CADT
IDT វិទ្យាស្ថានបច្ចេកវិទ្យាឌីជីថល
Institute of Digital Technology

SpotGen Track Visualization

DATA SCIENCE -INFORMATION VISUALIZATION USING POWER-BI

NOVEMBER 2023

AUTHOR:
NEANG ROTHMNY
EANG KIMCHHIV
GNEP VANDARA
THEAR SITHYSAK

Table of Contents

List of Figure

List of Tables

I. Introduction

1. Project Objective

The primary objective of this project is to analyze and visualize a dataset that combines Spotify and Genius track information. The dataset encompasses details about music tracks, including song attributes, lyrics, artist information, and popularity metrics. The goal is to gain comprehensive insights into the music streaming industry, exploring patterns, trends, and relationships between various elements. Specifically, we aim to investigate the characteristics of popular tracks, artist collaborations, the impact of lyrics on song popularity, and regional variations in music preferences. Additionally, we will delve into the correlation between Spotify metrics (e.g., danceability, energy) and the success of tracks based on Genius annotations. The outcomes of this analysis will be valuable for stakeholders in the music industry, including streaming platforms, record labels, and artists, providing actionable insights to enhance music discovery, user experience, and content creation.

2. Expected Delivery

A Comprehensive Music Insight gains a deep understanding of the music dataset, identifying key patterns, trends, and characteristics that define popular tracks and artists. Also, Explore how lyrics contribute to the popularity of songs. Analyze the relationship between lyrical content, sentiment, and music engagement.

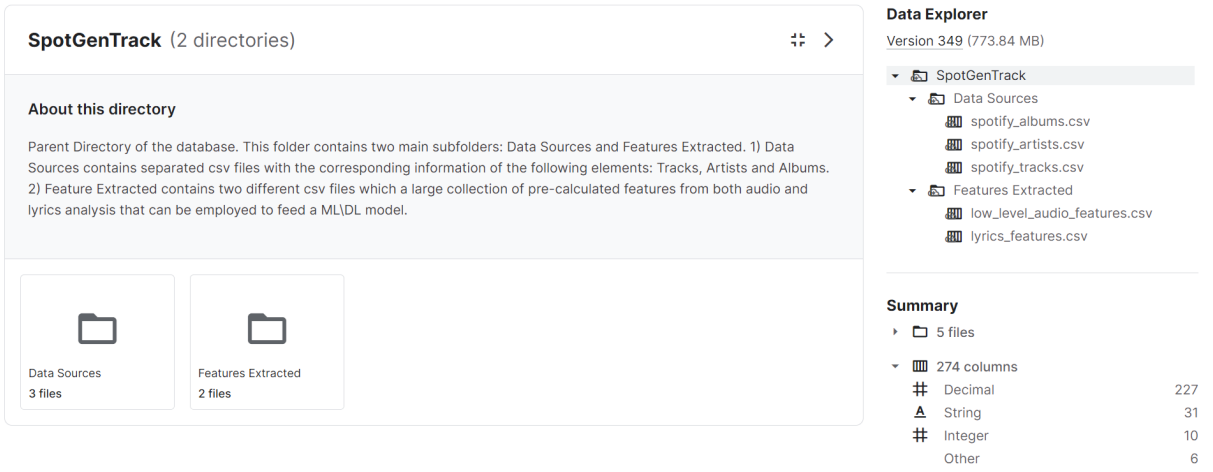
Moreover, it evaluates the correlation between Spotify metrics (e.g., danceability, energy) and the success of tracks based on Genius annotations. Identify which features contribute most to a track's popularity.

In addition, it examines geographic patterns in music preferences, identifying how different genres and artists resonate across various regions.

This analytic aims to provide actionable guidance for optimizing the music streaming experience, fostering artist collaboration, and tailoring content to user preferences, thus contributing to the ongoing evolution and innovation within the music streaming industry.

II. Data Description

1. Size



SpotGenTrack (773.84MB)		
Table	Total Row	Total Column
low_level_audio_features	101910	209
lyrics_features	94955	9
spotify_albums	75512	16
spotify_artists	56130	9
spotify_track	101938	32

Table 1

This table shows the size of this dataset which is 773.84 MB. It has a total of five tables.
Here are the table:

