SCHOOL OF COMPUTING, ENGINEERING AND DIGITAL TECHNOLOGY

MSc Computer Science

SPOTIFY MUSIC DATA ANALYSIS

Student Name: RUFUS OLUDARE

Student Number: C2676906

Submission date: 10/1/2024

Contents

Table Of Figures

BUSINESS REPORT

INTRODUCTION

Spotify is a digital music platform, which gives you access to millions of songs and other content from creators all over the world. The dashboards that has been implemented are used to analyze Spotify music data, specifically focusing on song streams and artist popularity over a specified period. The dashboard provides a comprehensive view of various metrics including average streams per year, artists with most songs, tracks by number of streams, and track details based on selected criteria. The dashboard allows users to filter data by date, track, artist, and year to gain insights tailored to specific queries or interests.

DATA SOURCE

The dataset being analysed was gotten from Kaggle(https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023). This data set contains 954 rows and 25 columns. The dataset at hand encapsulates a comprehensive collection of information pertaining to a diverse array of music tracks, presumably aggregated from various streaming platforms. This amalgamation of musical data encompasses an extensive range of attributes, each meticulously curated to offer a nuanced perspective on the characteristics and performance metrics associated with individual tracks.

artist(s)_name	artist_count	released_y	released_r	released_d	in_spotify	in_spotify	streams	in_apple_	in_apple_	in_deezer	in_deezer	in_shazam	bpm	key	mode	danceabili	valence_%	energy_%	acousticn	instrumen	liveness_%	speechine
Latto, Jung Kook	2	2023	7	14	553	147	141381703	43	263	45	10	826	125	B	Major	80	89	83	31	0	8	4
Myke Towers	1	2023	3	23	1474	48	133716286	48	126	58	14	382	92	C#	Major	71	61	74	7	0	10	4
Olivia Rodrigo	1	2023	6	30	1397	113	140003974	94	207	91	14	949	138	F	Major	51	32	53	17	0	31	6
Taylor Swift	1	2019	8	23	7858	100	800840817	116	207	125	12	548	170	A	Major	55	58	72	11	0	11	15
Bad Bunny	1	2023	5	18	3133	50	303236322	84	133	87	15	425	144	A	Minor	65	23	80	14	63	11	6
Dave, Central Ce	2	2023	6	1	2186	91	183706234	67	213	88	17	946	141	C#	Major	92	66	58	19	0	8	24
Eslabon Armado,	2	2023	3	16	3090	50	725980112	34	222	43	13	418	148	F	Minor	67	83	76	48	0	8	3
Quevedo	1	2023	7	7	714	43	58149378	25	89	30	13	194	100	F	Major	67	26	71	37	0	11	4
Gunna	1	2023	5	15	1096	83	95217315	60	210	48	11	953	130	C#	Minor	85	22	62	12	0	28	9
Peso Pluma, Yng	2	2023	3	17	2953	44	553634067	49	110	66	13	339	170	D	Minor	81	56	48	21	0	8	33
Bad Bunny, Grup	2	2023	4	17	2876	40	505671438	41	205	54	12	251	83	F#	Minor	57	56	72	23	0	27	5
NewJeans	1	2023	7	7	422	55	58255150	37	202	21	5	168	150	F	Minor	78	52	82	18	0	15	7
Miley Cyrus	1	2023	1	12	12211	115	1316855716	300	215	745	58	1,021	118		Major	71	65	68	6	0	3	7
David Kushner	1	2023	4	14	3528	98	387570742	80	156	182	24	1,281	130	D	Minor	51	32	43	83	0	9	3
Harry Styles	1	2022	3	31	23575	130	2513188493	403	198	863	46		174	F#	Minor	52	66	73	34	0	31	6
SZA	1	2022	12	8	8109	77	1163093654	183	162	161	12	187	89	G#	Major	64	43	73	5	17	16	4
Fifty Fifty	1	2023	2	24	2942	77	496795686	91	212	78	6	0	120	B	Minor	78	76	59	43	0	34	3
Billie Eilish	1	2023	7	13	873	104	30546883	80	227	95	24	1,173	78		Major	44	14	9	96	0	10	3
Feid, Young Miko	2	2023	3	31	2610	40	335222234	43	100	54	14	187	100	B	Major	86	67	66	14	0	12	16
Jimin	1	2023	3	24	596	68	363369738	8	104	23	2	29	120	G	Major	63	36	73	0	0	36	4
Gabito Ballestero	3	2023	6	22	332	26	86444842	11	163	10	4	0	140	F	Minor	65	87	74	22	0	42	4
Taylor Swift	1	2023	7	7	516	38	52135248	73	119	42	1	150	123	F#	Major	69	82	76	6	0	6	3
Arctic Monkeys	1	2013	1	1	12859	110	1297026226	24	98	582	2	73	135		Minor	48	44	42	12	2	11	3
Bizarrap, Peso Pl	2	2023	5	31	1313	40	200647221	17	152	32	11	139	133	F	Minor	85	81	67	26	0	12	5
The Weeknd, Ma	3	2023	6	2	1945	87	115364561	74	182	87	14	1,093	99	C#	Major	85	83	68	7	0	36	20
Fuerza Regida	1	2023	6	22	250	26	78300654	16	149	10	5	168	130	G	Minor	79	96	86	9	0	9	9
RÃ¯Â¿Â½Ã¯Â¿Â½	2	2022	3	25	7112	77	899183384	202	119	318	38	96	107	B	Major	80	82	80	43	0	14	4
Tainy, Bad Bunny	2	2023	6	29	859	40	61245289	35	109	41	14	211	122	F#	Minor	81	74	71	14	0	56	4
Morgan Wallen	1	2023	1	31	2420	19	429829812	52	107	15	1	325	204	F#	Major	52	52	68	46	0	15	4
Dua Lipa	1	2023	5	25	2988	101	127408954	0	0	143	38	0	110	B	Minor	67	78	85	2	0	33	5
Troye Sivan	1	2023	7	13	864	78	22581161	71	135	50	1	294	126	F	Minor	74	35	84	0	0	11	6

Figure 1: Dataset

Central to the dataset is the fundamental information about the music tracks themselves. The nomenclature of each composition is encapsulated in the column labeled track_name, providing a distinctive identifier for every entry. Complementing this, the artist(s)_name column enumerates the creators behind the musical endeavor, thereby affording due recognition to the individuals or groups responsible for the auditory creation.

In recognizing the collaborative nature of the music industry, the dataset incorporates a metric denoted as artist_count. This numerical representation elucidates the number of artists associated with each track, thereby acknowledging instances where multiple talents converge to produce a singular artistic expression.

Temporal aspects of the dataset are encapsulated in the trio of columns labeled released_year, released_month, and released_day. Collectively, these components delineate the chronological release date of each track, thereby providing insights into the temporal distribution of the dataset and potentially facilitating further analyses.

Moving beyond the temporal aspects, the dataset delves into the realm of track popularity across various streaming platforms. A notable feature is the in_spotify_playlists column, which enumerates the number of Spotify playlists that feature a given track. Correspondingly, the in_spotify_charts column provides a quantitative reflection of the number of charts a specific track is present on.

The cross-platform analysis extends to Apple Music and Deezer, with the in_apple_playlists, in_apple_charts, in_deezer_playlists, and in_deezer_charts columns. These metrics furnish a general view of a track's integration into playlists and the number of charts a specific track is present on, both on Apple Music and Deezer, enriching the dataset with insights into the cross-platform reception of the music.

Shazam, a platform synonymous with music discovery, finds representation in the dataset through the in_shazam_charts column. This metric pinpoints the number of charts a specific track is present on in Shazam’s charts, serving as an additional dimension to gauge its resonance and recognition among users leveraging music identification services.

