SQL ASSIGNMENT 5

Assignment - 5: SPOTIFY DATA ANALYSIS

Instructions :

- *Reference: Lectures In Snowflake & SQL folder (AWA - APP+WEBSITE)*
- *Due Date : 30th-Sept-2023 11:59 PM(Midnight)*
- *Late submissions will not be evaluated*
- *Its mandatory to do all questions*
- *Use SNOWFLAKE for the task submission while for practice one can execute in MySQL Workbench too.*
- *Proper comments should be given for the code explanation wherever required.*
- *Proper snippets should be attached of the output(mandatory) and write the code too.*
- *Don't do plagiarism*
- *Kindly don't USE JOINS or WINDOWS functions in any of the problems.*
- *Kindly upload the assignment by uploading it in the below GOOGLE DRIVE FOLDER as per the mentioned format(only pdf) as fullname_assignment_name_yyyy_mm_dd.pdf(anandjha_sql_assignment5_2023_09_20.pdf) :*
- *If possible , create a ppt and video and share across linkedin and tag ANALYTICSWITHANAND & Me(Anand Jha) in order to reach visibility*

https://drive.google.com/drive/folders/1BaoDQlFR_0yTRyfPDkRUwPn5alp-wgXN?usp=sharing

About Exploratory Data Analysis on "Top Songs of Spotify" between 2010-2022. This report explores on different data relations which can be formed from the given dataset.

SQL ASSIGNMENT 5

Introduction

Spotify is a digital music streaming service that provides users access to over 82 million songs, podcasts and audio books. The app was developed by Daniel Ek and Martin Lorenzton in 2006. This app has become a family name over the years and boasts over 457 million subscribers as of 2022, rivaling SoundCloud and Apple Music.

Spotify measures the popularity of its' artists based on their monthly listeners and number of streams they receive on songs produced. These streams are then multipled by (0.003) and paid to artists as "Royalties", it is a modernized system of monetizing digital sales from traditional album sales (100 streams = 1 album). Ed Sheeran was Spotify's most streamed artist in 2019, however, the rank placements change rapidly depending on album relases, EP's, mixtapes and so forth!

Spotify is a perfect dataset to measure the popularity of songs against various music elements, across a large set of songs throughout the decades. This analysis can be used to demonstrate how peoples music tastes have been translated throughout the past two decades!

I will be creating an exploratory analysis by creating data visualizations and conducting statistical analyses to investigate the relationship between the use of non-traditional musical elements and the popularity of Spotify hits from 2000 to 2019.

Track Metadata

column	description
track_name	Song title
artist_name	Song artist
artist_genre	Song genre category
year	Song Billboard chart entry year

SQL ASSIGNMENT 5

Audio Numerical Quantitive Data

column	description
loudness	Loudness - How loud a song is (db)
duration_ms	Duration - How long the song is (seconds)
tempo	Tempo - How fast a song is (bpm)

Audio Qualitative Data

column	description
energy	Energy level - How energetic the song is
danceability	How easy it is to dance to
valence	How positive the mood of the song is
acousticness	How accoustic sounding the song is
speechiness	How much of a song is spoken word
track_popularity	How popular a song is (as of time of data collection)

Link to Dataset: https://www.kaggle.com/datasets/josephinelsy/spotify-top-hit-playlist-2010-2022

Tech Stack

- Excel *(Data Cleaning)*
- SQL *(Exploratory Data Analysis, Case, Subqueries)*

SQL ASSIGNMENT 5

Summary of Analysis (need to perform this)

- Top Songs based on popularity
- Top songs from each year
- Analysis based on Tempo
- Analysis based on energy (db)
- Analysis based on Danceability
- Analysis based on Loudness
- Analysis based on Valency
- Analysis based on Speechiness
- Analysis based on Acousticness

Task :

1. Check the entire dataset
2. Number of songs on Spotify for each artist
3. Top 10 songs based on popularity
4. Total number of songs on spotify based on year
5. Top song for each year (2000-2022) based on popularity

