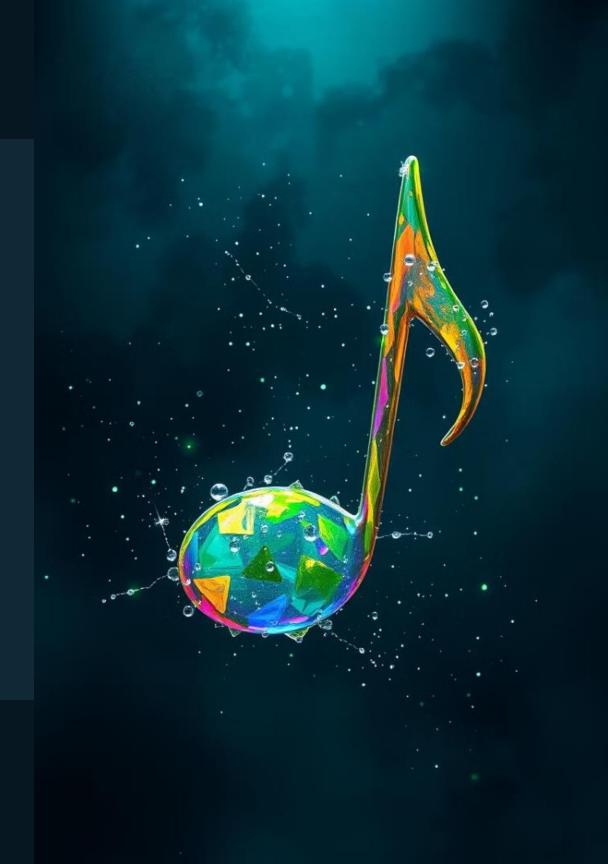


# Spotify Data Analysis

---by Gitanjali Pekamwar



## Introduction

In today's digital age, music streaming platforms like Spotify have revolutionized how we consume and discover music. This project delves into a dataset of Spotify audio features, aiming to uncover insights using SQL (Structured Query Language) into the characteristics of popular music and explore potential relationships between audio attributes and listener preferences.



# Table of Contents

1 Introduction

Setting the stage for our analysis and outlining the project's objectives.

3 Business Problem

Identifying the real-world business questions we aimed to address through data analysis.

5 Insights

Unveiling the key findings and trends discovered through our data exploration.

7 Tech Stack

A glimpse into the tools and technologies employed in this project.

4 Querying & Analysis

Diving deep into the data extraction and analysis techniques utilized.

6 Conclusion

Summarizing the key takeaways and highlighting the project's impact.

# **Technology Used**







Gamma



MS Power-Point



## Database & Source

## Spotify Dataset

The project utilizes a publicly available dataset containing millions of Spotify music tracks, encompassing information like artist, genre, release date, and popularity metrics.

#### Data Source

The dataset was sourced from Kaggle, a platform for data science and machine learning competitions. It provides a rich resource for exploring music trends and patterns.



# **Business Problem**

**Music Trends** 

Understanding current music trends, emerging artists, and popular genres.

Audience Insights

Identifying audience preferences, listening habits, and demographic patterns.

Marketing Strategies

Developing targeted marketing strategies based on datadriven insights.

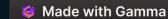
2

3



## ☐ Table Creation

```
DROP TABLE IF EXISTS spotify;
3
4 CREATE TABLE spotify (
         artist VARCHAR(255),
5
         track VARCHAR(255),
6
         album VARCHAR(255),
         album_type VARCHAR(50),
         danceability FLOAT,
         energy FLOAT,
10
         loudness FLOAT,
11
         speechiness FLOAT,
12
         acousticness FLOAT,
13
         instrumentalness FLOAT,
14
15
         liveness FLOAT,
         valence FLOAT,
16
17
         tempo FLOAT,
         duration_min FLOAT,
18
         title VARCHAR(255),
19
         channel VARCHAR(255),
20
         views FLOAT,
21
         likes BIGINT,
22
         comments BIGINT,
23
24
         licensed BOOLEAN,
         official_video BOOLEAN,
         stream BIGINT,
26
         energy_liveness FLOAT,
27
         most_played_on VARCHAR(50)
28
29
    );
```



