Spotify

By - Aman Sharma

Upgrade

Aman sharma

Welcome

This project focuses on analyzing Spotify data using SQL to uncover meaningful insights related to user behavior, music trends, and track performance. The primary objective was to explore how listeners engage with different genres, artists, and songs over time, and to derive patterns that could inform data-driven decisions in the music streaming industry. By leveraging SQL, the project emphasizes structured data exploration, trend identification, and the interpretation of listening habits across various dimensions. The analysis aims to provide a comprehensive understanding of what drives engagement on the Spotify platform.

Play

Follow

Upgrade

Aman sharma

Data Overview

The dataset used in this project contains comprehensive information on Spotify tracks, including metadata such as track name, artist, genre, release date, and popularity, along with technical audio features like danceability, energy, tempo, and Acousticness. It covers a broad spectrum of music, making it suitable for analyzing user preferences and streaming Trends. Prior to analysis, the data was cleaned and standardized to ensure consistency. This involved addressing missing values, removing duplicates, and formatting key fields. The prepared dataset enabled efficient querying and robust analysis using SQL.

Project Objective

This project is an end-to-end SQL-based analysis of Spotify's music streaming dataset, designed to uncover meaningful insights about tracks, artists, and listener trends. It begins with structuring and exploring a denormalized dataset and progresses throughwriting and optimizing SQL queries of increasing complexity—from basic aggregations to advanced window functions and CTEs.

The dataset includes detailed attributes such as track and artist names, audio features (e.g., energy, danceability, tempo),engagement metrics (streams, views, likes, comments), and platform-specific details like whether a track is an official video or Licensed.

Key objectives of the project include:

Practicing advanced SQL concepts in a practical setting.

Categorizing and solving 15 real-world music analytics questions.

Gaining hands-on experience with query performance optimization.

Drawing data-driven insights that could support strategic decisions in music Platforms.

The project showcases both technical proficiency and analytical thinking—essential skills for roles in data analysis, business intelligence, and product analytics.

-- Retrieve the names of all tracks that have more than 1 Billion Streams --

```
42
43  -- ------------------------------------------------------------
44  -- Data Analysis - Easy Category
45  -- ------------------------------------------------------------
46
47
48  -- Q1. Retrieve the names of all tracks that have more thaan 1 billion streams. --
49
50  select * from spotify
51  where stream > 1000000000
52
53
```

Data Output | Messages | Notifications

Showing rows: 1 to 385 | Page No: 1 of 1

	artist character varying (255)	track character varying (255)	album character varying (255)	album_type character varying (50)	danceability double precision	energy double p
1	Gorillaz	Feel Good Inc.	Demon Days	album	0.818	
2	Red Hot Chili Peppers	Californication	Californication (Deluxe Edition)	album	0.592	
3	Red Hot Chili Peppers	Under the Bridge	Blood Sugar Sex Magik (Deluxe Edition)	album	0.559	
4	50 Cent	In Da Club	Get Rich Or Die Tryin'	album	0.902	
5	Coldplay	Yellow	Parachutes	album	0.429	
6	Coldplay	Viva La Vida	Viva La Vida or Death and All His Friends	album	0.496	

Table rows counted: 20594

-- List all album along with their respective artists --

```
56
57  -- Q2. List all album along with their respective artists. --
58
59  select
60      distinct album,artist
61      from spotify
62      order by 1
63
64
65
```

Data Output | Messages | Notifications

SQL

	album character varying (255)	artist character varying (255)
1	'Justments	Bill Withers
2	'N Sync UK Version	*NSYNC
3	'The Sounds of Nightwish Reborn: Early Demos for "Dark Passion Play" and B-Sides'	Nightwish
4	- TRAGEDY +	$NOT
5	!Volare! The Very Best of the Gipsy Kings	Gipsy Kings
6	"Awaken, My Love!"	Childish Gambino
7	"Heroes" (2017 Remaster)	David Bowie
8	"Let's Rock"	The Black Keys
9	"Let Go" Dj Pack	Dina Rae
10	"Miguel"	Miguel Bosé