- Low_level_audio_features

	Chroma_1	Chroma_1	Chroma_1	Chroma_1	Chroma_2	Chroma_3	Chroma_4	Chroma_5	Chroma_6	Chroma_7	Chroma_8	Chroma_9	MEL_1	MEL_10	MEL_100	MEL_101	MEL_102	MEL_103
0	0.438296	0.472769	0.427441	0.436688	0.467697	0.493862	0.512244	0.568658	0.560524	0.513068	0.544648	0.544643	0.00014	0.667637	1.63E-05	8.69E-06	6.49E-06	5.49E-06
1	0.596605	0.368288	0.285263	0.302211	0.905805	0.510909	0.221708	0.311248	0.491277	0.416469	0.411171	0.553497	0.053749	20.36411	0.022766	0.01999	0.013676	0.015588
2	0.505224	0.50042	0.506773	0.488258	0.498356	0.573582	0.690761	0.742858	0.686282	0.657118	0.550437	0.491275	381.7365	50.09699	0.166018	0.159811	0.180781	0.169495
3	0.52569	0.666469	0.579492	0.49892	0.598528	0.631578	0.501693	0.500468	0.587101	0.546499	0.474716	0.557924	77.63077	13.2806	0.281888	0.280577	0.177236	0.211261
4	0.632214	0.503698	0.496942	0.611532	0.634613	0.697265	0.557012	0.530836	0.444279	0.466659	0.605773	0.546381	11.86479	27.35888	0.362137	0.237292	0.265146	0.346296
5	0.557243	0.484811	0.514799	0.541027	0.601414	0.626298	0.62787	0.579796	0.511928	0.466329	0.486777	0.496173	0.003337	6.828721	0.000164	0.000104	5.59E-05	2.31E-05
6	0.433017	0.1675	0.149971	0.306848	0.395572	0.422261	0.439999	0.574798	0.496	0.281756	0.242395	0.193941	0.001096	2.448581	7.12E-09	1.27E-08	5.22E-09	3.32E-09
7	0.479246	0.621048	0.485272	0.469533	0.556495	0.506786	0.512823	0.604547	0.486548	0.46058	0.487228	0.621503	27.43644	55.46504	0.182349	0.174922	0.139704	0.07979
8	0.689166	0.686932	0.534504	0.458062	0.801822	0.716508	0.616022	0.69947	0.63015	0.50748	0.54352	0.729082	91.77487	21.13086	0.001821	0.002041	0.001643	0.001327
9	0.538653	0.655673	0.571634	0.523218	0.600089	0.552792	0.556076	0.614316	0.535167	0.514874	0.522921	0.608575	149.3014	44.37619	0.171359	0.200451	0.189772	0.168346
10	0.474265	0.382225	0.459537	0.540791	0.503772	0.608711	0.504953	0.401289	0.360799	0.427362	0.51008	0.432267	0.12354	93.00485	5.26E-07	1.61E-06	1.73E-06	1.23E-06
11	0.330715	0.419457	0.358616	0.298958	0.351895	0.368575	0.440659	0.442566	0.464749	0.411754	0.410802	0.483531	17.63582	7.985379	0.02239	0.026255	0.028019	0.02516
12	0.360483	0.325977	0.46544	0.398544	0.444369	0.481422	0.519514	0.414698	0.605985	0.786623	0.526968	0.339225	0.359643	11.84443	0.006397	0.006613	0.005789	0.006933
13	0.493732	0.574134	0.521128	0.445576	0.422728	0.398856	0.481248	0.706493	0.775649	0.627658	0.696882	0.776313	99.61003	10.76	0.05228	0.050135	0.043568	0.048465
14	0.470844	0.444825	0.373913	0.41989	0.38609	0.337439	0.367221	0.51007	0.541657	0.437294	0.404412	0.420085	1.651544	22.12412	0.055377	0.045359	0.048508	0.032925
15	0.543407	0.488654	0.494951	0.500554	0.5631	0.573342	0.511611	0.490795	0.499816	0.514764	0.555958	0.534366	0.003489	0.580688	0.000138	0.000211	0.000317	0.00029
16	0.394121	0.545147	0.423545	0.42348	0.593704	0.647074	0.433854	0.408588	0.357044	0.356403	0.264195	0.388409	1.108978	13.54188	0.000831	0.00095	0.000857	0.000801
17	0.334222	0.277214	0.304116	0.284	0.325836	0.405081	0.417769	0.396376	0.460126	0.495104	0.616565	0.421234	11.81119	19.51696	0.019462	0.02273	0.021884	0.019256
18	0.510559	0.497537	0.459487	0.457975	0.588498	0.618203	0.678372	0.715285	0.754543	0.717141	0.632063	0.576887	175.7261	9.501726	0.122828	0.102212	0.034	0.026048
19	0.390787	0.582078	0.482396	0.351842	0.588884	0.694692	0.544835	0.533387	0.575839	0.462801	0.358703	0.443796	13.15731	173.1166	0.836487	0.951712	0.926344	0.779757

Table 2 Low level audio features

- Lyrics_features

	mean_syll	mean_wo	n_sentenc	n_words	sentence_	track_id	vocabulary_wealth
0	-1	-1	-1	-1	-1	5KIfHjHI5N	-1
1	1.1	5.65	31	326	0.043011	13keyz9ikI	0.45
2	1.37	4.77	74	532	0.050352	1WugzepX	0.59
3	1.95	3.38	72	430	0.02856	2MO6oEA	0.49
4	1.16	2.99	68	368	0.047849	1i4St7fmS	0.47
5	1.32	4.21	39	256	0.040486	3UyfvY3Gs	0.6
6	1.24	2.98	50	287	0.038367	2zhk0kypC	0.49
7	1.32	3.44	86	524	0.050068	3pjampEJV	0.59
8	1.16	3.11	38	194	0.083926	3xBGFVU2	0.58
9	1.86	2.94	34	187	0.048128	4JbTcgrkSv	0.62
10	1.4	3.53	53	320	0.056604	5wAhrduF	0.52
11	1.13	2.95	43	212	0.062016	5uPWILEX	0.54
12	1.13	3.25	76	398	0.076491	6jRNplNtd	0.36
13	1.18	3.88	48	306	0.062943	0x30gzRkb	0.42
14	2.09	5.52	23	226	0.071146	7rAeisif8x	0.69
15	1.4	3.53	53	320	0.056604	4vV5HmD'	0.52
16	1.2	3.65	40	234	0.101282	0E4PLIdbL	0.53
17	1.24	3.65	34	203	0.010695	28C4nmrd	0.72
18	1.36	5.04	85	574	0.054622	4BxfXnGbr	0.44
19	1.87	4.12	17	137	0.044118	24U7yMeI	0.62

