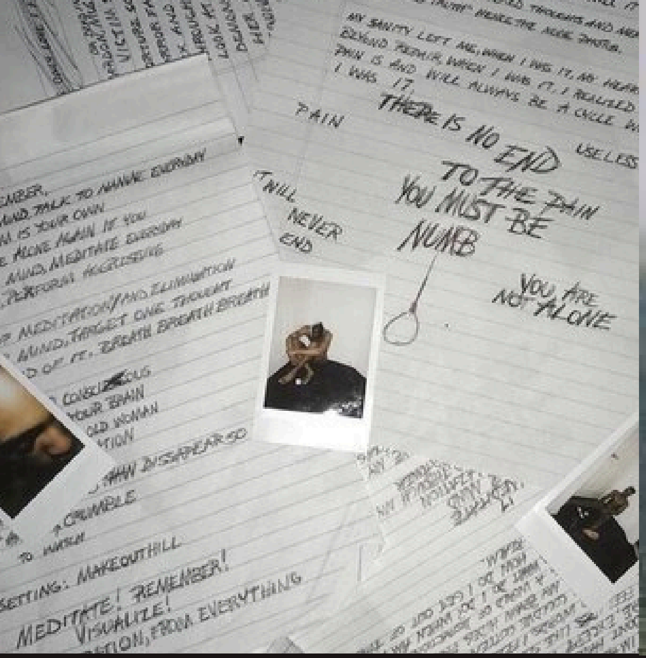

Spotify

Data Analysis

by Kanishk Mehra

Yourwebsite.com

- Welcome
- Our Album
- Playlist
- Artists
- Podcasts & Shows
- Your Library
- Trending
- Help
- Report History

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 Conclusion Tools & Technologies Used Learning Outcomes Contact Us Thank You 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

Yourwebsite.com

Welcome

Data Overview

Playlist

Artists

Podcasts & Shows

Your Library

Trending

Help

Report History

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

Yourwebsite.com

Welcome

Data Overview

Project Objective

Artists

Podcasts & Shows

Your Library

Trending

Help

Report History

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 through writing 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

Yourwebsite.co
m

Welcome

Data Overview

Project Objective

Project

Key Insights & Finding

Project

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

```sql
-- Q1 Retrieve the names of all tracks that have more than 1 billion streams.

SELECT * FROM SPOTIFY
WHERE stream > 1000000000
```

Yourwebsite.com

Welcome

Data Overview

Project Objective

Project

Key Insights & Finding

Project

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

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

SELECT DISTINCT artist , album
FROM SPOTIFY
ORDER BY 1 ;
```

Project

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

```
-- Q3 Get the total number of comments for tracks where licensed = TRUE.

SELECT SUM(comments) as total_comments
FROM SPOTIFY
WHERE licensed = 'true'
```

Yourwebsite.com

- Welcome
- Data Overview
- Project Objective
- Project
- Key Insights & Finding

Project

-- 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'
```

Yourwebsite.com

Project

- Welcome
- Data Overview
- Project Objective
- Project
- Key Insights & Finding

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

```sql
-- Q5 Count the total number of tracks by each artist.

SELECT artist , count(*) as total_no_songs
FROM SPOTIFY
GROUP BY artist
ORDER BY 2 DESC ;
```

Yourwebsite.com

- Welcome
- Data Overview
- Project Objective
- Project
- Key Insights & Finding

Project

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

```
-- Q6 Calculate the average danceability of tracks in each album.

SELECT album , AVG(danceability) as avg_danceability
FROM SPOTIFY
GROUP BY 1
ORDER BY 2 DESC ;
```

Yourwebsite.co
m

Project

Welcome

Data Overview

Project Objective

Project

Key Insights & Finding

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

```
-- Q7 Find the top 5 tracks with the highest energy values.

SELECT track , MAX(energy)
FROM SPOTIFY
GROUP BY 1
ORDER BY 2 DESC
LIMIT 5 ;
```

Yourwebsite.com

Welcome

Data Overview

Project Objective

Project

Key Insights & Finding

Project

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

```sql
-- Q8 List all tracks along with their views and likes where official_video = TRUE.

-- pick the highest single snapsot of views and likes per tracks.

SELECT track , MAX(views) AS max_views , MAX(likes) AS max_likes
FROM SPOTIFY
WHERE official_video = 'true'
GROUP BY track
ORDER BY max_views DESC ;

-- sum up all the snapshots of views and likes per tracks.

SELECT track , SUM(views) AS total_views , SUM(likes) AS total_likes
FROM SPOTIFY
WHERE Official_video = 'true'
GROUP BY track
ORDER BY total_views DESC ;
```

Project

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

```
-- Q9 For each album, calculate the total views of all associated tracks.

SELECT album , SUM(views) AS total_album_views
FROM SPOTIFY
GROUP BY album
ORDER BY total_album_views DESC ;
```

Yourwebsite.com

Project

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

```
-- Q10 Retrieve the track names that have been streamed on Spotify more than YouTube.

SELECT * FROM
(SELECT track , COALESCE(SUM(CASE WHEN most_played_on = 'Youtube' THEN stream END ),0) AS stream_on_youtube ,
               COALESCE(SUM(CASE WHEN most_played_on = 'Spotify' THEN stream END ),0) AS stream_on_spotify
FROM SPOTIFY
GROUP BY 1 ) AS T1
WHERE stream_on_spotify > stream_on_youtube
AND
stream_on_youtube != 0
```

Yourwebsite.com

Project

- Welcome
- Data Overview
- Project Objective
- Project
- Key Insights & Finding

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

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

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 ;
```

Yourwebsite.com

Project

Welcome

Data Overview

Project Objective

Project

Key Insights & Finding

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

```sql
-- Q12 Write a query to find tracks where the liveness score is above the average.

SELECT track , artist , liveness
FROM SPOTIFY
WHERE liveness > ( SELECT AVG(liveness)
                   FROM SPOTIFY )
```

Yourwebsite.com

Project

- Welcome
- Data Overview
- Project Objective
- Project
- Key Insights & Finding

-- 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 tracks in each album.

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_difference
FROM CTE
ORDER BY 2 DESC ;
```

Yourwebsite.com

Welcome

Data Overview

Project Objective

Project

Key Insights & Finding

Project

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

```
-- Q14 Find tracks where the energy-to-liveness ratio is greater than 1.2.

SELECT energy , liveness , energy/liveness AS ratio
FROM SPOTIFY
WHERE energy/liveness > 1.2
ORDER BY 3 DESC ;
```

Yourwebsite.com

Welcome

Data Overview

Project Objective

Project

Key Insights & Finding

Project

-- 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 functions.

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

Yourwebsite.com

- Welcome
- Data Overview
- Project Objective
- Project
- Key Insights & Finding

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: Conclusion Tools & Technologies Used Learning Outcomes Contact Us Thank You

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.

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

Yourwebsite.co
m

Conclusion

Welcome

Data Overview

Project Objective

Project

Key Insights & Finding

This project demonstrates the power of SQL in extracting meaningful insights from complex and high-volume datasets. By applying Conclusion Tools & Technologies Used Learning Outcomes Contact Us Thank You 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

Yourwebsite.co
m

Welcome

Data Overview

Project Objective

Project

Key Insights & Finding

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

Yourwebsite.com

Thankyou

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