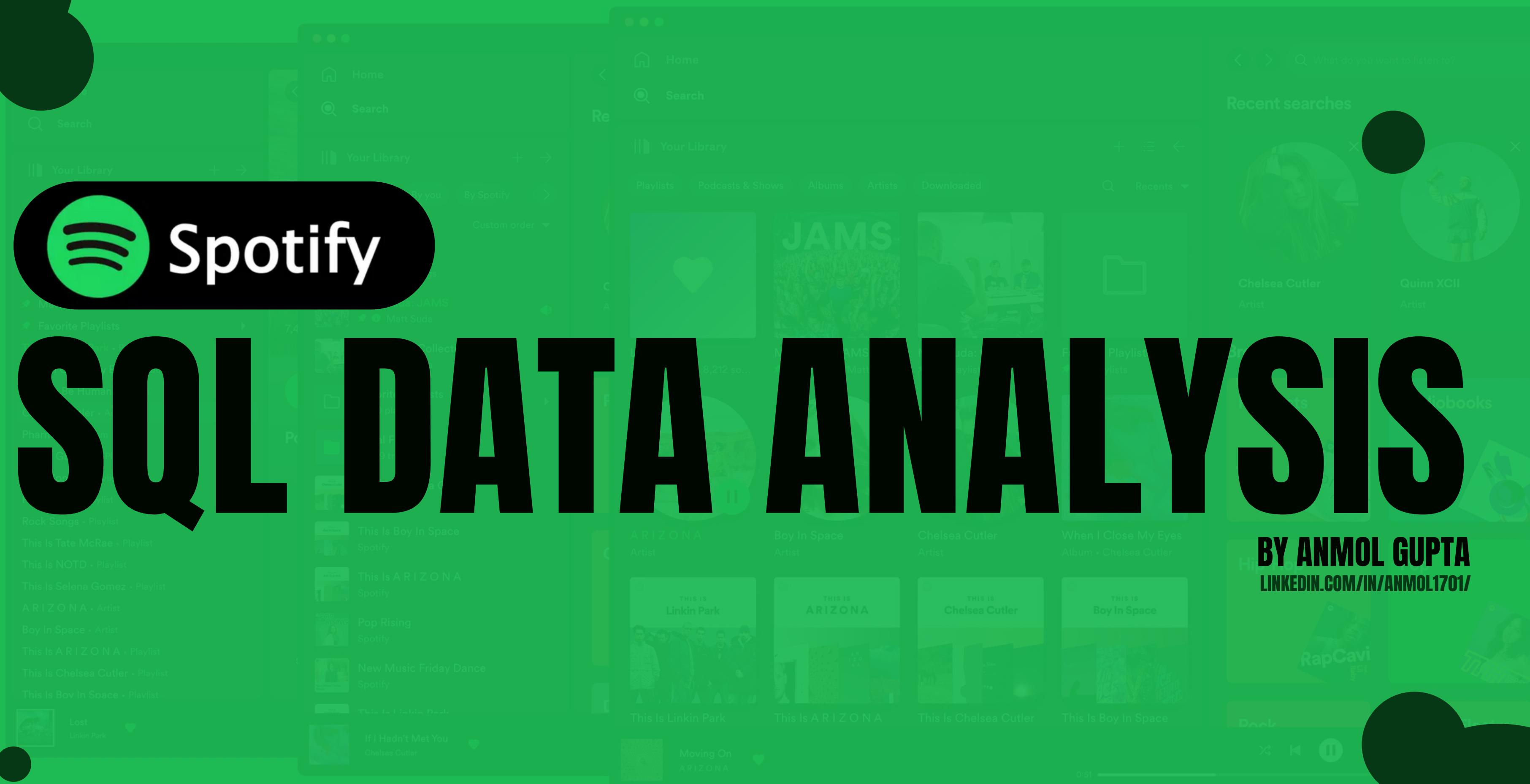




SQL DATA ANALYSIS

BY ANMOL GUPTA
[LINKEDIN.COM/IN/ANMOL1701/](https://www.linkedin.com/in/anmol1701/)