Table 3 Lyrics features

- Spotify_albums

	album_typ	artist_id	available_	external_u	href	id	images	name	release_date	release_da	total_trac	track_id	track_nam	uri	type
0	single	3DiDSECU	['AD', 'AE',	{'spotify':	https://ap	1gAM7M4	[{'height':	If I Ain't Got You	2/8/2019	day	6	2iejTMy9X	track_32	spotify:alt	album
1	album	6s1pCNXc	['AD', 'AE',	{'spotify':	https://ap	4KfJZV7W	[{'height':	Shostakovich Syn	3/1/2019	day	8	1WQfghEj	track_11	spotify:alt	album
2	single	5YjfNaHqC	['AD', 'AE',	{'spotify':	https://ap	7nLYY7uA	[{'height':	Take My Bass	3/14/2019	day	1	3jJKj4QTK	track_15	spotify:alt	album
3	single	2G9Vc16JC	['AD', 'AE',	{'spotify':	https://ap	6p20Rt4x2	[{'height':	Hypnotizing (Are	11/16/2016	day	1	1xGtDafU2	track_46	spotify:alt	album
4	single	2dwM9Oc	['AD', 'AE',	{'spotify':	https://ap	1XeoOqC1	[{'height':	Sunshine	7/20/2018	day	1	0gWtsXvX	track_10	spotify:alt	album
5	single	6bwXuNL4	['AD', 'AE',	{'spotify':	https://ap	0LFnBypm	[{'height':	Moderniste - Sor	3/1/2019	day	1	5rlcRoYi2F	track_8	spotify:alt	album
6	album	5y9NnD1A	['AD', 'AE',	{'spotify':	https://ap	4ozx3bix6	[{'height':	An Enduring Voic	3/1/2019	day	11	6XW8jOID	track_16	spotify:alt	album
7	single	0W03t1E2	['AD', 'AE',	{'spotify':	https://ap	5mwUwm	[{'height':	Quiet	3/15/2019	day	1	6BHHQJCj	track_19	spotify:alt	album
8	album	7GRrLcUx	['AD', 'AE',	{'spotify':	https://ap	2BlcdDBE	[{'height':	High Visceral, Pt.	3/19/2016	day	10	7sP57RtB3	track_49	spotify:alt	album
9	single	5j1NzKTo	['AD', 'AE',	{'spotify':	https://ap	2WYcjaqZ	[{'height':	AutÃ©ntica Llane	8/17/2018	day	1	2FCIggXNc	track_21	spotify:alt	album
10	album	2wOqMjp	['AD', 'AE',	{'spotify':	https://ap	4tpxeDnp	[{'height':	Beethoven: 6 Bag	3/1/2019	day	11	0vF8nd5yJ	track_10	spotify:alt	album
11	album	13iBt0Nnv	['AD', 'AE',	{'spotify':	https://ap	6ijzOqTgw	[{'height':	Au Long de la Loi	3/1/2019	day	17	5v6RFQXo	track_17	spotify:alt	album
12	single	0n255T84	['AD', 'AE',	{'spotify':	https://ap	2tgRVRHy2	[{'height':	I Still Miss U	3/15/2019	day	1	1XHUsvM	track_21	spotify:alt	album
13	album	3bLCTs7b	['AD', 'AE',	{'spotify':	https://ap	3vlQK6zfk	[{'height':	Kolmekymppiner	1980	year	12	3v4x9Hf9	track_5	spotify:alt	album
14	single	1CJIcvG0A	['AD', 'AE',	{'spotify':	https://ap	1zjJie8cQ9	[{'height':	Bruja	4/24/2018	day	1	45RBx5W	track_22	spotify:alt	album
15	album	3vrMRlMs	['AD', 'AE',	{'spotify':	https://ap	5FF5Bvr1C	[{'height':	Sonatas for two v	3/1/2019	day	14	2r3q57Fh	track_12	spotify:alt	album
16	album	4NsS37Ec	['AD', 'AE',	{'spotify':	https://ap	6sUSCnAh	[{'height':	SirÃ¨nes	3/1/2019	day	18	0eeQWYlr	track_18	spotify:alt	album
17	single	5jO3o2nsu	['AD', 'AE',	{'spotify':	https://ap	7FJ9me2sI	[{'height':	Light of Day	3/1/2019	day	2	4QZpZd7d	track_29	spotify:alt	album
18	single	1nh1tWk4	['AD', 'AE',	{'spotify':	https://ap	7zG1CY6e	[{'height':	Madame del Can	8/3/2015	day	1	69OItPzxD	track_39	spotify:alt	album
19	album	2wOqMjp	['AD', 'AE',	{'spotify':	https://ap	1J1GKjKa4	[{'height':	Beethoven: Pianc	3/1/2019	day	7	3fqR3ZwG	track_19	spotify:alt	album

Table 4 Spotify albums

- Spotify_artists

	artist_pop	followers	genres	id	name	track_id	track_nam	type
0	44	23230	['sertanejo	4mGnpjhqgx4RU	Juliano Cezar	0wmDmAILu	track_9	artist
1	22	313	[]	1dLnVku4VQUOL	The Grenadines	4wqwj0gA8q	track_30	artist
2	26	1596	['danish p	6YVY310fjfUzKi8h	Gangway	1bFqWDbvH	track_38	artist
3	31	149	['uk altern	2VElyouiCfoYPDJ	FADES	3MFSUBAidP	track_34	artist
4	21	11	['french ba	4agVy03qW8juSy	Jean-Pierre Guignon	2r3q57Fhxds	track_26	artist
5	43	81	[]	38VBjthd0szbS6v	Filhos	453KeZU566h	track_15	artist
6	34	8358	[]	36mHwYa65L0W	Eloq	6hC5Tl0S5aC	track_43	artist
7	7	158	[]	1jJyy00XfxjB4tMA	FravÃ¦r	5005eeCJ9KT	track_48	artist
8	21	30	[]	10A8WbBJ0zW8I	Camille PÃ©pin	56fBsCXwpBr	track_73	artist
9	25	992	['classic fir	5ijoPCUhV0dW8E	Pepe Willberg & The F	6WVlVPoesX	track_8	artist
10	4	44	[]	7dUWZW268V8q	Marie Bergman & Las	4AXl1EfaAgk	track_51	artist
11	53	373085	['pagode',	0ozdcqmICsOZDe	Rick & Renner	1x0nDKfYDec	track_1	artist
12	21	1	[]	0GUBxLwTer6gJZ	Fiona McGown	56fBsCXwpBr	track_74	artist
13	11	115	[]	60uB6JvZMMLXg	Silhuetit	2K9NF7Z0uw	track_25	artist
14	16	498	[]	4uQXIJKWCNEGu	Pligten Kalder	6xQZaoYaK7(	track_52	artist
15	43	1029	[]	2zYNTwc8lCUzPu	BAER	7E0vigeBuAlL	track_1	artist
16	21	3	[]	6m099EAiKxVRkr	Ensemble Polygones	56fBsCXwpBr	track_76	artist
17	46	30344	['abstract	4sH7SLGP6BLOYI	P.O.S	5TXuAkgAYyc	track_3	artist
18	41	24922	['sertanejo	0wWNGmLoY3KE	MilionÃ¡rio & Marcian	2V68QXKzrgF	track_7	artist
19	34	644	[]	69APTdcqAgiLyR	Lydia Liza	5TXuAkgAYyc	track_9	artist