-- Get the total number of comments for tracks where licensed = True --

```
64  -- Q3. Get the total number of comments for tracks where licensed = True --
65
66  select
67        sum(comments) as total_comments
68        from spotify
69  where licensed = 'true'
70
71
72
73
74
75
```

Data Output | Messages | Notifications

SQL

	total_comments numeric
1	497015695

-- Find all tracks that belong to the album type SINGLE --

```sql
-- Q4. Find all tracks that belong to the album type single. --

select * from spotify
where album_type = 'single'
```

Data Output | Messages | Notifications

Showing rows: 1 to 1000 | Page No: 1 of 5

	artist character varying (255)	track character varying (255)	album character varying (255)
1	Gorillaz	New Gold (feat. Tame Impala and Bootie Brown)	New Gold (feat. Tame Impala and Bootie Brown)
2	Gorillaz	New Gold (feat. Tame Impala and Bootie Brown) - Dom Dolla Remix	New Gold (feat. Tame Impala and Bootie Brown) [Dom D
3	Gorillaz	Cracker Island (feat. Thundercat)	Cracker Island (feat. Thundercat)
4	50 Cent	Best Friend - Remix	Best Friend (Remix)
5	Metallica	Lux Æterna	Lux Æterna
6	Metallica	Screaming Suicide	Screaming Suicide
7	Daft Punk	Get Lucky (feat. Pharrell Williams & Nile Rodgers) - Radio Edit	Get Lucky (feat. Pharrell Williams & Nile Rodgers) [Radio
8	Linkin Park	Numb / Encore	Numb / Encore: MTV Ultimate Mash-Ups Presents Collisi
9	Black Eyed Peas	Pump It Louder	Pump It Louder

Table rows counted: 20594

Total rows: 4973 | Query complete 00:00:00.224 | CRLF | Ln 75, Col 28

-- Count the total number of tracks by each artist --

```
80  -- Q5. Count the total number of tracks by each artist. --
81
82  select
83      artist,
84      count(*) as total_no_songs
85  from spotify
86  group by artist
87  order by 2 desc
88
```

Data Output | Messages | Notifications

SQL

	artist character varying (255)	total_no_songs bigint
1	Marisela	10
2	Ray Charles	10
3	Joey Bada$$	10
4	La Mosca Tse-Tse	10
5	TheFatRat	10
6	Leo Dan	10
7	Tyga	10
8	Grupo Marca Registrada	10
9	Tyler, The Creator	10
10	NCT DREAM	10

-- Calculate the average Danceability of tracks in each album --

```
93
94
95  -- Q6. Calculate the average danceability of tracks in each album. --
96
97  select
98     album,
99     avg(danceability) as avg_danceability
100 from spotify
101 group by 1
102 order by 2 desc
```

Data Output | Messages | Notifications

SQL

	album character varying (255)	avg_danceability double precision
1	FOR CERTAIN (Deluxe)	0.975
2	Funky Friday	0.975
3	The House Is Burning [homies begged]	0.971
4	Quality Control: Control The Streets Volume 2	0.97
5	N9NA	0.97
6	Aka Entre el Humo	0.967
7	Peru (Remix)	0.966
8	Street Gossip	0.964
9	Imdat (feat. Hadise)	0.964
10	Freddie's Inferno - Ghost Slowed	0.963

Total rows: 11853 Query complete 00:00:00.158

-- Find the top 5 tracks with the highest energy values --

```
105  -- Q7. Find the top 5 tracks with the highest energy values. --
106
107  Select
108      track,
109      max(energy)
110  from spotify
111  group by 1
112  order by 2 desc
113  limit 5
114
115
116
```

Data Output | Messages | Notifications

SQL

	track character varying (255)	max double precision
1	Rain and Thunderstorm, Pt. 7	1
2	Rain and Thunderstorm, Pt. 33	1
3	Rain and Thunderstorm, Pt. 4	1
4	Rain and Thunderstorm, Pt. 6	1
5	Gentle Piano Melodies	1

