

Izvorni znanstveni rad  
UDK: 78:004.738.5  
15. V. 2025.

PETRA HABJANEC\*  
MARINA BAGIĆ BABAC\*\*

# QUANTITATIVE ANALYSIS OF SONGS WITH OVER A BILLION STREAMS

## ABSTRACT

This research paper conducts an in-depth quantitative analysis of songs surpassing one billion streams on Spotify to uncover trends and insights shaping the music industry. Utilizing statistical methods, including linear and polynomial regression, the study examines song attributes such as duration, energy, release year, popularity, explicitness, genres, and artist popularity. The findings reveal that billion-stream songs span diverse genres, with pop, rap, and hip hop being prominent, and highlight the role of collaborations, alongside trends in song duration, energy, and explicitness. By clarifying factors contributing to streaming success, this study offers insights into listener preferences and evolving music industry dynamics, informing marketing strategies, song creation, and playlist curation for artists, record labels, and streaming platforms.

**Keywords** – quantitative analysis, music streaming, songs, popularity, genres.

\* University of Zagreb, Zagreb, Croatia, [petra.habjanec@fer.hr](mailto:petra.habjanec@fer.hr)

\*\* University of Zagreb, Zagreb, Croatia, [marina.bagic@fer.hr](mailto:marina.bagic@fer.hr)

## INTRODUCTION

The popularisation of music streaming services has changed the way people consume music (Wu, 2019), giving people access to a vast library of songs at their fingertips (Vinnicombe and Sou, 2017). Among the platforms, Spotify has established itself as a global leader, providing a platform where both mainstream and smaller independent artists can show off their music (Hoskins and Griffin, 2023). In recent years, the streaming industry has reached such heights that the most popular songs are songs getting over a billion streams on Spotify (Coronel and Irimiás, 2023). The analysis of highly streamed songs on music streaming platforms has been a topic of interest for researchers and music industry professionals (Sweeney and Wyber, 2002).

Several studies have explored the musical characteristics of highly streamed songs (Weinberger and Bouhnik, 2021). Nijkamp (2018) analysed the 1000 songs from Spotify API from different genres using regression, to build a prediction model, and concluded that audio features from Spotify have little to moderate explanatory power for higher stream count. Also, (2018) analysed the audio features and genres of top-ranking songs on Spotify in 2017, using machine learning that predicts genres for songs based on the song's audio features. In addition, Nugroho et al. (2023) analysed the trend of Spotify API's song analysis of the most popular songs over the years of the COVID-19 pandemic, compared to before it.

The purpose of this research paper is to conduct an in-depth data analysis of songs that have achieved this significant milestone on Spotify. By examining various attributes and characteristics of these songs, we aim to gain insight into factors contributing to their popularity.

## RESEARCH FOCUS

In the analysis of highly streamed songs, we examined the time it takes for songs to reach a billion streams, studied popularity distribution,

and investigated collaborations among artists (Hutcherson and Haenfler, 2010). Additionally, we explored the sequencing of songs on albums, trends in song duration and energy, the ongoing popularity of associated artists, and the change in the explicitness of songs through time (Lam, 2011). With this analysis, we aim to contribute to a better understanding of the characteristics and dynamics of songs that have achieved tremendous streaming success on Spotify. The research is guided by the following questions:

- RQ1. Which songs through time got onto the list and how long it took them to get there?
- RQ2. Which record labels are the most present?
- RQ3. How do the songs with multiple artists compare to those with only one?
- RQ4. How has the average duration and energy of popular songs changed over time?
- RQ5. How popular are the artists currently?
- RQ6. How do the "one-hit wonders" compare to artists with multiple popular songs?
- RQ7. How has the explicitness of songs changed over time?
- RQ8. How many of the songs are singles compared to those that are on albums?
- RQ9. Which genres are the most prevalent?
- RQ10. Which artists are the most present and how many songs do they have on the list?

In this analysis, an array of different metrics about individual songs was extracted from Spotify API. Popularity, with the value from 0 to 100, which is mostly determined by algorithm based on the total number of plays and how recent the plays are. Acousticness, spanning from 0.0 to 1.0, is a confidence measure, offering insights into whether the song has electrical amplification. Danceability, in contrast, explains the rhythmic realm, accessing how danceable the track is depending on the tempo, rhythm stability, and beat strength, where 0.0 represents the least danceable and 1.0 the most danceable. Energy, measured on a scale of 0.0 (least energetic) to 1.0 (most energetic), provides a measure intensity and activity of the track. It considers the dynam-

ic range and perceived loudness, whereas energetic tracks are fast-paced, and loud and have dynamic musical elements. Instrumentalness indicates the likelihood of vocal content on the track. Values close to 1.0 signify a greater probability of instrumental dominance, distinguishing the track as predominantly devoid of vocals. The key is represented by integers following Pitch Class notation, while mode, indicates major (1) or minor (0). Liveness describes whether the track is a live performance, with values above 0.8, suggesting a strong likelihood of a live recording. Loudness, expressed in dB, provides an averaged perspective across the entire track. Speechiness indices the presence of spoken words, with values of 0.66 signalling predominantly spoken content. Tempo reflects the estimated beats per minute, defining the pace and rhythm of the track. Time signature, ranging from 3 to 7, estimates, the number of beats in each bar. Finally, valence, ranging from 0.0 and 1.0, paints how emotional the track is, conveying its musical positiveness. Higher valence indicates a more positive emotional tone. Together, these metrics will provide us with a good understanding of the auditory journey within a track.

## DESCRIPTIVE ANALYSIS OF HIGH-STREAMING SONGS

Our data was retrieved in November of 2023 from the Spotify playlist BILLION CLUB which contains all of the songs that have more than one million. At the time the playlist contained 509 songs. Our data set contains tracks with the timestamp of when they got added to the playlist, the name of the album the track is on, the release date of the album, total number of tracks on the album, genres, a list of the artists on the track, list of country codes where the track is available, duration of the track in ms, name of the song, the popularity of the song, track number, a measure of danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, and time signature. Based on this dataset, we aim to answer questions ( $RQ1-RQ1$ ) and test the hypotheses ( $H_a-H_h$ ).