Table 5 Spotify artists

- Spotify_track

	acousticne	album_id	analysis_u	artists_id	available_	country	danceabili	disc_numl	duration_	energy	href	id	instrumen	key	liveness	loudness	lyrics	mode	na
0	0.294	0D3QufeC	https://ap	['3mxJuHF	['AD', 'AE',	BE	0.698	1	235584	0.606	https://ap	5qljLQuKn	2.69E-06	10	0.151	-7.447		0	Blc
1	0.863	1bcqsH5U	https://ap	['4xWMev	['AD', 'AE',	BE	0.719	1	656960	0.308	https://ap	3VAX2MJc	0	6	0.253	-10.34		1	Th
2	0.75	4tKijjmxG	https://ap	['3hYaK5F	['GB']	BE	0.466	1	492840	0.931	https://ap	1L3YAhsEI	0	4	0.938	-13.605		0	Jin
3	0.763	6FeJF5r8r	https://ap	['2KQsUB9	['AD', 'AE',	BE	0.719	1	316578	0.126	https://ap	6aCe9zzoz	0	3	0.113	-20.254		0	Th
4	0.77	4tKijjmxG	https://ap	['3hYaK5F	['GB']	BE	0.46	1	558880	0.942	https://ap	1Vo802A3	0	7	0.917	-13.749		1	Se
5	0.971	7CCwkPw	https://ap	['3kzwYV3	['AE', 'AR',	BE	0.367	1	183653	0.349	https://ap	4PrAZpH9	0.296	11	0.633	-7.74		1	Al
6	0.824	51g5viCaY	https://ap	['4xWMev	['AD', 'AE',	BE	0.688	1	29240	0.304	https://ap	1WJzRtl1A	0	10	0.142	-9.96		1	Th
7	0.719	4tKijjmxG	https://ap	['3hYaK5F	['GB']	BE	0.513	1	591160	0.889	https://ap	6ileHaSaG	0	3	0.722	-13.47		0	Th
8	0.752	1UMRkrYS	https://ap	['5uNTrZ1.	['AD', 'AE',	BE	0.629	1	753947	0.194	https://ap	6rlEcNrUC	0	5	0.144	-26.588			

Table 6 Spotify track

2. Data modeling

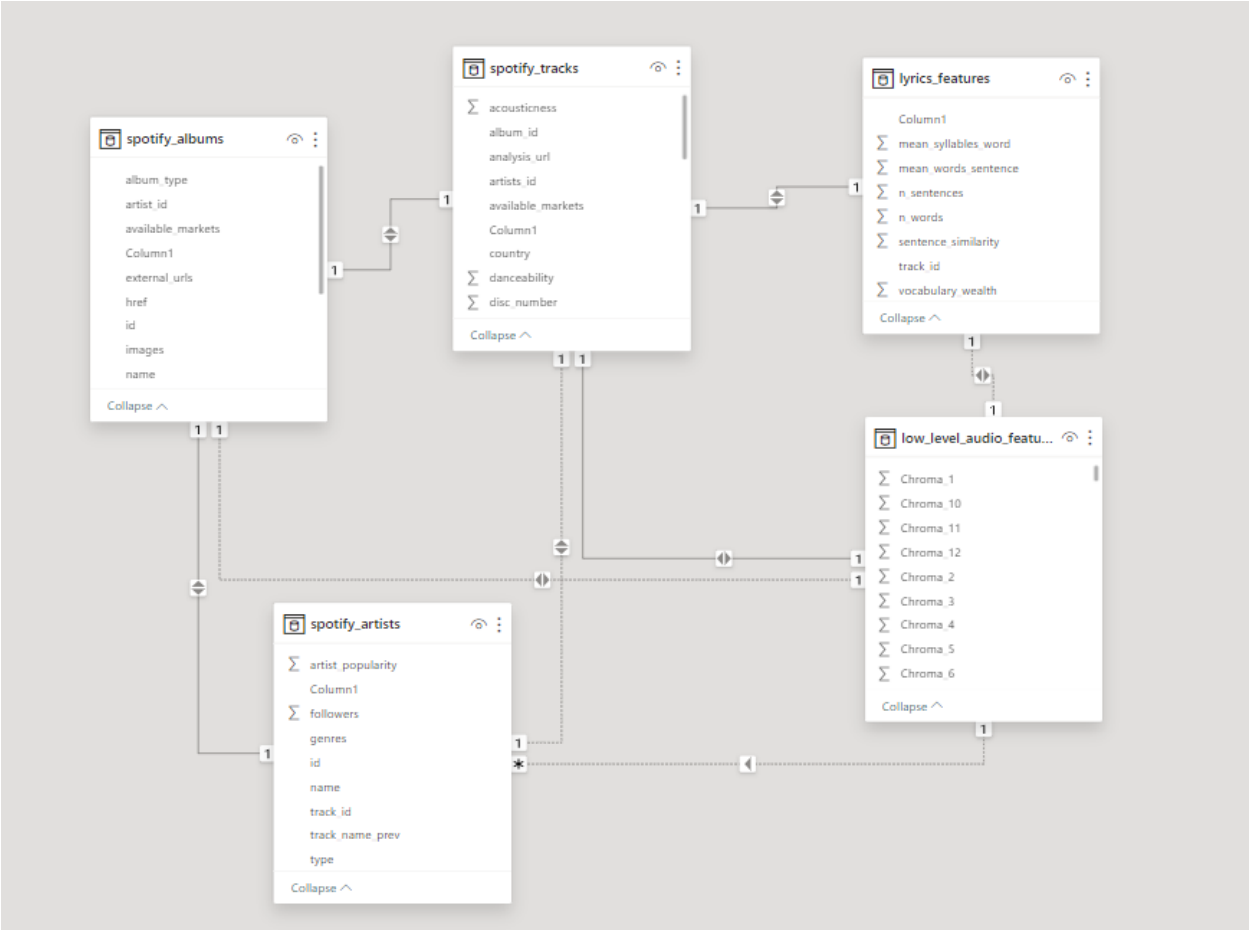


Figure 1

III. Implementation

1. Connect Dataset

There are some steps to connect Power BI with MySQL:

