**Analyzing Spotify Dataset Using SQL**

**Project Overview**

This project analyzes a publicly available Spotify dataset containing various attributes of tracks, albums, and artists. The aim is to derive insights related to artist popularity, song duration, audio features, and listener engagement through advanced SQL queries and data analysis techniques.

**Dataset Overview**

* **Source**: [Spotify Dataset on Kaggle](https://www.kaggle.com/datasets/sanjanchaudhari/spotify-dataset)
* **Key Features**:
  + **Artist**: Name of the artist.
  + **Track**: Name of the song.
  + **Album**: Album containing the track.
  + **Duration\_min**: Length of the song in minutes.
  + **Danceability**, **Energy**, **Loudness**: Various audio features.
  + **Stream**, **Views**, **Likes**, **Comments**: Listener engagement metrics.

**Step-by-Step Process for Analyzing the Spotify Dataset**

**1. Data Understanding**

* **Objective**: Explore the structure and key features of the dataset.
* **Key Features**: Track, artist, album, duration, popularity, audio features.

**2. Data Cleaning and Preparation**

* Handle missing values and remove duplicates.
* Standardize formats for consistency.
* Perform descriptive statistics on numerical columns.

**3. Data Normalization**

* **Objective**: Optimize query performance.
* **Process**: Transform denormalized tables into a normalized structure by creating separate tables for artists, albums, and tracks.

**4. Exploratory Data Analysis (EDA)**

* Conduct bivariate analysis (e.g., Popularity vs. Tempo).
* Analyze trends based on release years.
* Use multivariate analysis to explore factors influencing song popularity.

**5. Data Exploration**

* Gain insights into artist trends and listener behavior.
* Count tracks per artist and analyze release years.

**6. SQL Queries**

* **Easy Queries**: Count total tracks, albums, and unique artists.
* **Intermediate Queries**: Analyze top artists based on popularity.
* **Advanced Queries**: Analyze song duration and its impact on popularity.

**7. Key Business Problems to Solve**

* Identify top artists based on the number of songs and popularity.
* Analyze how song length impacts popularity.
* Study the effects of collaborations on track performance.

**8. Visualization & Insights**

* **Tools**: Use Power BI or Tableau for visualization.
* **Charts**: Total Sales by Fat Content (Donut chart), Most Played Tracks (Bar Chart), Stream Trends by Month (Line Chart).
* **Insights**: Provide actionable insights for music industry stakeholders.

**9. Optimization Techniques**

* Implement indexing on frequently queried columns (e.g., artist, track).
* Optimize query performance using JOIN operations and CTEs.

**10. Real-Life Applications**

* **Music Recommendations**: Build a recommendation system based on popularity and genre.
* **Business Decision Support**: Provide insights into optimizing song releases for record labels.

**Analysis**

**Easy Level Questions**

1. Top 5 Songs by Energy
2. Most Played Songs by Platform
3. Duration and Popularity of Songs by Artist
4. Average energy\_liveness & Average liveness by each Track
5. Songs with Highest Danceability and Acousticness
6. Retrieve tracks with more than 1 billion streams
7. List all albums along with respective artists
8. Total number of comments for licensed tracks
9. Tracks belonging to album type single
10. Total number of tracks by each artist
11. Total number of tracks in the dataset
12. Distinct artists in the dataset
13. Top 10 most popular songs
14. Total number of albums in the dataset
15. Tracks with duration greater than 5 minutes
16. List all songs by a specific artist

**Medium Level Questions**

1. Average danceability of tracks in each album
2. Top 5 tracks with the highest energy values
3. Tracks along with views and likes where official\_video = TRUE
4. Total views of all tracks per album
5. Track names streamed more on Spotify than YouTube
6. Total Views and Likes by Platform
7. Songs with High Popularity and Low Acousticness
8. Top 5 Most Lively Songs by Track
9. Top 5 artists with the most songs

**Advanced Level Questions**

1. Top 5 Artists by Total Stream Count for Official Videos
2. Songs with the Highest Difference Between Views and Likes
3. Correlating Popularity (stream) and Danceability Across All Songs
4. Most played tracks in specific channels (top 10)
5. Cluster songs based on audio features (e.g., danceability, energy)

**Conclusion**

This structured approach allows for comprehensive analysis of the Spotify dataset, leading to meaningful insights into music trends, artist performance, and listener behavior. By leveraging SQL and visualization tools, this project aims to provide valuable recommendations for stakeholders in the music industry.