

## Project Title:

### Spotify Data Analysis using SQL



---

## Project Objective:

The objective of this project is to analyze music streaming data from Spotify using SQL in order to extract valuable insights about artists, tracks, albums, and listener behavior. This analysis enables data-driven decision-making for content creators, music marketers, and business stakeholders in the digital music industry.

---

## Project Goals:

1. **Explore** the structure of the Spotify dataset (artists, tracks, streams, likes, views, etc.).
2. **Perform descriptive analytics** such as averages, totals, maximums, and counts across different dimensions.
3. **Use SQL queries to uncover trends** like most streamed tracks, high-energy music, track popularity, and more.
4. **Apply advanced SQL concepts** like CTEs, window functions (RANK, ROW\_NUMBER), aggregation with HAVING, and subqueries.
5. **Classify music content** into custom buckets (e.g., energy levels) for better segmentation.
6. **Support marketing, content planning, and listener engagement strategies** through actionable data insights.

---

### Dataset Columns Used:

- artist, track, album, album\_type
- danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, tempo
- duration\_min, views, likes, comments, stream, channel, official\_video, licensed, most\_played\_on, energy\_liveness

---

### SQL Concepts Applied:

- Basic Queries (SELECT, WHERE, DISTINCT)
- Aggregations (SUM, AVG, MAX, COUNT)
- Grouping & Filtering (GROUP BY, HAVING)
- Sorting (ORDER BY)
- Subqueries
- Pattern Matching (LIKE)
- Window Functions (RANK, DENSE\_RANK, ROW\_NUMBER)
- CTEs (Common Table Expressions)
- Conditional Logic (e.g., CASE for classification)

---

### Sample Insights (Based on Questions):

- The track with the **highest likes** is ...
- Most **viewed tracks** are by artists such as ...
- **High-energy music** is typically found in the album type ...
- Tracks with **speechiness > 0.5 and acousticness < 0.3** might represent podcast-like content.
- Some artists have **over 1M views** on multiple tracks, indicating viral or trending popularity.
- A **classification** of tracks into **High, Medium, Low energy** was performed using a CTE.

---

### Tools Used:

- **Database:** Postgres SQL
- **Platform:** Localhost / Cloud SQL

- **Language:** SQL
  - **Data Source:** Cleaned and structured Spotify dataset
  - **Optional:** Power BI/Excel for visualization after SQL analysis (if needed)
- 

### **Conclusion:**

This project demonstrates how raw Spotify data can be transformed into actionable business insights using SQL. By answering 40 business-relevant questions, we've explored performance metrics, user preferences, content classification, and artist-level trends. This kind of analysis is directly useful for data-driven marketing, playlist curation, and artist strategy on music platforms.