1. Open Wamp server

Figure 2

2. Open MySQL Workbench

Figure 3

3. Create a new schema

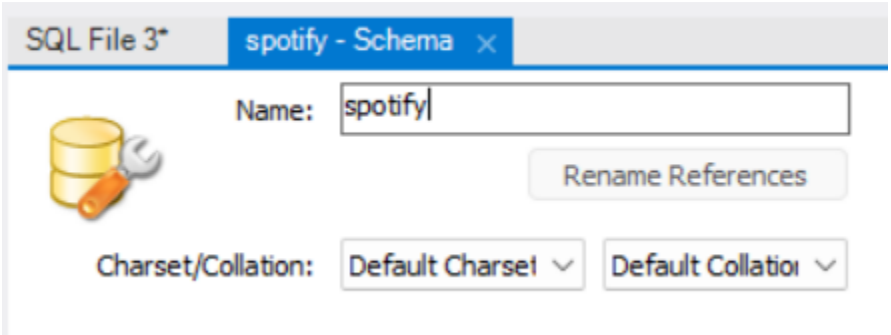


Figure 4

4. Import all CSV files of our dataset in the table of schema above

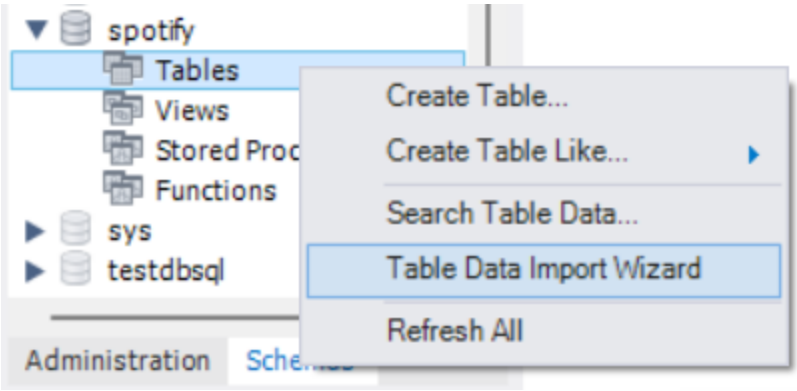


Figure 5

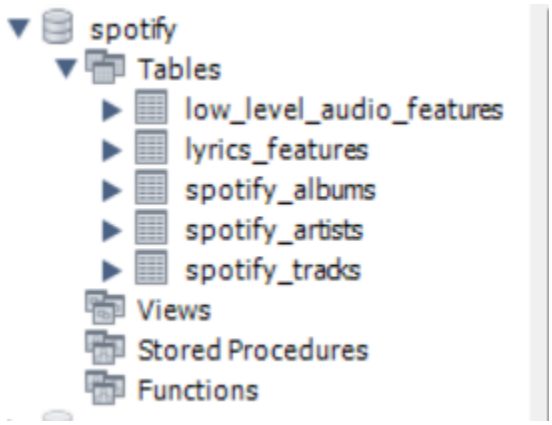


Figure 6

5. After we got all the datasets we needed to open Power BI

Figure 7

6. Click on Get Data

Figure 8

7. Search MySQL then click connect

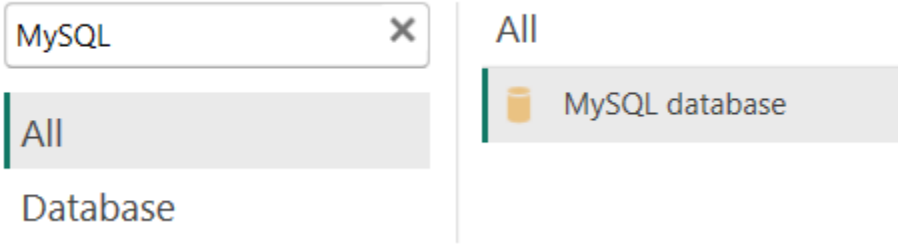


Figure 9

8. Enter the Server and Database name

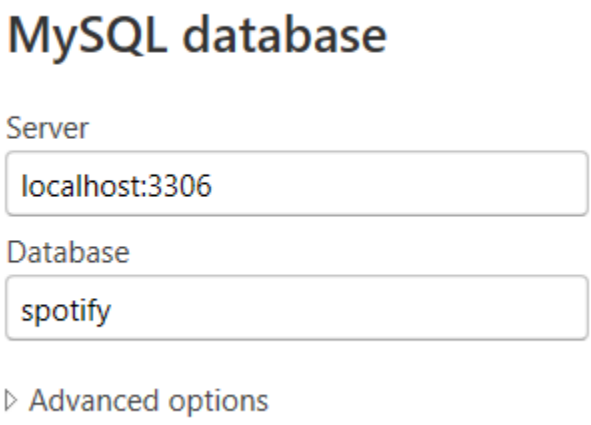


Figure 10

9. Then click OK and load our data.

2. Data Preprocessing

Column1	mean_syllables_word	mean_words_sentence	n_sentences	n_words	sentence_similarity
0	-1	-1	-1	-1	-1
36	-1	-1	-1	-1	-1
38	-1	-1	-1	-1	-1
41	-1	-1	-1	-1	-1
43	-1	-1	-1	-1	-1
45	-1	-1	-1	-1	-1
46	-1	-1	-1	-1	-1
48	-1	-1	-1	-1	-1
50	-1	-1	-1	-1	-1

Figure 11

We use the MICE method to clean by inputting the wrong values of data by average value. But for column “sentence_similarity” we change from -1 to 0 because similarity is a percentage so it must be between 0 and 1.

