BIG DATA VISUAL ANALYTICS (CS661)
PROJECT PROPOSAL REPORT
GROUP-01

Audio-Aura: Music Data Visualization

1. Introduction:

Music is a fundamental aspect of human culture, language, and borders to connect people from all around the world. In today's digital age, the way we listen and interact with music has undergone a revolution, thanks to various streaming platforms like Spotify, iTunes, Wynk Music, Amazon Music, and many more. These platforms provide access to a large library of music and also generate huge amounts of data that can offer deep insights into listening habits, musical preferences, and trends.

In our project, we aim to utilize the power of Spotify's data to create an immersive and informative web dashboard for music data visualization. The dashboard will serve as a dynamic tool for exploring and understanding the insights of music, offering users a visual journey through genres, artists, and trends.

We aim to provide a unique and engaging experience for music enthusiasts, researchers, and industry professionals via a variety of interactive visualizations, including charts, graphs, and various creative graphics, to help them explore different aspects of music. In this project, we hope to find out hidden insights and patterns within Spotify's data, shedding light on the evolving landscape of music consumption and production.

2. Data Sources:

Various sources of raw data for our project:

- **Spotify Web API:** Provides access to a vast collection of music metadata, including song attributes, artist information, and user preferences.
- **RapidAPI (Glavier):** Offers additional song data including lyrics, album information, and user-generated content.
- **Zenodo:** A repository providing access to openly available datasets related to music research and analysis, which also consists of country-wise data for songs and instrumental data.

Data Description

Among the data that we have come across, few of the many attributes that we found are as follows: track name, artist name, genre of songs, genre of artists, explicit content in song, duration, artist info, artist popularity, song properties (viz. danceability, valence, instruments, loudness, speechiness, acousticness, liveness, key, mode, tempo, energy, etc.), playlists info, albums info, top playlists, track popularity, available markets for tracks, release_data (aka. date-month-year), category, etc. There are many other features as well.

3. Specific Tasks:

For this project we will perform the following main tasks:

- **Data Aggregation**: Retrieve song data from multiple sources and perform data aggregation to get new important features.
- **Data Preprocessing**: Perform data cleaning, removing noise, handle missing values, and data transformation for analysis. This also includes identifying important features for visualization.
- **Data Visualization:** Develop interactive visualizations to explore song attributes, trends, and patterns.
- **Insights Generation**: Implement algorithms for deriving insights from the data, such as genre classification, mood analysis, and popularity trends.
- **User Interaction:** Enable users to interact with the visualizations, customize views, and save insights.

Our wishful thinking is to find such **insights** from the available data:

- Change in music genre over the years.
- Similar kinds of song recommendation based on properties of songs.
- Region-wise taste in music.
- Recommendation of similar types of artists based on a particular artist.
- Patterns in top-songs, such as attribute-wise factor domination.

4. Overall Solution:

Our Visual Analytics System will be in the form of a Web Application to ensure accessibility and ease of use. It will feature an intuitive user interface with interactive visualizations, powered by modern web technologies. For this purpose, we hereby propose 5 major tasks that will help us gain the hidden patterns, and insights about songs, their popularity, etc. Nonetheless, as we progress through the project, we may attempt to find other insights as well.

Visualization Tasks

1) Genre Evolution over Years: Performing this visualization helps us explore the dynamic shifts in music genre preferences across the years, unveiling the evolutionary trajectory of user tastes. It helps us understand emerging trends and fading genres, allowing us to adapt its content strategy to the ever-changing musical landscape. Main attributes for this viz. will be track_id, track_audio_features, track_genre, track_release_date, track_explicit_content, track_duration etc.

2) Song Recommendations based on Intrinsic Properties: Applying various recommendation-based algorithms to recommend songs with similar properties, going beyond genre labels. To create this visualization, we will utilize attributes such as tempo, mood, and instrumentation, as they offer users a personalized experience that aligns with their musical preferences. The main attributes for this viz. will be track_id, track_audio_features (like danceability, valence, instruments, loudness, speechiness, acousticness, liveness, key, mode, tempo, energy etc.), track_genre, track_explicit_content, track_duration, track_artist etc.

3) Regional Variations in Music Taste: Using this visualization, we can explore diverse regional music preferences to understand the unique tastes across cultures. Customizing content and promotions based on these insights will help us strengthen connections with a varied audience. The main attributes for this viz. will be track_id, instruments_used in that track track_audio_features, track_genre, track_explicit_content etc.

4) Artist Recommendations Anchored in Similarity: In this particular task, we intend to build a clever system that recommends artists by looking at simple things like how they sing, the topics in their songs, and how complex their music is. This way, one can easily find new artists whose style matches what they already enjoy, making music discovery more enjoyable and personalized. Similarly, the attributes will be the same and to make this visualization we will try to add more features along with the mentioned above.

5) Top-Song Patterns Analysis: For this visualization, we will examine the essential elements within chart-topping songs, analyzing factors like lyrics, instrumental nuances, and production techniques. By dissecting these elements, we will identify the core aspects that contribute to a song's success. This visualization and analysis will guide the development of content that aligns with these winning elements, ensuring a strong resonance with diverse global audiences.

And many more as we delve into this BIG data.

5. Tech Stack:

Backend Development: The backend development will be powered by Python, utilizing frameworks such as TensorFlow for machine learning tasks, and deployment will be managed through Heroku.

Frontend Development: We have decided to use JavaScript and TypeScript to create dynamic and responsive interfaces, maintaining the flexibility to choose from popular frameworks such as React, Vue, or Angular to ensure adaptability and scalability of our final web application, which shall then be deployed and hosted on Netlify (or a suitable platform, as per evolving requirements).

Data Visualization: Visualization needs will be met using D3.js (particularly, Observable Plot) for creating interactive and data-driven graphics to gain insights from the underlying dataset.

6. Team Members and Responsibilities:

Following are the assignments of roles and responsibilities expected within our team; however, everyone shall contribute as per the requirements so as to maintain a balanced workload and fair share of responsibilities.

Backend Development and Database Management:

- Akash Shivaji Varude (231110006)
- Darshan Jain (231110009)
- Pranjal Maroti Nandeshwar (231110035)

Machine Learning and Visualization Module:

- Deepen Shrestha (231110061)
- Pankaj Siddharth Nandeshwar (231110034)
- Shaurya Agarwal (231110046)

Frontend Development and Application Hosting:

- Indraneel Rajeevan (231110403)
- Nitish Kumar (231110033)

An endeavor to merge sound and sight, creating a symphony for the eyes.