The summary statistics of our data set is shown in Table 1. In the table, we can see that most of the song durations lie between 3mins 15s and 4mins. Where the 3 shortest ones are *Everybody Dies In Their Nightmares*, *Hope* and *Jocelyn Flores* all by XXXTENTACION, and the three longest ones are *Te Boté* by Nio Gacia, Casper Magico, Bad Bunny, Darell, Ozuna and Nicky Jam, *Hotel California - 2013 Remastered* by Eagles and *Nothing Else Matters(Remastered)* by Metallica. We can also see that the popularity of most of the songs lies between 78 and 86, with *Cruel Summer* by Taylor Swift, *Seven(feat. Latto) (Explicit Ver.)* by Jung Kook and Latto, and *I Wanna Be Yours* by Arctic Monkeys as the most popular, and *Bohemian Rhapsody - 2011 Remastered* by Queen, *Roar* by Katy Perry, and *Bad Romance* by Lady Gaga as least popular. For danceability majority of the tracks are situated between 0,578 and 0,755. The most danceable ones are *The Real Slim Shady* by Eminem, *WAP* by Cardi B and Mega Thee Stallion, *Another One Bites The Dust - Remastered 2011* by Queen, and with least danceable ones being *All I Want* by Kodakly, *Fix You* by Coldplay and *Dusk Till Dawn - Radio Edit* by ZAYN and Sia. The energy of the most popular songs falls between 0,533 and 0,773. Where *Welcome To The Jungle* by Guns N' Roses, *Hey Ya!* By Outkast, and *Promiscuous* by Nelly and Timbaland are the most energetic, and *when the party's over* by Billie Eilish, *Say Something* by A Great Big World and Christina Aguilera, and *White Christmas -1947 Version* by Bing Crosby, Ken Darby Singers, and John Scott Trotter & His Orchestra as least energetic. For key most of the songs are between D and G#. Loudness of most streamed songs falls between -7,463 and -6,283. *Africa* by Toto, *White Christmas - 1947 Version* by Bing Crosby, Ken Darby Singers, and John Scott Trotter & His Orchestra, and *Revenge* by XXXTENTACION stand out as the quietest, and *Hey Ya!* by Outkast, *FRIENDS* by Marshmellow and Anne-Marie, and *What Makes You Beautiful* by One Direction as loudest. For speechiness majority of the songs are situated between 0,038 and 0,090. Some of the most speechy songs are *Life Is Good* by Future and Drake, *Youngblood* by 5 Seconds of Summer, and *Panda* by Desiigner, and some of the least speechy songs are *Perfect*

by Ed Sheeran, *The Scientist* by Coldplay, and *Set Fire to the Rain* by Adele. In the table, we can see that the acousticness mostly falls between 0,035 and 0,310. *Smells Like Teen Spirit* by Nirvana, *Thunderstruck* by AC/DC, *Come As You Are* by Nirvana being the least acoustic, *when the party's over* by Billie Eilish, *The Night We Met* by Lord Huron, and *Bruises* by Lewis Capaldi being most acoustic. For instrumentalness, most of the values fall into 0,000, which means not instrumental at all. The ones with the greatest values of instrumentalness are *everything i wanted* by Billie Eilish, *White Christmas - 1947 Version* by Bing Crosby, Ken Darby Singers, John Scott Trotter & His Orchestra, and *Better* by Khalid. Values of liveness lie between 0,093 and 0,193. The tracks with the lowest values of liveness are *Flowers* by Miley Cyrus, *Uptown Funk* by Mark Ronson and Bruno Mars, and *Cake By The Ocean* by DNCE, and the tracks with highest values of liveness *Rap God* by Eminem, *Dancing Queen* by ABBA, and *The Box* by Roddy Ricch where their values of liveness are less than 0,799 which is less than 0,8 what is a Spotify defined limit of track actually being live. The valence of most of the tracks falls between 0,327 and 0,665. With *September* by Earth, Wind & Fire, *There's Nothing Holdin' Me Back* by Shawn Mendes, *Pumped Up Kicks* by Foster The People having the highest value of valence, and *Falling* by Harry Styles, *HIGHEST*

*IN THE ROOM* by Travis Scott, and *Lose Yourself* by Eminem having the lowest value. The tempo of most of the tracks is situated between 98,007 and 136,041. But the ones that pop out are *changes* by XXXTENTACION, *Make You Feel My Love* by Adele, and *I Wanna Be Yours* by Arctic Monkeys with the lowest values, and *FourFiveSeconds* by Rihanna, Kanye West, and Paul McCartney, *Animals* by Maroon 5, and *Back In Black* by AC/DC with highest values.

It should be noted that in this analysis the Spotify API does not provide a release date for a specific song, so the release date of an album is used as also the release date of a song. We also got the data for all the artists and albums for songs. From there we got the data for labels, popularity of artists, and genres. Since the data was taken from the playlist, which was made on the 21st of July 2021, for the dates before that we can't know when the songs achieved a billion streams, therefore those are excluded from any analysis including the time of when a billion streams was achieved.

To address RQ1, we analysed the temporal distribution of songs reaching one billion streams, focusing on the frequency of new entries over time. Starting from October 2021, approximately ten songs have been added to the list each month. Notably, this growth rate has accelerat-

Table 1: Summary statistics of the dataset

Variable	Min.	1st Quartile	Median	Mean	3rd Quartile	Max
Duration(in ms)	95466,000	195373,000	19573,000	219344,306	241693,000	417920,000
Popularity	45,000	78,000	83,000	81,782	86,000	99000,000
Danceability	0,188	0,578	0,674	0,660	0,755	0,949
Energy	0,111	0,533	0,662	0,642	0,773	0,987
Key	0,000	2,000	5,000	5,267	8,000	11,000
Loudness	-18,064	-7,463	-5,866	-6,283	-4,662	-2,261
Speechiness	0,023	0,038	0,054	0,090	0,099	0,481
Acousticness	0,000	0,035	0,117	0,219	0,310	0,978
Instrumentalness	0,000	0,000	0,000	0,008	0,000	0,657
Liveness	0,023	0,093	0,116	0,169	0,193	0,799
Valence	0,059	0,327	0,479	0,498	0,665	0,979
Tempo	64,934	98,007	117,996	118,961	136,041	205,846
Time signature	1,000	4,000	4,000	3,961	4,000	5,000

Table 2: Songs how have a shortest time from release date and reaching billion streams

Name	Artist	Added at	Album release date	Difference in days
Flowers	Miley Cyrus	2023-05-04	2023-08-18	-106
MONTERO (Call Me By Your Name)	Lil Nas X	2021-10-04	2021-09-17	17
As It Was	Harry Styles	2022-07-28	2022-05-20	69
drivers license	Olivia Rodrigo	2021-08-11	2021-05-21	82
Seven (feat. Latto) (Explicit Ver.)	Jung Kook, Latto	2023-10-30	2023-07-14	108
STAY (with Justin Bieber)	The Kid LAROI, Justin Bieber	2021-11-04	2021-07-09	118
Bad Habits	Ed Sheeran	2022-03-31	2021-10-29	153
good 4 u	Olivia Rodrigo	2021-10-26	2021-05-21	158
Kill Bill	SZA	2023-06-05	2022-12-09	178
Enemy (with JID) - from the series Arcane League of Legends	Imagine Dragons, JID, Arcane, League of Legends	2022-12-27	2022-07-01	179

ed in recent months, suggesting an increasing trend in the production or consumption of globally popular music.

We have also analysed how long it takes for songs to get the achievement. Out of 349 songs (the ones that got a billion streams after the 21st of July 2021). On average it took 3,913.994 days (10 years and 264 days) for songs to reach the achievement but a standard deviation of 4,496.741 (12 years and 116 days). The median stands at 2,268 days (6 years and 78 days), 1st quartile at 1,214 (3 years and 119 days) and 3rd quartile at 4,427 (12 years and 47 days).

