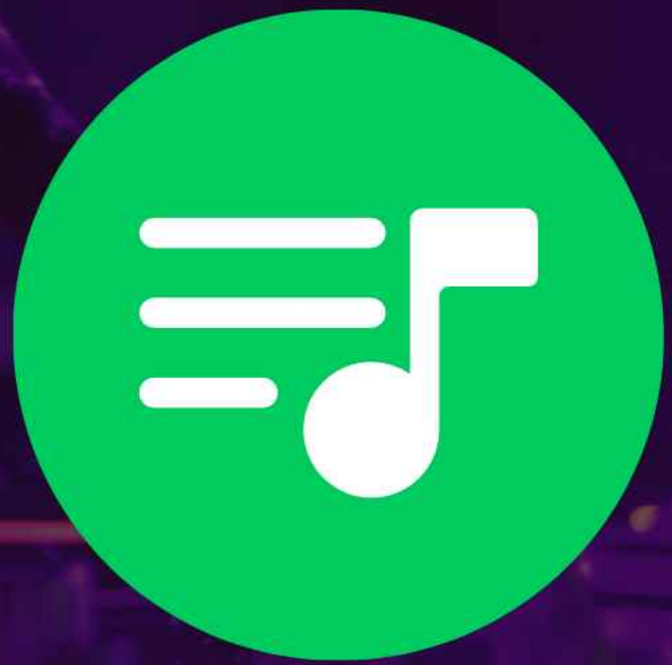
Home About Playlist

SQL PROJECT

Spotify

DATA MEETS MUSIC: SPOTIFY ANALYSIS USING SQL

LISTEN ON



SPOTIFY ANALYSIS

Presented by: [Shilpi Kumari]

Project Objective

- To analyze Spotify's music dataset for trends in music popularity, artist performance, and genre insights using SQL.
- Showcase the value of SQL for extracting insights from large datasets.