Beyond the realm of platform-specific metrics, the dataset incorporates musical attributes that contribute to the characterization of each track. The bpm (beats per minute) column encapsulates the tempo of the music, the key and mode columns denote the musical key and mode of each track, respectively.

Further enriching the musical characterization are a set of percentage-based metrics such as danceability_%, valence_%, energy_%, acousticness_%, instrumentalness_%, liveness_%, and speechiness_%. These values offer insights into the danceability, emotional valence, energy, acoustic nature, instrumental characteristics, liveness, and speechiness of each track, respectively, painting a vivid picture of the diverse musical landscape encapsulated in the dataset.

KPI REQUIREMENTS AND QUESTIONS

QUESTIONS

- Which artists has the most songs ?
- Which tracks and artist have the most streams ?
- How has the number of song releases changed over time?
- What day of the week or month in the year do songs get released?
- What are the characteristics of a popular track ?
- Are there certain attributes that are more common in highly streamed tracks?
- Which tracks are the most popular on playlists across different platforms?

KPI

1 **Track Popularity Metrics**:
   - Inclusion in Playlists: Measure the impact of a track by analysing its presence in Spotify, Apple Music, and Deezer playlists.
   - Chart Performance: Evaluate the track's standing in platform-specific charts on Spotify, Apple Music.

2 **Temporal Trends**:
   - Release Dynamics: Examine the distribution of track releases over time using metrics such as released year, month, and day.
   - Temporal Popularity Changes: Identify temporal patterns in track popularity across platforms.

3 **Musical Characteristics**:
   - Musical Attributes: Explore trends in danceability, energy.

4 **Cross-Platform Analysis**:

- Platform-Specific Metrics: Compare the track's performance on Spotify, Apple Music, Deezer, and Shazam based on playlist performance and chart performance.

BUSINESS PROCESSES

- **Content Curation**: Improve playlist recommendations by understanding popular tracks across platforms.
- **Artist Collaboration Strategy**: Identify successful collaborations and potential partnerships.
- **Marketing Strategy**: Tailor marketing efforts based on platform-specific trends and user engagement.

USER GROUPS

- **Music Analysts**: Utilize analytics to understand industry trends and recommend content.
- **Artists and Managers**: Gain insights into factors contributing to a track's success for strategic decision-making.
- **Marketers**: Tailor promotional activities based on platform-specific insights.

DATA PRE-PROCESSING

DATA CLEANSING

As soon as the data was downloaded from Kaggle we had to manually create a new column called cover_url and the purpose of this column is so that we can visualize the track's album cover to make the data visualization of this analysis more attractive. The url for all the tracks album cover was manual gathered and added to the new column. This column does not impact the original data in any way as it is meant to just create a more appealing visual.

We loaded the data into power bi and we decided to create an index column called "Index" so we can split this dataset into smaller tables and connect them together to make up an efficient data model. After creating the index column, we then reordered the column arrangement so that index column is the first column on the main table. Proceeding to cleansing the data, we decided to check for null and empty values and we replaced null values that were present in "in_shazam_charts" column with "0" and we also replaced null values that were present in the "key" column with "A".

The next step was to create a calculated column using DAX. We had to create this calculated column because the date in our dataset is spread across three columns which are "released_year, released_month, released_day". We named the new calculated column DATE, and we used the DAX code shown below to calculate it.

```
1  Date = DATE([released_year], [released_month], [released_day])
```

Figure 2: DATE Column DAX Code

After calculating the new date column, we then changed the data type to "DATE" and we changed the format to "short date". We used an external tool called BRAVO to create a calendar table that we can use to connect to the new date column we just created using a many to one relationship.

DATA ANALYSIS AND KEY FINDINGS

- ARTIST POPULARITY: From our analysis we can deduce that a number of artists have more than one hit song like the weeknd, ed Sheeran, harry styles etc. They have over one billion streams which is a huge number, and it signifies popularity. This means that these artists have a huge following and the artist can be crucial in marketing strategies and promotions.

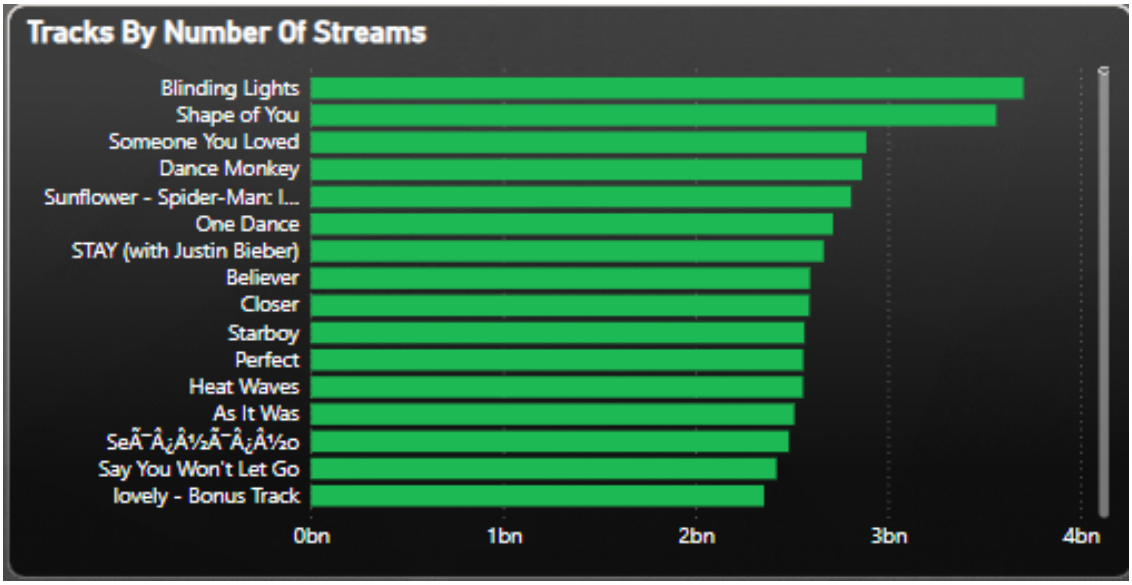


Figure 3: Tracks By number Of Stream

- TRACK RELEASE FREQUENCY: From our analysis we can see that the release frequency of tracks skyrocketed in 2020 which means there is a growing audience for

music listeners as artist release lots of music. From this data we can deduce that streams are definite going to increase and users will search for platforms where they can stream their favourite new music.

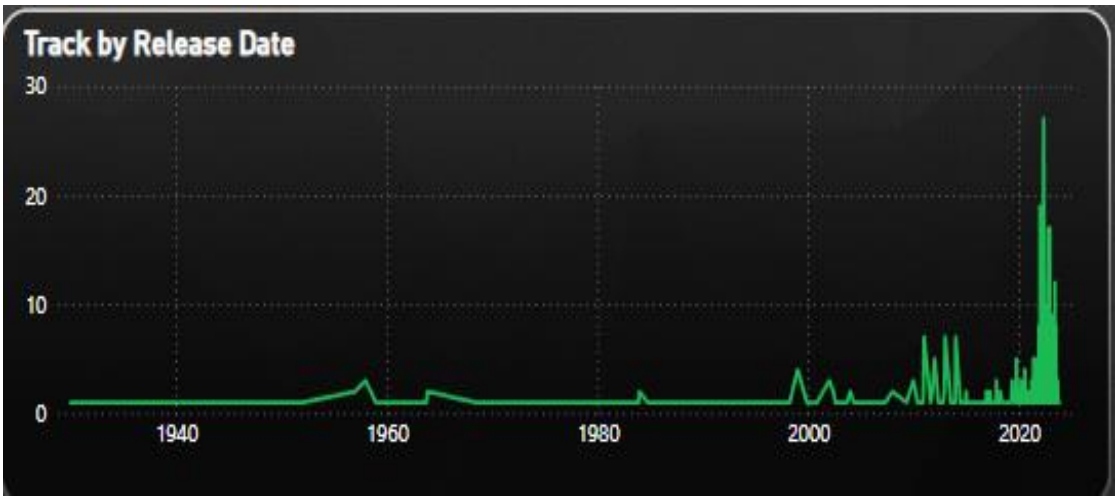


Figure 4: Track By release Date

- PLAYLISTS: From our analysis we can see that Spotify and Deezer feature more tracks on playlists which means there are more active users on Spotify and Deezer as this playlist are more likely going to be created by users rather than the platform themselves. This also means apple music will need to work on promoting playlist to catch up with the competition.