Column1	mean_syllables_word	mean_words_sentence	n_sentences	n_words	sentence_similarity
0	1.05	3.6	42.3	317	0
36	1.05	3.6	42.3	317	0
38	1.05	3.6	42.3	317	0
41	1.05	3.6	42.3	317	0
43	1.05	3.6	42.3	317	0
45	1.05	3.6	42.3	317	0
46	1.05	3.6	42.3	317	0
48	1.05	3.6	42.3	317	0
50	1.05	3.6	42.3	317	0

Figure 12

3. The cards show the Total of (track, album_type, and Artists)

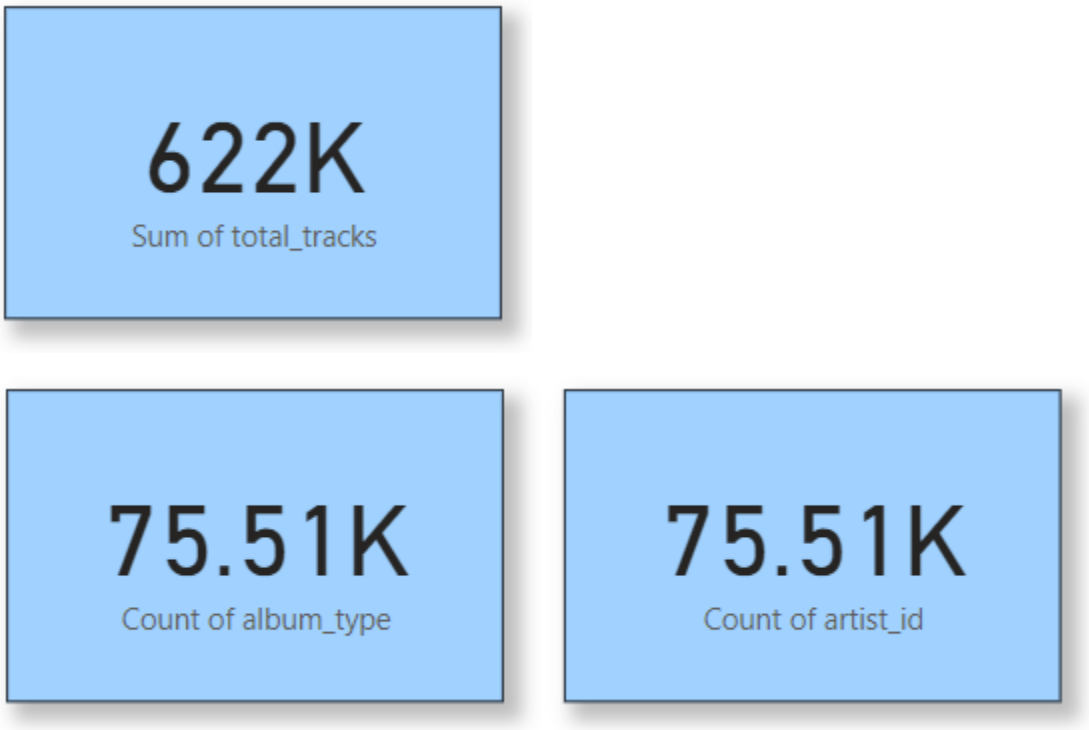


Figure 13

In this dashboard, we include the card type to show the total amount of tracks, album_type, and artist. We use card types to illustrate the amount of sum numbers to the audience more easily and also to make sure that it is easy to analyze the data accurately. The first card shows the total amount of tracks that are available in spotify. Second card illustrates the total album of all artists on the platform. The last card tells us about the amount of artists that cooperate with spotify.

4. Top 5 Popular Artists

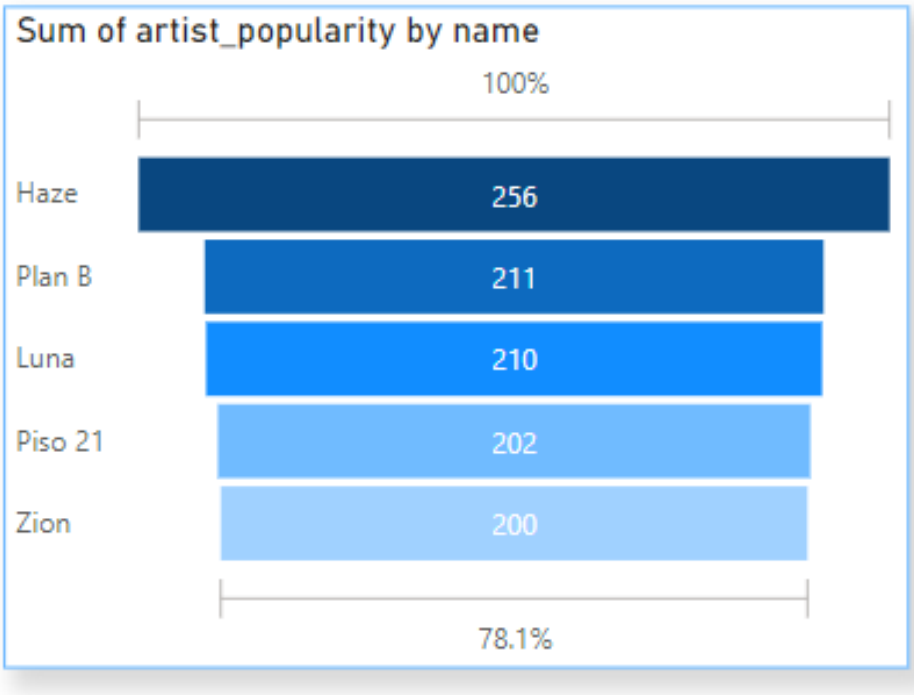


Figure 14

The graph above displays an artist's popularity and reputation by collecting data from external platforms based only on the artist's real-world popularity, without getting involved with follower amounts. The graph illustrates that the most popular artists are Haze, Plan B, Luna, Piso 21, and Zion which are the top 5 popular artists.

