Spotify 2023 Data Analysis Project

The goal of this project is to better understand the top spotify songs in 2023 using the Top_Spotify_Songs_2023 database available within Kaggle. This is my first personal SQL project and would like to showcase some of my understandings while further understanding the data, in hopes to gain some important insights regarding the top spotify songs of the year.

Data Exploration

First we want to make sure our data imported correctly and we also want to familiarize with all column names and the idea of the row contents:

```
select *
from spotify2023
```

Analyzing the Data

What are the top 5 songs on spotify in 2023?

```
select *
from spotify2023
ORDER by streams desc
limit 5;
```

What are all the top 5 songs with multiple artists?

```
select artists_name, track_name
from spotify2023
where artist_count > 1
ORDER by streams desc
limit 5;
```

How many top songs are in more than 500 spotify playlsts and more than 1 artist?

```
SELECT COUNT(*)
from spotify2023
where artist_count > 1
and in_spotify_playlists > 500
```

Output:

There are 312 songs that are in more than 500 spotify playlists AND have more than 1 artist

How many songs have multiple artists AND over 1 billion streams?

```
SELECT COUNT(*) as num_artist
from spotify2023
where artist_count > 1
and streams > 1000000000;
```

Output:

There are 42 songs that have multiple artists and over 1 billion streams.

What are the top 5 songs that have the word "Love" in the track name?

```
SELECT track_name
from spotify2023
where track_name like '%Love%'
order by streams desc
limit 5 ;
```

What are the average streams per day on spotify for each song?

```
SELECT *,
 DATEDIFF(day, release_date, GETDATE())
AS elapsed_time_in_days,
 ROUND
(streams / NULLIF(DATEDIFF(day, release_date, GE
AS avg_streams_per_day
FROM spotify2023
ORDER BY avg_streams_per_day DESC
LIMIT 5;
```

group the songs into quartiles to determine which songs are the most popular?

```
SELECT NTILE (4) OVER (
    ORDER BY streams DESC
  ) AS 'top_stream_groups',
  track_name, artists_name, streams
FROM spotify2023
```

What is the running total amount of streams for Ed Sheeran?

```
SELECT artists_name, track_name, released_date,
sum(streams) over
 (partition by artists_name order by released_da
as running_total_stream
from spotify2023
where artists_name = 'Ed Sheeran'
```

What are the average streams based on 'Key'? Which key has the most streams?

```
SELECT key, avg(streams) as key_stream
from spotify2023
where key is not null
GROUP by key
ORDER by key_stream DESC
```

Determine the 'grooviness' of each song?

```
SELECT track_name, artist_s_name, CASE
 WHEN danceability > 80 THEN 'The Grooviest'
 WHEN danceability > 60 THEN 'Groovy'
 WHEN danceability > 40 THEN 'Somewhat Groovy'
 WHEN danceability > 20 THEN 'Barely Groovy'
 ELSE 'Not Groovy'
  END AS 'The_Groovy_Scale'
FROM spotify2023
ORDER BY The_Groovy_Scale;
```

Output: The output of the query above leaves us with information as to how groovy a song is. It lists each songs as well as their respective 'grooviness'

Note: this could be used to figure out which songs to play depending on the venue, audience, occasion, etc.

What is the most streamed single for each artist?

```
select max(streams)as most_stream,
 artists_name,
 track_name
from spotify2023
where artist_count = 1 and streams is not null
GROUP by artists_name,track_name
ORDER by most_stream DESC
```

Which artist has the most songs released in 2023 that are in the top spotify charts?

```
select  artists_name, count(*) total_released
from spotify2023
where released_year = 2023
group by artists_name
ORDER by total_released DESC
```

Output: Morgan Wallen.

Despite Taylor Swift's tour generating enough revenue to boost the American economy, it was not enough to take the top spot from Morgan Wallen for most top songs released in 2023 (8 total songs!).

What percentage of songs in the top 2023 spotify songs were actually released in 2023?

```
with cte as (select count(*) total_count,
sum(case when released_year = '2023' then 1 else
as total_2023
from spotify2023)

SELECT (total_2023 * 100/ total_count)
as percentage_released_in_2023
from cte
```

Conclusions

After completing my first personal SQL project, I was able to formulate some interesting conclusions. Firstmost, The majority of songs that were present in the top spotify songs of 2023 were not even released in 2023. Out of close to the top 1000 songs, only 18.363064008394% of them were released in 2023 which suggests that most people still like to listen to songs released in previous years opposed to new songs. Another interesting finding of the data is that despite the popularity of Taylor Swift's 2023 tour (especially in regards to the tour generating enough revenue to kickstart the American economy), Morgan Wallen took the throne with 8 of his songs released in 2023 being included in the top spotify charts. It was also extremely interesting to see which key has the highest average streams as this provides us with consumer preference information as to which musical 'key' listeners enjoy. Please feel free to leave a comment with some feedback for me as this was my first personal project and wish to continue improving in the future.