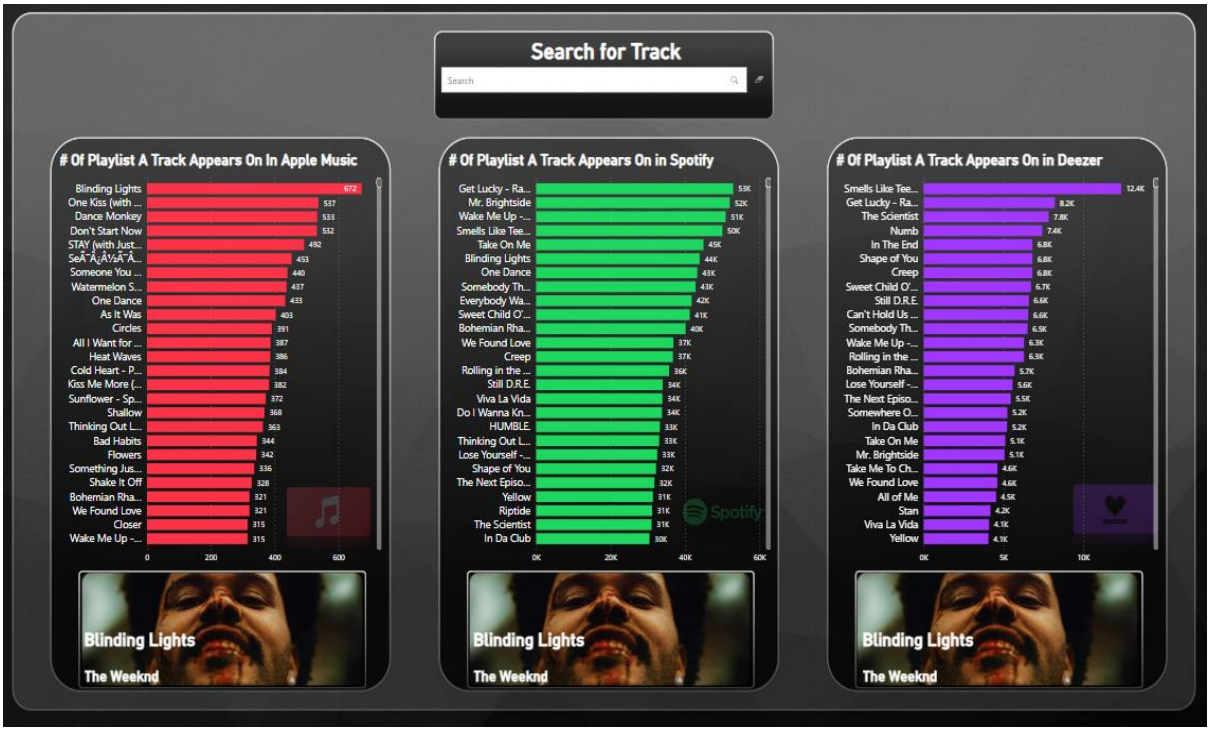


Figure 5: playlist Chart

- CHARTS: From our analysis we can see that a few a few tracks appear on all three platform charts which signifies popularity of those tracks and the track being loved by most people regardless of the platform. Some of these tracks are Flowers, what was I made for and Cruel summer and Sprinter. Artists that composed these songs will be popular due to this data and collaborating with this artiste can be used to promote business as they have a massive audience.

Figure 6: Number Of Charts song Feature in

- SONG ATTRIBUTES

  From our analysis we were able to deduce that most songs that are popular usually have danceability and energy of 50 to 80 percent and sometimes even higher.

Figure 7: Sprinter Song Attributes

Figure 8: Cruel Summer Song Attributes

CHALLENGES AND LIMITATIONS

The challenges that were faced during the analysis of this data was getting album covers of the tracks and visualizing them on the dashboard. Another challenge was implementing the DENEB code to visualizer the songs attributes.

The Limitations to this analysis are the lack of more data like video streams, and more songs data that have been released in past years like song ratings which would have helped to determine the market trend for music in this present age of the music industry.

CONCLUSION AND RECOMMENDATIONS

Artists like The Weeknd, Ed Sheeran, and Harry Styles have demonstrated their influence with multiple hit songs, indicating their strong appeal to listeners. The increase in track releases in 2020 suggests a growing demand for new music, reflecting an expanding music audience. Spotify and Deezer, with more user-generated playlists, show higher user engagement compared to Apple Music. Certain tracks enjoy universal popularity across all platforms, highlighting their broad appeal. Most popular songs tend to have a danceability and energy percentage between 50 and 80 percent, suggesting these attributes resonate with listeners. These insights are invaluable for shaping effective marketing strategies, collaborating strategies, and guiding music production to align with listener preferences.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my lecturer Bianca Ogbo for their invaluable guidance and support throughout the process of this analysis. Their expertise and insights have been instrumental in shaping my understanding and approach towards music data analysis. Their patience and encouragement have inspired me to explore, learn, and grow in this field. Thank you for making this a rewarding and enriching experience.

REFERENCES

Horne, I. (2020) 'Section 1: Introduction to DAX for the BI Pro', in 'Section 1: Introduction to DAX for the BI Pro', *Hands-On Business Intelligence with DAX*. United Kingdom: Packt Publishing, Limited.

Bravo for Power BI - SQLBI Documentation. Available at: docs.sqlbi.com/bravo/ (Accessed: 7 Jan. 2024).

APPENDIX

INTRODUCTION

In an era dominated by the digital revolution, the music industry has undergone significant transformations in the way artists produce, release, and engage with their audience. With the advent of streaming platforms like Spotify, Apple Music, Deezer, and Shazam, the dynamics of music consumption have evolved, offering both challenges and opportunities for artists and industry stakeholders alike. This report delves into a comprehensive analysis of a diverse set of musical tracks released in recent years, leveraging data gathered from various streaming platforms combined.

The dataset under consideration encompasses a variety of artists, ranging from chart-toppers like Taylor Swift and The Weeknd to rising stars like Myke Towers and Olivia Rodrigo. Each track is characterized by a multitude of features, including release details, playlist and chart appearances, streaming statistics, and musical attributes such as tempo, key, and mode. Through a meticulous examination of this dataset, we aim to extract valuable insights that shed light on the factors influencing the success of a song in today's digital music landscape.

Our analysis will encompass several key aspects:

1. **Trends in Streaming Platforms:** We will explore the prevalence of tracks in popular playlists and charts across Spotify, Apple Music, Deezer, and Shazam, providing an overview of the platforms that significantly contribute to an artist's visibility and success.

2. **Musical Attributes and Success:** Examining the correlation between musical attributes (e.g., energy, danceability) and a track's performance on streaming platforms, we aim to identify patterns that resonate with contemporary audience preferences.

3. **Brand Collaborations:** Explore any partnerships or collaborations opportunities between artists and brands based on the outcome of our analysis.

4. **Temporal Patterns:** Considering the time-based features such as release month and day, we will identify any temporal patterns or seasonal influences on a track's performance.

5. **Platform-Specific Analysis:** Break down the data according to different streaming platforms. Compare the performance of songs on platforms like Spotify, Apple Music, and others, and assess whether there are platform-specific trends or audience preferences.