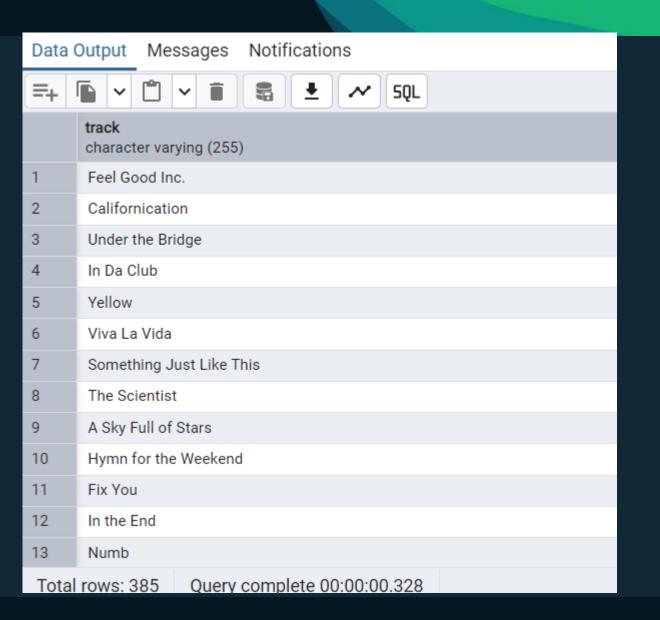
## ☐ Data Exploration (EDA)

```
-- EDA
33
    select count(*) from spotify;
34
35
     select count( Distinct artist) from spotify;
36
37
     select count( Distinct album) from spotify;
38
39
     select Distinct album_type from spotify;
40
41
     select max(duration_min) from spotify;
42
43
    select min(duration_min) from spotify;
44
45
    select * From spotify
     where duration_min=0 ;
47
48
    Delete from spotify
     where duration_min = 0;
50
51
     select count( Distinct channel) from spotify;
52
53
     select Distinct most_played_on from spotify;
54
55
     select count( Distinct title) from spotify;
56
```



• Task1) Retrieve the names of all tracks that have more than 1 billion streams.

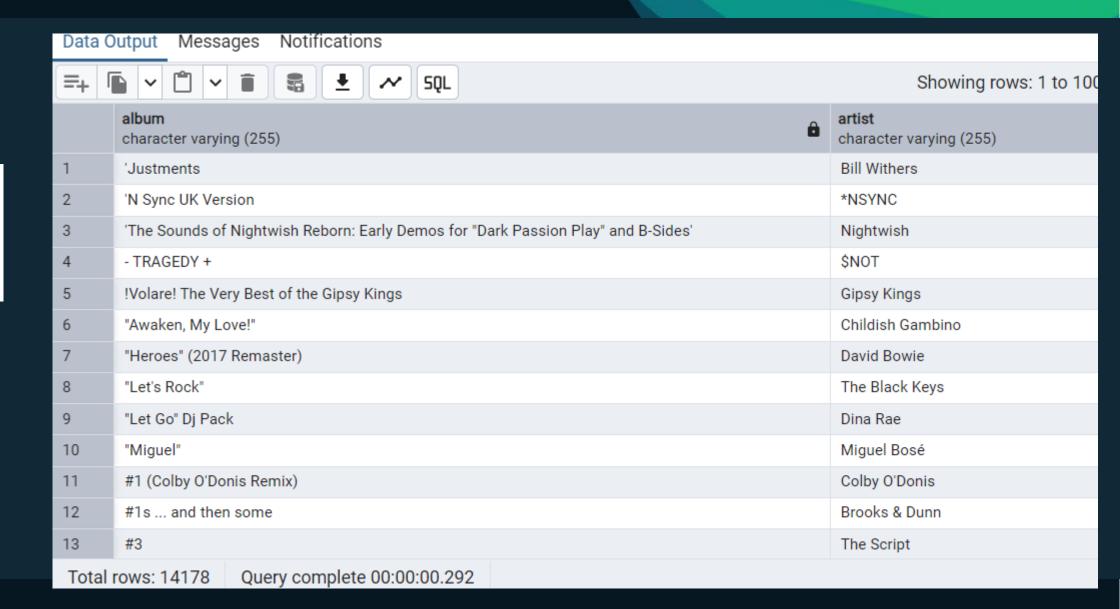
```
61
62 select track
63 from spotify
64 where stream > 1000000000
65 ;
66
```



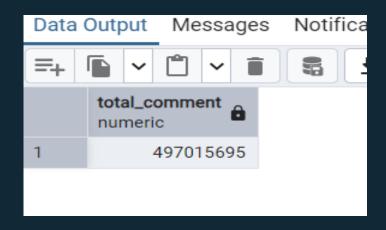


## Task2)List all albums along with their respective artists.

```
69 v select Distinct album,
70 artist
71 from spotify
72 order by 1;
73
```



Task3)Get the total number of comments for tracks where licensed = TRUE.