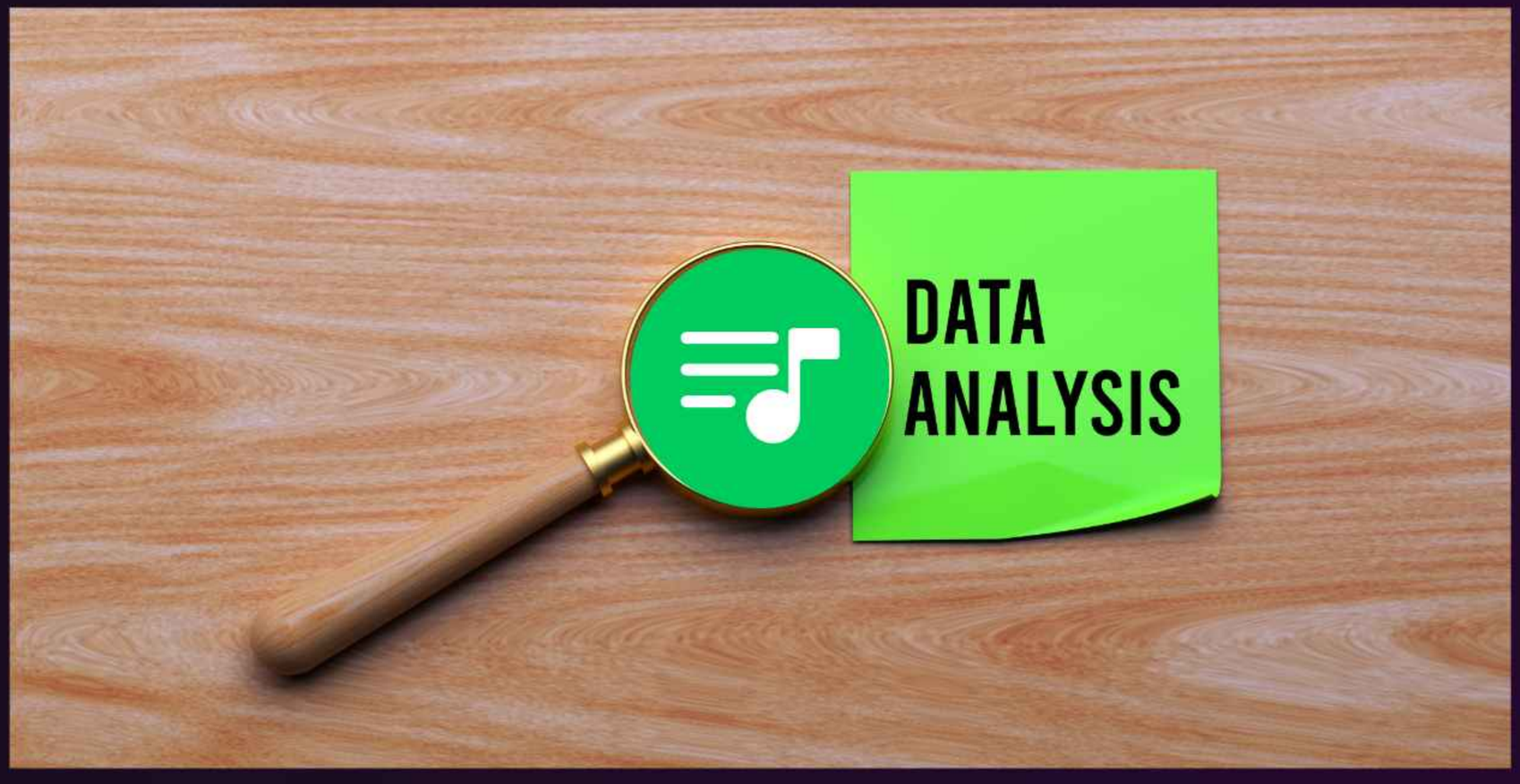
Dataset Columns Overview

Columns Used in the Project

- title: Name of the song
- artist: Name of the artist
- top genre: Genre of the song
- year: Release year of the song
- bpm: Beats per minute (tempo)
- nrgy: Energy level
- dnce: Danceability score
- dB: Loudness in decibels
- live: Live performance score
- val: Positivity of the song
- dur: Duration of the song (seconds)
- acous: Acoustic level
- spch: Speechiness level
- pop: Popularity score

ANALYSIS BY BASIC LEVEL QUERIES

QUESTIONS-1 TO 5



SPOTIFY ANALYSIS

Query Levels Introduction

Queries are divided into three levels:

- Basic Queries
- Moderate Queries
- Advanced Queries

Let's Showcase them into basic, moderate, and advanced levels with Its syntax with Its Insights

1. Basic Level Queries (With Insights)

Q-1) Write a query to display all columns from the dataset.

Input:-

```
Select * From spotify_data
```

Output :-

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

Serial_Number	title	artist	top genre	year	bpm	nrgy	dnce	dB	live	val	dur	acous	spch	pop
1	Hey, Soul Sister	Train	neo mellow	2010	97	89	67	-4	8	80	217	19	4	83
2	Love The Way You Lie	Eminem	detroit hip hop	2010	87	93	75	-5	52	64	263	24	23	82
3	TiK ToK	Kesha	dance pop	2010	120	84	76	-3	29	71	200	10	14	80
4	Bad Romance	Lady Gaga	dance pop	2010	119	92	70	-4	8	71	295	0	4	79
5	Just the Way You Are	Bruno Mars	pop	2010	109	84	64	-5	9	43	221	2	4	78
6	Baby	Justin Bieber	canadian pop	2010	65	86	73	-5	11	54	214	4	14	77
7	Dynamite	Taio Cruz	dance pop	2010	120	78	75	-4	4	82	203	0	9	77
8	Secrets	OneRepublic	dance pop	2010	148	76	52	-6	12	38	225	7	4	77
9	Empire State of Mind (Part II) Broken Down	Alicia Keys	hip pop	2010	93	37	48	-8	12	14	216	74	3	76
10	Only Girl (In The World)	Rihanna	barbadian pop	2010	126	72	79	-4	7	61	235	13	4	73
11	Club Can't Handle Me (feat. David Guetta)	Flo Rida	dance pop	2010	128	87	62	-4	6	47	235	3	3	73
12	Marry You	Bruno Mars	pop	2010	145	83	62	-5	10	48	230	33	4	73

potify_data15 ×

Insight of this query - It provides a complete view of the dataset's structure and contents, serving as the foundation for deeper analysis.

Q-2) Retrieve all columns for songs released in a 2010, 2015, 2019 year

Input -:

```
select *
from
Spotify_data
where year in ('2010','2015','2019')
```

Output -:

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

Serial_Number	title	artist	top genre	year	bpm	nrgy	dnce	dB	live	val	dur	acous	spch	pop
1	Hey, Soul Sister	Train	neo mellow	2010	97	89	67	-4	8	80	217	19	4	83
2	Love The Way You Lie	Eminem	detroit hip hop	2010	87	93	75	-5	52	64	263	24	23	82
3	TiK ToK	Kesha	dance pop	2010	120	84	76	-3	29	71	200	10	14	80
4	Bad Romance	Lady Gaga	dance pop	2010	119	92	70	-4	8	71	295	0	4	79
5	Just the Way You Are	Bruno Mars	pop	2010	109	84	64	-5	9	43	221	2	4	78
6	Baby	Justin Bieber	canadian pop	2010	65	86	73	-5	11	54	214	4	14	77
7	Dynamite	Taio Cruz	dance pop	2010	120	78	75	-4	4	82	203	0	9	77
8	Secrets	OneRepublic	dance pop	2010	148	76	52	-6	12	38	225	7	4	77
9	Empire State of Mind (Part II) Broken Down	Alicia Keys	hip pop	2010	93	37	48	-8	12	14	216	74	3	76
10	Only Girl (In The World)		barbadian pop	2010	126	72	79	-4	7	61	235	13	4	73
11	Club Can't Handle Me (feat. David Guetta)		dance pop	2010	128	87	62	-4	6	47	235	3	3	73

Insight of this query- This query retrieves all songs from the years 2010, 2015, and 2019, allowing for focused analysis of tracks released in these specific years.

