--Niyo final Projects - Spotify EDA

--Retrieving the DataSet, Top 10

SELECT \* FROM `niyo-sql-project.Spotify.Analysis`

LIMIT 10;

-- Also shows top 10 streamed songs

--Data set is going to be a recommendation for TWICE so need to see their position and peak rank

SELECT \* FROM `niyo-sql-project.Spotify.Analysis`

WHERE Artist\_Name = 'TWICE '

LIMIT 100;

-- Top 10 most streamed artists

SELECT Artist\_Name, SUM(Total\_Streams)

FROM `niyo-sql-project.Spotify.Analysis`

GROUP BY Artist\_Name

ORDER BY SUM(Total\_Streams) DESC

LIMIT 10;

-- Drake, Post Malone, Juice WRLD, The Weeknd, XXXTENTACION, Taylor Swift, Ariana Grande, Billie Eilish, Lil Uzi Vert, Bad Bunny.

SELECT Artist\_Name, SUM(Total\_Streams)

FROM `niyo-sql-project.Spotify.Analysis`

WHERE Artist\_Name LIKE '%BTS%'

GROUP BY Artist\_Name;

-- BTS has 456 million total streams in this dataset, compared to current 31 billion

-- BTS is also a KPOP group, by seeing their data and trends, this may suggest how to increase TWICE's Streaming numbers

--Retriving Artist name, song and how many times they have been in the top 10

SELECT Position, Artist\_Name, Song\_Name, Top\_10\_\_xTimes\_

FROM `niyo-sql-project.Spotify.Analysis`

ORDER BY Top\_10\_\_xTimes\_ DESC;

-- Sunflower Spiderman's (by Post Malone) current position is 1st

-- been in the top 10 the most times (302)

-- Stay (The Kid LAROI) and INDUSTRY BABY (Lil Nas X) have positions not within top 20 of the dataset, however, the songs have been in top 10 the 2nd and 3rd most times

--Retrieve Unique values from artist name column

SELECT DISTINCT Artist\_Name

FROM `niyo-sql-project.Spotify.Analysis`;

-- This lets us know that there are 1612 Different artists in the top 10,000

-- How many times has Post Malone been in top 10,000

SELECT \*

FROM `niyo-sql-project.Spotify.Analysis`

WHERE Artist\_Name = 'Post Malone ';

-- Out of 10,000 songs, Post Malone's songs account for 69 of them

-- How many songs are released within 20 days

SELECT \* FROM `niyo-sql-project.Spotify.Analysis`

WHERE Days BETWEEN 0 AND 20

ORDER BY Position ASC;

-- 7473 songs meet this Criteria, However, the highest they ranked higher was 1029 in top 10,000, where 11 of them where SZA songs.

-- Does the number of days the song has been released have any impact on number of streams

SELECT Position, Artist\_Name, Total\_Streams, Days

FROM `niyo-sql-project.Spotify.Analysis`

WHERE Position = 1 OR Position = 1029

ORDER BY Position ASC;

-- Can see that although the first position has 20 times more streams, it has been released for much longer (60 times more days) which may be a contributing factor to its high streams and therefore position

-- Does the number of days the song has been released have any impact on number of streams

SELECT Position, Song\_Name, Total\_Streams, Days

FROM `niyo-sql-project.Spotify.Analysis`

ORDER BY Days DESC

LIMIT 10;

-- Looking at the songs with the top 10 highest days since release, 7 of them have positions within the top 10 of 10,000

-- Looking for null values in Song or Artist name

SELECT \* FROM `niyo-sql-project.Spotify.Analysis`

WHERE Artist\_Name IS NULL OR Song\_Name IS NULL;

-- 4 songs (Position 5507, 7178, 8216, 6218) without names. No Artist names missing

-- How many times, on average, has each artist been in the top 10, and who are the top 3

SELECT Artist\_Name, ROUND(AVG(Top\_10\_\_xTimes\_),2) AS AVG\_Times\_in\_Top\_10

FROM `niyo-sql-project.Spotify.Analysis`

GROUP BY Artist\_Name

ORDER BY AVG(Top\_10\_\_xTimes\_) DESC;

-- On average, The top 3 artists whose songs have been in the top 10 the most are G-Eazy, Sheck Wes and Carolina Gaitan. But this may be because they have less songs in the top 10,000

SELECT \* FROM `niyo-sql-project.Spotify.Analysis`

WHERE Artist\_Name IN ('G-Eazy', 'Sheck Wes ', 'Carolina Gaitán ');

-- Shows that these artists only have one song each, and so will have a higher average than post malone (for example) who has 69 songs in top 10,000.