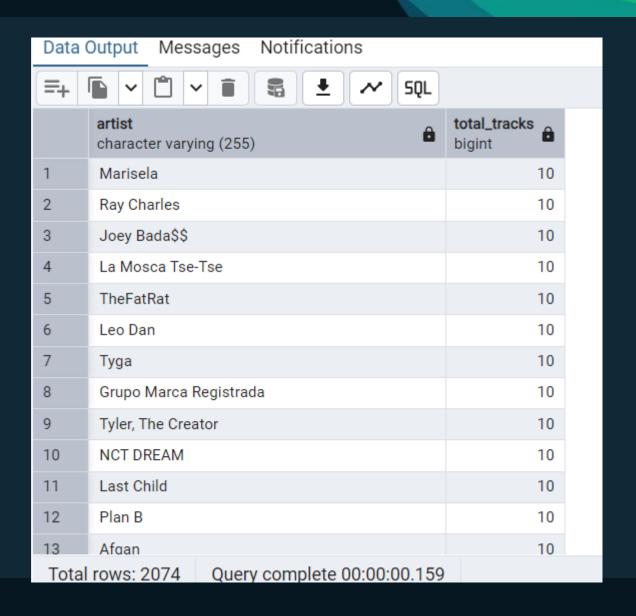


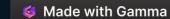


Task4)Find all tracks that belong to the album type single.

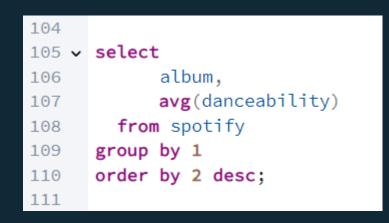


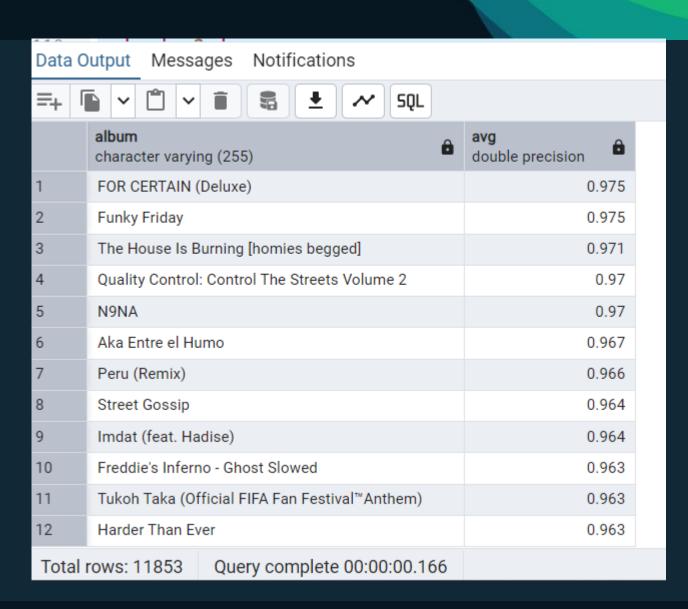
Task5)Count the total number of tracks by each artist





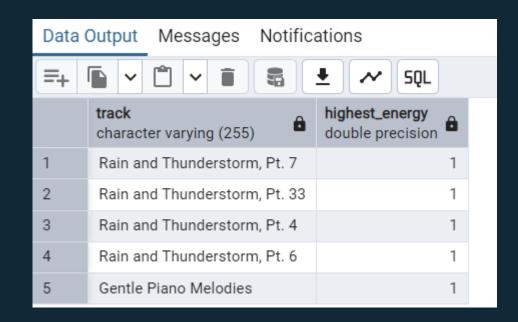
Task6)Calculate the average danceability of tracks in each album.