Q-3) Count the number of songs per year

Input-:

```
SELECT year, COUNT(*) AS song_count
FROM Spotify_data
GROUP BY year;
```

Output-:

Result Grid | Filter Rows:

	year	song_count
▶	2010	51
	2011	52

Insights of this query-:Identifies the most active years for song releases.

Q-4) Retrieve the top 5 songs with the highest popularity score?

Input-:

```
SELECT title, artist, pop
FROM Spotify_data
ORDER BY pop DESC
LIMIT 5;
```

Output-:

Result Grid | Filter Rows:

	title	artist	pop
▶	Hey, Soul Sister	Train	83
	Love The Way You Lie	Eminem	82
	A Thousand Years	Christina Perri	81
	TiK ToK	Kesha	80
	Someone Like You	Adele	80

Insights of this query-: Highlights the most successful and well-liked songs

Q-5) Display songs where the acoustic level is greater than 80?

Input:-

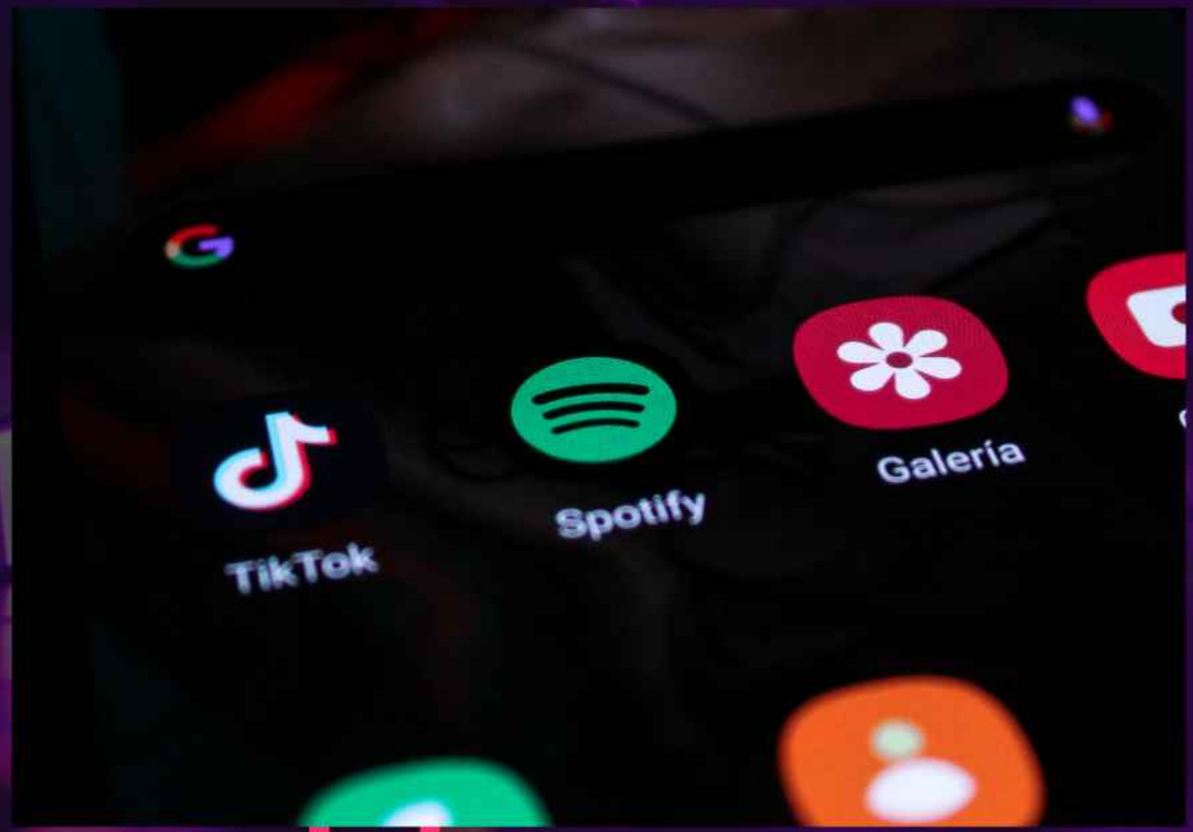
```
SELECT title, artist, acous
FROM Spotify_data
WHERE acous > 80;
```