Table 2 presents the outliers with the shortest time span between the album release date and the point at which the song reached one billion streams on Spotify. Most notable is Flowers with a negative value, but as mentioned before we are looking at album release dates, not song release dates, and nowadays artists release one or a few songs from the album before the release of the actual album. So, Miley Cyrus' Flowers reached a billion streams even before the release of the album it's on. It should also be noted that Olivia Rodrigo has two songs with some of the shortest times between release date and reaching billion streams. BBoth songs, *drivers licence* and *good 4*

*you* are on the same album *SOUR*. It is also interesting that only two songs that reached a billion streams in 2023 are also on the list, those being *Flowers* by Miley Cyrus and *Seven(feat. Latto)(Explicit Ver.)* by Jung Kook and Latto.

Table 3 displays the songs with the longest duration between their album release and the point at which they surpassed one billion streams on Spotify. It is fascinating how the oldest song with a billion streams is a Christmas song from the 40's, and that the song achieved a billion streams just in the begging of autumn. We can also see how the older songs that have achieved the milestone are mostly from 'classic' bands or artists like Queen, Fleetwood Mac, ABBA, The Beatles, and The Rolling Stones. The performers we maybe don't listen to usually but still know either their names or even songs. Those are the artists and songs that have become timeless and established themselves as everlasting.

Our analysis of songs getting over a billion streams on Spotify reveals evolving trends in music releases and highlights the enduring appeal of timeless classics. The monthly distribution showcases a recent surge in additions. Notable outliers, including pre-album releases and entries from 2023, add nuance to our findings.

Table 3: Songs how have a longest time from release date and reaching billion streams

Name	Artist	Added at	Album release date	Difference in days
We Will Rock You - Remastered 2011	Queen	2023-02-24	1977-10-28	16555
Dreams - 2004 Remaster	Fleetwood Mac	2022-06-27	1977-02-04	16579
Dancing Queen	ABBA	2023-07-11	1976-10-11	17074
Sweet Home Alabama	Lynyrd Skynyrd	2022-11-07	1974-04-15	17738
Have You Ever Seen The Rain	Creedence Clearwater Revival	2023-03-02	1970-12-07	19078
Fortunate Son	Creedence Clearwater Revival	2023-06-13	1969-11-02	19581
Here Comes The Sun - Remastered 2009	The Beatles	2023-05-09	1969-09-26	19583
Ain't No Mountain High Enough	Marvin Gaye, Tammi Terrell	2023-05-17	1967-08-29	20350
Paint It, Black	The Rolling Stones	2023-10-19	1966-04-15	21006
White Christmas - 1947 Version	Bing Crosby, Ken Darby Singers, John Scott Trotter & His Orchestra	2023-09-27	1942-01-01	29854

QUANTITATIVE ANALYSIS OF SONGS WITH OVER A BILLION STREAMS

To address RQ2, we identified the most prominent record labels based on the number of songs surpassing one billion streams on Spotify. Columbia Records leads by a substantial margin, followed by Atlantic Records, Atlantic Records UK, and Republic Records—each contributing approximately half as many songs to the list. To address RQ3, we investigate the prevalence of songs featuring multiple artists and compares

their popularity to songs performed by a single artist. The comparison of mean popularity between the two groups over time is shown in Figure 1. The analysis reveals that, from 2010 onwards, songs by a single artist have generally been more popular. The largest observed difference occurred in 2019, where single-artist songs exceeded their multi-artist counterparts in average popularity by 7.004 points. In contrast, the

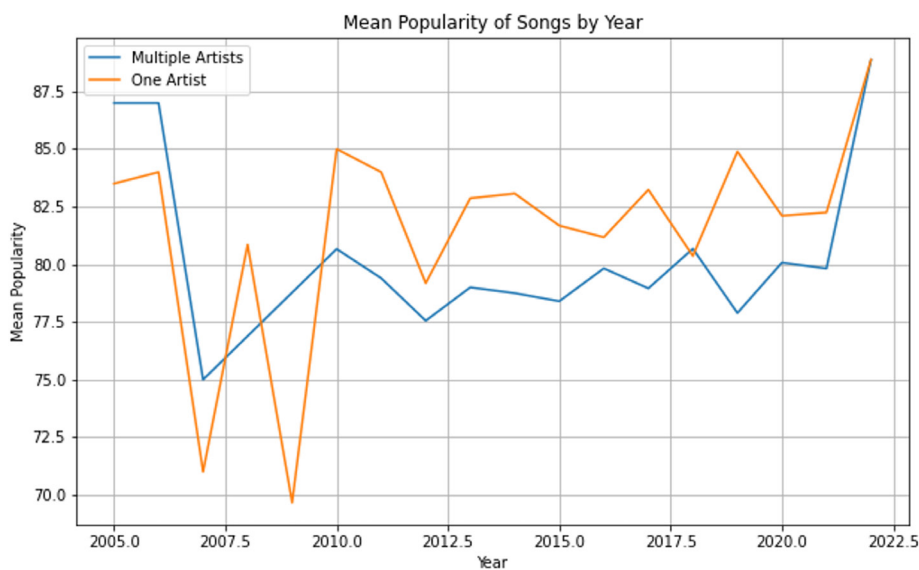


Figure 1: Mean popularity of songs with multiple artists and one artist

smallest difference was recorded in 2018, with multi-artist songs showing a slight advantage of 0.315 points.

These findings challenge the intuitive assumption that collaborations would naturally lead to higher popularity due to combined fan bases and broader exposure (Šimić and Bađić Babac, 2024). One possible explanation is that collaborations between artists from different genres may produce songs that do not strongly resonate with either audience, thus diluting their appeal. It is also important to note that the year 2023 was excluded from the analysis due to the limited number of qualifying songs (only two) released that year.

### INFERENCE STATISTICAL ANALYSIS OF SONG CHARACTERISTICS

To assess whether there are statistically significant differences in song characteristics depending on the number of contributing artists (Ivezić and Bađić Babac, 2023), we conducted a series of non-parametric tests (Mattsson et al., 2010). Specifically, we applied the Mann–Whitney U test to compare the distributions of key audio and popularity features between songs performed by a single artist and those featuring multiple artists (Negro et al., 2022; Santini, 2011). The following null hypotheses ( $H_a - H_b$ ) were tested:

- **Ha:** There is no significant difference in the distribution of the duration of songs with multiple artists and songs with only one artist.

- **Hb:** There is no significant difference in the distribution of popularity of songs with multiple artists and songs with only one artist.
- **Hc:** There is no significant difference in distribution of danceability of songs with multiple artists and songs with only one artist.
- **Hd:** There is no significant difference in the distribution of energy of songs with multiple artists and songs with only one artist.
- **He:** There is no significant difference in the distribution of loudness of songs with multiple artists and songs with only one artist.
- **Hf:** There is no significant difference in the distribution of speechiness of songs with multiple artists and songs with only one artist.
- **Hg:** There is no significant difference in the distribution of valence of songs with multiple artists and songs with only one artist.

The test results are summarised in Table 4. For duration and valence, we failed to reject the null hypotheses, suggesting no significant differences in distribution between the two groups. However, for all other features—popularity, danceability, energy, loudness, and speechiness—the null hypotheses were rejected, indicating that the presence of multiple artists does have a statistically significant effect on these song characteristics.

