How to Make Music Feature Explorer and Song Similarity Calculator in Streamlit

Executive Summary

Welcome to the Music Feature Explorer and Song Similarity Calculator, an innovative data visualization project that dives deep into the world of music. This report provides a comprehensive "behind-the-scenes" look at the project, targeted at the data visualization community.

Here is the [link](#) to my dashboard.

Project Context

Music is a universal language, stirring emotions, triggering memories, and shaping collective sentiments. The fusion of lyrics and chord progressions forms the essence of a song's ambiance, making it vital for academic study and the music industry. This project bridges the gap between the subjective interpretation of music and quantitative analysis.

Our project employs interactive visualizations to explore and understand the complex musical and lyrical elements of songs. We examine key features such as acousticness, speechiness, loudness, key, energy, danceability, tempo, valence, liveness, instrumentalness, and time signature. These features offer a nuanced perspective on the characteristics of music, from its acoustic nature to its emotional impact.

In essence, this project aims to uncover the science behind the art of music, shedding light on how various elements converge to influence listeners.

Target Audience

Our primary audience includes Record Labels, Musicians, AI-Powered Music Creation platforms, and music enthusiasts.

- **Record Labels**: They benefit from data-driven insights into music trends and listener preferences, aiding in scouting and signing artists whose music resonates emotionally.

- **Musicians**: Artists can explore music to find inspiration, understand patterns, and stay connected with evolving trends.

- **AI-Powered Music Creation**: This sector can use our rich data repository to develop algorithms for creating targeted music content, aligning with specific moods and themes.

- **Music Enthusiasts**: Casual listeners can explore and gain a deeper understanding of their favorite songs.

This app caters to a diverse audience at the intersection of technology, creativity, and the music business, providing insights into the interplay between musical elements and emotional impact.

Data Sources

Source
Raw data comes from two open datasets: [Spotify's million song dataset](#) and [chord progression assistant](#).

Here is the [data](#) used for this blog.

Collection Methods
The data collection process involved merging two datasets together using the song names, after combining two datasets, there are 2673 section of songs, which include ***text*** data ('lyrics') and other song features in ***numeric*** format ('chord', 'tempo', 'key'), ***object*** format ('type'), and characteristics assigned by sportify algorithm ('danceability' 'energy' 'loudness' 'mode' 'speechiness' 'acousticness' 'instrumentalness' 'liveness' 'valence').

Biases/Sampling
While extensive efforts were made to assemble a diverse collection of music samples, it's important to acknowledge that the dataset may still carry certain biases. These biases could be influenced by factors such as the popularity of specific music genres or the availability of music data for certain songs. It's worth noting that the dataset's size, while significant, may not comprehensively represent the entire spectrum of music distribution. Users should exercise caution and consider these potential biases when interpreting the results generated by the Song Similarity Calculator.

In addition to analyzing audio features, you can also use algorithms to automatically assign lyrics to different sections of a song, such as verses and choruses. This process involves segmenting the lyrics based on patterns in the text or by leveraging natural language processing techniques. The resulting sections can provide valuable insights into the song's structure, making it easier to identify and analyze specific parts of the composition. Keep in mind that these assignments are algorithmic and may not always align perfectly with the artist's intended structure. Nonetheless, they offer a useful way to explore songs in more detail.

Descriptive Statistics
The data collection process involved merging two datasets together using the song names, after combining two datasets, there are 2673 section of songs, which include ***text*** data ('lyrics') and other song features in ***numeric*** format ('chord', 'tempo', 'key'), ***object*** format ('type'), and characteristics assigned by sportify algorithm ('danceability' 'energy' 'loudness' 'mode' 'speechiness' 'acousticness' 'instrumentalness' 'liveness' 'valence'). The distribution of features varies across different tracks, enabling a comprehensive analysis of music characteristics. More detailed definition of each variable can be found [here](#).

Data Cleaning Methods & Identified Issues
Data cleaning methods were applied to ensure data accuracy and consistency. Missing values were handled, outliers were addressed, and data types were standardized. Despite these efforts, some issues, such as missing lyrics for certain tracks, were identified.

Privacy
Privacy and data protection were paramount considerations in this project. Personal and sensitive information was not collected, ensuring compliance with privacy regulations.

Technologies/Platforms Used
- **Streamlit**: Used for creating interactive web applications with Python.
- **Pandas**: Employed for data manipulation and analysis.
- **Scikit-learn**: Used for machine learning tasks, including cosine similarity calculations.
- **Plotly and Matplotlib**: Utilized for data visualization.
- **Natural Language Toolkit (NLTK)**: Applied for text processing and analysis.

Summary of Analysis

The goal of the Music Feature Explorer and Song Similarity Calculator is to provide a holistic view of music through data-driven insights. Key variables, including acousticness, speechiness, loudness, key, energy, danceability, tempo, valence, liveness, instrumentalness, and time signature, offer a comprehensive understanding of a song's characteristics.

