SQL Data Analysis Project on Spotify's most streamed songs of 2023

★ Dataset on Spotify's most streamed songs of 2023 is obtained from Kaggle Website: [Most Streamed Spotify Songs 2023]

★ The Key Features of the Dataset includes:

- **track_name**: *Name of the song*
- **artist(s)_name**: *Name of the artist(s) of the song*
- **artist_count**: *Number of artists contributing to the song*
- **released_year**: *Year when the song was released*
- **released_month**: *Month when the song was released*
- **released_day**: *Day of the month when the song was released*
- **in_spotify_playlists**: *Number of Spotify playlists the song is included in*
- **in_spotify_charts**: *Presence and rank of the song on Spotify charts*
- **streams**: *Total number of streams on Spotify*
- **in_apple_playlists**: *Number of Apple Music playlists the song is included in*
- **in_apple_charts**: *Presence and rank of the song on Apple Music charts*
- **in_deezer_playlists**: *Number of Deezer playlists the song is included in*
- **in_deezer_charts**: *Presence and rank of the song on Deezer charts*
- **in_shazam_charts**: *Presence and rank of the song on Shazam charts*
- **bpm**: *Beats per minute, a measure of song tempo*
- **key**: *Key of the song*
- **mode**: *Mode of the song (major or minor)*
- **danceability_%**: *Percentage indicating how suitable the song is for dancing*
- **valence_%**: *Positivity of the song's musical content*
- **energy_%**: *Perceived energy level of the song*
- **acousticness_%**: *Amount of acoustic sound in the song*
- **instrumentalness_%**: *Amount of instrumental content in the song*
- **liveness_%**: *Presence of live performance elements*
- **speechiness_%**: *Amount of spoken words in the song*

★ The questions I am going to answer based on this Project are:

➔ Which artist has the most songs in the top streamed list of 2023?
➔ What is the average number of streams for songs released in each month of 2023?
➔ Are songs with higher danceability percentages generally more popular (i.e. have more streams)?
➔ What percentage of songs in the 'Top Spotify Songs 2023' were actually released in 2023?
➔ How many songs are released in 2023 among different age groups(minor or major)?
➔ What is the most used 'Key of the song'?

The project mainly includes four steps:
1.Data Exploration
2.Data Cleaning
3.Data Analysis
4.Conclusion

➢ Step 1:

Selecting the contents of the table obtained from kaggle which is named as ‘music’ and exploring the data.

```
SELECT *
FROM music
LIMIT 100;
```

➢ Step 2:

Cleaning the data involves removing null values from artists name,streams and track name.

```
7  •  SELECT COUNT(*)
8     FROM music
9     WHERE
10    track_name IS NULL OR `artist(s)_name` IS NULL OR streams IS NULL;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

COUNT(*)
0

WOW!..As there are no null values,I can proceed with further process.

Now I am going to convert all the uppercase column names to lowercase and also change some column names as per convenience

```
11
12 •  SELECT track_name,
13    LOWER(`artist(s)_name`) AS artists_name,
14    artist_count,
15    released_year,
16    released_month,
17    released_day,
18    in_spotify_playlists,
19    in_spotify_charts,
20    streams,
21    `key`,
22    `mode`,
23    `danceability_%` AS danceability
24    FROM music
25    LIMIT 100;
26
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

track_name	artists_name	artist_count	released_year	released_month	released_day	in_spotify_playlists	in_spo
Seven (feat. Latto) (Explicit Ver.)	latto, jung kook	2	2023	7	14	553	147
LALA	myke towers	1	2023	3	23	1474	48
vampire	olivia rodrigo	1	2023	6	30	1397	113
Cruel Summer	taylor swift	1	2019	8	23	7858	100

.

➢ Step 3:

Analysing the data:

➔ Which artist has the most songs in the top streamed list of 2023?

```
26 •  SELECT artists_name,COUNT(*) AS song_count
27    FROM music
28    GROUP BY artists_name
29    ORDER BY song_count DESC
30    LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wr

artists_name	song_count
taylor swift	33

Thus,Taylor swift has the most songs in the top streamed list of 2023.

➔ What is the average number of streams for songs released in each month of 2023?

```
32 •  SELECT released_month,ROUND(AVG(streams),0) AS streams_avg
33    FROM music
34    WHERE released_year=2023
35    GROUP BY released_month
36    ORDER BY released_month;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

released_month	streams_avg
1	230456561
2	231191625
3	170092262
4	138967171
5	94420746
6	68862912
7	42347572

Thus,February is the month with the highest avg number of streams

➔ Are songs with higher danceability percentages generally more popular (i.e. have more streams)?

```sql
SELECT
CASE
WHEN danceability<=25 then '0-25%'
WHEN danceability<=50 then '26-50%'
WHEN danceability<=75 then '51-75%'
ELSE '75-100%'
END
AS danceability_group,
AVG(streams) AS streams_avg
FROM music
GROUP BY danceability_group
ORDER BY streams_avg;
```

Result Grid | Filter Rows: | Export

danceability_group	streams_avg
75-100%	320367372.5535
51-75%	387339111.3303
26-50%	388881790.4860
0-25%	452250817.6667

Thus,Stream average is more in 0-25% danceability.So danceability do not affect more streams.

➔ What percentage of songs in the 'Top Spotify Songs 2023' were actually released in 2023?

```sql
WITH songcount AS
(SELECT COUNT(*) AS total,SUM( CASE WHEN released_year=2023 THEN 1 ELSE 0 END) AS 2023_released
FROM music)
SELECT (2023_released*100/total) AS RELEASED_IN_2023 FROM songcount;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

RELEASED_IN_2023
20.5811

Thus,only 20% of songs were released in 2023.

➔ How many songs are released in 2023 among different age groups(minor or major)?

```sql
SELECT `mode`,COUNT(*)
FROM music
WHERE released_year=2023
GROUP BY `mode`;
```

Result Grid | Filter Rows:

mode	COUNT(*)
Major	85
Minor	85

➔ How many songs are released in years other than 2023 among different age groups(minor or major)?

```
58 •  SELECT `mode`,COUNT(*)
59    FROM music
60    WHERE released_year!=2023
61    GROUP BY `mode`;
```

Result Grid | Filter Rows:

mode	COUNT(*)
Major	386
Minor	270

➔ What is the most used ‘Key of the song’?

```
63 •  SELECT `key`,COUNT(`key`) AS count_key
64    FROM music
65    GROUP BY `key`
66    ORDER BY count_key desc;
```

Result Grid | Filter Rows: | Export: | W

key	count_key
C#	102
G	83
F	81
	81
G#	79
B	71
D	68
A	66
F#	63
E	53
A#	52
D#	27

Thus,C# key is mostly used.

➢ Step 4:

Conclusion:

Thus,analysed some of the topics in Spotify's most streamed songs of 2023 using SQL With Mysql.