To further investigate the positioning of successful tracks within albums, we isolated the subset of 414 songs that are part of album releases. For each track, we calculated the relative position by

Table 4. Results of Mann–Whitney U Tests Comparing Songs with Single vs. Multiple Artists

Variable	U statistic	p-value	Decision	Interpretation
Duration ( $H_a$ )	26,847.5	0.209	Confirmed	Same dispersion
Popularity ( $H_b$ )	20,490.0	$9.90 \times 10^{-8}$	Rejected	Single-artist tracks more popular
Danceability ( $H_c$ )	35,433.0	$2.35 \times 10^{-5}$	Rejected	Distributions differ
Energy ( $H_d$ )	32,684.5	0.013	Rejected	Distributions differ
Loudness ( $H_e$ )	33,519.5	0.003	Rejected	Distributions differ
Speechiness ( $H_f$ )	35,040.0	$6.97 \times 10^{-5}$	Rejected	Distributions differ
Valence ( $H_g$ )	31,794.0	0.057	Confirmed	Same dispersion

dividing its track number by the total number of tracks on the album. The resulting distribution shows a clear skew toward early positions: more than half of the billion-stream songs are located within the first third of their respective albums. The median track ratio is 0.300, with the first quartile at 0.154 and the third at 0.571.

To statistically evaluate this observation, we tested the following hypothesis:  
 $H_h$ : The average track ratio is equal to or greater than 1/3.

Using a one-sample t-test, we obtained a test statistic of 3.597 and a p-value of 0.00036 (df = 413). At the standard significance level ( $\alpha = 0.05$ ), the null hypothesis was rejected. We therefore conclude that billion-stream songs tend to appear significantly earlier than one-third of the way through an album. This result supports the notion that track sequencing plays a role in streaming success, possibly due to higher exposure of early tracks.

In answer to RQ4, we examine temporal trends in two key audio features: song duration and energy. To explore how these characteristics have evolved over time, we applied regression analysis to aggregated yearly averages.

A linear regression model was applied to examine how release year predicts song duration. The analysis, visualised in Figure 2, reveals a significant downward trend: on average, song duration decreases by 2,576.37 milliseconds per year. The predictor variable in the model is the release year, while the dependent variable is the average duration of songs.

The model yielded a coefficient of determination ( $R^2$ ) of 0.487, meaning that 48.7% of the variance in song duration is explained by the year of release. This indicates a moderately strong relationship, although it also implies that other unobserved factors account for the remaining 51.3% and should be considered in future research. The downward trend is statistically meaningful and reflects a broader industry shift toward shorter track formats over time.

To model the trajectory of average energy across time, we fitted a 4th-degree polynomial regression, whose coefficients are: [5.64040642e-06, -4.52949648e-02, 1.36401309e+02, -1.82558675e+05, 9.16253299e+07]. This model achieves an  $R^2$  value of 0.595, meaning it explains nearly 60% of the observed variation in energy levels. As shown in Figure 3, the average energy of songs declined steadily until 2018, which marks the local minimum (0.605). From

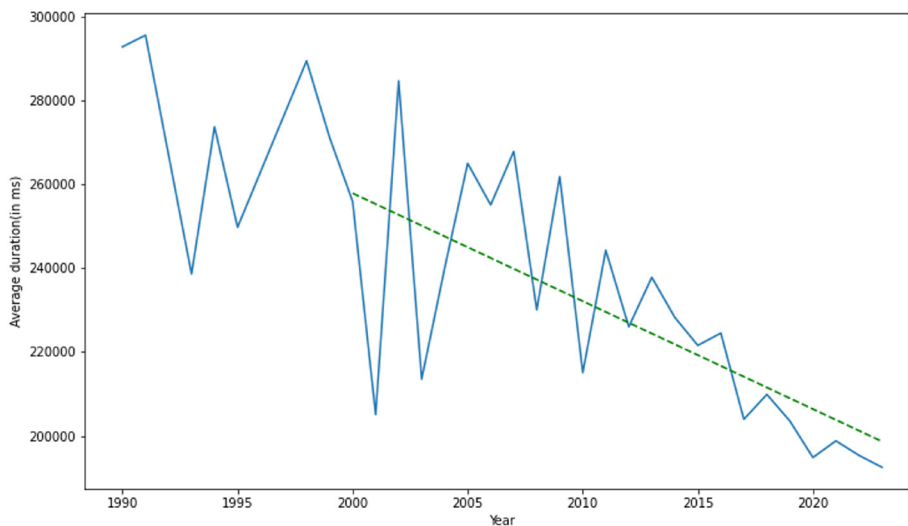


Figure 2: Trend of average duration of songs through years, with a linear regression line

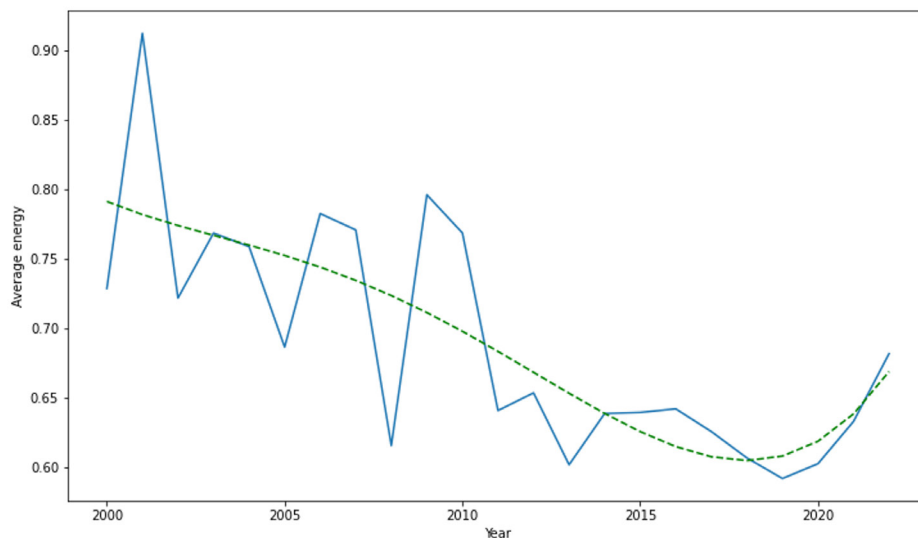


Figure 3: Trend of average energy of songs through years, with a linear regression line

that point onward, energy values exhibit an upward trend, suggesting a shift in musical production or listener preference.

These results indicate that, over the past two decades, songs have become progressively shorter and experienced a non-linear change in energy, with recent years showing a rebound in intensity.