Through this comprehensive analysis, we seek to provide a nuanced understanding of the factors driving success in the digital music landscape, empowering both artists and industry stakeholders to navigate the evolving industry with informed strategies.

DATA SOURCE

The dataset being analysed was gotten from Kaggle(https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023). This data set contains 954 rows and 25 columns. The dataset at hand encapsulates a comprehensive collection of information pertaining to a diverse array of music tracks, presumably aggregated from various streaming platforms. This amalgamation of musical data encompasses an extensive range of attributes, each meticulously curated to offer a nuanced perspective on the characteristics and performance metrics associated with individual tracks.

artist(s)_name	artist_count	released_	released_	released_	in_spotify	in_spotify	streams	in_apple_	in_apple_	in_deezer	in_deezer	in_shazam	bpm	key	mode	danceabili	valence_%	energy_%	acousticn	instrumen	liveness_%	speechine
Latto, Jung Kook	2	2023	7	14	553	147	141381703	43	263	45	10	826	125	B	Major	80	89	83	31	0	8	4
Myke Towers	1	2023	3	23	1474	48	133716286	48	126	58	14	382	92	C#	Major	71	61	74	7	0	10	4
Olivia Rodrigo	1	2023	6	30	1397	113	140003974	94	207	91	14	949	138	F	Major	51	32	53	17	0	31	6
Taylor Swift	1	2019	8	23	7858	100	800840817	116	207	125	12	548	170	A	Major	55	58	72	11	0	11	15
Bad Bunny	1	2023	5	18	3133	50	303236322	84	133	87	15	425	144	A	Minor	65	23	80	14	63	11	6
Dave, Central Ce	2	2023	6	1	2186	91	183706234	67	213	88	17	946	141	C#	Major	92	66	58	19	0	8	24
Eslabon Armado,	2	2023	3	16	3090	50	725980112	34	222	43	13	418	148	F	Minor	67	83	76	48	0	8	3
Quevedo	1	2023	7	7	714	43	58149378	25	89	30	13	194	100	F	Major	67	26	71	37	0	11	4
Gunna	1	2023	5	15	1096	83	95217315	60	210	48	11	953	130	C#	Minor	85	22	62	12	0	28	9
Peso Pluma, Yng	2	2023	3	17	2953	44	553634067	49	110	66	13	339	170	D	Minor	81	56	48	21	0	8	33
Bad Bunny, Grup	2	2023	4	17	2876	40	505671438	41	205	54	12	251	83	F#	Minor	57	56	72	23	0	27	5
NewJeans	1	2023	7	7	422	55	58255150	37	202	21	5	168	150	F	Minor	78	52	82	18	0	15	7
Miley Cyrus	1	2023	1	12	12211	115	1316855716	300	215	745	58	1,021	118		Major	71	65	68	6	0	3	7
David Kushner	1	2023	4	14	3528	98	387570742	80	156	182	24	1,281	130	D	Minor	51	32	43	83	0	9	3
Harry Styles	1	2022	3	31	23575	130	2513188493	403	198	863	46		174	F#	Minor	52	66	73	34	0	31	6
SZA	1	2022	12	8	8109	77	1163093654	183	162	161	12	187	89	G#	Major	64	43	73	5	17	16	4
Fifty Fifty	1	2023	2	24	2942	77	496795686	91	212	78	6	0	120	B	Minor	78	76	59	43	0	34	3
Billie Eilish	1	2023	7	13	873	104	30546883	80	227	95	24	1,173	78		Major	44	14	9	96	0	10	3
Feid, Young Miko	2	2023	3	31	2610	40	335222234	43	100	54	14	187	100	B	Major	86	67	66	14	0	12	16
Jimin	1	2023	3	24	596	68	363369738	8	104	23	2	29	120	G	Major	63	36	73	0	0	36	4
Gabito Ballestero	3	2023	6	22	332	26	86444842	11	163	10	4	0	140	F	Minor	65	87	74	22	0	42	4
Taylor Swift	1	2023	7	7	516	38	52135248	73	119	42	1	150	123	F#	Major	69	82	76	6	0	6	3
Arctic Monkeys	1	2013	1	1	12859	110	1297026226	24	98	582	2	73	135		Minor	48	44	42	12	2	11	3
Bizarrap, Peso Pl	2	2023	5	31	1313	40	200647221	17	152	32	11	139	133	F	Minor	85	81	67	26	0	12	5
The Weeknd, Ma	3	2023	6	2	1945	87	115364561	74	182	87	14	1,093	99	C#	Major	85	83	68	7	0	36	20
Fuerza Regida	1	2023	6	22	250	26	78300654	16	149	10	5	168	130	G	Minor	79	96	86	9	0	9	9
RÃ¯Â¿Â½Ã¯Â¿Â½	2	2022	3	25	7112	77	899183384	202	119	318	38	96	107	B	Major	80	82	80	43	0	14	4
Tainy, Bad Bunny	2	2023	6	29	859	40	61245289	35	109	41	14	211	122	F#	Minor	81	74	71	14	0	56	4
Morgan Wallen	1	2023	1	31	2420	19	429829812	52	107	15	1	325	204	F#	Major	52	52	68	46	0	15	4
Dua Lipa	1	2023	5	25	2988	101	127408954	0	0	143	38	0	110	B	Minor	67	78	85	2	0	33	5
Troye Sivan	1	2023	7	13	864	78	22581161	71	135	50	1	294	126	F	Minor	74	35	84	0	0	11	6

Figure 9: Dataset

Central to the dataset is the fundamental information about the music tracks themselves. The nomenclature of each composition is encapsulated in the column labeled track_name, providing a distinctive identifier for every entry. Complementing this, the artist(s)_name

column enumerates the creators behind the musical endeavor, thereby affording due recognition to the individuals or groups responsible for the auditory creation.

In recognizing the collaborative nature of the music industry, the dataset incorporates a metric denoted as artist_count. This numerical representation elucidates the number of artists associated with each track, thereby acknowledging instances where multiple talents converge to produce a singular artistic expression.

Temporal aspects of the dataset are encapsulated in the trio of columns labeled released_year, released_month, and released_day. Collectively, these components delineate the chronological release date of each track, thereby providing insights into the temporal distribution of the dataset and potentially facilitating further analyses.

Moving beyond the temporal aspects, the dataset delves into the realm of track popularity across various streaming platforms. A notable feature is the in_spotify_playlists column, which enumerates the number of Spotify playlists that feature a given track. Correspondingly, the in_spotify_charts column provides a quantitative reflection of the number of charts a specific track is present on.

The cross-platform analysis extends to Apple Music and Deezer, with the in_apple_playlists, in_apple_charts, in_deezer_playlists, and in_deezer_charts columns. These metrics furnish a general view of a track's integration into playlists and the number of charts a specific track is present on, both on Apple Music and Deezer, enriching the dataset with insights into the cross-platform reception of the music.

Shazam, a platform synonymous with music discovery, finds representation in the dataset through the in_shazam_charts column. This metric pinpoints the number of charts a specific track is present on in Shazam’s charts, serving as an additional dimension to gauge its resonance and recognition among users leveraging music identification services.

Beyond the realm of platform-specific metrics, the dataset incorporates musical attributes that contribute to the characterization of each track. The bpm (beats per minute) column encapsulates the tempo of the music, the key and mode columns denote the musical key and mode of each track, respectively.

Further enriching the musical characterization are a set of percentage-based metrics such as danceability_%, valence_%, energy_%, acousticness_%, instrumentalness_%, liveness_%, and speechiness_%. These values offer insights into the danceability, emotional valence,

energy, acoustic nature, instrumental characteristics, liveness, and speechiness of each track, respectively, painting a vivid picture of the diverse musical landscape encapsulated in the dataset.