5. Top 5 By The MostArtists

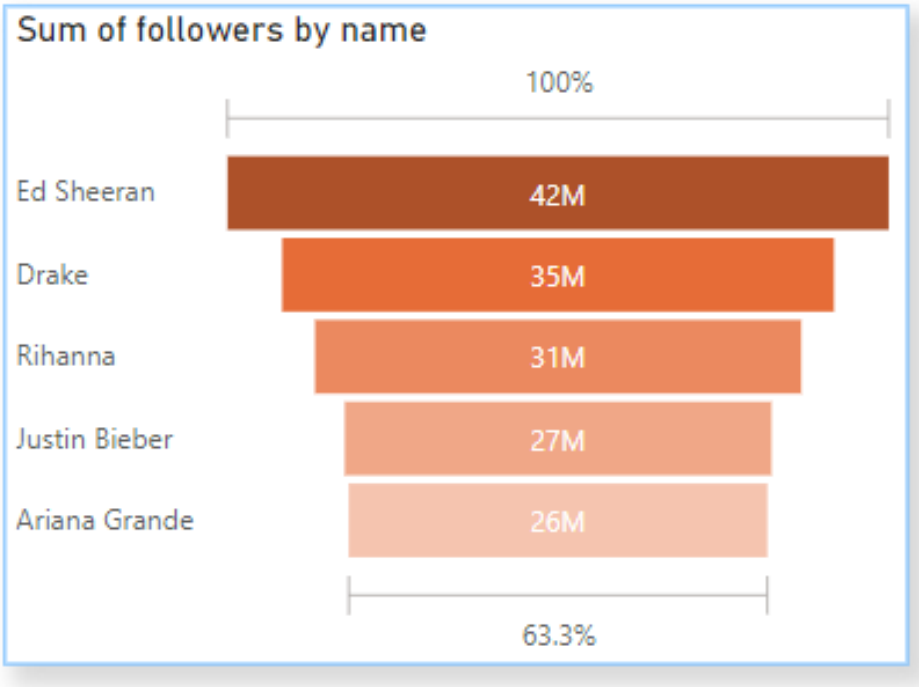


Figure 15

This graph is similar to the last graph which shows the popularity of artists. However, this graph shows the popularity of artists as well but it calculates the amount of followers that each artist has. This graph shows the top 5 followers on spotify platform. As we can see, Ed Sheeran is the artist that has the most followers with more than 40 million followers.

6. Number Albums by Year

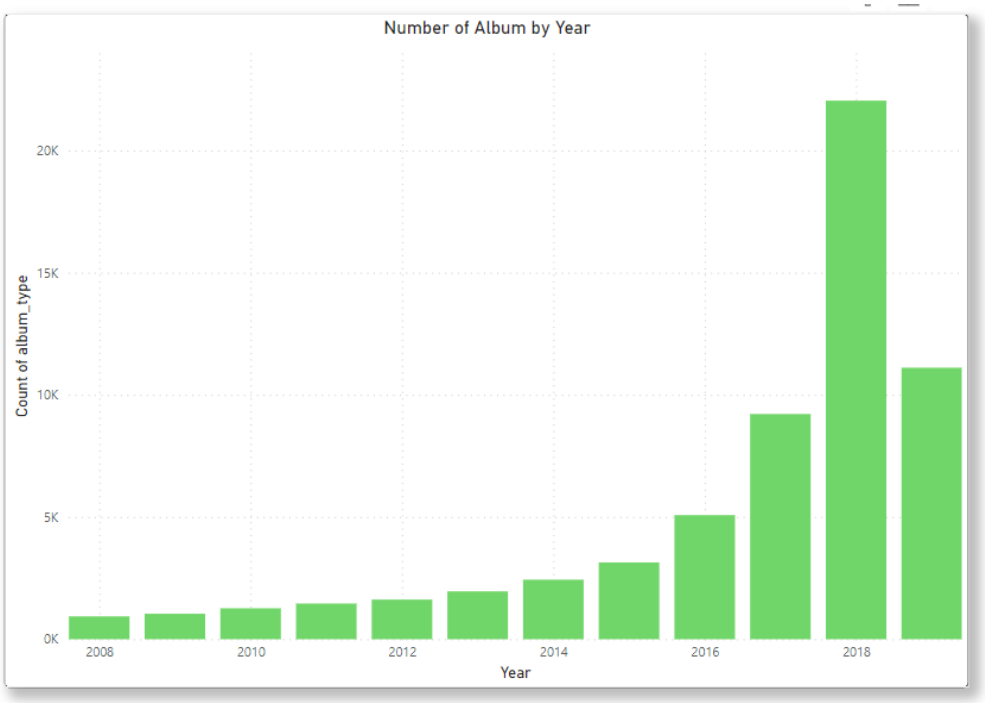


Figure 16

This bar chart illustrates the number of albums that have been released each year in Spotify since 2008 until 2019. The above bar chart shows that the number of albums have rise dramatically since 2016 and reach the peak of around 20k albums in 2018. The reason that we use bar charts to illustrate this data is because we think that by using bar charts it will be easy for us to visualize and analyze the data as we want to find the correlation between time and album.