-- List all tracks along with their views and likes where official_video = TRUE --

```
115  -- Q8. List all tracks along witht their views and likes where official_video = true
116
117  Select
118      track,
119      sum(views) as total_views,
120      sum(likes) as total_likes
121  from spotify
122  where official_video ='true'
123  group by 1
124  order by 2 desc
125  limit 5
126
```

Data Output | Messages | Notifications

SQL

	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
6	Swalla (feat. Nicki Minaj & Ty Dolla $ign)	5162403752	28864162
7	Wheels on the Bus	4898831101	14396841
8	Taki Taki (with Selena Gomez, Ozuna & Cardi B)	4828403435	37552053
9	Uptown Funk (feat. Bruno Mars)	4821016218	20067879
10	Thrift Shop (feat. Wanz)	4508467245	30242165

-- For each album, calculate the total views of all associated tracks. --

```
127  -- Q9. For each album, calculate the total views of all associated tracks. --
128
129  select
130     album,
131     track,
132     sum(views)
133  from spotify
134  group by 1, 2
135  order by 3 desc
136
137
138
```

Data Output | Messages | Notifications

Showing rows: 1 to 1000 | Page No:

	album character varying (255)	track character varying (255)	sum 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 & Cardi B)	Taki Taki (with Selena Gomez, Ozuna & Cardi B)	4828403435
10	Uptown Special	Uptown Funk (feat. Bruno Mars)	4821016218

Total rows: 18680 Query complete 00:00:00.147

-- Retrieve the track names that have been streamed on spotify more than youtube --

```sql
-- Q10. Retrieve the track names that have been streamed on spotify more than youtube. --

SELECT * FROM
(
    SELECT
        track,
        COALESCE(SUM(CASE WHEN LOWER(most_played_on) = 'youtube' THEN stream END), 0) AS streamed_on_youtube,
        COALESCE(SUM(CASE WHEN LOWER(most_played_on) = 'spotify' THEN stream END), 0) AS streamed_on_spotify
    FROM spotify
    GROUP BY 1
) AS t1
WHERE streamed_on_spotify > streamed_on_youtube
and
streamed_on_youtube <> 0
```

Data Output | Messages | Notifications

	track character varying (255)	streamed_on_youtube numeric	streamed_on_spotify numeric
1	Usted	30059201	137916795
2	21 Hungarian Dances, WoO 1: Hungarian Dance No. 5 in G Minor. Allegro (Orch. Schmeling)	39575743	79151486
3	Mientes Tan Bien	6915455	224299945
4	Have You Ever Seen The Rain	61903001	975300588
5	Dream A Little Dream Of Me	157256901	495674374
6	When I Grow Up	231236307	260959663
7	Me Hace Tanto Bien	56694580	187498268
8	What You Want (feat. Total)	13099909	85458315

Total rows: 155 Query complete 00:00:00.164

-- Find the top 3 most viewed tracks for each artist using window function. --

```sql
-- Q11. Find the top 3 most-viewed tracks for each artist using window function. --

with ranking_artist
as
(select
  artist,
  track,
  sum(views) as total_view,
  dense_rank() over(partition by artist order by sum(views) desc) as rank
from spotify
group by 1, 2
order by 1, 3 desc
)
select * from ranking_artist
where rank <= 3
```

Data Output | Messages | Notifications

SQL | Showing rows: 1 to 1000

	artist character varying (255)	track character varying (255)	total_view double precision	rank bigint
9	(G)I-DLE	Oh my god	191587399	3
10	*NSYNC	Bye Bye Bye	315874581	1
11	*NSYNC	This I Promise You	252377383	2
12	*NSYNC	It's Gonna Be Me	172368673	3
13	070 Shake	Escapism. - Slowed Down	25149399	1
14	070 Shake	Escapism.	25149399	1
15	070 Shake	Escapism. - Sped Up	25149399	1
16	070 Shake	Guilty Conscience	10148280	2

