QMSS GR 5063: Data Visualization
Final Project Proposal
Team 3: Burui Chen, Hanyi Xu, Min Zhuang, Xiaorong Yu

Title: Data Visualization of Musical Patterns and Preferences

Abstract:
Music is an integral part of human culture, influencing emotions, behaviors, and personal preferences. We will explore and visualize musical patterns and preferences using a dataset that captures various aspects of songs, such as genre, tempo, popularity, and other musical attributes. We want to uncover trends in musical consumption, identify correlations between song characteristics and listener preferences, and provide interactive visual representations for intuitive analysis. We will use python packages to visualize our results. Our project will utilize charts, heatmaps, and clustering techniques to present insights in an accessible and engaging manner. The findings may offer valuable perspectives for music recommendation systems, industry professionals, and enthusiasts interested in understanding the evolving landscape of music preferences.

Here are some potential research questions we will answer:
- How do different musical attributes (e.g., tempo, energy, valence) vary across genres?
- What are the most common characteristics of popular songs?
- Visualize musical trends evolved over time based on available features.

For some advanced research questions, we may consider:
- Can we predict a song's popularity based on its musical attributes?
- What factors contribute most to determining whether a song will be a hit?

Dataset: Kaggle. (n.d.). *Spotify Dataset 1921-2020 (160k+ Tracks)*. Retrieved from https://www.kaggle.com/datasets/varunmohanty/spotify-dataset-19212020-160k-tracks

Technics:
- We will use **Python**
- Python packages: plotly, plotnine(ggplot2), matplotlib, seaborn, altair, networkx, wordcloud, Streamlit

Description:
The dataset contains 2,000 songs with 18 attributes, providing a detailed overview of various musical and metadata features. It includes basic information such as the artist, song title, release year, and whether the track has explicit content. Additionally, it covers quantitative

aspects like duration (in milliseconds), popularity score, and several audio features, including danceability, energy, key, loudness, and tempo. Other attributes, such as speechiness, acousticness, instrumentalness, liveness, and valence, help characterize the sonic and emotional qualities of the songs. The mode indicates whether the song is in a major or minor key, while the genre provides categorical classification. This dataset is well-suited for analyzing musical trends, examining the relationship between audio features and popularity, and exploring genre-based differences in song characteristics.

The primary data source for this project is the Spotify Dataset 1921-2020 (160k+ Tracks), originally retrieved from Kaggle. This dataset provides comprehensive information on a wide range of songs, including attributes such as tempo, danceability, energy, and release year, enabling in-depth analysis of musical trends over time.

In the next phase of the project, we plan to expand the dataset by integrating additional data through API access and web scraping techniques. Specifically, we will retrieve real-time music metadata, track popularity metrics, and user-generated playlists using the Spotify Web API, which offers robust access to song details, artist information, and playlist curation data.

To complement this, we will leverage web scraping tools to gather music reviews, listener sentiment, and trending song data from music websites, blogs, and online forums. Potential target websites for scraping include:

- Billboard (https://www.billboard.com/) – for weekly charts and music trend analysis.
- Genius (https://genius.com/) – for song lyrics, annotations, and user discussions.
- AllMusic (https://www.allmusic.com/) – for album reviews, genres, and release histories.
- Reddit (https://www.reddit.com/r/popheads/) – for real-time listener opinions and emerging trends.

This combined approach will enrich the existing dataset, providing both quantitative attributes from Spotify and qualitative insights from fan communities and critics, enabling a more holistic understanding of musical trends and audience preferences.

Web Scraping tool
https://github.com/AliAkhtari78/SpotifyScraper
The SpotifyScraper tool is a Python library designed to extract detailed information from Spotify, such as track details, artist information, and playlist specifics. It allows users to download track previews and album covers, making it useful for data analysis and research on music streaming behaviors.