KPI REQUIREMENTS AND QUESTIONS

QUESTIONS

- Which artists has the most songs ?
- Which tracks and artist have the most streams ?
- How has the number of song releases changed over time?
- What day of the week or month in the year do songs get released?
- What are the characteristics of a popular track ?
- Are there certain attributes that are more common in highly streamed tracks?
- Which tracks are the most popular on playlists across different platforms?

KPI

5 **Track Popularity Metrics**:

- Inclusion in Playlists: Measure the impact of a track by analysing its presence in Spotify, Apple Music, and Deezer playlists.
- Chart Performance: Evaluate the track's standing in platform-specific charts on Spotify, Apple Music.

6 **Temporal Trends**:

- Release Dynamics: Examine the distribution of track releases over time using metrics such as released year, month, and day.
- Temporal Popularity Changes: Identify temporal patterns in track popularity across platforms.

7 **Musical Characteristics**:

- Musical Attributes: Explore trends in danceability, energy.

8 **Cross-Platform Analysis**:

- Platform-Specific Metrics: Compare the track's performance on Spotify, Apple Music, Deezer, and Shazam based on playlist performance and chart performance.

BUSINESS PROCESSES

- **Content Curation**: Improve playlist recommendations by understanding popular tracks across platforms.
- **Artist Collaboration Strategy**: Identify successful collaborations and potential partnerships.
- **Marketing Strategy**: Tailor marketing efforts based on platform-specific trends and user engagement.

USER GROUPS

- **Music Analysts**: Utilize analytics to understand industry trends and recommend content.
- **Artists and Managers**: Gain insights into factors contributing to a track's success for strategic decision-making.
- **Marketers**: Tailor promotional activities based on platform-specific insights.

DATA PRE-PROCESSING

DATA CLEANSING

As soon as the data was downloaded from Kaggle we had to manually create a new column called cover_url and the purpose of this column is so that we can visualize the track's album cover to make the data visualization of this analysis more attractive. The url for all the tracks album cover was manual gathered and added to the new column. This column does not impact the original data in any way as it is meant to just create a more appealing visual.

We loaded the data into power bi and we decided to create an index column called "Index" so we can split this dataset into smaller tables and connect them together to make up an efficient data model. After creating the index column, we then reordered the column arrangement so that index column is the first column on the main table. Proceeding to cleansing the data, we decided to check for null and empty values and we replaced null values that were present in "in_shazam_charts" column with "0" and we also replaced null values that were present in the "key" column with "A".

The next step was to create a calculated column using DAX. We had to create this calculated column because the date in our dataset is spread across three columns which are "released_year,

released_month, released_day". We named the new calculated column DATE, and we used the DAX code shown below to calculate it.

```
1  Date = DATE([released_year], [released_month], [released_day])
```

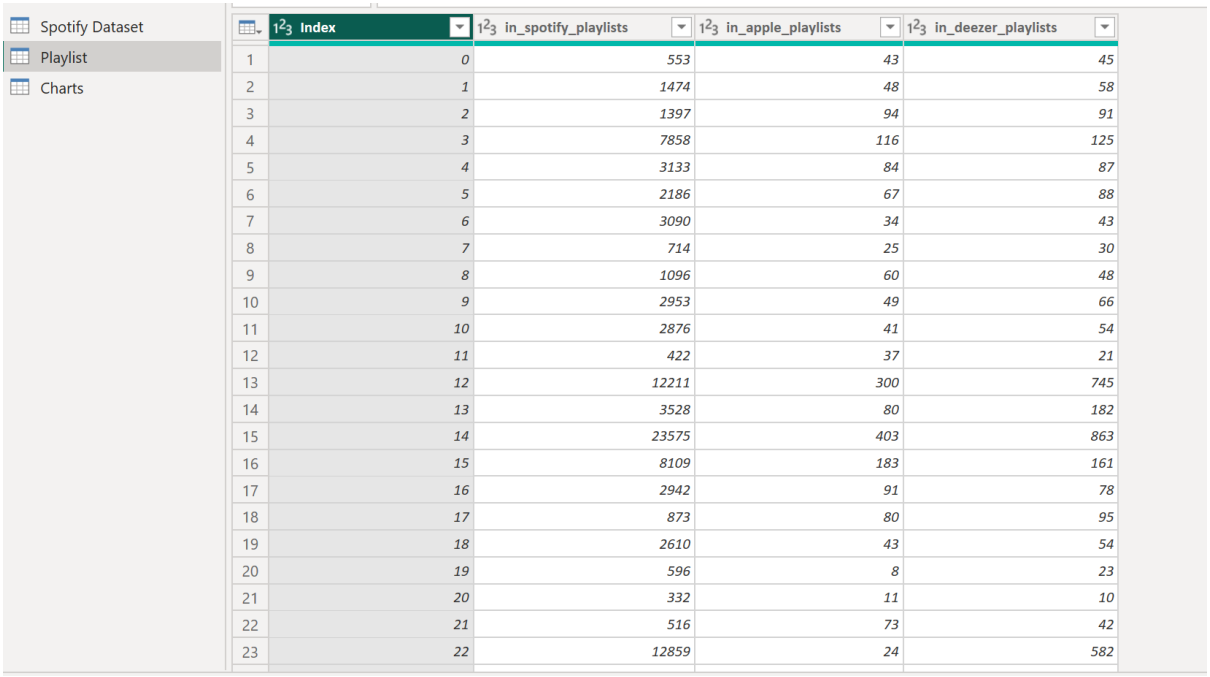
Figure 10: DATE Column DAX Code

After calculating the new date column, we then changed the data type to "DATE" and we changed the format to "short date". We used an external tool called BRAVO to create a calendar table that we can use to connect to the new date column we just created using a many to one relationship.

DATA MODELLING

MODEL

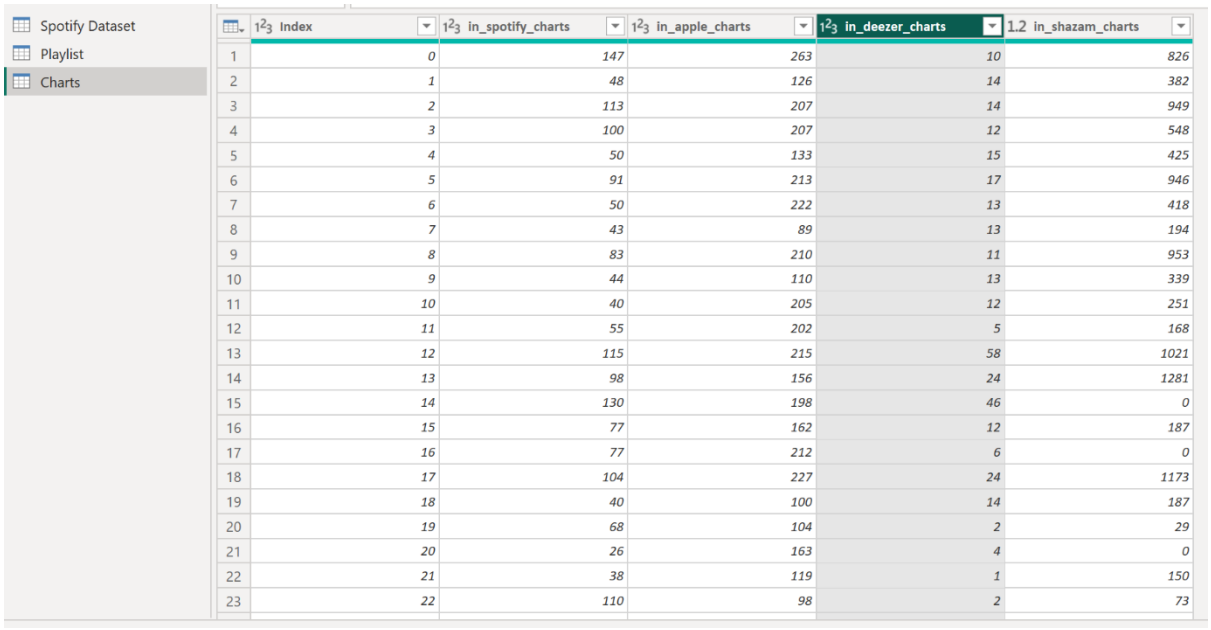
Creating the model was relatively easy now that we have an index in our main table. We created new tables called playlist which consist of index, in_spotify_playlists, in_apple_playlists and in_deezer_playlists.