-- Write a query to find tracks where the liveness score is above the average. --

```
178
179  -- Q12. Write a query to find tracks where the liveness score is above the average. --
180
181  select
182    track,
183    artist,
184    liveness
185  from spotify
186  where liveness > (select avg(liveness) from spotify)
187
188  select avg(liveness) from spotify -- 0.19
189
190
191
```

Data Output | Messages | Notifications

SQL | Showin

	track character varying (255)	artist character varying (255)	liveness double precision
1	Feel Good Inc.	Gorillaz	0.613
2	DARE	Gorillaz	0.298
3	New Gold (feat. Tame Impala and Bootie Brown) - Dom Dolla Remix	Gorillaz	0.325
4	Cracker Island (feat. Thundercat)	Gorillaz	0.325
5	Dirty Harry	Gorillaz	0.672
6	Dani California	Red Hot Chili Peppers	0.346
7	Candy Shop	50 Cent	0.38
8	Just A Lil Bit	50 Cent	0.315
9	Disco Inferno	50 Cent	0.279

-- Use a with clause to calculate the difference between the highest and lowest energy values for track in each.--

```sql
-- Q13. Use a with clause to calculate the difference between the highest and lowest energy values for track in each. --

with cte
as
(select
   album,
   max(energy) as highest_energy,
   min(energy) as lowest_energy
   from spotify
group by 1
)
select
   album,
   highest_energy - lowest_energy as energy_diff
   from cte
   order by 2 desc
```

Data Output | Messages | Notifications

SQL

Showing rows: 1 to 10

	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)	0.8049999999999999
7	If I Can Dream: Elvis Presley with the Royal Philharmonic Orchestra	0.787

Total rows: 11853 | Query complete 00:00:00.115

-- Find tracks where the energy-to-liveness ratio is greater than 1.2 --

```
209
210  -- Q14. Find tracks where the energy-to-liveness ratio is greater than 1.2. --
211
212  SELECT
213      track,
214      energy,
215      liveness,
216      (energy / liveness) AS energy_to_liveness_ratio
217  FROM spotify
218  WHERE liveness IS NOT NULL AND liveness <> 0
219    AND (energy / liveness) > 1.2;
220
221
```

Data Output | Messages | Notifications

SQL | Showing rows:

	track character varying (255)	energy double precision	liveness double precision	energy_to_liveness_ratio double precision
1	Rhinestone Eyes	0.703	0.0463	15.183585313174945
2	New Gold (feat. Tame Impala and Bootie Brown)	0.923	0.116	7.956896551724138
3	On Melancholy Hill	0.739	0.064	11.546875
4	Clint Eastwood	0.694	0.0698	9.94269340974212
5	DARE	0.891	0.298	2.9899328859060406
6	New Gold (feat. Tame Impala and Bootie Brown) - Dom Dolla Remix	0.897	0.325	2.76
7	She's My Collar (feat. Kali Uchis)	0.815	0.112	7.2767857142857135
8	Cracker Island (feat. Thundercat)	0.913	0.325	2.809230769230769
9	Dirty Harry	0.877	0.672	1.3050595238095237

Total rows: 18797 | Query complete 00:00:00.128

-- Calculate the cumulative sum of likes for tracks ordered by the number of views, using window function.--

```
-- Q15. Calculate the cumulative sum of likes for tracks ordered by the number of views, using window function.--

SELECT
    track,
    views,
    likes,
    SUM(likes) OVER (ORDER BY views) AS cumulative_likes
FROM spotify;
```

Data Output | Messages | Notifications

SQL | Showing row

	track character varying (255)
1	96 (Theme) - From "96"
2	The Boxer
3	Yesterday
4	White Noise
5	Besos Moja2
6	Mayor Que Usted
7	Vapor
8	Si Te Pillo
9	Noche De Entierro
10	Algo Me Gusta De Ti
11	Noche De Sexo

Total rows: 20592 | Query complete 00:00:00.153

Key Insights & Finding

Through structured SQL analysis and query optimization, the project delivered several actionable insights into user behavior, content performance, and strategic trends on Spotify:

High-Energy Tracks Perform Better: Tracks with elevated energy and danceability scores were strongly associated with higher stream counts, highlighting the correlation between upbeat audio features and listener preference.