Output:-

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

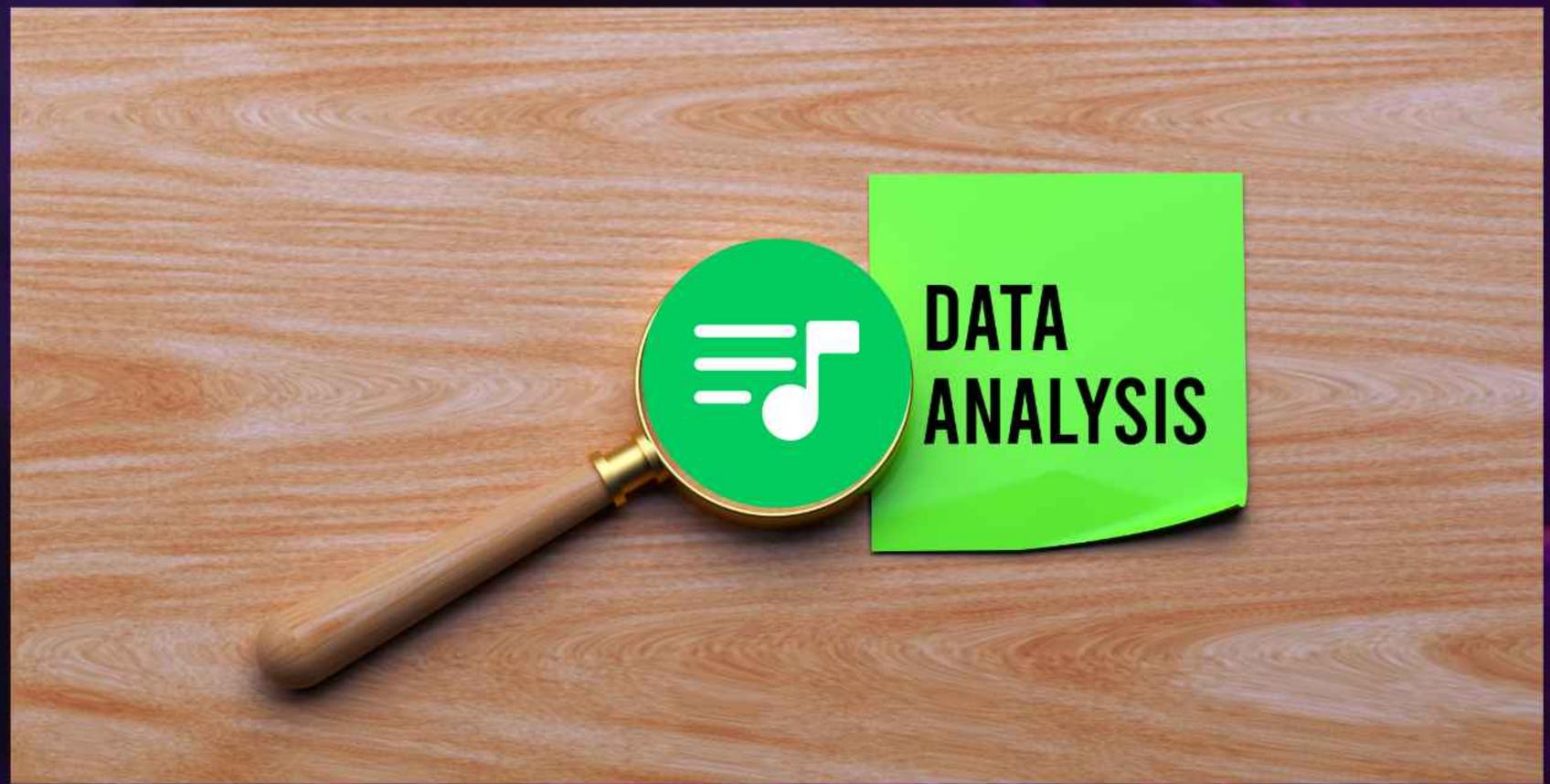
title	artist	acous
Bound To You - Burlesque Original Motion Pictur...	Christina Aguilera	83
You Lost Me	Christina Aguilera	85
Someone Like You	Adele	89
Turning Page	Sleeping At Last	91

Insight of this query -Identifies songs with prominent acoustic features for relaxing playlists.



Analysis by Medium level Queries

Questions-6 to 10



SPOTIFY ANALYSIS

Q-6) What is the average BPM for each genre?

Input -:

```
SELECT 'top genre', AVG(bpm) AS avg_bpm
FROM Spotify_data
GROUP BY 'top genre';
```

Output -:

Result Grid | Filter Rows:

top genre	avg_bpm
top genre	120.7573

Insight of this query- Reveals tempo trends across different genres.

Q-7) Find songs with an energy score above 70 and danceability above 80?

Input -:

```
SELECT title, artist, nrgy, dnce
FROM Spotify_data
WHERE nrgy > 70 AND dnce > 80;
```

Output -:

Result Grid | Filter Rows: | Export:

title	artist	nrgy	dnce
Telephone	Lady Gaga	83	83
The Time (Dirty Bit)	The Black Eyed Peas	81	82
S&M Remix	Rihanna	81	82

Insight of this query-Identifies high-energy and highly danceable songs, making it ideal for curating upbeat and engaging playlists.

Q-8) Calculate the total duration of songs for each artist?

