Capstone project

Antonio Muro Parra

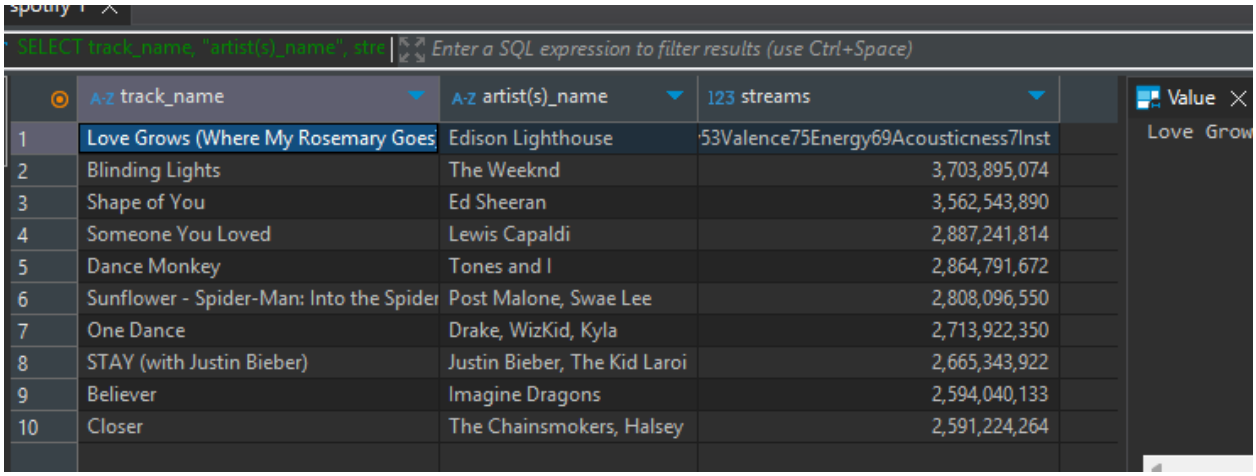
Spotify Song Analysis

Spotify remains one of the most influential platforms for music streaming globally. With over 500 million active users, analyzing Spotify data offers insight into consumer behavior, music trends, and artist performance. This report for my capstone project focuses on an analysis of the "Top Spotify Songs 2023" dataset which I will be analyzing with SQL queries.

. As well as show a dashboard with visualizations showing trends that I found. I used SQL to organize the data from the cvs file I got off Kaggle.
I will be using SQlite and will be doing all this on Dbeaver. For my dashboard I have made it in Power BI. As well have uploaded all my files to Github. For the first order of business I did some EDA. Making sure my dataset worked for me as well checked for any null or missing values. Made sure that all the columns were present when I transferred my cvs file into Dbeaver. Which thankfully did work, also found another cvs file to supplement incase of anything. Here it is, its more spotify tracks that have helped me identify more trends. As for the purpose of this project, I simply just wanted to use SQL as a means of identifying trends from a dataset. My questions regarding this project are the following. Which artists were the most popular during 2023? What song was the most streamed?

What songs were more found to be “dancing” songs than others? Does the popularity of artists have to do with how catchy the song is or how many songs were in rotation?

Now for my first question, what I could possibly get from this dataset was, how popular some tracks were during this year, as shown here



From the query itself, We can see that Blinding Lights has 3.7 millions of streams during this year, making it one of the most popular streams of this year. Second to that is Shape of You, coming in at 3.5 million streams. This shows what the general “vibe” was around this year which seems to be more towards pop/hip pop. This will be the main trend for the rest of our dataset, the most popular songs were songs that come from the genres hip pop and pop. As for the next question I had was, the danceability of these top songs we had in the first query. To determine what makes these top songs so popular. Here is the query I committed:

```sql
--Danceability of the most popluar songs
SELECT track_name, "artist(s)_name", streams, "danceability_%"
FROM (
    SELECT *
    FROM spotify
    ORDER BY streams DESC
    LIMIT 10
) AS top_tracks
ORDER BY "danceability_%" DESC;
```

Now with this we can see how “danceability” these songs really are and if this is what makes them popular. As shown down below, some of the top songs with tons and tons of streams aren't that “danceable”. Blinding light by The Weekend was one of the top songs but its has such a low score of “danceability”

track_name	artist(s)_name	danceability_%
Shape of You	Ed Sheeran	83
Dance Monkey	Tones and I	82
One Dance	Drake, WizKid, Kyla	77
Believer	Imagine Dragons	77
Sunflower - Spider-Man: Into the Spider-Verse	Post Malone, Swae Lee	76
Closer	The Chainsmokers, Halsey	75
STAY (with Justin Bieber)	Justin Bieber, The Kid Laroi	59
Love Grows (Where My Rosemary Goes)	Edison Lighthouse	53
Blinding Lights	The Weeknd	50
Someone You Loved	Lewis Capaldi	50

Therefore meaning these top songs weren’t top songs just because of how they made people move. Meaning there must be another reason as to why so many people decided to listen to these songs. Mind you the avg danceability of all these songs are 64-65% meaning most songs are actually making people dance

and move around. Okay so why else could these top songs be so popular and that's when I realized, what if it's the amount of songs that these artists had in rotation that made them so popular. Using this query I found out that some of the top artist had tons of tracks playing in 2023

```sql
--Artist Saturation
SELECT "artist(s)_name", COUNT(*) AS track_count
FROM spotify s
GROUP BY "artist(s)_name"
HAVING COUNT(*) > 3;
```

Which resulted in, The weekend having around 22 tracks, Taylor Swift had 34, and SZA had 19 tracks. This had me wondering if the saturation of music is why these artists with an average danceability of 50 and below had millions of streams. This is most likely because of the saturation of music that was playing during this year. Which just so happens to be from these artists. Now I decided to check which genres were most popular during this time out of these top artists we saw from our previous query.

```sql
SELECT s.track_name, s."artist(s)_name", s.streams, d.track_genre
FROM spotify s
JOIN dataset d ON s.track_name = d.track_name
ORDER BY s."artist(s)_name" DESC
LIMIT 20;
```

With this query, it'll show us the 20 songs and their ranks. Most of these songs are kpop and pop songs. Giving us the theory that trending songs are majority pop songs. Making pop a high demand for most consumers. This is pretty much why most songs are pop in trending due to the high amount of songs and artists. In conclusion, we can demonstrate that even if these songs are trending,

popularity soars high due to its high volume. Despite most of these songs not being the most “danceable” songs to dance to. Understandable so, furthermore these songs are overrated and played on repeat in commercial malls, shoe stores, and other public outing places. This has been my research on the topic. Thank you for your time, farewell. :)