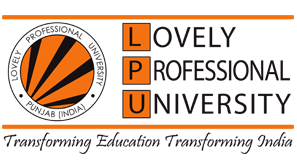
****

**CSE 353 EDA PROJECT REPORT**

**ON**

**DATA SET: Top 10000 songs on Spotify 1960-Now**

**Submitted by:**

**Komal pasumarthy**

**Reg No: 12106222**

**Section: K21UG**

**Roll Number: RK21UGB38**

**Domain Knowledge**

In today's fast-paced and interconnected world, music has cemented its position as a universal and indispensable medium of human expression. It transcends linguistic and cultural boundaries, touching the hearts and souls of people across the globe. From soothing melodies that offer solace in times of distress to pulsating beats that ignite euphoria on dance floors, music is an art form that resonates deeply with us all.

At the forefront of this musical revolution stands Spotify, a true pioneer in the realm of music streaming. In an age where access to virtually any song, artist, or album is just a few clicks away, Spotify has emerged as a dominant force, redefining how we consume and engage with music. Its remarkable prowess lies not only in its extensive catalogue but also in its ability to understand, adapt, and cater to the unique tastes and preferences of its users.

What sets Spotify apart is its unparalleled recommendation system, a technological marvel that seamlessly curates and tailors music experiences for each individual listener. Drawing upon a vast trove of data and employing cutting-edge machine learning algorithms, Spotify has perfected the art of suggesting songs, playlists, and artists based on an intricate web of factors. These factors encompass everything from language preferences to artist affinities, moods, occasions, and even intricate listening patterns. It's this ability to understand and predict what a listener desires at any given moment that has left music enthusiasts utterly captivated.

Yet, in the world of technology and data-driven innovation, the pursuit of perfection is an ongoing journey. Spotify, like any visionary entity, recognizes that there is always room for enhancement, evolution, and refinement. The quest to elevate the user experience, to immerse listeners in a world of music that resonates with them on a profound level, continues unabated.

This brings us to the heart of our project—a comprehensive Exploratory Data Analysis (EDA) of Spotify's Top 10,000 Songs. Spanning from the musical landscape of 1960 to the present day, this dataset encapsulates a rich tapestry of sonic experiences, offering a window into the evolution of musical preferences and trends over the decades.

Through this EDA, we embark on a journey to unearth hidden gems of insight and understanding within the vast universe of Spotify's musical data. Our goal is twofold: first, to gain a deeper understanding of the dynamic interplay between songs, artists, and their listeners; and second, to harness this newfound knowledge to further elevate user satisfaction and enrich the overall music listening experience on the platform.

As we delve into the dataset, we will explore the multifaceted world of music, examining attributes such as tempo, danceability, energy and many more. These dimensions, often analyzed by Spotify's recommendation engine, shape the musical landscape and influence the songs that resonate most with individual listeners. Understanding these attributes is essential in deciphering the intricate art of song recommendation.

**Data Understanding**

Track URI: A unique identifier for each track on Spotify, often used for linking to specific songs. Data Type: String (URI).

Track Name: The name or title of the song. Data Type: String.

Artist URI(s): Unique identifiers for the artist(s) associated with the track, typically used for linking to artist profiles. Data Type: String (URI).

Artist Name(s): The name(s) of the artist(s) who performed the track. Data Type: String.

Album URI: A unique identifier for the album that contains the track, often used for linking to album details. Data Type: String (URI).

Album Name: The title of the album where the track is featured. Data Type: String.

Album Artist URI(s): Unique identifiers for the artist(s) who contributed to the album. Data Type: String (URI).

Album Artist Name(s): The name(s) of the artist(s) responsible for the album. Data Type: String.

Album Release Date: The date when the album containing the track was released. Data Type: Date or String (depends on the format).

Album Image URL: A URL pointing to an image representing the album cover. Data Type: String (URL).

Disc Number: Indicates which disc of a multi-disc album the track belongs to. Data Type: Integer.

Track Number: The position of the track within its album. Data Type: Integer.

Track Duration (ms): The duration of the track in milliseconds. Data Type: Integer.

Track Preview URL: A URL that allows users to preview a short segment of the track. Data Type: String (URL).

Explicit: Indicates whether the track contains explicit content (e.g., explicit lyrics). Data Type: Boolean (or Integer with binary values).

Popularity: A measure of the track's popularity on Spotify, typically represented as a numerical value. Data Type: Integer or Float.

ISRC: International Standard Recording Code, a unique identifier for sound recordings. Data Type: String.

Added By: The user or source that added the track to the dataset. Data Type: String or User ID.

Added At: The date and time when the track was added to the dataset. Data Type: Date and Time.

Artist Genres: Genres associated with the artist(s) of the track. Data Type: String or List of Strings.