Task7)Find the top 5 tracks with the highest energy values.

```
select
    track,
    max(energy) as highest_energy
  from spotify
group by 1
order by 2 desc
limit 5;
```





 Task8) List all tracks along with their views and likes where official\_video = TRUE.

```
select track,
sum(views) as total_views,
sum(likes) as total_likes
from spotify
where official_video = 'true'
group by 1
order by 2 desc
limit 5;
```

Data Output Messages Notifications					
=+					
	track character varying (255)	total_views double precision	total_likes numeric		
1	Despacito	16159296273	101577278		
2	See You Again (feat. Charlie Puth)	11547595554	80295292		
3	Shape of You	5908398479	31047780		
4	Calma - Remix	5322011392	25649519		
5	This Is What You Came For	5252059812	21207312		

• Task9) For each album, calculate the total views of all associated tracks.

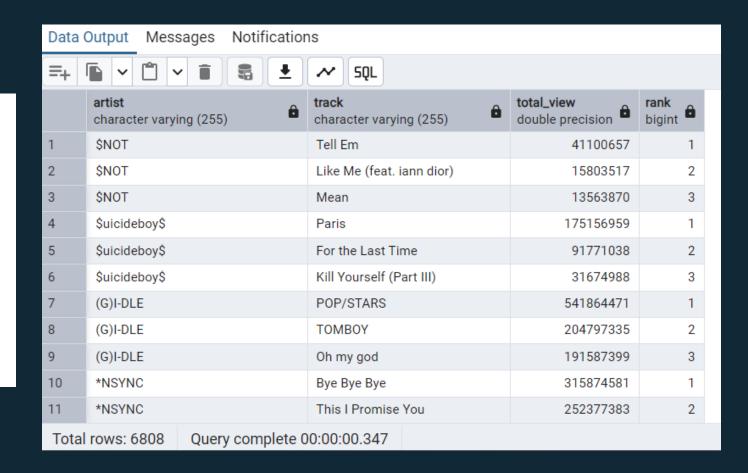
Data Output Messages Notifications						
=+						
	album character varying (255)	track character varying (255)	total_views double precision			
1	VIDA	Despacito	16159296273			
2	See You Again (feat. Charlie Puth)	See You Again (feat. Charlie Puth)	11547595554			
3	Peace Is The Mission (Extended)	Lean On	9974504694			
4	÷ (Deluxe)	Shape of You	5908398479			
5	MUNAY	Calma - Remix	5322011392			
6	This Is What You Came For	This Is What You Came For	5252059812			
7	Swalla (feat. Nicki Minaj & Ty Dolla \$ign)	Swalla (feat. Nicki Minaj & Ty Dolla \$ign)	5162403752			
8	CoComelon Kids Hits, Vol. 1	Wheels on the Bus	4898831101			
9	Taki Taki (with Selena Gomez, Ozuna & Card	Taki Taki (with Selena Gomez, Ozuna & Cardi B)	4828403435			
Total rows: 18680 Query complete 00:00:00.165						

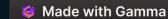
• Task10) Retrieve the track names that have been streamed on Spotify more than YouTube.

Data Output Messages Notifications			
=+	<b>□ ∨ □ ∨ ■ ■ ★  ✓ SQL</b>		
	track character varying (255)		
1	Usted		
2	21 Hungarian Dances, WoO 1: Hungarian Dance No. 5 in G Minor. Allegro (Orch. Schmeling		
3	Mientes Tan Bien		
4	Have You Ever Seen The Rain		
5	Dream A Little Dream Of Me		
6	When I Grow Up		
7	Me Hace Tanto Bien		
8	What You Want (feat. Total)		
9	Suavemente		
10	Emotions		
Total	rows: 155 Query complete 00:00:00.127		

