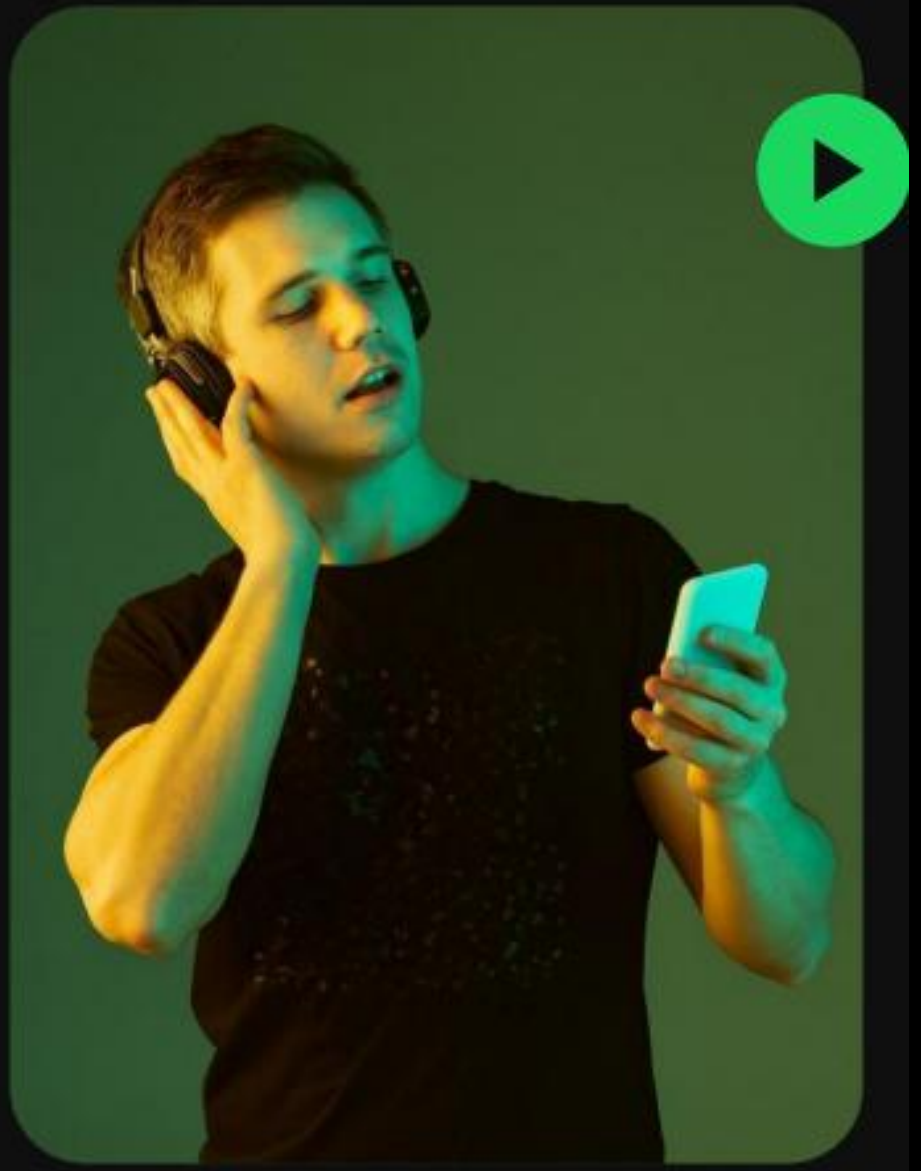
Spotify

Spotify Data Analysis Using SQL

Presented by:
Syed Zaki Ali

Project Overview

The Spotify SQL Analysis project aims to explore and analyze key performance metrics of Spotify tracks, albums, and artists using SQL. The analysis is structured across three levels of complexity: Easy, Medium, and Advanced. Through this project, we will derive insights into streaming patterns, track popularity, and listener engagement metrics. The insights gathered will help in understanding music trends, artist performance, and factors contributing to track popularity.



Project Objectives

Data Exploration and Basic Insights (Easy Level)

-Identify tracks with more than 1 billion streams.
-List all albums along with their respective artists.
-Calculate the total number of comments for tracks where licensing is true.
-Filter tracks that belong to the album type "Single."
-Count the total number of tracks produced by each artist.

Intermediate Analysis for Deeper Insights (Medium Level)

-Calculate the average danceability of tracks in each album.
-Identify the top 5 tracks with the highest energy values.
-Retrieve tracks with views and likes where the official video is available.
-Calculate the total views of tracks for each album.
-Identify tracks streamed more on Spotify than on YouTube.

Advanced Analytical Insights (Advanced Level)

-Use window functions to find the top 3 most-viewed tracks for each artist.
-Identify tracks where the liveness attribute exceeds the average.
-Apply a WITH clause to calculate the difference between the highest and lowest energy values within each album.
-Determine tracks with an energy-to-liveness ratio greater than 1.2.
-Calculate the cumulative sum of likes for tracks based on the order of views using window functions.

Data Exploration and Basic Insights (Easy Level)

Intermediate Analysis for Deeper Insights (Medium Level)

Intermediate Analysis for Deeper Insights (Medium Level)

Tools and Techniques

Excel:

Data cleaning: Handling missing values, removing duplicates, and standardizing formats.

