**Music Trend Analysis of Million Songs Dataset**

By: Madison L Rundell, Michael A Byrd, Sudha Rani Seeli, Jheel Jagani.

MIS 548: Milestone1-Project Report

**Dataset Selection:**

Our group selected the Million Song Dataset for Music Trend Analysis due to its extensive coverage of contemporary music tracks, offering a diverse range of genres and artist information. The comprehensive dataset allows us to explore evolving music characteristics, identify emerging patterns, and understand the factors influencing listener preferences over time.

**Business Problem:**

With **Music Trend Analysis** we aim to uncover insights into the dynamic and evolving landscape of the music industry.

By leveraging this rich dataset, we aim to provide valuable insights that inform strategic decisions for artists, producers, record labels, and streaming platforms, ultimately contributing to a deeper understanding of current trends within the music industry.

By analyzing historical data and identifying patterns, this project seeks to gain an understanding of trends that can inform strategic decisions, guide marketing efforts, and support the creation of music that resonates with current and emerging audience preferences.

**Summary of the steps taken so far:**

1. Import Libraries: We began by importing essential libraries.
2. Establish Directory: We checked the current working directory, changed the working directory to the desktop, and verified the change in the working directory.
3. Track MetaData: Here, we established a connection to the SQLite database named "track\_metadata.db" and created a cursor to execute SQL commands. We Followed by executing an SQL query to fetch the names of tables in the database. All table names were fetched and printed. The data from the 'songs' table is read into a Pandas DataFrame and then exported to a CSV file named 'track\_metadata.csv'.

Finally, the SQLite connection is closed.

A screenshot of a computer

Description automatically generated

1. Artist Similarity: Similarly, the code connects to another SQLite database named 'artist\_similarity.db'. It retrieves and prints the table names in the database.

Data from 'artists' and 'similarity' tables are read into Pandas DataFrames (df1 and df2), both DataFrames are then exported to CSV files ('artists.csv' and 'similarity.csv'), and the SQLite connection is closed.

1. Artist Term: In this section the code connects to the 'artist\_term.db' SQLite database, and retrieves and prints the table names. Data from various tables ('artists', 'terms', 'artist\_term', 'mbtags', 'artist\_mbtag') is read into separate Pandas DataFrames. Each DataFrame is exported to a corresponding CSV file and the SQLite connection is closed.

**Conclusion:**

The milestones achieved include establishing connections to the databases, understanding their schema through the exploration of table names, and successfully transforming the data into a more flexible and shareable format.

As we move forward in our project, our next milestone will involve leveraging these CSV files for deeper analysis, exploring relationships within the data, and potentially integrating additional sources to enhance the richness of our insights.