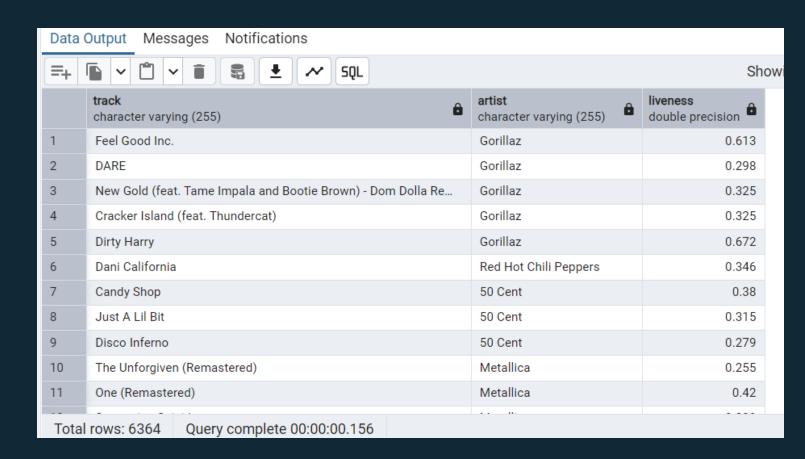


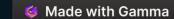
# Task11) Find the top 3 most-viewed tracks for each artist using window functions.





• Task12) Write a query to find tracks where the liveness score is above the average.





 Task13) Use a WITH clause to calculate the difference between the highest and lowest energy values for tracks in each album

```
204 v with cte
205
206
      select album,
207
             max(energy) as highest_energy,
208
             min(energy) as lowest_energy
209
      from spotify
210
      group by 1
211
212
213
      select album,
214
             highest_energy - lowest_energy as energy_diff
215
      from cte
      order by 2 desc ;
216
217
```

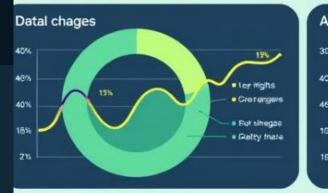
Data O	Data Output Messages Notifications				
=+ • V • SQL					
	album character varying (255) €	energy_diff double precision			
1	White Noise	0.9067500000000001			
2	Spotify Singles - Holiday	0.8360000000000001			
3	Spotify Singles	0.8232			
4	UNDERTALE Soundtrack	0.816			
5	Making Mirrors	0.8109000000000001			
6	Everytime We Touch (Premium Edition)	0.804999999999999			
7	If I Can Dream: Elvis Presley with the Royal Philharmonic Orchestra	0.787			
8	Fallen	0.7649999999999999			
9	CeeLo's Magic Moment	0.762			
10	Greatest Hits	0.741			
11	Safe Haven	0.738			
Total rows: 11853 Query complete 00:00:00.152					

# Insights

- ✓ Firstly, I explored the dataset, Found that there is 20592 of total content on spotify.
- ✓ There are 385 tracks having 1 billion stream.
- ✓ There are 6673 channels on spotify.
- ✓ 'Despacito' track has the highest views and likes.
- ✓ There are 155 tracks have streamed on spotify more than YouTube.
- ✓ There are 6364 tracks having the liveness score above the average.
- ✓ I used CTE, window functions, subqueries and joins to find advanced analysis.

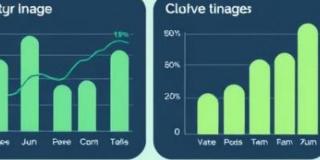
#### beaiztmetts



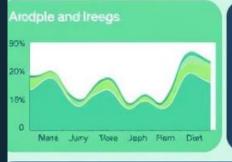


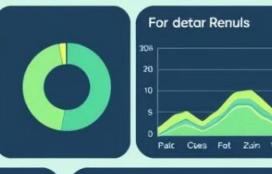


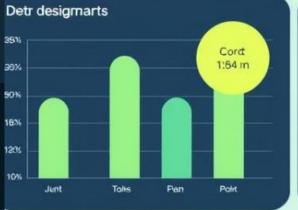


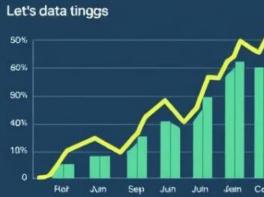












ay 201 Stritefalot.com

## Conclusion

1

#### Data-Driven Insights

Data analysis provides valuable insights into music trends, audience preferences, and marketing strategies.

2

## **Enhanced Decision-Making**

These insights empower music industry stakeholders to make informed decisions regarding artist development, content creation, and marketing campaigns.

3

## **Future Opportunities**

Further analysis can explore the impact of social media, streaming platform algorithms, and user behavior on music trends.





# Thank You

Thank you for your time and interest in this presentation. I hope it provides valuable insights into the power of data analysis in understanding music trends and driving business success.