DATA DICTIONARY

1. **artist**: Name of the artist(s) who performed the track, *VARCHAR(255)*
2. **track**: Title of the track, *VARCHAR(255)*
3. **album**: Name of the album the track belongs to (if applicable), *VARCHAR(255)*
4. **album_type**: Type of album (e.g., "album," "single"), *VARCHAR(50)*
5. **danceability**: A value between 0 and 1 indicating how suitable a track is for dancing based on tempo, rhythm stability, etc., *FLOAT*
6. **energy**: A value between 0 and 1 indicating the intensity and activity of a track, *FLOAT*
7. **loudness**: The average decibel level of the track, typically ranging from -60 to 0 dB, *FLOAT*
8. **speechiness**: A value between 0 and 1 indicating the presence of spoken words in the track, *FLOAT*
9. **acousticness**: A value between 0 and 1 that indicates the likelihood of the track being acoustic, *FLOAT*
10. **instrumentalness**: A value between 0 and 1 indicating the probability that the track contains no vocals, *FLOAT*
11. **liveness**: A value between 0 and 1 indicating the presence of an audience or live performance in the recording, *FLOAT*
12. **valence**: A value between 0 and 1 describing the musical positivity of the track, *FLOAT*
13. **tempo**: The overall estimated tempo of the track in beats per minute (BPM), *FLOAT*
14. **duration_min**: The length of the track in minutes, *FLOAT*
15. **title**: The full title of the track as shown on the platform, *VARCHAR(255)*
16. **channel**: Name of the YouTube or streaming channel associated with the track's video, *VARCHAR(255)*
17. **views**: The total number of views the track's video has received, *FLOAT*
18. **likes**: The total number of likes the track's video has received, *BIGINT*
19. **comments**: The total number of comments the track's video has received, *BIGINT*
20. **licensed**: Indicates whether the track/video is licensed (TRUE/FALSE), *BOOLEAN*
21. **official_video**: Indicates whether the video is the official version released by the artist (TRUE/FALSE), *BOOLEAN*
22. **stream**: The total number of streams the track has received on Spotify or similar platforms, *BIGINT*
23. **energy_liveness**: A derived metric combining energy and liveness values (Energy * Liveness), *FLOAT*
24. **most_played_on**: The platform where the track is most frequently played (e.g., Spotify, YouTube), *VARCHAR(50)*

Problem 1

Find the top 5 most-liked tracks for each album type.

```
WITH RankedTracks AS (
    SELECT
        artist,
        track,
        album,
        album_type,
        likes,
        DENSE_RANK() OVER (PARTITION BY album_type ORDER BY likes DESC) AS rank_position
    FROM spotify
)
SELECT artist, track, album, album_type, likes
FROM RankedTracks
WHERE rank_position <= 5;
```

artist	track	album	album_type	likes
Luis Fonsi	Despacito	VIDA	album	50788652
Daddy Yankee	Despacito	VIDA	album	50788626
BTS	Dynamite	BE	album	35892575
Ed Sheeran	Shape of You	÷ (Deluxe)	album	31047780
Halsey	Boy With Luv (feat. Halsey)	MAP OF THE SOUL : PERSONA	album	27588224
Shakira	Waka Waka (This Time for Africa) [The Official 201...	Listen Up! The Official 2010 FIFA World Cup Album	compilation	20327883
Ellie Goulding	Love Me Like You Do	Fifty Shades Freed (Original Motion Picture Soundt...	compilation	12418131
PSY	Gentleman	Tik Tock Hits 2020	compilation	8084738
George Michael	Careless Whisper	Ladies And Gentlemen... The Best Of George Michael	compilation	6693888
Dave Stewart	Sweet Dreams (Are Made of This)	Greatest Hits	compilation	6430729
Charlie Puth	See You Again (feat. Charlie Puth)	See You Again (feat. Charlie Puth)	single	40147674
Wiz Khalifa	See You Again (feat. Charlie Puth)	See You Again (feat. Charlie Puth)	single	40147618
PSY	Gangnam Style (?????)	Gangnam Style (?????)	single	26399133
Billie Eilish	lovely (with Khalid)	lovely (with Khalid)	single	24426202
Khalid	lovely (with Khalid)	lovely (with Khalid)	single	24425840

Problem 2

Identify no. of artists who have at least two tracks where the valence is higher than 0.75.

```
SELECT COUNT(*)  
FROM (  
    SELECT artist  
    FROM spotify  
    WHERE valence > 0.75  
    GROUP BY artist  
    HAVING COUNT(track) >= 2  
) as subquery;
```

