# **Spotify Music Dataset Analysis using SQL**

### Overview:

This project involves analyzing a spotify dataset having various attributes about albums, tracks, streams, views, likes, artists and other components using SQL. It covers the end-to-end process of normalizing a denormalized dataset, performing SQL queries of varying complexity (easy, medium, and advanced), and optimizing query performance. The primary goal of the project is to practice advanced SQL skills and generate valuable insights from the dataset.

## **Project Steps**

## 1. Data Exploration

Before diving into SQL, it's important to understand the dataset thoroughly. It is said 80% is data exploration & understanding and 20% is data analysis. The dataset contains attributes such as:

- Artist
- Track
- Album
- Album Type
- Duration
- Platform
- Various metrics such as danceability, liveness, energy, Loudness and more.

## 2. Data cleaning

During data exploration, I found some inconsistency in the dataset which results in dropping rows for a more consistent and result-driven dataset.

### 3. Querying the Data

After the data is cleaned, various SQL queries can be written to explore and analyze the data. Queries are categorized into easy, medium and advanced levels to help progressively develop SQL proficiency.

### **Easy Queries**

- Simple data retrieval, filtering and basic aggregations.

#### Medium Queries

- More complex gueries involving grouping, aggregation functions,

## **Advanced Queries**

- Nested subqueries, window functions, CTEs, and performance optimization

## 15 Practice Questions

# **Easy Level**

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

```
-- Q1. Retrieve the names of all tracks that have more than 1 billion streams.
  From project..spotify data
    Where Stream >= 1000000000
    Order BY Stream DESC;
00 % - 4
■ Results  Messages
    Track
                                        Stream
    Blinding Lights
                                        3386520288
                                        3362005201
2
    Shape of You
    Someone You Loved
                                        2634013335
    rockstar (feat. 21 Savage)
                                        2594926619
    Sunflower - Spider-Man: Into the Spider-Verse
                                        2538329799
    Sunflower - Spider-Man: Into the Spider-Verse
                                        2538329799
```

List all albums along with their respective artists.

```
-- Q2. List all albums along with their respective artists.

| Select Distinct Album, Artist | from project..spotify_data;
```



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

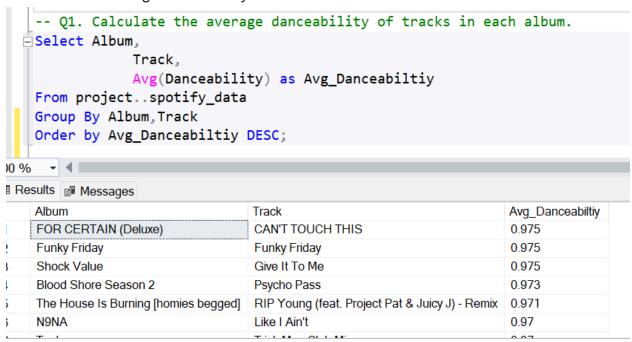
```
-- Q3. Get the total number of comments for tracks where licensed = TRUE.
 select Track, sum(Comments) as total_comment
   From project..spotify_data
   Where Licensed = 1
   Group by Track;
)% - 4 ■
Results Messages
   Track
                         total_comment
   !ly (feat. Coez)
                         1628
   #1 - Colby O'Donis Remix
                         12359
   #41
                         1248
   #NAME?
                         70172
   #thatPOWER
                         301321
   #Ysya2020 Vol. 5 - Silbando 521
```

- 4. Find all tracks that belong to the album type single.
- 5. Count the total number of tracks by each artist.

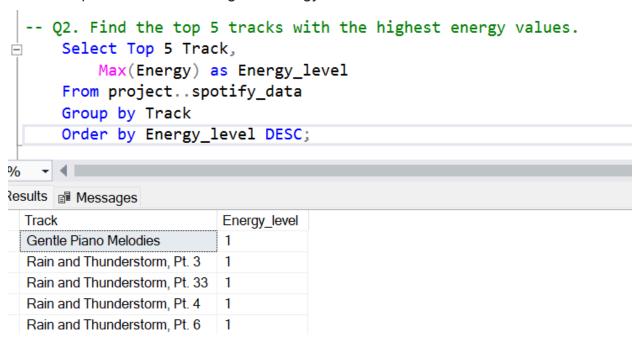
```
-- Q5. Count the total number of tracks by each artist.
 select Artist, COUNT(Track) as total_tracks
   from project..spotify_data
   Group by Artist
   Order by total_tracks;
) %
Results Messages
   Artist
                    total_tracks
                    1
   back number
   Bootie Brown
                    1
   Stars Music Chile
                    1
   Mrs. GREEN APPLE
                    2
   Vaundy
                    2
                    3
   Jimin
```