Data preparation: Structuring the dataset for SQL integration.

SQL Server:

Basic Queries: SELECT, WHERE, GROUP BY, ORDER BY.

Aggregation: COUNT, SUM, AVG, MAX, MIN.

Filtering: Using WHERE and HAVING clauses for conditional queries.

Advanced Techniques: Window functions (dense rank, RANK, OVER), subqueries, and WITH clauses for complex analysis

Conditional Expression:Nullif,Coalesce

Data Exploration and Basic Insights (Easy Level)

Q.1 Identify tracks with more than 1 billion streams.

```
select track from spotify
where stream >= 1000000000
```

Q2.list all albums along with their respectives artists.

```
select album,artist
from spotify
```

Data Exploration and Basic Insights (Easy Level)

Q.3 Get The Total Number Of Comments For Tracks Where Licensed =True.

```
select sum(comments)as total_number from
spotify where licensed =1
```

Q4. Find All The Tracks That Belong To The Album Type Single.

```
select * from spotify
where Album_type='single'
```

Data Exploration and Basic Insights (Easy Level)

Q.5 Count The Total Number Of Tracks By Each Artists

```
select count(track)as Count_of_track ,artist from spotify
group by Artist
```

Intermediate Analysis for Deeper Insights (Medium Level)

Q.6 Calculate the average danceability of tracks in each album.

```
select AVG(danceability)as Avg_Danceability,album
from spotify
group by Album
```

Q.7 Find The Top 5 Tracks With The Highest Energy Values.

```
select top 5 track,max(energy) as highest_energy from spotify
group by track
order by highest_energy desc
```

Q.8 List All The Tracks Along With Their Views And Likes Where Official Video Is True

```
select track,sum(views)as total_views,sum(likes) as total_likes
from spotify
where official_video =1
group by track
order by total_likes desc
```

Q.9 For Each Album ,Calculate The Total Views Of All Associated Tracks

```sql
select track,album,sum(views) as total_views from spotify
group by track,album
order by total_views desc
```

Q.10 Retrieve The Track Names That Have Been Streamed On Spotify More Than Youtubes

```sql
select * from
(select track,
coalesce(sum(case when most_playedon='Youtube' then stream end),0) as Streamed_on_youtube,
coalesce(sum(case when most_playedon='Spotify' then stream end),0) as Streamed_on_spotify
from spotify
group by track
)as t1

where Streamed_on_spotify> Streamed_on_youtube
and
Streamed_on_youtube <> 0
```

Advanced Analytical Insights (Advanced Level)

Q.11 Find The Top 3 Most Viewed Tracks For Each Artist Using Window Function

```
with ranking_artist
as
(select track,artist,sum(views) as total_views,
DENSE_RANK()over(partition by artist order by sum(views) desc) as ranktrack
from spotify
group by track,artist
)
select * from ranking_artist
where ranktrack<=3
order by artist,total_views desc
```

Q.12 Write A Query To Find Tracks Where The Liveness Is Above The Average

```
select track,liveness
from spotify
where Liveness >(select AVG(liveness) from spotify)
```

Q.13 Apply a WITH clause to calculate the difference between the highest and lowest energy values within each album.

```sql
with cte
as(
select album,max(energy) as highest_energy
,min(energy) as lowest_energy
from spotify
group by album
)

select album,highest_energy-lowest_energy as energy_difference
from cte
order by energy_difference desc
```

Q.14 Determine tracks with an energy-to-liveness ratio greater than 1.2.

```sql
select track,energy,liveness,(energy/coalesce(nullif(liveness,0),1)) as energy_liveness_ratio
from spotify
where (Energy/coalesce(nullif(liveness,0),1))>1.2
```

Q.15 Calculate the cumulative sum of likes for tracks based on the order of views using window functions.

```sql
SELECT
    track,
    artist,
    views,
    likes,
    SUM(likes) OVER (ORDER BY views DESC ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) AS cumulative_likes
FROM spotify
ORDER BY views DESC;
```

Key Insights

1. Tracks like *Feel Good*, *Californication*, *Under the Bridge*, *In Da Club*, and others have surpassed 1 billion streams.

2. There are *497,023,475* total comments on tracks where the license status is set to true.

3. Most artists have released around *10 tracks*.

4. Gentle Piano Melodies, *Rain and Thunderstorm, Pt. 3*, *Rain and Thunderstorm, Pt. 33*, *Rain and Thunderstorm, Pt. 4*, and *Rain and Thunderstorm, Pt. 6* are among the tracks with high energy levels.

5. Albums and tracks with significant views include *VIDA* with *16,159,296,273* views, *See You Again (feat. Charlie Puth)* with *11,547,595,554* views, and *Peace Is The Mission (Extended)* with *9,974,504,694* views.

6. Old Town Road - Remix, *Silence*, *Photograph*, *Stay*, and *Perfect Strangers* are among the tracks with more streams on Spotify compared to YouTube.

"Music gives a soul to the universe, wings to the mind, flight to the imagination, and life to everything."

— Plato

Thanks !