COUNT(*)
1095

Problem 3

Find albums where all tracks have an instrumentalness score greater than 0.8.

```
SELECT album  
FROM spotify  
GROUP BY album  
HAVING COUNT(*) = COUNT(CASE When instrumentalness > 0.8 Then 1 END);
```

Problem 4

Identify albums where the loudest track also has the lowest valence.

```
WITH AlbumLoudness AS (
    SELECT
        album,
        track,
        loudness,
        valence,
        RANK() OVER (PARTITION BY album ORDER BY loudness DESC) AS loudest_rank,
        RANK() OVER (PARTITION BY album ORDER BY valence ASC) AS lowest_valence_rank
    FROM spotify
)
SELECT album, track, loudness, valence
FROM AlbumLoudness
WHERE loudest_rank = 1 AND lowest_valence_rank = 1;
```

album	track	loudness	valence
!Volare! The Very Best of the Gipsy Kings	Hotel California (Spanish Mix)	-9.069	0.531
"Heroes" (2017 Remaster)	Heroes - 2017 Remaster	-6.491	0.435
"Let Go" Dj Pack	Let Go - Acapella	-11.733	0.565
"Let's Rock"	Lo/Hi	-7.445	0.707
"Miguel"	Te Amaré	-17.023	0.381
#1 (Colby O'Donis Remix)	#1 - Colby O'Donis Remix	-4.686	0.608
#FREESANTANA	Bloody Mary	-4.443	0.35

Problem 5

Identify the most-played platform (Spotify or YouTube) for each artist based on total plays.

```
WITH platformplay AS (
    SELECT
        artist,
        SUM(stream) AS total_spotify_plays,
        SUM(views) AS total_youtube_views,
        CASE
            WHEN SUM(stream) > SUM(views) THEN 'Spotify'
            ELSE 'YouTube'
        END AS most_played_platform
    FROM spotify
    GROUP BY artist)
SELECT artist, most_played_platform
FROM platformplay;
```

artist	most_played_platform
\$NOT	Spotify
\$uicideboy\$	Spotify
(G)I-DLE	YouTube
*NSYNC	Spotify
070 Shake	Spotify
112	Spotify
11:11 Music Group	Spotify
1nonly	Spotify
2 Chainz	Spotify
21 Savage	Spotify
24kGoldn	Spotify
2Pac	Spotify
3 Doors Down	YouTube
42 Dugg	Spotify
5 Seconds of Summer	Spotify

Problem 6

Rank artists based on the total number of tracks they have that exceed the average tempo.

```
WITH Avgtempo AS (
    SELECT AVG(tempo) AS overall_avg_tempo FROM spotify), Artisttrackcount AS (
        SELECT
            artist,
            COUNT(*) AS high_tempo_tracks
        FROM spotify, Avgtempo
        WHERE tempo > overall_avg_tempo
        GROUP BY artist)
SELECT
    artist,
    high_tempo_tracks,
    DENSE_RANK() OVER (ORDER BY high_tempo_tracks DESC) AS artist_rank
FROM Artisttrackcount;
```



artist	high_tempo_tracks	artist_rank
Scooter	10	1
ATB	10	1
La K'onga	10	1
Oliver Heldens	10	1
PSY	10	1
MEDUZA	10	1
Nightcrawlers	10	1
Rae Sremmurd	10	1
Afrojack	10	1
Riton	10	1
Hardwell	10	1
PXNDX	10	1
Vintage Culture	10	1
Grupo Frontera	10	1
Soft Cell	10	1
Chris Lake	10	1
King Von	10	1
Jax Jones	10	1
Martin Solveig	10	1
Inspector	10	1
Miel San Marcos	10	1
Moira Dela Torre	10	1
Benny Benassi	10	1
Cedric Gervais	10	1
James Hype	10	1
Mr.Kitty	10	1
Raí Saia Rodada	10	1
LMFAO	10	1
DENNIS	10	1
Aventura	10	1
Bailey Zimmerman	10	1
Gotye	9	2
Kylie Minogue	9	2
Sorriso Maroto	9	2
Steve Edwards	9	2
Clayton & Romário	9	2
David Guetta	9	2
Vengaboys	9	2
MARINA	9	2
Placebo	9	2
Loud Luxury	9	2
LMFAO	9	2

THANK YOU

What's Next ?