7. Sentence That Contains N-Word by Artist

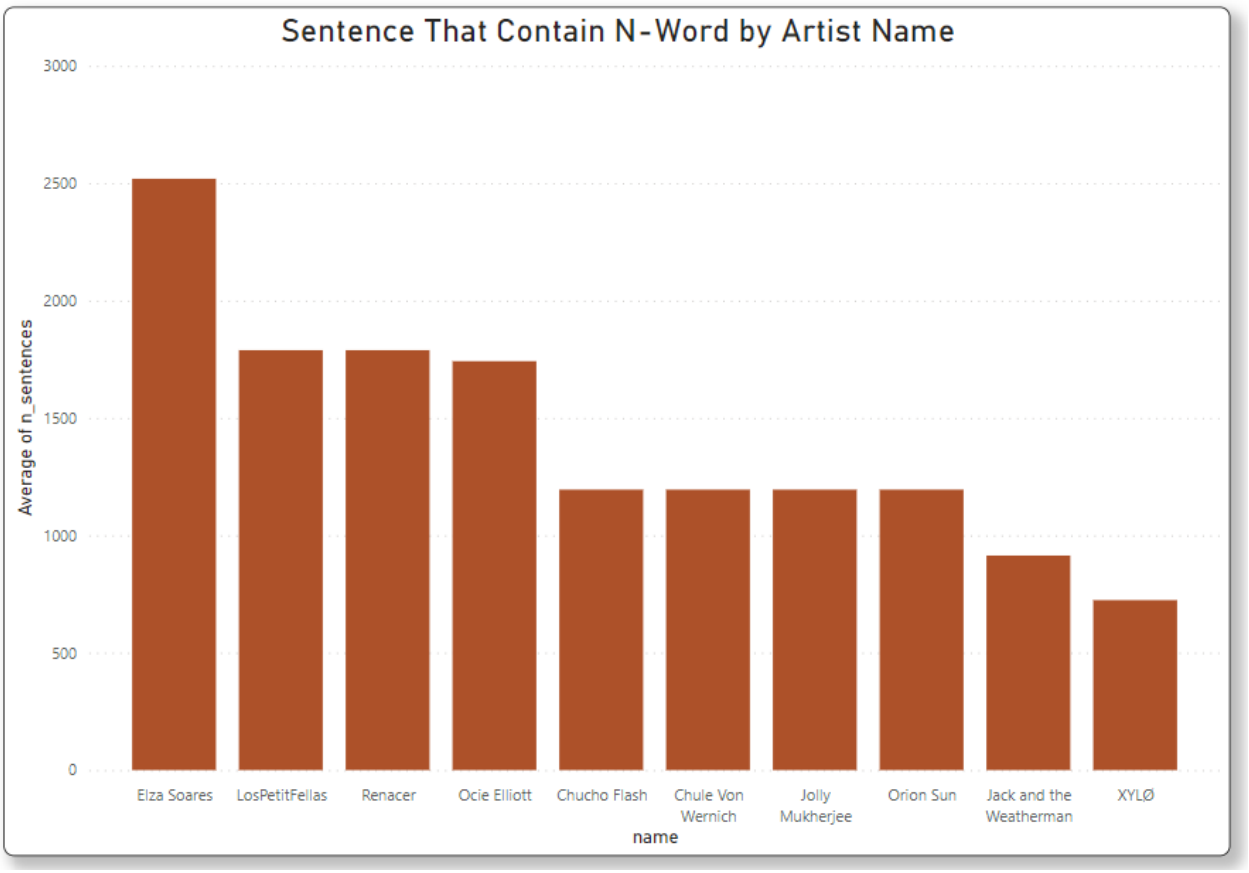


Figure 17

This chart shows the number of n-word that used by each artist. As a result, we can see that Eiza Soares is the artist that used the most n-word. The reason that we want to detect the amount of n-word that each artist used because we want to create a function which can provide information for customers to acknowledge about the content of each music and artist.

8. Top 10 Artists by Total Duration in Song

Artist Name	Duration (ms)
Danny Avila	5505831
Ernesto Amstein	4694040
Cor Veleno	4667397
Santhanam	4508318
Yall	4500037
Familia Torres	4462995
Deeb	4437892
Malk De Koijn	4423000
Zuma Dionys	4373004
	216000
Total	**216000**

Figure 18

This table represents the duration of time that each artist had in spotify platform. It means that the more amount of artist had the more music and time that artist involve more with the platform. This will shows that the company have to make a new strategies to cooperate and work more with those who had least time on the platform to improve the quality and satisfy customer needed.

9. Regions that Spotify Spread

Figure 19

Above map illustrates the total number of available markets of spotify around the globe. According to the data, it shows that nowadays spotify has work and cooperate with 180 countries which means that spotify is available and accessible in those 180 countries.

IV. Conclusion and Recommendation

1. Conclusion

As technology continues to evolve, the future of music streaming platforms will likely bring about significant improvements in user experience, music discovery, and content management. These advancements will contribute to a more personalized and enjoyable music streaming experience for users. The music industry, as a whole, will also experience changes as these platforms gain more influence and prominence. However, there are challenges that these platforms must overcome to fully harness their potential.

In conclusion, the future of music streaming platforms will likely be marked by continued innovation and improvement, benefiting both users and the music industry. However, these platforms must overcome various challenges to fully realize their potential and deliver a seamless, enjoyable, and personalized music streaming experience for users.

2. Recommendation

For the recommendation we would like to enhance the user experience and promote artist discovery, we recommend implementing the following features on the music streaming platform such as:

1. Machine Learning-based Recommendations: Implement machine learning algorithms to recommend artists and songs to users based on their mutual listening habits. This feature will broaden the musical horizons of listeners and increase exposure for lesser-known musicians.

2. Explicit Mode: Provide a mode for users to access explicit-free content, demonstrating a commitment to inclusivity and ensuring a comfortable listening experience for all users.

3. User Preferences and Filters: Allow users to specify their preferences, favorite genres, and artists, and enable them to filter out songs they do not wish to hear. This level of customization enhances user engagement and satisfaction, as it ensures that users are presented with content that aligns with their individual tastes and preferences.

4. Genre-based Artist Promotion: Expand the popularity of artists based on their genre by featuring and promoting artists within specific genres, thereby increasing their visibility and introducing users to new music within their preferred genres.

Reference

https://www.kaggle.com/datasets/saurabhshahane/spotgen-music-dataset