In answer to RQ6, we analysed of artists' popularity, distinguishing between "one-hit wonders" and those with multiple hit songs. For "one-hit

wonders", having 248 artists, the mean popularity score stands at approximately 70.83, with a standard deviation of 6.72. The range extends from a minimum of 56 to a maximum of 89, and the interquartile range (IQR) spans from the 25th percentile (66) to the 75th percentile (75). In contrast, artists with more than one hit song, totalling 125 artists, exhibit a notably higher mean popularity of around 80.15, accompanied by a standard deviation of 6.02. The popularity scores for this category range from a minimum of 63 to a maximum of 100, with an IQR between the 25th percentile (76) and the 75th percentile

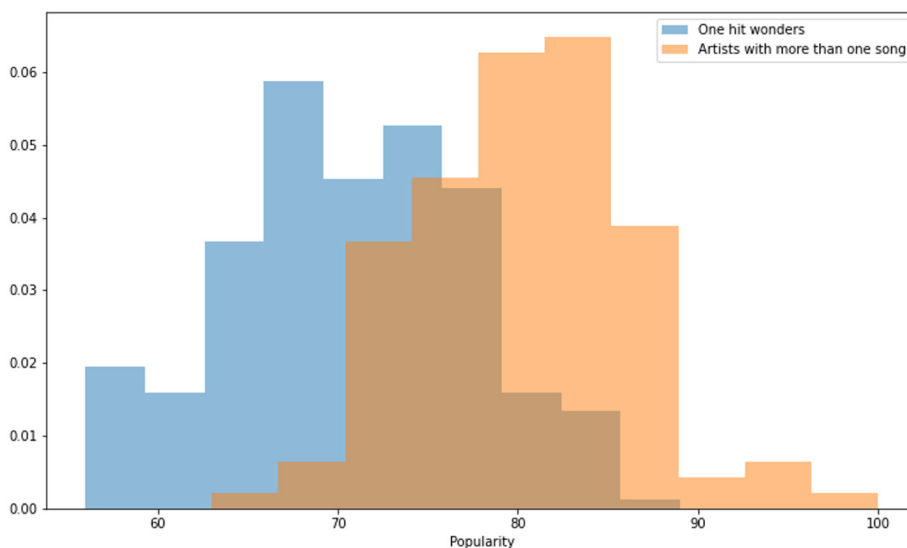


Figure 4: Popularity distribution comparison between "one-hit wonders" and multi-hit artists

(84). This analysis, coupled with visualisation in the plot in Figure 4, provides insights into the distinct popularity patterns between “one-hit wonders” and artists with lasting success, helping us understand the dynamics of longevity in the music industry.

To answer RQ7, we took a deeper look into involving prevalence of explicit songs over the years, examining the percentage representation of explicit and non-explicit content within the most streamed songs. The data reveals fluctuating trends, providing a view into social attitudes and musical preferences. In the early 2000s, explicit songs held a relatively lower percentage, with notable increases in the following years. The explicit percentage peaked in 2021 at 61.29%, signifying a potential shift in the acceptance or production of explicit content. On the other hand, the percentage of non-explicit songs displayed a dynamic pattern, reaching its peak in 2006, 2008 and 2009 at 100%, suggesting a contrasting trend.

Taking a closer look into the yearly production of songs on albums versus the single (RQ8), the data reveals a dynamic pattern in the number of songs from albums, hitting its peak in 2018 and 2017 with 47 songs released each year. Yet, the count of singles demonstrates more variability, reaching its peak at 15 in 2017. To better under-

stand these trends from 2000 to 2022, refer to Figure 5.

Genres, as defined by Spotify for songs with over a billion streams, give us a peek into the different kinds of popular music (RQ9). Figure 6 shows that pop is the most prevalent genre, with a lead of 298 songs falling into this category. Right behind it, we have genres like rap, dance pop, and hip hop, all contributing to the mix. There are some more specific genres like Canadian contemporary R&B, Miami hip-hop, and Barbadian pop, but also some main ones like rock, pop, rap, and hip-hop.

Then, we decided to look at the presence of ‘main’ genres, so we counted the presence of pop for example as well as pop dance, and dance pop, for hip-hop we counted in Miami hip-hop, etc. That change is shown in Figure 7, where pop is even more prevalent with 385 songs. Here, we can see metal, soul, k-pop, country, grunge, and indie, which were not present before. Trap has surpassed EDM, with 65 songs, while EDM has only 39. In addition, we can explore the diversity of the music on the streaming platform and different tastes, that all combine into a unique worldwide music scene.