6. Analysis based on Tempo :
   tempo > 121.08 -> 'Above Average Tempo'
   tempo = 121.08 -> 'Average Tempo'
   tempo < 121.08 -> 'Below Average Tempo'
7. Songs with Highest Tempo

8. Number of Songs for different **Tempo** Range : **track_name, energy**
   Modern_Music -> tempo BETWEEN 60.00 AND 100.00
   Classical_Music -> tempo BETWEEN 100.001 AND 120.00
   Dance_Music -> tempo BETWEEN 120.001 AND 150.01
   HighTempo_Music -> tempo > 150.01

9. Energy Analysis : **TOP 10 track_name, danceability, track_popularity**
   energy > 0.64 -> 'Above Average Energy
   energy = 0.64 -> 'Average Energy′
   energy < 0.64 -> 'Below Average Energy′
   energy BETWEEN 0.1 AND 0.3 -> 'Calm Music'

energy BETWEEN 0.3 AND 0.6 -> 'Moderate Music'
Energy >0.6 -> 'Energetic Music'

10. Number of Songs for different **energy** ranges(above)

11. Danceability Analysis : **Top 20 track_name, danceability**
    danceability BETWEEN 0.69 AND 0.79 -> 'Low Danceability'
    (danceability BETWEEN 0.49 AND 0.68) OR (danceability BETWEEN 0.79 AND 0.89) -> 'Moderate Danceability'
    (danceability BETWEEN 0.39 AND 0.49) OR (danceability BETWEEN 0.89 AND 0.99) -> 'High Danceability'
    danceability < 0.39 OR danceability > 0.99 -> 'Cant Dance on this one'

12. Number of Songs for different **danceability** ranges(above)

13. Loudness Analysis : **Top 20 track_name, loudness,**
    loudness BETWEEN -23.00 AND -15.00 ->'Low Loudness'
    loudness BETWEEN -14.99 AND -6.00 -> 'Below Average Loudness'
    loudness BETWEEN -5.99 AND -2.90   -> 'Above Average Loudness'
    loudness BETWEEN -2.89 AND -1.00   -> 'Peak Loudness'

14. Number of Songs for different **loudness** ranges(above)

15. Valence Analysis : **Top 20 track_name, valence, track_popularity,**
    valence > 0.535 -> Above Avg Valence
    valence = 0.535 -> Avg Valence
    valence < 0.535 -> Below Average'

16. Number of Songs for different **valence** ranges(above)

17. Speechiness Analsis : **Top 20 track_name, speechiness, tempo,**
    speechiness > 0.081-> Above Avg Speechiness
    speechiness = 0.081-> Avg Speechiness
    speechiness < 0.081-> Below Speechiness

18. Acoustic Analysis : **DISTINCT TOP 25 track_name, album, artist_name, acousticness**
    (acousticness BETWEEN 0 AND 0.40000 -> 'Not Acoustic'
    (acousticness BETWEEN 0.40001 AND 0.80000) ->'Acoustic'
    (acousticness BETWEEN 0.80001 AND 1) ->'Highly Acoustic'

SQL ASSIGNMENT 5

Conclusion

In conclusion, this analysis demonstrates the dynamic and evolving nature of musical preferences on Spotify from 2010 to 2022. While there are general trends that suggest what makes a song popular, there are always exceptions and outliers that capture the hearts of listeners. This dataset serves as a testament to the ever-changing landscape of music, driven by innovation, creativity, and the evolving tastes of music enthusiasts.

Understanding these trends can be invaluable for artists, record labels, and music producers looking to create music that resonates with a broad audience on streaming platforms like Spotify. Additionally, this analysis provides music enthusiasts with a glimpse into the fascinating world of data-driven music exploration, where numbers and statistics harmonize with melodies and lyrics to create a richer listening experience.

*******************************THANK YOU !!*******************************