Input-:

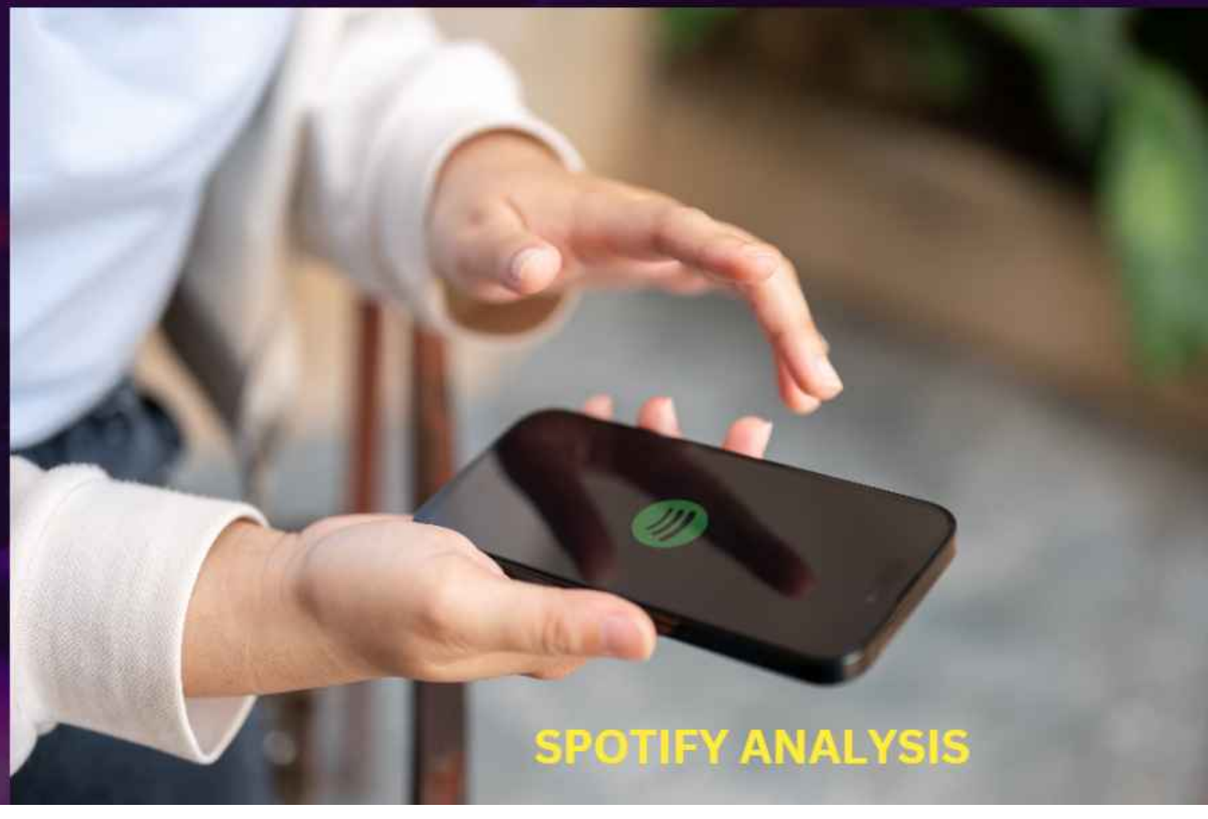
```
SELECT artist, SUM(dur) AS total_duration
FROM Spotify_data
GROUP BY artist
ORDER BY total_duration DESC;
```

Output-:

Result Grid | Filter Rows: | Export:

artist	total_duration
Lady Gaga	2192
Rihanna	1483
The Black Eyed Peas	1338
Kesha	1199
Katy Perry	1152
Bruno Mars	1125
BeyoncÃ©	1031
Jennifer Lopez	909
Christina Aguilera	892
Britney Spears	890
Alicia Keys	738
OneRepublic	716

Insights of this query-Highlights prolific artists with long total durations.



Q-9) What is the song from the Spotify dataset with the highest and lowest "top genre" in alphabetical order?

Input-:

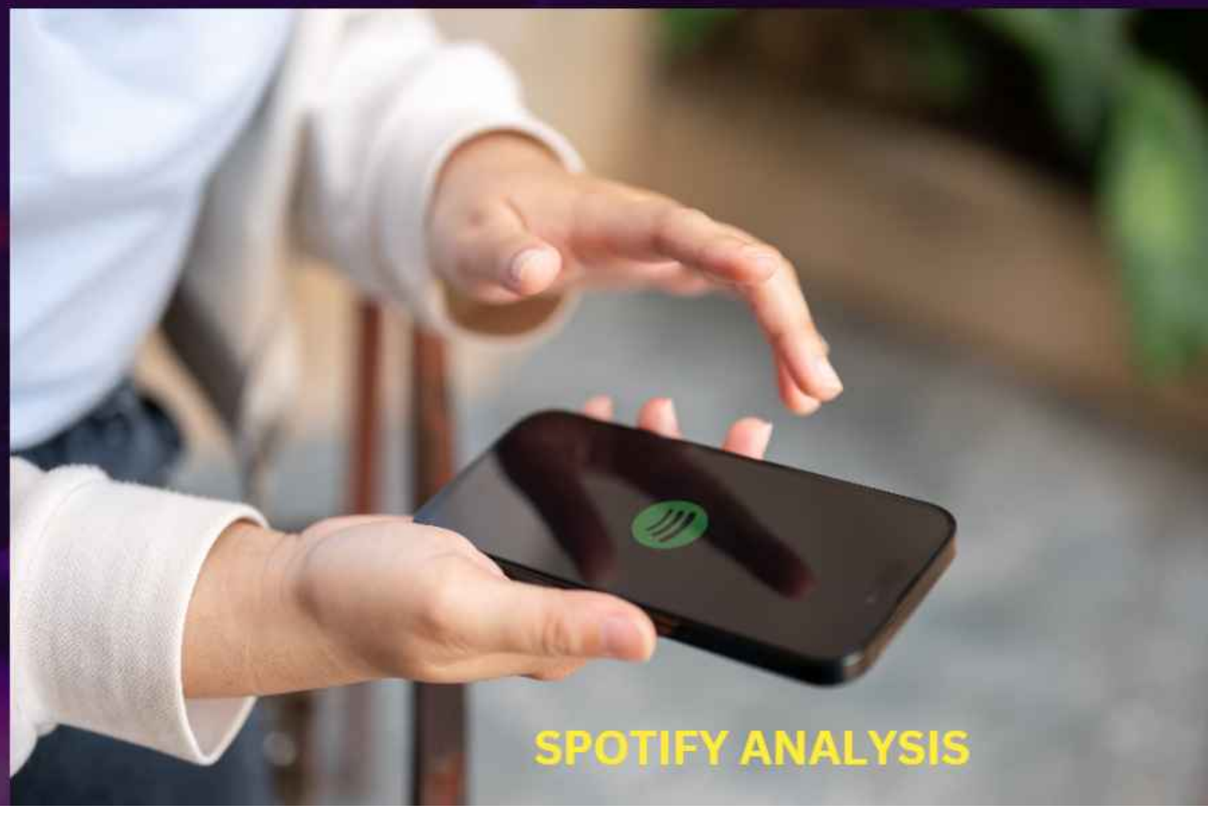
```
SELECT title, artist, pop ,'top genre'
FROM spotify_data ORDER BY 'top genre'  DESC LIMIT 1;
SELECT title, artist, pop , 'top genre' FROM spotify_data ORDER BY 'top genre' ASC LIMIT 1;
```

Output-:

Result Grid | Filter Rows:

	title	artist	pop	top genre
▶	Hey, Soul Sister	Train	83	top genre

Insights of this query-: This query identifies the songs with the highest and lowest alphabetical "top genre" values, along with their popularity, providing insights into genre distribution and its potential impact on song popularity.



Q-10) What are the top 5 genres with the highest average 'DNCE' values in the Spotify dataset?

Input-:

```
SELECT 'top genre',
SUM(DNCE) / COUNT('top genre') AS AVERAGE_DNCE
FROM spotify_data
GROUP BY 'top genre'
ORDER BY AVERAGE_DNCE DESC
LIMIT 5;
```