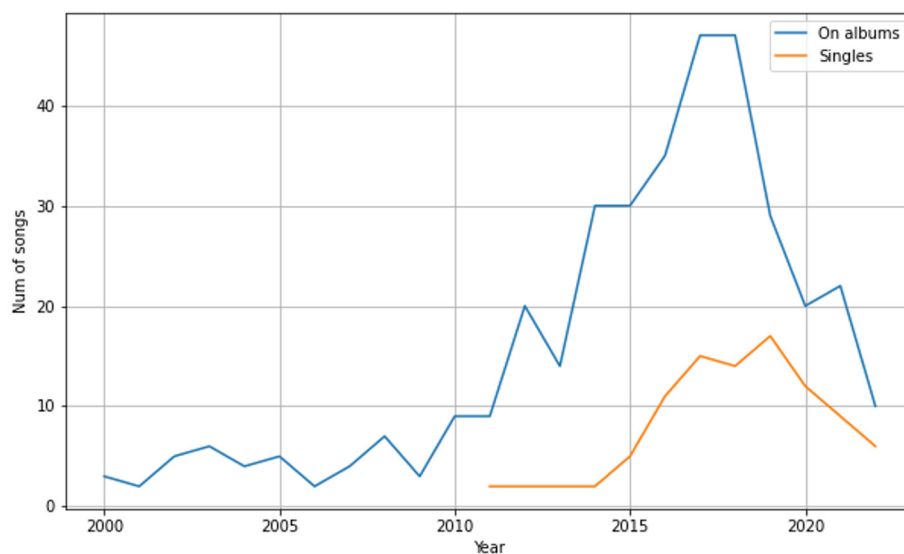


Figure 5: Count of singles and songs from albums released through years

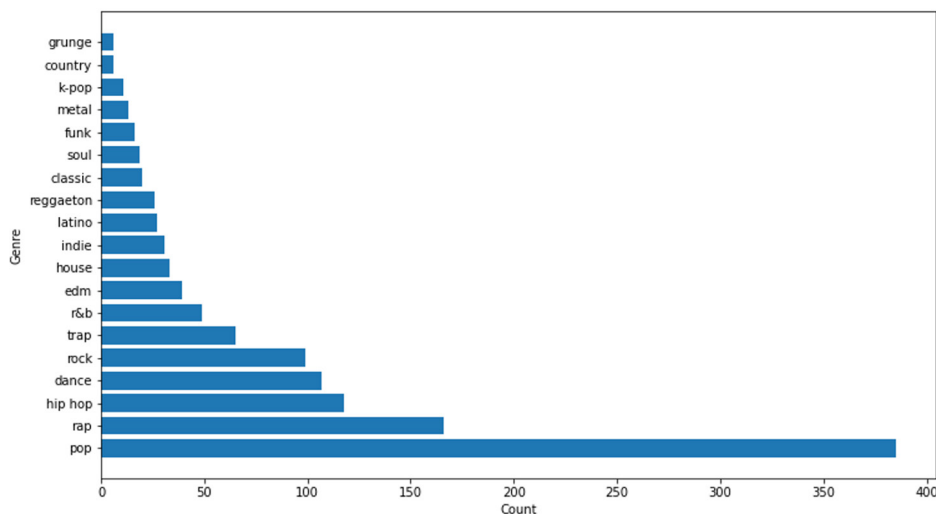


Figure 6: Top 30 Spotify genres and their count

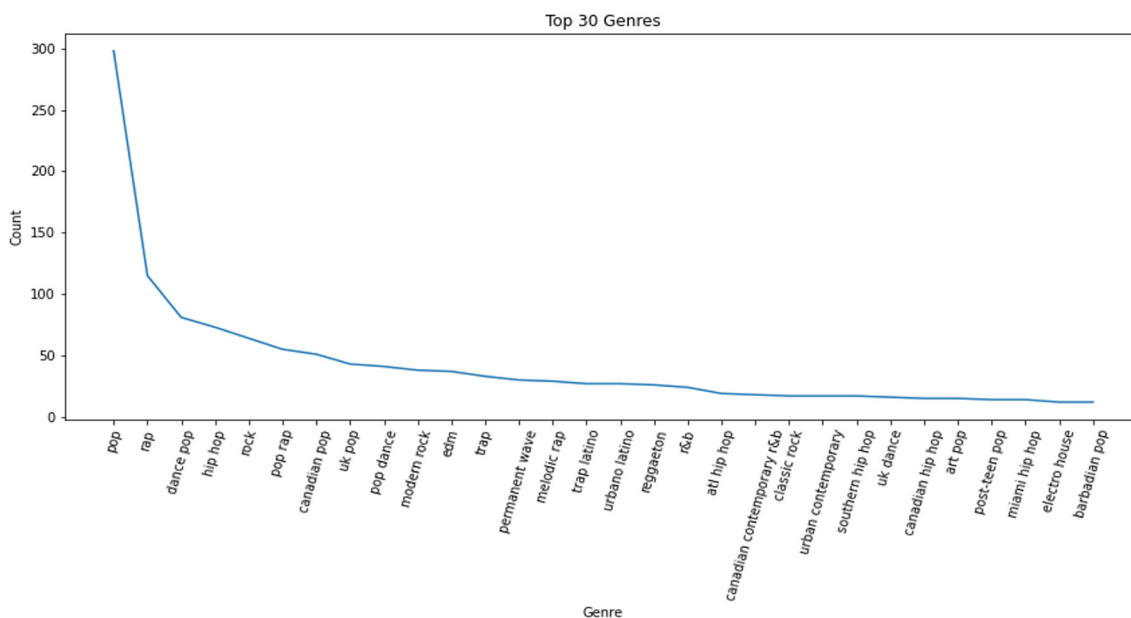


Figure 7: Most prevalent 'main' genres and their count

Figure 8 shows the artists who have achieved remarkable success in the realm of billion-stream hits on Spotify (*RQ10*), showcasing the most prolific contributors to this exclusive club. Justin Bieber leads with 14 songs that have surpassed a billion streams, followed closely by Drake with 13, Rihanna and Bad Bunny with 12 each, and Ariana Grande and The Weeknd with 11 each. The

list further includes renowned artists such as Ed Sheeran, Post Malone, Bruno Mars, and XXXTENTACION, each boasting 10 songs that have garnered over a billion streams. Diverse in genres and styles, these artists have left a significant mark on the global music scene, attaining widespread popularity and recognition for their contributions to the billion-stream club on Spotify.

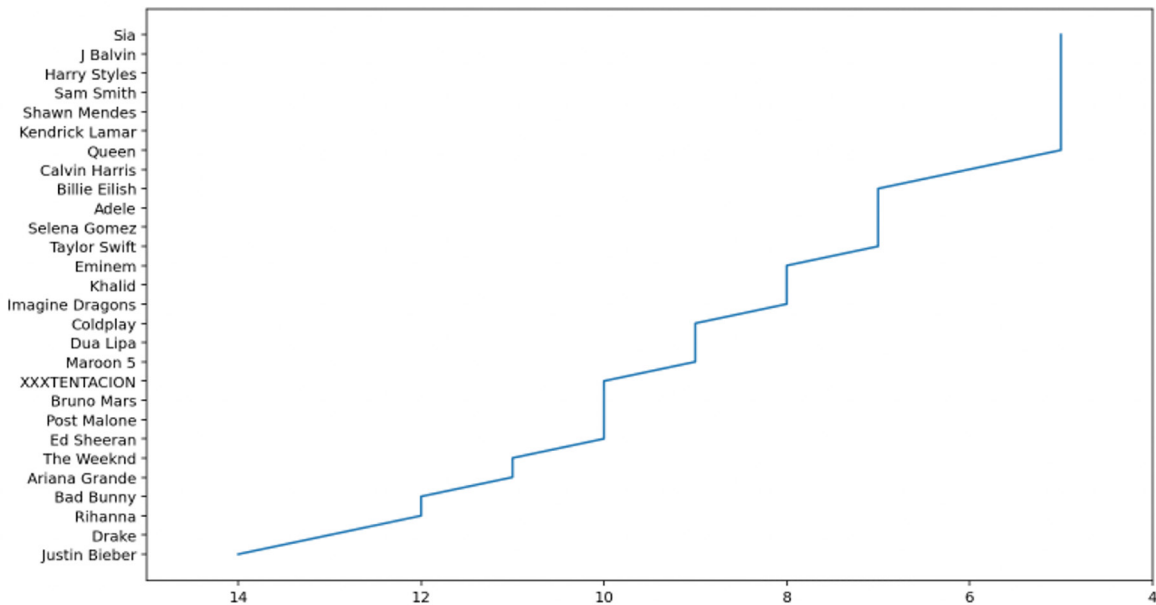


Figure 8: Most prevalent artists and the number of their songs with over billion streams(with more than 5 songs)

In answer to RQ5, we present a statistical description of the popularity of artists based on a dataset of popularity scores. The data reveals a diverse distribution, with a mean popularity score of approximately 73.95. Figure 9 represents the spread of artist popularity, showcasing how scores are distributed across the

dataset. The minimum popularity is 56, the maximum is 100, and most artists fall within the interquartile range (IQR) between the 25th and 75th percentiles. The standard deviation of 7.85 indicates a moderate level of variability around the mean.

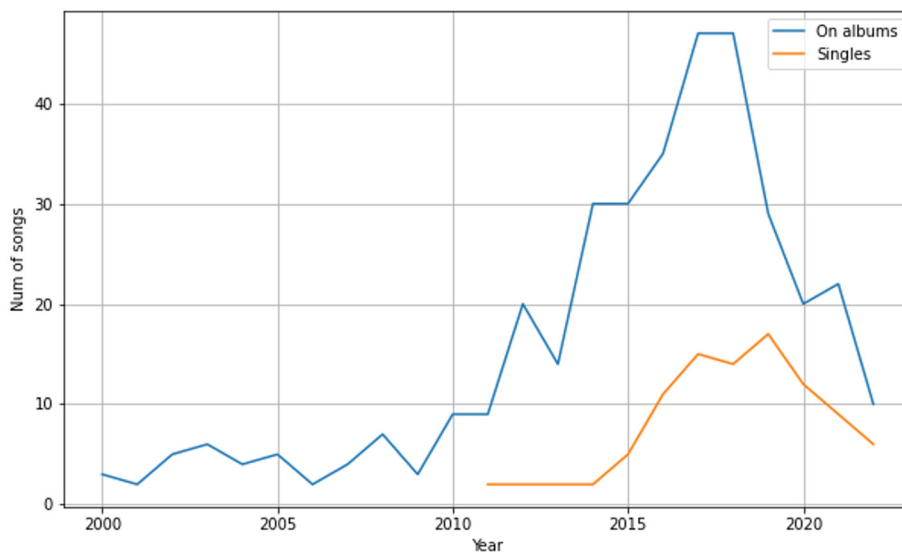


Figure 9: Popularity distribution of artists

Some of the most popular artists with songs streamed more than a billion times include Taylor Swift(100), Drake(95), Bad Bunny(95), The Weekend(93), Travis Scott(90), contrary to some of the least popular artists are Jack Ü(56), Bipolar Sunshine(57), Nyla(57), Jawsh 685(57) and WATT(57).