## **Medium Level**

1. Calculate the average danceability of tracks in each album.

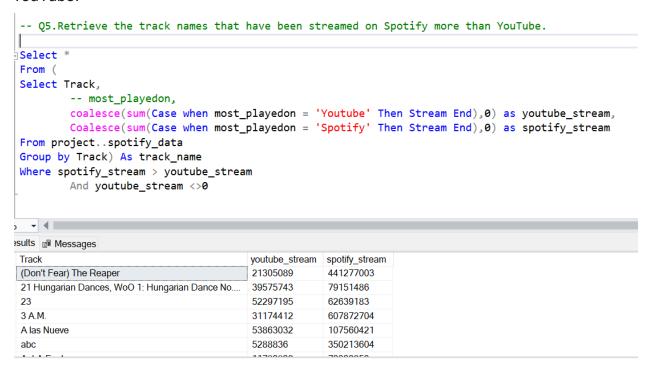


2. Find the top 5 tracks with the highest energy values.



- 3. List all tracks along with their views and likes where official\_video = TRUE.
- 4. For each album, calculate the total views of all associated tracks.

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



## **Advanced Level**

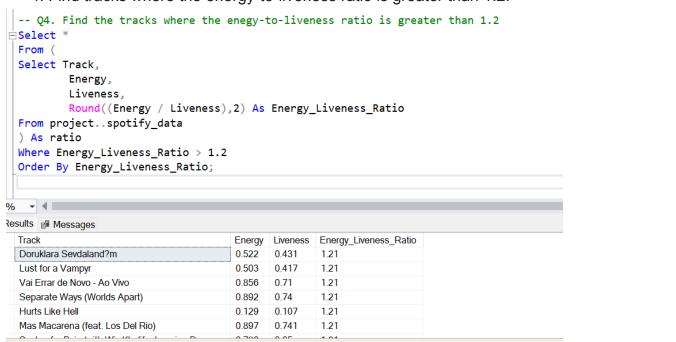
1. Find the top 3 most-viewed tracks for each artist using window functions.

```
-- Q1. Find the top 3 most-viewed tracks for each artist using window functions.
Select *,
          Dense_Rank() over(Partition By Artist Order By total_views DESC) as view_rank
 From (
 Select Artist, Track, sum(Views) as total_views
 From project..spotify_data
 Group By Artist, Track
 -- Order by Artist, total_views DESC
 ) As track_view
 Select *
       --Artist, Track, total views
 From most_view_tranks
 Where view_rank <=3;
% - ∢ ■
lesults Messages
 Artist
           Track
                             total_views view_rank
 $NOT
           Tell Em
                             41100657 1
           Like Me (feat. iann dior) 15803517 2
  $NOT
 $NOT
           Mean
                             13563870 3
                             175156959 1
 $uicideboy$ Paris
                             91771038 2
 $uicideboy$ For the Last Time
                             31674988 3
 $uicideboy$ Kill Yourself (Part III)
```

- 2. Write a query to find tracks where the liveness score is above the average.
- 3. Use a WITH clause to calculate the difference between the highest and lowest energy values for tracks in each album.

```
-- Q3. Use a WITH clause to calculate the difference between the highest and lowest energy values for tracks in each album.
With energy_difference As(
Select Album,
         MAX(Energy) Max_Energy,
         MIN(Energy) As Min_Energy
From project..spotify_data
Group by Album
 -- Order by Album
Select *.
    (Max_Energy - Min_Energy) as energy_diff
From energy_difference
Order By energy_diff DESC;
- - 4 ■
sults Messages
Album
                                      Max_Energy Min_Energy energy_diff
White Noise
                                     0.908
                                                0.00125
                                                           0.90675
 Spotify Singles - Holiday
                                      0.904
                                                 0.068
                                                           0.836
                                                 0.0708
Spotify Singles
                                      0.894
                                                           0.8232
UNDERTALE Soundtrack
                                      0.978
                                                 0.162
                                                           0.816
                                                 0.0891
                                                           0.8109
Making Mirrors
                                      0.9
Everytime We Touch (Premium Edition)
                                      0.978
                                                 0.173
                                                           0.805
```

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



5.

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

