VIS CSE 564 - Project Proposal

Spotify Pulse: Mapping Global Music Trends

Anurag Varma Penmetsa ID: #115603446

Suhas Guptha Maddi ID: #115694891

Spotify Dataset :

Source: Top Spotify Songs in 73 Countries (Daily Updated)

Link: Top Spotify Songs in 73 Countries (Daily Updated) | Kaggle

This dataset presents the top songs currently trending for over 70 countries.

The top 50 songs for each country are updated daily to provide the most up-to-date information on the popularity of songs in the world.

The dataset selected is in the range starting from 2023-10-18 till 2024-02-02. It contains approximately 400,000 rows.

Dataset Columns:

The dataset contains the following columns which are described as follows:

1. **Daily Movement:** Represents the change in song rankings compared to the previous day, indicating daily popularity fluctuations.
2. **Weekly Movement:** Indicates the change in song rankings compared to the previous week, offering insights into weekly popularity shifts.
3. **Is Explicit:** Binary indicator (True/False) signaling whether the song contains explicit lyrics, aiding in categorizing content based on explicitness.
4. **Duration (ms):** Length of the song in milliseconds, providing the time commitment required for listening.
5. **Album Release Year:** Denotes the release year of the album to which the song belongs, placing it within the artist's discography timeline.

6. **Danceability:** Measure of a song's suitability for dancing, aiding in categorizing its dance appeal.
7. **Energy:** Measure of a song's intensity and activity level, offering insights into its overall energy.
8. **Key:** Represents the musical key of the song, contributing to its musical characteristics.
9. **Mode:** Indicates whether the song is in a major (1) or minor (0) key, aiding in classifying its emotional tone.
10. **Speechiness:** Measure of the presence of spoken words in the song, distinguishing between instrumental and vocal-heavy tracks.
11. **Acousticness:** Measure of the acoustic quality of the song, ranging from least to most acoustic.
12. **Liveness:** Measure of the presence of a live audience in the recording, indicating a live performance atmosphere.
13. **Valence:** Measure of the musical positiveness conveyed by the song, indicating its emotional tone or mood.
14. **Tempo:** The speed of the song in beats per minute (BPM), influencing its overall feel and energy.
15. **Time Signature:** The estimated overall time signature of the song, describing its rhythmic structure.

Analysis of Data :

The 'Top Spotify Songs in 73 Countries' dataset lacks a column specifying the genre of the songs. To address this gap, we propose the following methodology:

- **Dataset Integration:** We will leverage another dataset, the 'Top 50 Spotify songs BY EACH COUNTRY', which includes genre information but has fewer countries and lacks time series data. This additional dataset provides a crucial genre attribute that our main dataset misses.
- **Machine Learning Model Development:**
    - **Feature Utilization:** We will develop a machine-learning model capable of predicting a song's genre. The model will use features such as Danceability, Energy, Key, Mode, Speechiness, Acousticness, Liveness, Valence, and Tempo.
    - **Training Process:** The model will be trained on the 'Top 50 Spotify Songs by Each Country' dataset to learn the correlation between these features and the song's genre.

- **Data Processing for Efficiency:**
    - **Data Cleaning:** We will clean the data to ensure accuracy and relevance by removing outliers and handling missing values.
    - **Random Sampling:** To enhance processing speed and scalability, we will perform random sampling of the data. This step will reduce the computational load while maintaining a representative subset of the dataset, thus optimizing our analysis efficiency.
- **Dataset Modification:** Throughout the process, we will make necessary modifications to the dataset to meet the specific requirements of our analysis and model training.

This approach ensures a comprehensive analysis, leveraging machine learning to bridge the data gap and enhance our understanding of global music trends.

Problem Statement:

In the rapidly evolving music industry, stakeholders from artists and record labels to streaming services face challenges in understanding and responding effectively to dynamic shifts in listener preferences and market demands. There is a need for a nuanced analysis of how global and regional music tastes evolve over time, influenced by a complex interplay of temporal, geographic, and cultural factors. Without insights into these trends, it is difficult to:

- **Optimize Music Promotion:** Tailoring marketing strategies to target specific regions or demographic groups based on their unique music preferences.
- **Forecast Trends:** Predicting future shifts in music tastes to stay ahead of market curves and prepare for upcoming changes in listener behavior.
- **Artist Development and Representation:** Identifying emerging artists or genres gaining traction to make informed decisions about artist signings and promotions.
- **Content Curation:** Curating music content that resonates with the diverse preferences of listeners in different geographic areas.
- **Cultural Relevance:** Ensuring that music offerings align with and respect cultural trends and events, maximizing relevance and engagement.

Approach with different plots:

To get insights into the trends in music and to solve the problems mentioned in the problem statement, we can use the following different graphs to get more insights into the music dataset:

- **Time Series & Line Graph:** Tracks global trends in song popularity over time, revealing seasonality, the impact of cultural events, and the lifecycle of hit songs.

- **Geo map area chat:** Visualizes the geographic distribution of song popularity, highlighting regional preferences and the distinction between global hits and local favorites.
- **Parallel Coordinates Plot:** Explores correlations between musical attributes (like danceability, and energy) and popularity, identifying what characteristics are common in popular music across different genres.
- **Sunburst Chart:** Shows the distribution of music genres over selected periods, providing insights into genre trends and preferences, and how they evolve over time.
- **Word Cloud:** Displays the influence of artists by region, identifying which artists are most popular and how their popularity spans across different areas.

Interactions:

The following are the interactions taking place in each of the graphs:

- **Time Series Graph (Top Left):** This graph allows users to select a specific time range to view data. It features a brush filter to easily adjust the timeframe and displays a line graph showing the frequency of the top 10 songs globally on a specific day across 72 countries.
- **Geo map area chat (Top Right):** This map displays countries colored according to the frequency of song plays for the timeframe selected in the first graph. Hovering over a country will reveal its name and the associated data number. Users can click on one or multiple countries to refine the displayed data in all other graphs based on the selected regions.
- **Parallel Coordinates Plot (Bottom Left):** The PCP allows for dynamic filtering and rearrangement of axes. Users can interact with each axis to filter data or change their positions, which alters the visualization of lines representing the data.
- **Sunburst Chart (Bottom Middle):** This chart presents music genres as segments in a circular ring. Clicking on a segment focuses the display on data related to that particular genre, providing specific insights about its prevalence and characteristics.
- **Word Cloud (Bottom Right):** This visualization displays the names of popular artists in varying colors and orientations, reflecting their frequency and popularity in the region. The size of each artist's name indicates their popularity level, with more frequently mentioned artists appearing larger. This graph is non-interactive and serves solely as a visual display of data.