Through our data analysis, we have researched several questions to gain a deeper understanding of songs that have achieved over a billion streams on Spotify. We examined the time it took for songs to reach a billion streams and identified the distribution of songs over months and years. The average time it takes to get a billion streams is 10 years and 264 days, but the median stands at 6 years and 78 days. There are many 'outliers' in the form of older songs which make the distribution skew to the right. We also analysed the record labels with the most songs that have achieved over a billion streams. Columbia Records emerging as the front-runner, highlighting its significance in the music industry.

Contrary to popular belief (Šandor and Bagić Babac, 2024), we concluded that songs with only one artist tend to be more popular than songs with multiple artists. We also concluded that songs with multiple artists and songs with only one have no significant difference in the distribution of their durations and valence, but there is a significant difference in the distribution of popularity, danceability, energy, loudness, and speechiness. Furthermore, we have seen a trend of decreasing average song duration over time. We also observed variations in the average energy levels of songs, with a decrease until 2018, and an uprise in recent years. Additionally, we examined that artists with more than one hit song tend to have a higher mean popularity score than "one-hit wonders", suggesting a lasting impact and longevity in the music industry. The percentage of explicit songs has increased over time, reflecting changing social attitudes. By exploring the genre presence in the "BILLIONS CLUB" playlist we can see pop as the most prevalent genre, followed by rap, dance pop, and hip hop. The analysis of "main" genres further highlights the prevalence of pop.

Moreover, we've seen that Justin Bieber leads the list of artists with the most songs surpassing a billion streams, closely followed by Drake, Rihanna, and Bad Bunny. The current popularity of artists is diverse with a mean of 73.95. Some of the most popular are Taylor Swift, Drake, and Bad Bunny, while some of the least popular ones are Jack Ü, Bipolar Sunshine, and Nyla.

We also analysed the most popular words in song titles, with them indicating the presence of a lot of features on songs and the presence of modified and updated versions of songs (Bagić Babac, 2023). Some words suggest themes of conveying personal experiences and emotions through songs (Hitl et al, 2025). Our analysis provides valuable insights into the enduring appeal of timeless classics, the impact of collaborations on popularity, and the diverse representation of genres among highly streamed songs.

## CONCLUSION

Our study challenges common beliefs in the music industry by offering new insights into the journey of songs that reach a billion streams on Spotify. The findings contribute to a clearer understanding of the temporal dynamics of music popularity (RQ1), showing that achieving such a milestone typically takes much longer than widely assumed. The dominance of Columbia Records (RQ2) highlights the continuing influence of major labels in the digital age, challenging narratives of decentralised and democratised music distribution. The observed popularity of solo artist performances (RQ3), supported by hypothesis testing ( $H_b$ ), contradicts prevailing assumptions about the inherent advantage of collaborations in achieving mainstream success. Moreover, the analysis of musical attributes - including duration, energy, danceability, loudness, speechiness, and valence - through a series of hypothesis tests ( $H_a-H_g$ , RQ4) provides a more detailed view of how sonic features contribute to audience engagement. These results offer meaningful implications for music producers and marketers seeking to align with evolving listener preferences.

The study further shows that artists with multiple high-streaming songs tend to be significantly more popular than “one-hit wonders” (RQ5, RQ6), reinforcing the value of sustained artistic presence. Trends in explicit content (RQ7) suggest shifting audience norms, while the positioning of songs within albums (RQ8), supported by hypothesis Hh, confirms the strategic importance of track sequencing. Genre analysis (RQ9) affirms the continued prevalence of pop while also pointing to diversification through emerging subgenres. Finally, artist-level analysis (RQ10) reveals that a small group of performers dominate the billion-stream space, underlining the concentration of success within a relatively narrow field.

While our study provides valuable insights, it is limited to songs that have surpassed a billion streams and thus may not capture broader music trends. Future research could extend the scope to include songs with lower stream counts, offering a more complete picture of music consumption (Askin and Mol, 2018; Grgurević and Bagić Babac, 2025; Brzić et al., 2023, Šporčić and Pešek, 2020). Additionally, integrating qualitative methods (Gegač et al., 2025) – such as listener interviews or surveys—could deepen our understanding of the emotional and psychological dimensions of music popularity.

## REFERENCES

- Askin, N. and Mol, J. (2018), "Institutionalizing Authenticity in the Digitized World of Music", Jones, C. and Maoret, M. (Ed.) *Frontiers of Creative Industries: Exploring Structural and Categorical Dynamics (Research in the Sociology of Organizations, Vol. 55)*, Emerald Publishing Limited, Leeds, pp. 159-202.
- Bagić Babac, M. (2023), "Emotion analysis of user reactions to online news", *Information Discovery and Delivery*, Vol. 51 No. 2, pp. 179-193. <https://doi.org/10.1108/IDD-04-2022-0027>
- Bagić Babac, M. and Podobnik, V. (2016), "A sentiment analysis of who participates, how and why, at social media sport websites: How differently men and women write about football", *Online information Review*, Vol. 40, No. 6, pp. 814-833. <https://doi.org/10.1108/OIR-02-2016-0050>
- Brzić, B., Botički, I. and Bagić Babac, M. (2023), "Detecting Deception Using Natural Language Processing and Machine Learning in Datasets on COVID-19 and Climate Change // Algorithms", Vol. 16, No. 221. <https://doi.org/10.3390/a16050221>
- Coronel, M. and Irimiás, A. (2023), "“A very special song from Queen to you!” The role of music in destination promotional videos", *Consumer Behavior in Tourism and Hospitality*, Vol. 18 No. 1, pp. 110-124. <https://doi.org/10.1108/CBTH-11-2021-0265>
- Draganova, A. (2019), "The Soundtrack of Transition: Exploring the Cultural Politics of Rock and Pop-folk", *Popular Music in Contemporary Bulgaria: At the Crossroads (Emerald Studies in Alternativity and Marginalization)*, Emerald Publishing Limited, Leeds, pp. 99-137.
- Gegač, J., Greb, N. and Bagić Babac, M. (2025), "Exploring the VIA classification of human strengths and virtues using topic modeling", *Information Discovery and Delivery*, <https://doi.org/10.1108/IDD-10-2023-0116>
- Grgurević, Ante; Bagić Babac, Marina (2025), "Transformer based approach for solving mathematical problems using automatic speech recognition", *IEEE Access*, Vol. 13, pp. 79845-79859, doi: 10.1109/ACCESS.2025.3564121.
- Hitl, M., Greb, N. and Bagić Babac, M. (2025), "Quantitative analysis of the relationship between expressing gratitude and forgiveness and user sentiment on social media", *Global Knowledge, Memory and Communication*, Vol. 74 No. 11, pp. 42-62. <https://doi.org/10.1108/GKMC-03-2024-0138>
- Hoskins, J.D. and Griffin, A. (2023), "Market selection and product positioning decisions' implications for short- and long-term performance: evidence from the US music industry", *Journal of Product & Brand Management*, Vol. 32 No. 4, pp. 582-599. <https://doi.org/10.1108/JPBM-01-2022-3835>
- Hutcherson, B. and Haenfler, R. (2010), "Musical genre as a gendered process: Authenticity in extreme metal", Denzin, N.K. (Ed.) *Studies in Symbolic Interaction (Studies in Symbolic Interaction, Vol. 35)*, Emerald Group Publishing Limited, Leeds, pp. 101-121.