	Index	in_spotify_playlists	in_apple_playlists	in_deezer_playlists
1	0	553	43	45
2	1	1474	48	58
3	2	1397	94	91
4	3	7858	116	125
5	4	3133	84	87
6	5	2186	67	88
7	6	3090	34	43
8	7	714	25	30
9	8	1096	60	48
10	9	2953	49	66
11	10	2876	41	54
12	11	422	37	21
13	12	12211	300	745
14	13	3528	80	182
15	14	23575	403	863
16	15	8109	183	161
17	16	2942	91	78
18	17	873	80	95
19	18	2610	43	54
20	19	596	8	23
21	20	332	11	10
22	21	516	73	42
23	22	12859	24	582

Figure 11: Playlist table

We then created a new table called Charts which consist of index, in_spotify_charts, in_apple_charts, in_deezer_deezer



Spotify Dataset
Playlist
Charts

	Index	in_spotify_charts	in_apple_charts	in_deezer_charts	in_shazam_charts
1	0	147	263	10	826
2	1	48	126	14	382
3	2	113	207	14	949
4	3	100	207	12	548
5	4	50	133	15	425
6	5	91	213	17	946
7	6	50	222	13	418
8	7	43	89	13	194
9	8	83	210	11	953
10	9	44	110	13	339
11	10	40	205	12	251
12	11	55	202	5	168
13	12	115	215	58	1021
14	13	98	156	24	1281
15	14	130	198	46	0
16	15	77	162	12	187
17	16	77	212	6	0
18	17	104	227	24	1173
19	18	40	100	14	187
20	19	68	104	2	29
21	20	26	163	4	0
22	21	38	119	1	150
23	22	110	98	2	73

Figure 12:Charts table

The creation of these new tables led to the creation of our data models we now have a one-to-one relationship between the main table, playlist and charts tables.

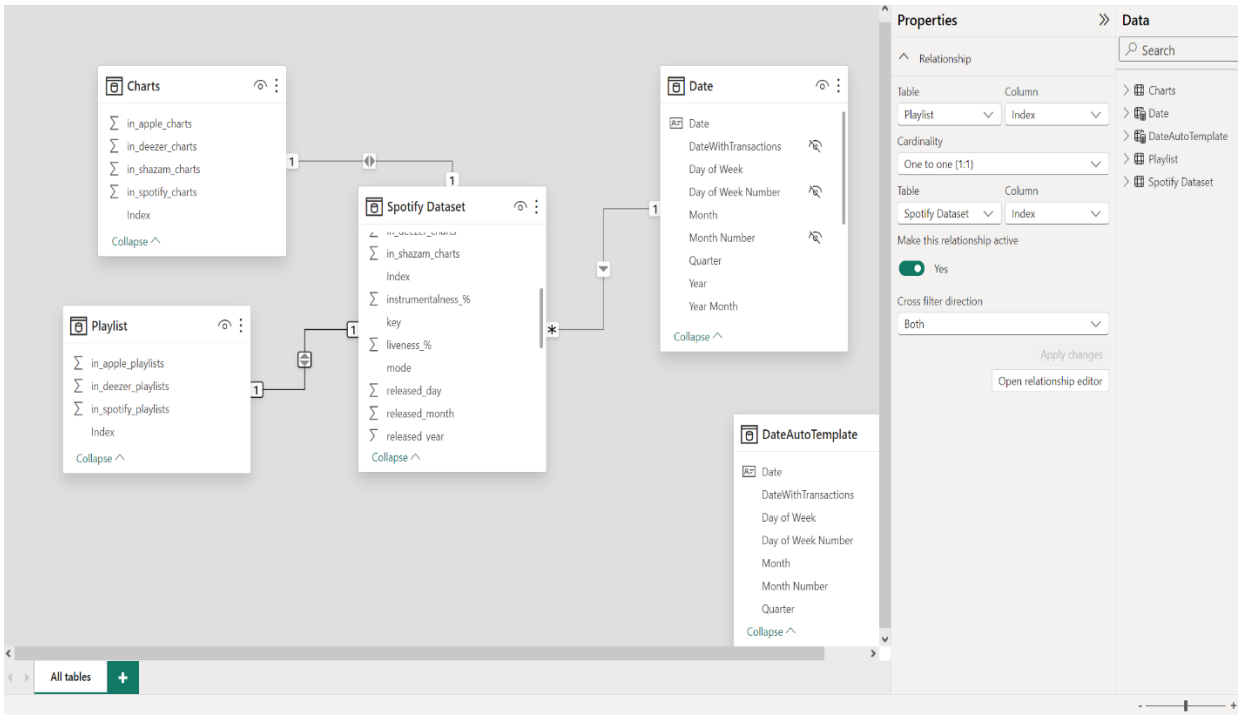


Figure 13: Data Model

DAX AND M LANGUAGE

DAX AND M LANGUAGE

To create some of our power BI visuals, we made use of several DAX codes as show below. We will discuss some of these DAX code in detail in this section. The first DAX CODE which was implemented was used to create a new measure for the purpose of counting the number of appearances of each artist in the artist(s)_name column.

```
1  _ArtistAppearanceCount = COUNT('Spotify Dataset'[artist(s)_name])
```

Figure 14: Artist Appearances

ArtistAppearanceCount is the name assigned to the calculated column or measure. It represents the result of the calculation, which is the count of artist appearances. COUNT is an aggregation function in DAX that is used to count the number of rows in a table that meet a specified condition. 'Spotify Dataset'[artist(s)_name]: This is the column reference from the 'Spotify Dataset' table that contains the names of artists. The expression is counting the occurrences of artist names in this column.

The second DAX code that was created was used to create a new measure that calculates the average number of streams per year, considering only the filters related to the 'Year' column in the 'Date' table. This measure is useful for analysing the streaming trends over different years in the context of the Spotify dataset.

```
_Average Stream per year = CALCULATE(AVERAGE('Spotify Dataset'[streams]),ALLEXCEPT('Spotify Dataset','Date'[Year]))
```

Figure 15: Streams per year

_Average Stream per year is the name assigned to the calculated measure. It represents the result of the calculation, which is the average number of streams per year. CALCULATE is a powerful function in DAX that allows you to modify the context in which an expression is evaluated. It is often used to apply filters and conditions. AVERAGE('Spotify Dataset'[streams]) is the expression to be evaluated. It calculates the average of the 'streams' column in the 'Spotify Dataset' table. It gives you the average number of streams for the specified context. ALLEXCEPT('Spotify Dataset', 'Date'[Year]) is a table function that removes all filters from the 'Spotify Dataset' table except for the 'Year' column in the 'Date' table. It ensures that the average is calculated per year, disregarding other filters that might be present.

The third DAX code implemented is meant to calculates the average danceability percentage based on the values in the 'danceability_%' column.

```
_dance_Percent_Val =
AVERAGE([danceability_%])
```

Figure 16: Dance Percentage

_dance_Percent_Val is the name assigned to the calculated measure. It represents the result of the calculation, which is the average value of the 'danceability_%' column. AVERAGE([danceability_%]) is the expression to be evaluated. It calculates the average of the 'danceability_%' column. It gives you the average danceability percentage for the specified context.

