots further illustrated how certain audio features correlated with the number of streams, with speechiness and instrumentalness showing negative correlations.

Our methodology, driven by Python with Pandas, Matplotlib, Seaborn, and Scikit-Learn, enabled efficient data manipulation, exploration, and visualization. Data pre-processing steps, feature engineering, and statistical analysis contributed to a robust understanding of the dataset.