Official Content Drives Engagement: Official videos and licensed tracks consistently attracted more views, likes, and comments, suggesting that verified and authorized content improves trust and engagement across platforms.

Artists with Diverse Portfolios Excel: Artists with a larger volume of tracks and albums tended to maintain high overall engagement, indicating the value of consistent content delivery and brand loyalty.

Singles Dominate Streaming Trends: A significant share of popular tracks belonged to the "single" album type, aligning with modern release strategies that prioritize shorter, frequent releases over full-length albums.

Platform-Specific Behavior Noted: Several tracks recorded higher streaming activity on Spotify compared to YouTube, emphasizing differences in user consumption patterns and the importance of platform-specific marketing.

Liveness and Authenticity Matter: Tracks with above-average liveness scores—suggestive of live or acoustic elements—exhibited stronger interaction, pointing toward user appreciation for authentic, concert-like experiences.

These insights not only showcase the practical application of advanced SQL techniques but also reflect how data can guide strategic decision-making in music production, content curation, and platform engagement strategies.

Conclusion

This project demonstrates the power of SQL in extracting meaningful insights from complex and high-volume datasets. By applying advanced querying techniques—including joins, aggregations, window functions, and CTEs—alongside performance optimization strategies, the analysis provided a comprehensive view of how various musical attributes and content formats influence listener engagement across platforms.

Beyond technical execution, the project reflects real-world applications of data analytics in the music and streaming industry. From identifying high-performing content to understanding platform-specific trends, the findings can support strategic decisions related to content production, artist management, and user experience enhancement.

This project not only reinforced my SQL proficiency but also deepened my ability to approach business questions with a data-driven mindset—an essential skill for roles in analytics, business intelligence, and digital strategy.