Another DAX code was created for a new measure to calculate the average energy percentage based on the values in the 'energy_%' column.

```
_Energy_Percent_Val =
AVERAGE([energy_%])
```

Figure 17: Energy Percentage

_Energy_Percent_Val is the name assigned to the calculated measure. It represents the result of the calculation, which is the average value of the 'energy_%' column. AVERAGE([energy_%]) is the expression to be evaluated. It calculates the average of the 'energy_%' column. It gives you the average energy percentage for the specified context.

The Fourth DAX code that was created for a Power BI dynamically display of album covers based on the album with the highest number of streams, presenting them in a visually appealing manner within a defined container.

```
_Image html =
Var x =
CALCULATE(
MAX('Spotify Dataset'[cover_url]),
'Spotify Dataset'[streams] = MAX('Spotify Dataset'[streams])
)
return

"


<!DOCTYPE html>
<html lang='en'>
<head>
  <meta charset='UTF-8'>
  <title>Image Cropping</title>
  <style>
  .image-container {
    width: 444px; /* Width of the container */
    height: 165px; /* Height of the container */
    overflow: hidden; /* Hide parts of the image that don't fit */
    border-radius: 14px; /* Rounded corners */
    position: relative; /* Relative positioning for the child element */
   }

   .image {
    object-fit: cover; /* Cover the entire container */
    object-position: center; /* Center the image */
    width: 100%; /* Full width */
    height: 100%; /* Full height */
   }
  </style>
</head>
<body>
  <div class='image-container'>
```

Figure 18: Album cover

Image html: This is a calculated column or measure that generates an HTML snippet.

VAR x: A variable that stores the cover URL for the album with the maximum number of streams.

CALCULATE(MAX('Spotify Dataset'[cover_url]), 'Spotify Dataset'[streams] = MAX('Spotify Dataset'[streams])): It retrieves the cover URL for the album with the maximum number of streams.

RETURN: Specifies the beginning of the HTML content to be returned.

img src='" & x & "': Dynamically inserts the cover URL obtained from the DAX calculation.

class='image': Applies the specified styling for the image.

The fifth DAX code that was implemented was to create a new measure to provide the maximum number of streams present in the 'Spotify Dataset' table.

```
_Max streams = MAX('Spotify Dataset'[streams])
```

Figure 19:Streams

_Max streams is the name of the measure being created. MAX('Spotify Dataset'[streams]) uses the MAX function to find the maximum value of the 'streams' column in the 'Spotify Dataset' table.

Another DAX code was implemented to calculate the total number of streams for the song that have the maximum number of streams in the 'Spotify Dataset' table.

```
_Top Song streams = CALCULATE(SUM('Spotify Dataset'[streams]),'Spotify Dataset'[streams] = MAX('Spotify Dataset'[streams]))
```

Figure 20: Top Songs Streams

_Top Song streams: This is the name of the measure being created.

CALCULATE: It is a DAX function that is used to evaluate an expression in a context modified by filters.

SUM('Spotify Dataset'[streams]): This is the expression to be evaluated. It calculates the sum of the 'streams' column in the 'Spotify Dataset' table.

'Spotify Dataset'[streams] = MAX('Spotify Dataset'[streams]): This part of the formula is a filter condition within the CALCULATE function. It filters the data to include only rows where the 'streams' column is equal to the maximum value of the 'streams' column in the 'Spotify Dataset' table.

The seventh DAX CODE was used to create a new measure to visually represent whether the top song's performance (as measured by "_Top song vs avg val") is above or below the average. It does so by displaying a percentage change along with an arrow symbol (up or down) to indicate the direction of the change.

```
_Top song vs AVG =
VAR x = [_Top song vs avg val] RETURN

IF(x > 0,
FORMAT(x,"#.0%") & " " & UNICHAR ( 9650),
FORMAT(x, "#.0%") & " " & UNICHAR( 9660 ))
```

Figure 21: Top Songs Against Average

_Top song vs AVG: This is the name of the measure being created.

VAR x = [_Top song vs avg val]: This line declares a variable x and assigns it the value of another measure or calculated column named "_Top song vs avg val."

RETURN: This keyword is used to specify the value that the measure should return.

FORMAT(x, "#.0%"): This part formats the value of x as a percentage with one decimal place.

& " " & UNICHAR(9650): If the value of x is greater than 0, it appends a space, an upward-pointing triangle character (▲), indicating an increase.

& " " & UNICHAR(9660): If the value of x is less than or equal to 0, it appends a space, a downward-pointing triangle character (▼), indicating a decrease.

The Eight DAX code was used to create a measure and the purpose of this measure is to calculate the relative difference between the streams of the top song and the average stream per year. It expresses this difference as a ratio, allowing for a standardized comparison.

```
_Top song vs avg val = DIVIDE([_Top Song streams] - [_Average Stream per year] , [_Average Stream per year])
```

Figure 22: Top Songs vs Average stream per Year

_Top song vs avg val: This is the name of the measure being created.

DIVIDE([_Top Song streams] - [_Average Stream per year] , [_Average Stream per year]): This is the main calculation. It calculates the relative difference between the streams of the top song and the average stream per year.

[_Top Song streams]: This is assumed to be another measure or column representing the total streams of the top song.

[_Average Stream per year]: This is assumed to be another measure or column representing the average stream per year.

DIVIDE(): This function is used to perform division. It takes two arguments: the numerator and the denominator.

The final measure was created using a DAX code that counts the number of distinct track names in the 'Spotify Dataset' table. It provides information about the total variety or number of unique tracks present in the dataset.

```
_Track = COUNT('Spotify Dataset'[track_name])
```

Figure 23: Heat Map

DASHBOARD

The first dashboard we created is organized into various sections, each displaying different types of data related to Spotify music data analytics. It includes graphical representations like pie charts, bar graphs heat maps, and line charts to visualize the data effectively. Here are some of the main features of the dashboard:

MAIN DASHBOARD

Figure 24: Main Dashboard

- On the top left, there is a section showing the average number of streams per year for Spotify, along with the percentage change from the previous year. This indicates how popular Spotify is among its users and how its growth rate is changing over time.
- Below this section, there is a pie chart titled “Artists with Most Songs”, which shows the top five artists who have the greatest number of songs on Spotify. This gives an idea of which artists are the most prolific and influential in the music industry.
- Underneath the pie chart, there is a line chart titled “Track by Release Date”, which shows the number of tracks released on Spotify over the years. This shows the trend of music production and consumption over time, and how it varies across different genres and seasons.
- In the top centre part of the dashboard, there is a section dedicated to the album cover of any song in focus, this section shows the details of the song as well like the name of the song and the name of the artiste.
- To the right of this central section, there is a bar graph titled “Tracks by Number Of Streams”, which shows the top ten songs on Spotify by the number of streams they have received. This shows which songs are the most popular and listened to by the Spotify users, and how they compare to each other in terms of popularity.
- At the bottom right corner, there is another detailed analysis section which shows the same information in a more detailed and numerical way. This section also shows a comparison of the song’s audio features with the average values for all songs on Spotify, which shows how the song differs from the norm in terms of its sound and mood.
- At the middle we have a heatmap that is used to visualize the release day and month of songs .
- At the very top we have different filters like date, artist, year. This filter was implemented so we can analyse our data better based on the song, artist or the year it was released.