Key Features and Code

1. **Numerical Feature Exploration Page**: Users can explore variables like key, tempo, and danceability through interactive pie charts, providing insights into the distribution of these features across different songs.
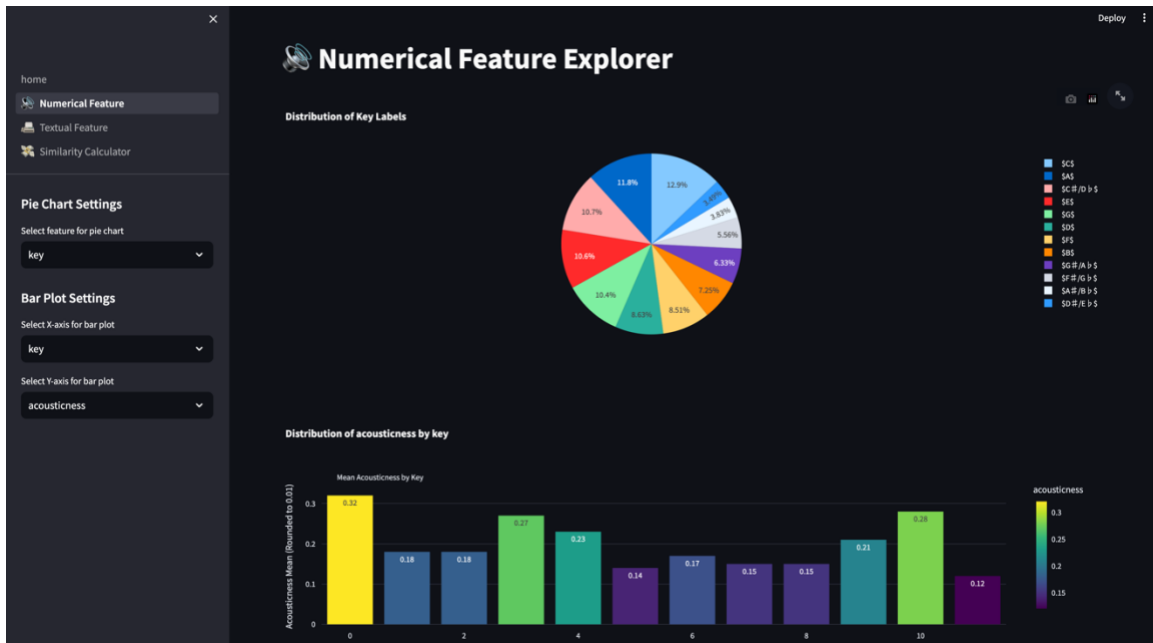
- Pie Chart: for pie chart, we have 3 variables that can be selected.
- Bar Plot: for the bar plot, we can see how different variables are distribution against each other.

2. **Textual Feature Exploration Page**: This section offers textual similarity analysis, sentimental analysis, and word clouds, allowing users to delve into the thematic and emotional content of lyrics.

3. **Song Similarity Feature Calculator Page**: Users can calculate and visualize the similarity between two songs using a sophisticated algorithm and radar graphs. This feature assists in understanding how different songs relate to each other.
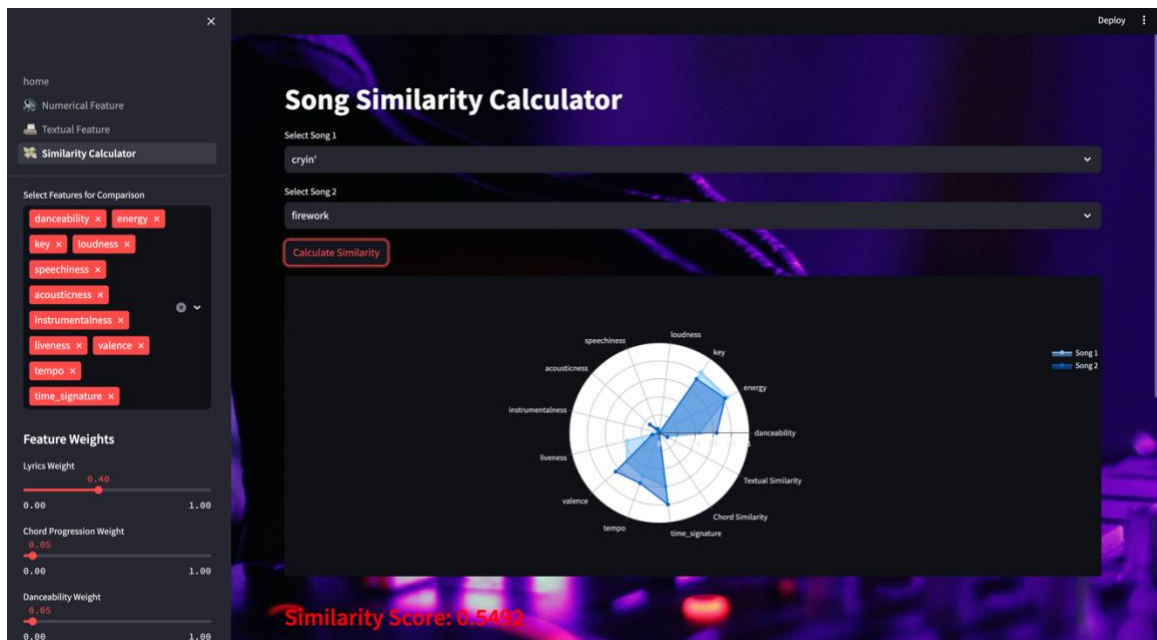
These features cater to a wide range of users, from casual listeners to industry professionals, seeking to unravel the intricate tapestry of music.

Conclusion

The Music Feature Explorer and Song Similarity Calculator project offer a unique and insightful perspective on the world of music. By quantifying and visualizing various musical and lyrical elements, this project helps individuals and professionals alike gain a deeper understanding of the emotional and structural aspects of songs. Whether you're exploring music for inspiration or making data-driven decisions in the music industry, our project empowers you to appreciate the art and science of music in a whole new way.