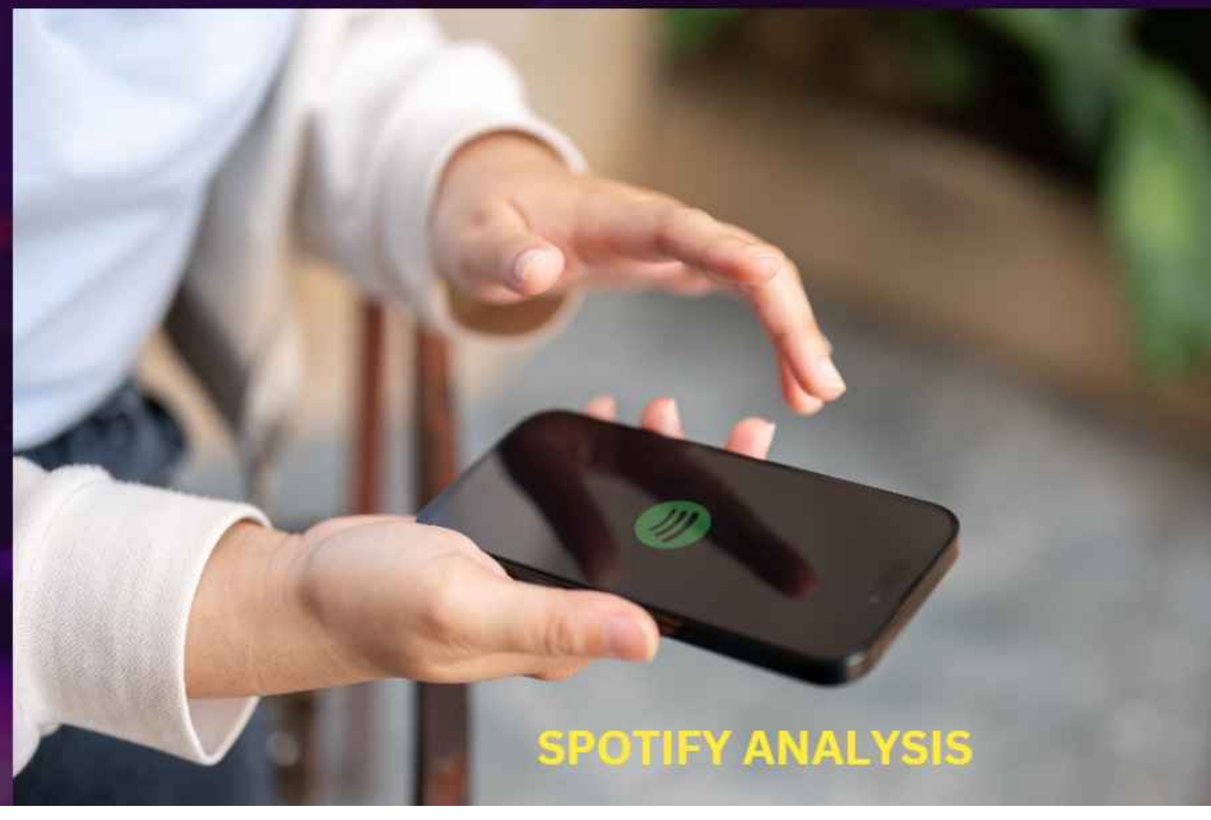
DOWNLOAD

Output-:

Result Grid | Filter Rows: | Export:

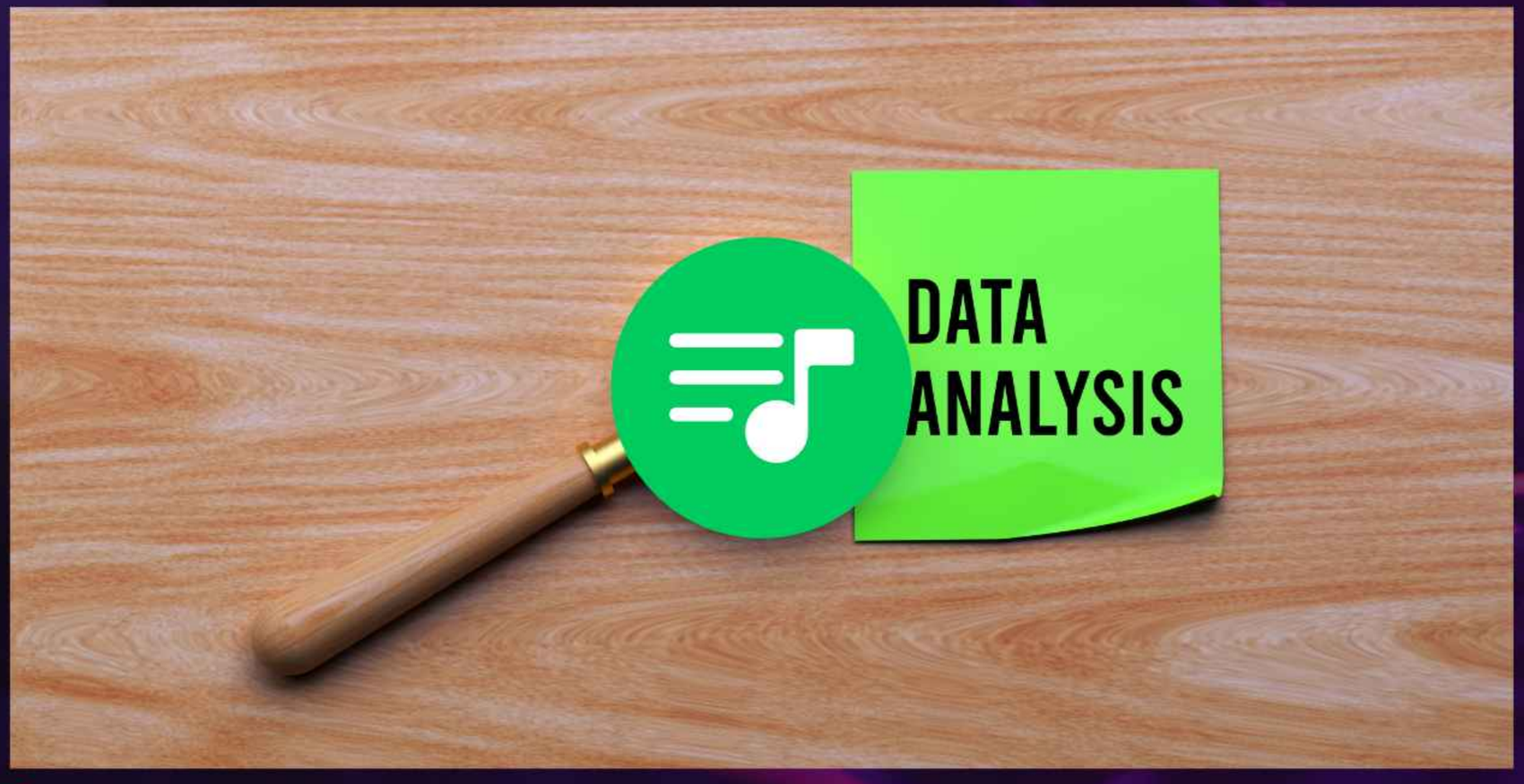
top genre	AVERAGE_DNCE
top genre	63.9223

Insights of this query-This query identifies the top 5 genres with the highest average danceability (DNCE) in the Spotify dataset.