Tools & Technologies used

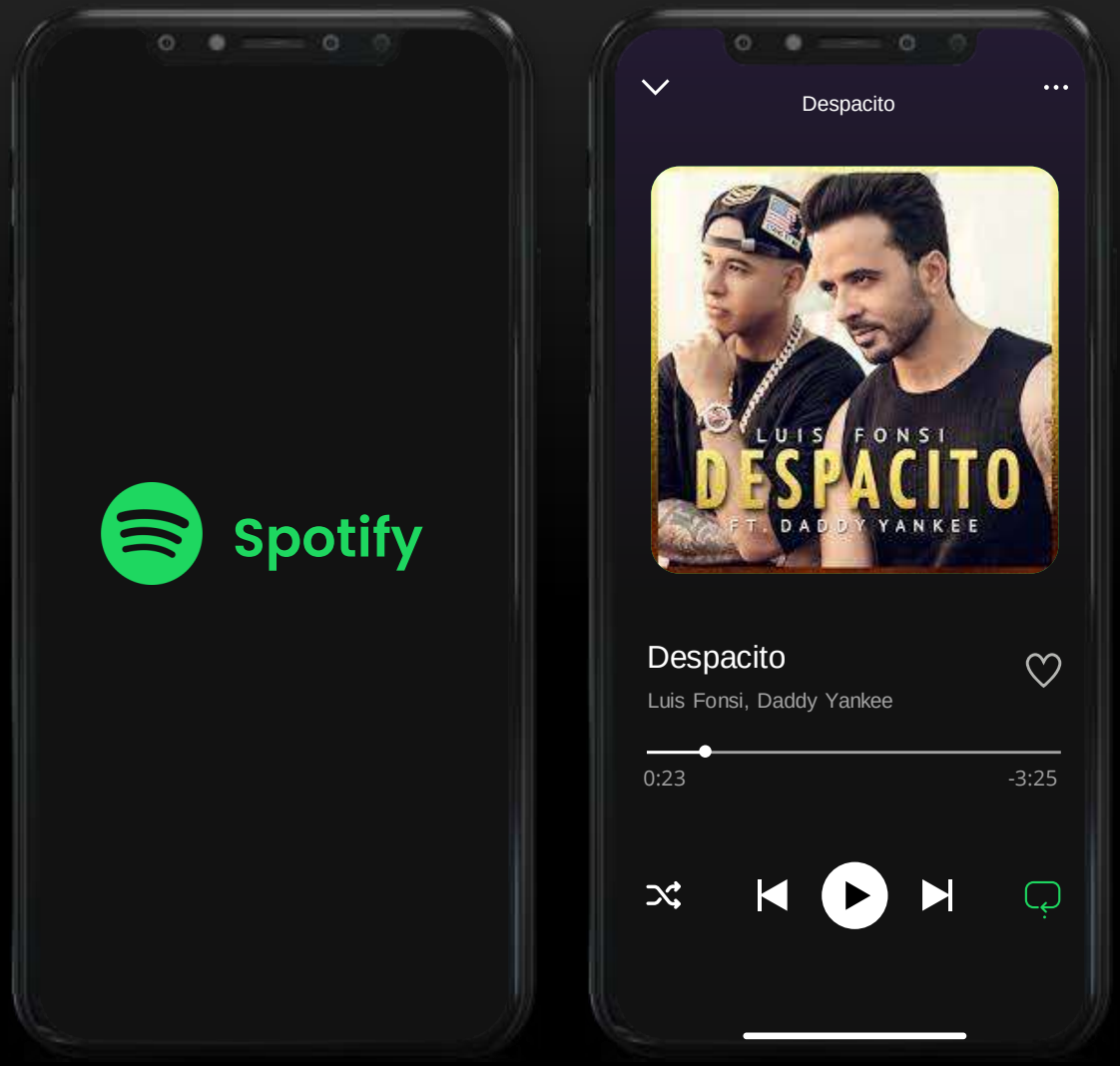
- SQL (PostgreSQL/MySQL): Utilized for writing complex queries, data manipulation, and extracting meaningful insights from the dataset.
- SQL Query Optimization: Applied optimization techniques such as EXPLAIN ANALYZE, indexing, and subquery restructuring to enhance query performance.
- Google Sheets/Excel: Used for initial data exploration, quick validations, and understanding dataset structure.
- GitHub: Managed version control, shared code, and documented the project for collaboration and public access.
- Database Management Tools (pgAdmin, MySQL Workbench):For seamless database interaction and query execution.
- Notion/Docs (Optional): Organized and planned SQL queries, tracked progress, and maintained detailed project documentation.

Learning outcomes

- Advanced SQL Skills: Acquired in-depth knowledge and practical experience in executing complex SQL queries, including advanced joins, aggregations, subqueries, window functions, and common table expressions (CTEs).
- Query Optimization Expertise: Gained proficiency in optimizing query performance through techniques such as indexing, query restructuring, and using EXPLAIN ANALYZE to fine-tune execution plans.
- Data Exploration & Transformation: Developed expertise in exploring large datasets, identifying key trends, and transforming raw data into structured, actionable insights.
- Performance Tuning Techniques: Strengthened ability to evaluate and optimize query execution plans, enhancing overall system performance and reducing query response time.
- Strategic Data Analysis: Enhanced capability to derive meaningful insights from data, supporting informed decision-making in the context of digital content streaming, audience behavior, and artist management.
- Data-Driven Decision Making: Gained valuable experience in leveraging data insights to support business decisions, including content strategy and user engagement optimization.

Contact Us

E-mail – aamansharma027@gmail.com



Thank You

Thank you for taking the time to explore my project. I appreciate any feedback or insights you may have. If you have any questions or would like to discuss the project further, feel free to reach Out. I look forward to connecting with other data enthusiasts and professionals to share knowledge and grow together.

By - Aman Sharma
aamansharma027@gmail.com

Despacito

Luis Fonsi, Daddy Yankee

0:23 -3:25