Danceability: A measure of how suitable the track is for dancing, typically represented as a numerical value. Data Type: Float.

Energy: A measure of the track's energy level, often represented as a numerical value. Data Type: Float.

Key: The key in which the track is composed, represented as a musical key (e.g., C, D, G). Data Type: String.

Loudness: The overall loudness of the track, typically represented as a numerical value. Data Type: Float.

Mode: Indicates whether the track is in a major or minor key. Data Type: Integer (0 for minor, 1 for major).

Speechiness: A measure of how much speech (spoken words) is present in the track. Data Type: Float.

Acousticness: A measure of the track's acoustic qualities (acoustic vs. electronic), typically represented as a numerical value. Data Type: Float.

Instrumentalness: A measure of whether the track is instrumental (no vocals), typically represented as a numerical value. Data Type: Float.

Liveness: A measure of the presence of a live audience in the track, typically represented as a numerical value. Data Type: Float.

Valence: A measure of the track's positivity or happiness, often represented as a numerical value. Data Type: Float.

Tempo: The tempo or beats per minute (BPM) of the track, typically represented as a numerical value. Data Type: Float.

Time Signature: The time signature of the track (e.g., 4/4), indicating the number of beats in each bar. Data Type: String.

Album Genres: Genres associated with the album containing the track. Data Type: String or List of Strings.

Label: The record label that released the track or album. Data Type: String.

Copyrights: Information regarding copyright and ownership of the track. Data Type: String.

**Key Questions**

In this EDA project on the Spotify Top 10,000 Songs dataset, I have outlined some key questions and strategies to guide our exploration and gain valuable insights:

1. Dataset Overview:

Our journey begins with a comprehensive overview of the dataset, including the size, structure, and any missing data. We'll decide on how to address any data gaps.

2. Unveiling Top Artists and Songs:

Let's dive into the world of music by identifying the standout artists and songs, and examining them through various lenses like popularity, track count, or user engagement.

3. Time Travel Through Music:

We'll embark on a musical time-travel adventure by exploring how the music landscape has evolved over the years. This will include an analysis of song durations and album release dates.

4. Mapping Musical Genres:

A crucial part of our exploration involves mapping the vast array of songs to their respective music genres. We'll uncover genre popularity trends and pinpoint the most and least explored genres.

5. Popularity Under the Microscope:

Popularity is a central theme, and we'll scrutinize it closely. This includes dissecting popularity scores and understanding what factors contribute to a song's popularity.

6. Dance Moves and Energy Levels:

We'll groove to the rhythm by examining danceability and energy levels in songs. Are there patterns that link these attributes to a song's popularity?

7. Decoding Keys and Modes:

Musical keys and modes add unique flavors to songs. We'll decode their presence and assess whether they're more prevalent in specific genres.

8. Vocals vs. Instruments:

We'll distinguish between songs with vocals and instrumental tracks, unravelling their distribution across genres and their impact on listeners.

9. Acoustic or Electric?

Acousticness is a fascinating attribute to explore. We'll analyze its distribution and how it varies across different genres and eras.

10. Riding the Tempo Waves:

- The tempo (BPM) of a song can set the mood. We'll visualize tempo trends and look for any noteworthy patterns.

11. Time Signatures:

- Time signatures are the heartbeats of music. We'll uncover their distribution and their relationship with genre and popularity.

12. The Explicit Factor:

- Exploring explicit content is intriguing. We'll investigate its prevalence and whether it has an impact on a song's popularity.

13. User Engagement Insights:

- If user engagement data is available, we'll gain insights into user behaviors such as plays, likes, and playlist additions. This helps identify highly engaging songs.

14. Musical Geography:

- If geographical data exists, we'll explore how regional preferences manifest in genres and artists, uncovering geographical music trends.

15. Crafting New Features:

- We'll get creative by engineering new features based on existing data to uncover deeper insights. An example could be creating an "Energy-to-Danceability" ratio.

16. Listening to Our Users:

- User segmentation based on listening habits and preferences is a fascinating avenue. We'll uncover common patterns among user groups.

17. Testing Hypotheses:

- We'll use statistical tests to explore hypotheses about music trends or relationships within the data. For example, we might test if key changes affect song popularity.

18. Visualizing Insights:

- Utilizing an array of data visualization techniques, we'll bring our findings to life. Expect histograms, scatter plots, heat maps, and line charts.

19. Crafting the Story:

- We'll transform our discoveries into a compelling narrative, providing actionable insights and recommendations to enhance the music listening experience.

20. Beyond EDA:

- Lastly, we'll discuss potential directions for future research or analysis based on our EDA findings and conclude with key takeaways from our exploration.

Throughout this journey, we'll document our methods, share code and visuals, and ensure our insights are presented clearly to resonate with our audience. Remember, EDA isn't just about analyzing data; it's about telling a captivating, data-driven story.