Analysis by Advanced level Queries

Questions- 11 to 15



SPOTIFY ANALYSIS

Q-11) What are the most popular songs (with a popularity score above 77) along with their artist names and genre, ordered by popularity in descending order?

Input-:

```
SELECT s.title, s.artist, s.pop, g.genre_name
FROM Spotify_data s
JOIN Genres g ON s.`top genre` = g.genre_id
WHERE s.pop > 77
ORDER BY s.pop DESC;
```

Output-:

Result Grid | Filter Rows: | Export:

title	artist	pop	genre_name
Just the Way You Are	Bruno Mars	78	Pop
Just the Way You Are	Bruno Mars	78	Pop

Insight of this query - This query identifies songs with a popularity score greater than 77, retrieves the song title, artist, popularity score, and genre name, and sorts the results to highlight the most popular songs first.

Q-12) Find the artists who have released the longest song in the dataset.

Input :-

```
SELECT artist, title, dur
FROM Spotify_data
WHERE dur = (SELECT MAX(dur) FROM Spotify_data);
```

Output:-

Result Grid | Filter Rows:

	artist	title	dur
▶	Kanye West	Monster	379

Insight of this query -: This subquery determines the longest song duration and retrieves the respective artist and song title, showcasing the use of a single-value subquery.

Q-13) Determine the trend of speechiness over the years.

Input:-

```sql
SELECT
YEAR,
AVG(SPCH) AS AVG_SPEECHINESS,
MIN(SPCH) AS MIN_SPEECHINESS,
MAX(SPCH) AS MAX_SPEECHINESS
FROM
Spotify_data
GROUP BY
YEAR
ORDER BY
YEAR;
```

Insight if this query -:This query analyzes the average, minimum, and maximum speechiness of songs for each year, grouped by YEAR and ordered chronologically.

Output:-

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

YEAR	AVG_SPEECHINESS	MIN_SPEECHINESS	MAX_SPEECHINESS
2010	8.8824	3	45
2011	9.2115	3	39

Q-14) Identify any correlations between BPM and energy levels across different genres? .

Input-:

```sql
SELECT
`top genre`,
correlation,
CASE
WHEN correlation > 0 THEN 'Positive Correlation'
WHEN correlation < 0 THEN 'Negative Correlation'
WHEN correlation = 0 THEN 'Zero Correlation'
END AS correlation_status
FROM (
SELECT
`top genre`,
(COUNT(*) * SUM(bpm * nrgy) - SUM(bpm) * SUM(nrgy)) /
(SQRT((COUNT(*) * SUM(bpm * bpm) - SUM(bpm) * SUM(bpm))
* (COUNT(*) * SUM(nrgy * nrgy) - SUM(nrgy) * SUM(nrgy)))) AS
correlation
FROM
spotify_data
GROUP BY
`top genre`
) AS subquery alias;
```

Output-:

Result Grid | Filter Rows: | Export:

top genre	correlation	correlation_status
neo mellow	NULL	NULL
detroit hip hop	NULL	NULL
dance pop	0.36116711272813506	Positive Correlation
pop	0.24657252350954725	Positive Correlation
canadian pop	1	Positive Correlation
hip pop	0.8414609872967534	Positive Correlation
barbadian pop	-0.6836136484660403	Negative Correlation
atl hip hop	-0.6970966755769257	Negative Correlation
australian pop	-0.8740045982274169	Negative Correlation
indie pop	NULL	NULL
art pop	NULL	NULL
colombian pop	NULL	NULL

Insight of this query -: This query calculates the correlation between BPM and energy levels (nrgy) for each music genre (top genre) in the Spotify dataset, categorizing the correlation as positive, negative, or zero.

Q-15) Indentify any outliers in the "dur" (duration) column using statistical methods?

Input -:

```
WITH Stats AS (
SELECT
AVG(dur) AS Mean,
STDDEV_POP(dur) AS StdDev
FROM
spotify_data
)
SELECT
dur,
CASE
WHEN ABS(dur - Mean) > 3 * StdDev THEN 'Outlier'
ELSE 'Not an Outlier'
END AS Outlier_Status
FROM
spotify_data, Stats
having Outlier_Status='Outlier';
```

Output-:

dur	Outlier_Status
329	Outlier
379	Outlier
329	Outlier
354	Outlier
424	Outlier
353	Outlier
403	Outlier
341	Outlier

Insight of this query -: This query identifies songs in the Spotify dataset as outliers if their duration (dur) deviates by more than three standard deviations from the mean duration.

KEY TAKEAWAYS

1)SQL queries enable powerful insights into Spotify's music data.

2)The project showcased various aspects such as genre analysis, artist trends, and song popularity.

3) Data-driven exploration can assist in music curation, marketing, and industry decisions.

Contact Information

LinkedIn Profile:

https://www.linkedin.com/in/shilpi-kumari-146826241/

Email Address:

shilpikumari9897@gmail.com

Resume:

https://shilpi-resume.tiiny.site/


THANK
YOU