-- Which artists have songs that have been in the top 10?. The total number of times that all the artists songs have been in the top 10 must be more than 1000

SELECT Artist\_Name, SUM(Top\_10\_\_xTimes\_) AS Total\_In\_Top\_10

FROM `niyo-sql-project.Spotify.Analysis`

WHERE Peak\_Position <= 10

GROUP BY Artist\_Name

HAVING SUM(Top\_10\_\_xTimes\_)>1000;

-- Post Malone, The Weeknd and Drake's songs have appeared more than 1000 times in the top 10

-- Categorise Total\_Streams to see distribution (make 9 Buckets)

SELECT COUNT(Total\_Streams),

CASE

WHEN Total\_Streams < 100000000 THEN 'Less than 100'

WHEN Total\_Streams > 800000000 THEN 'More than 800'

WHEN Total\_Streams > 700000000 THEN 'Between 700 and 800'

WHEN Total\_Streams > 600000000 THEN 'Between 600 and 700'

WHEN Total\_Streams > 500000000 THEN 'Between 500 and 600'

WHEN Total\_Streams > 400000000 THEN 'Between 400 and 500'

WHEN Total\_Streams > 300000000 THEN 'Between 300 and 400'

WHEN Total\_Streams > 200000000 THEN 'Between 200 and 100'

WHEN Total\_Streams > 100000000 THEN 'Between 100 and 200'

END AS TotalStreamsInMillions

FROM `niyo-sql-project.Spotify.Analysis`

GROUP BY TotalStreamsInMillions

ORDER BY COUNT(Total\_Streams) DESC;

-- Majority of the songs have less than 100 million streams

-- Hints that dataset may need to be split for clearer visuals

-- Viewing new dataset

SELECT \* FROM `niyo-sql-project.Spotify.Short`;

-- Inner Joins of Analysis and Short tables by song name

SELECT Position, Artist\_Name, Song\_Name, Peak\_Position, weeks\_on\_chart, Total\_Streams, Top\_10\_\_xTimes, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, tempo, duration\_ms

FROM `niyo-sql-project.Spotify.Short`

INNER JOIN `niyo-sql-project.Spotify.Analysis`

ON `niyo-sql-project.Spotify.Short`.track\_name = `niyo-sql-project.Spotify.Analysis`.Song\_Name

ORDER BY Position ASC;

-- Inner Joins on just the song name column would not be suitable as there as multiple songs with the same names but different artists. Will try to use Left\_on and right\_on in python.

--Looking only at rows with position in top 100

SELECT Position, Artist\_Name, Song\_Name, Peak\_Position, weeks\_on\_chart, Total\_Streams, Top\_10\_\_xTimes, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, tempo, duration\_ms

FROM `niyo-sql-project.Spotify.Short`

INNER JOIN `niyo-sql-project.Spotify.Analysis`

ON `niyo-sql-project.Spotify.Short`.track\_name = `niyo-sql-project.Spotify.Analysis`.Song\_Name

WHERE Position < 100

ORDER BY Position ASC;

-- Shows 27 results

-- Effect of duration on position

SELECT Position, Artist\_Name, Song\_Name, Peak\_Position, weeks\_on\_chart, Total\_Streams, Top\_10\_\_xTimes, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, tempo, duration\_ms

FROM `niyo-sql-project.Spotify.Short`

INNER JOIN `niyo-sql-project.Spotify.Analysis`

ON `niyo-sql-project.Spotify.Short`.track\_name = `niyo-sql-project.Spotify.Analysis`.Song\_Name

WHERE Position < 100

ORDER BY duration\_ms ASC;

-- Can't see much of a correlation between duration and position. Would need visual

SELECT Position, Artist\_Name, Song\_Name, Peak\_Position, weeks\_on\_chart, Total\_Streams, Top\_10\_\_xTimes, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, tempo, duration\_ms

FROM `niyo-sql-project.Spotify.Short`

INNER JOIN `niyo-sql-project.Spotify.Analysis`

ON `niyo-sql-project.Spotify.Short`.track\_name = `niyo-sql-project.Spotify.Analysis`.Song\_Name

WHERE Position < 100

ORDER BY danceability ASC;

-- Can’t see much of correlation between danceability and position.