- Ivezić, D. Bagić Babac, M. (2023), Trends and Challenges of Text-to-Figure Generation: Sustainability Perspective, *Croatian Regional Development Journal*, Vol. 3, No. 1, <https://hrcak.srce.hr/file/448285>
- Lam, M. (2011), "Towards a "musicianship model" for music knowledge organization", *OCLC Systems & Services: International digital library perspectives*, Vol. 27 No. 3, pp. 190-209. <https://doi.org/10.1108/1065075111164560>
- Luo, K. (2018). Machine Learning Approach for Genre Prediction on Spotify Top Ranking Songs. School of Information and Library Science of the University of North Carolina. Available: [https://cdr.lib.unc.edu/concern/masters\\_papers/ns064961b](https://cdr.lib.unc.edu/concern/masters_papers/ns064961b)
- McDonald, C., Foster, A.E. and Rafferty, P. (2025), "Playlists and genre: the role of music genre in Spotify's playlists", *Journal of Documentation*, Vol. 81 No. 1, pp. 1-23. <https://doi.org/10.1108/JD-08-2023-0152>
- Mattsson, J.T., Peltoniemi, M. and Parvinen, P.M.T. (2010), "Genre-deviating artist entry: the role of authenticity and fuzziness", *Management Decision*, Vol. 48 No. 9, pp. 1355-1364. <https://doi.org/10.1108/00251741011082107>
- Negro, G., Kovács, B. and Carroll, G.R. (2022), "Bus-tin' Out: The Evolution of Novelty and Diversity in Recorded Music", Cattani, G., Deichmann, D. and Ferriani, S. (Ed.) *The Generation, Recognition and Legitimation of Novelty (Research in the Sociology of Organizations, Vol. 77)*, Emerald Publishing Limited, Leeds, pp. 51-87.
- Nijkamp, R. (2018). Prediction of product success: explaining song popularity by audio features from Spotify data. University of Twente. Available: [http://essay.utwente.nl/75422/1/NIJKAMP\\_BA\\_IBA.pdf](http://essay.utwente.nl/75422/1/NIJKAMP_BA_IBA.pdf)
- Nopas, D.-s. (2025), "Melodies of empowerment: exploring the impact of Taylor Swift's music on lifelong learning in Thailand", *Qualitative Research Journal*, Vol. 25 No. 2, pp. 180-202. <https://doi.org/10.1108/QRJ-02-2024-0041>
- Nugroho, A., Manongga, D., & Purnomo, H. D. (2023). Analysis of Spotify Top Songs During the Covid-19 Pandemic. Universitas Semarang, Indonesia: Universitas Kristen Satya Wacana, Indonesia.
- Puh, K. and Bagić Babac, M. (2023), "Predicting sentiment and rating of tourist reviews using machine learning", *Journal of Hospitality and Tourism Insights*, Vol. 6 No. 3, pp. 1188-1204. <https://doi.org/10.1108/JHTI-02-2022-0078>
- Santini, R.M. (2011), "Collaborative classification of popular music on the internet and its social implications", *OCLC Systems & Services: International digital library perspectives*, Vol. 27 No. 3, pp. 210-247. <https://doi.org/10.1108/1065075111164579>
- Sweeney, J.C. and Wyber, F. (2002), "The role of cognitions and emotions in the music-approach-avoidance behavior relationship", *Journal of Services Marketing*, Vol. 16 No. 1, pp. 51-69. <https://doi.org/10.1108/08876040210419415>
- Šandor, D. and Bagić Babac, M. (2024), "Sarcasm detection in online comments using machine learning", *Information Discovery and Delivery*, Vol. 52 No. 2, pp. 213-226. <https://doi.org/10.1108/IDD-01-2023-0002>
- Šimić, A. and Bagić Babac, M. (2024), Artificial Intelligence in Classifying and Creating Art: a Survey, *International Journal of Student Project Reporting*, Vol.2 No.1, pp.59 - 89.
- Šporčič, A. and Pesek, G. (2020), "Local Folk Tales, Legends, and Slavic Mythology in Slovenian Heavy Metal Lyrics: A Quantitative Analysis", Valijärvi, R.-L., Doesburg, C. and Digioia, A. (Ed.) *Multilingual Metal Music: Sociocultural, Linguistic and Literary Perspectives on Heavy Metal Lyrics (Emerald Studies in Metal Music and Culture)*, Emerald Publishing Limited, Leeds, pp. 263-282. <https://doi.org/10.1108/978-1-83909-948-920200022>
- Vinnicombe, T. and Sou, P.U.J. (2017), "Socialization or genre appreciation: the motives of music festival participants", *International Journal of Event and Festival Management*, Vol. 8 No. 3, pp. 274-291. <https://doi.org/10.1108/IJEFM-05-2016-0034>
- Weinberger, M. and Bouhnik, D. (2021), "Various information aspects following the emergence of music streaming applications", *Online Information Review*, Vol. 45 No. 1, pp. 118-137. <https://doi.org/10.1108/OIR-04-2020-0118>
- Wu, H. (2019), "Music's role in peacebuilding", *Journal of Aggression, Conflict and Peace Research*, Vol. 11 No. 2, pp. 138-143. <https://doi.org/10.1108/JACPR-08-2018-0372>

## KVANTITATIVNA ANALIZA PJESAMA S VIŠE OD MILIJARDU STREAMOVA

U ovom radu provedena je detaljna kvantitativna analiza pjesama koje su na platformi *Spotify* premašile milijardu streamova, s ciljem otkrivanja trendova i uvida koji oblikuju glazbenu industriju. Korištenjem statističkih metoda, uključujući linearnu i polinomsku regresiju, istražuju se karakteristike pjesama kao što su trajanje, energija, godina izdanja, popularnost, eksplicitnost, žanrovi i popularnost izvođača. Rezultati pokazuju da pjesme s više od milijardu streamova obuhvaćaju raznolike žanrove, pri čemu dominiraju *pop*, *rap* i *hip-hop*, te ističu ulogu suradnji, kao i trendove u trajanju, energiji i eksplicitnosti pjesama. Pojasnivši čimbenike koji doprinose uspjehu na streaming platformama, ovo istraživanje nudi uvide u preferencije slušatelja i promjenjivu dinamiku glazbene industrije, pružajući smjernice za marketinške strategije, kreiranje pjesama i sastavljanje ili organizaciju glazbenih popisa za izvođače, diskografske kuće i streaming platforme.

KLJUČNE RIJEČI: kvantitativna analiza, glazbeni streaming, pjesme, popularnost, žanrovi