



SQL For Data Analyst

Presented by Wildan Zakaria



Wildanzakariaa@gmail.com

<https://www.linkedin.com/in/wildanzakaria>



INTRODUCTION

Hello everyone, I'm Wildan Zakaria and I am a final year student of Animal Husbandry at Padjadjaran University. Now I want to share **my mini portofolio** about data analysis using SQL. The dataset currently used is "**Most Streamed Songs (All Time) on Spotify**" taken from **Kaggle**. Spotify is a digital music, podcast, and video service that gives you access to millions of songs and other content from creators around the world.

Features.csv



DATASET

- **Id** : Unique ID given to each song on spotify
- **Name** : name of the song
- **Duration** : duration of the song in minutes
- **Energy** : energetic tracks feel fast, loud, and noisy.
- **Key** : The key the track is in.
- **Loudness** : The overall loudness of a track in decibels (dB).
- **Mode** : The type of scale from which its melodic content is derived.
- **Speechiness** : Speechiness detects the presence of spoken words in a track.
- **Acousticness** : A confidence measure of whether the track is acoustic.
- **Instrumentalness** : Predicts whether a track contains no vocals.
- **Liveness** : Detects the presence of the audience in the recording.
- **Valence** : Describes the musical positivity conveyed by a track.
- **Tempo** : Estimated overall track tempo in beats per minute (BPM).
- **Danceability** : Musical elements including tempo, rhythmic stability, beat strength, and overall regularity.



DATASET

Streams.csv

- **Song** : name of the song
- **Artist** : name of the artist
- **Streams_billions** : Streams in Billions
- **Release_date** : Release Date of the Song

ER Diagram

Field
Song
Artist
streams_billion
release_date



Field
id
name
duration
energy
key
loudness
mode
speechiness
acousticness
instrumentalness
liveness
valence
tempo
danceability

Case -1

What song titles by artist Ed Sheeran are included in the "Most Streamed Songs of All Time on Spotify"?

Query :

```
SELECT song  
FROM streams  
WHERE Artist = 'Ed Sheeran';
```

Output :

	song
▶	Shape of You
	Perfect
	Thinking Out Loud
	Photograph

Case -2

Song titles that start with the letter 'A' in table streams

Query :

```
SELECT song  
FROM streams  
WHERE song LIKE 'A%';
```

Output :

Result Grid	
	song
▶	As It Was
	All of Me

Case -3

Top 5 longest duration songs.

Query :

```
SELECT name AS song,  
       duration  
FROM features  
ORDER BY duration DESC  
LIMIT 5;
```

Output :

	song	duration
▶	Don't Stop Me Now	7.05
	Bohemian Rhapsody	5.91
	Lose Yourself	5.44
	Sicko Mode	5.21
	Smells Like Teen Spirit	5.03

Case -4

Top 5 shortest duration songs.

Query :

```
SELECT name AS song,  
       duration  
FROM features  
ORDER BY duration ASC  
LIMIT 5;
```

Output :

	song	duration
▶	Can't Hold Us	1.52
	Jocelyn Flores	1.99
	Mood	2.16
	Moonlight	2.25
	Montero (Call Me by Your Name)	2.3

Case -5

Song titles and artist on "Most Streamed Songs of All Time on Spotify," which has more than 1.5 billion streams and was released in 2019.

Query :

```
SELECT song, artist
FROM streams
WHERE streams_billion > 1.5
      AND release_date LIKE '%-19';
```

Output :

	song	artist
►	Blinding Lights	The Weeknd
	Dance Monkey	Tones And I
	Señorita	Shawn Mendes & Camila Cabello
	Bad Guy	Billie Eilish
	Don't Start Now	Dua Lipa
	Circles	Post Malone
	7 Rings	Ariana Grande

Case -6

The average duration of songs by artist Ed Sheeran that are included in the "Most Streamed Songs of All Time on Spotify."

Query :

```
SELECT artist,  
       ROUND(AVG(duration), 2) AS avg_duration  
FROM streams AS s  
INNER JOIN features AS f  
ON s.song = f.name  
WHERE Artist = 'Ed Sheeran';
```

Output :

	artist	avg_duration
▶	Ed Sheeran	3.97

Case -7

Top 3 songs for dancing. Where danceability is 0.0, at least it can be used for dancing, and where it is 1.0, it can be used for dancing the most.

Query :

```
SELECT name AS song,  
       danceability  
FROM features  
ORDER BY 2 DESC  
LIMIT 3;
```

Output :

	song	danceability
▶	Moonlight	0.921
	Humble	0.908
	Without Me	0.908

Case -8

Determine the average duration and energy for each artist.

Query :

```
SELECT artist,  
       ROUND(AVG(duration), 2) AS avg_duration,  
       ROUND(AVG(energy), 2) AS avg_energi  
FROM streams AS s  
INNER JOIN features AS f  
ON s.song=f.name  
GROUP BY artist  
ORDER BY artist ASC;
```

Output :

	artist	avg_duration	avg_energi
▶	24kGoldn & Iann Dior	2.16	0.74
	Alan Walker	3.54	0.63
	Arctic Monkeys	4.54	0.53
	Ariana Grande	3.22	0.48
	Avicii	4.12	0.72

Case -9

Determine the songs released in 2019 and determine the ranking based on the most streams. (The number after the comma consists of 2 digits)

Query :

```
SELECT song,  
       ROUND(streams_billion, 2) AS streams,  
       RANK() OVER(ORDER BY streams_billion DESC) AS rank_streams  
FROM streams  
WHERE release_date LIKE '%-19';
```


Output :

	song	streams	rank_streams
►	Blinding Lights	3.45	1
	Dance Monkey	2.77	2
	Señorita	2.37	3
	Bad Guy	2.22	4
	Don't Start Now	2.17	5

Case -10

Create a new field with the name "new_mode" where mode 1 is the same as major and 0 is the same as minor.

Query :



```
SELECT name AS song,  
       mode,  
       CASE WHEN mode = 1 THEN 'major'  
            WHEN mode = 0 THEN 'minor'  
            ELSE NULL END AS 'new_mode'  
FROM features  
ORDER BY name ASC;
```

Output :

	song	mode	new_mode
▶	Till I Collapse	1	major
	7 Rings	0	minor
	7 Years	1	major
	All of Me	1	major
	As It Was	0	minor

Case -11

Sum the songs with the 'minor' scale category from the previous query.

Query :

```
SELECT COUNT(new_mode) AS total_minor
FROM (SELECT name AS song,
             mode,
             CASE WHEN mode = 1 THEN 'major'
                  WHEN mode = 0 THEN 'minor'
                  ELSE NULL END AS 'new_mode'
      FROM features
      ORDER BY name ASC) AS new_features
WHERE new_mode = 'minor';
```

Output :

	total_minor
▶	35

35



THANK YOU!

Wildanzakariaa@gmail.com
<https://www.linkedin.com/in/wildanzakaria>