PLAYLIST DASHBOARD

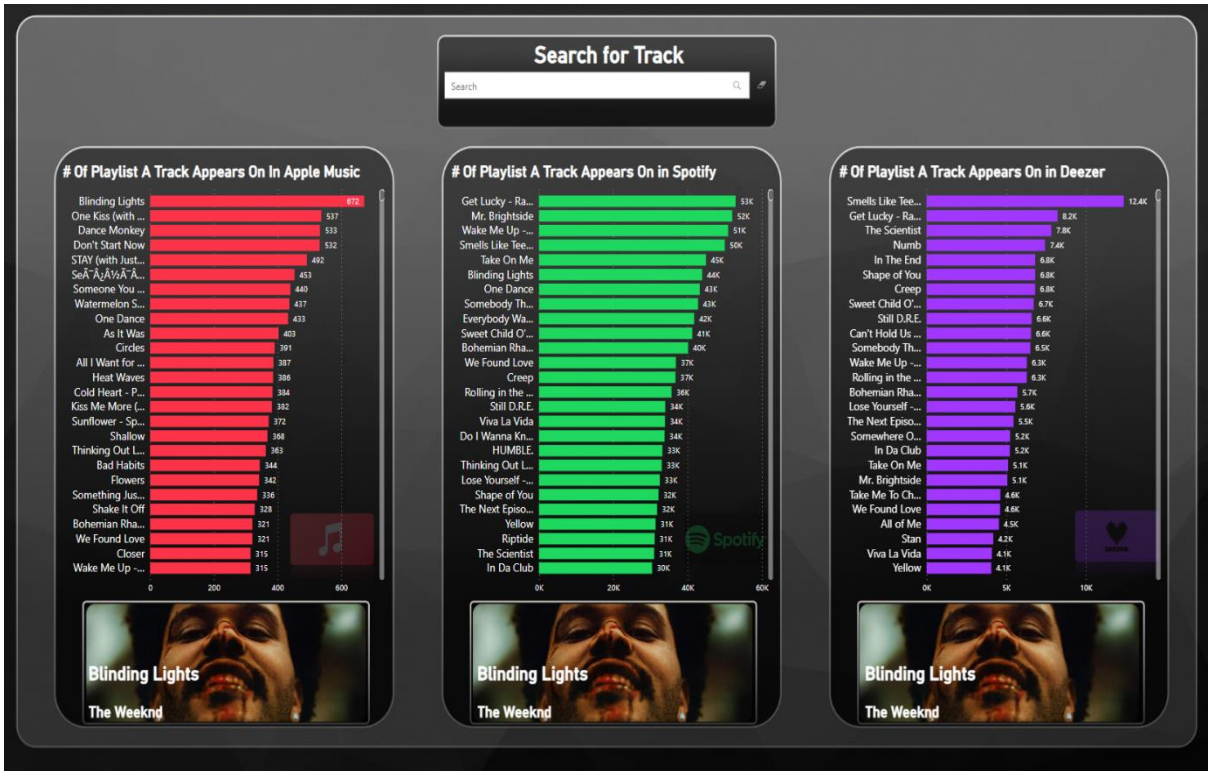


Figure 25: Playlist Dashboard

This dashboard is meant to compare and visualizing data from three different music streaming platforms: Apple Music, Spotify, and Deezer. The dashboard is organized into three main sections, each representing a different platform. Here are some of the main features of the dashboard:

Each section contains a bar chart displaying the number of playlists a track appears on within the respective platform. This shows how popular and influential a song is among the users of each service, and how they differ across platforms.

Each song is listed by name along with the corresponding number of playlists it appears on; these numbers are represented both numerically and visually with coloured bars. The colours match the theme of each platform: pinkish red for Apple Music, green for Spotify, and purple for Deezer.

At the bottom of each bar chart, we have images at the bottom of each section which is just meant to allow the user of the dashboard see the album cover of any song selected.

There is a search section at the middle of this which is meant to allow a user search for a specific so and see the number of playlists the song is featured in.

CHARTS DASHBOARD

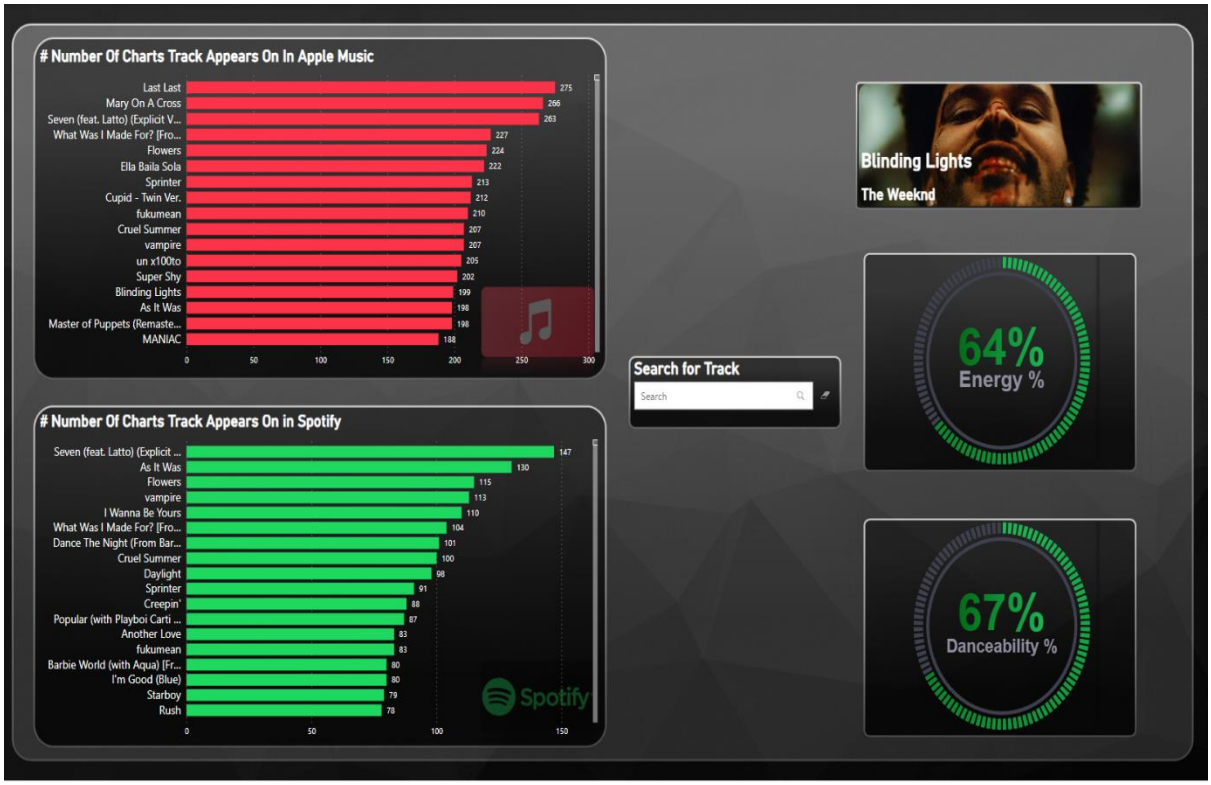


Figure 26: Charts Dashboard

This Chart dashboard is organized into several sections displaying various types of data. Here are some of the main features of the dashboard:

On the left, there are two bar graphs showing the number of charts a track appears on in Apple Music and Spotify, respectively. This shows how popular and influential a song is among the users of each service, and how they differ across platforms.

On the right, we have an HTML visual that shows the album cover of a song clicked on for a better visual appeal. We also have DENEB visuals showing the energy and danceability percentages displayed in circular graphs.

There is also a search bar labelled “Search for Track” in the middle of the dashboard allowing users to look up specific tracks and see their data across all platforms.

DATA SOURCES

- https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023 (DATASET)
- https://www.bing.com/ (ALBUM IMAGES)

GLOSSARY

DAX	Data Analysis Expressions is a formula language used in Power BI to create custom calculations and aggregations for data analysis. It manipulates and analyses data from different sources, creates new calculated columns and measures, and performs complex calculations and analyses.
DENEB	Deneb is a custom visual for Microsoft Power BI, which allows creators to use the declarative JSON syntax of the Vega or Vega-Lite languages to build their own bespoke data visualizations, without having to learn web development.