Anushka Churi (ac2224)
Meeti Dixit (md2335)
Neha Kulkarni (nk4349)
Shreya Sakpal (ss7807)

BIGDATA PROJECT PHASE - I
GROUP 34 - (SPOTIFY 12-M SONGS DATASET)

DATASET LINK
https://www.kaggle.com/datasets/rodolfofigueroa/spotify-12m-songs

DATASET DESCRIPTION
The Spotify 12M Song Dataset is a large collection of audio features extracted from songs on the popular music streaming platform, Spotify. The dataset contains data for over 12 million songs, and is available for download on Kaggle.
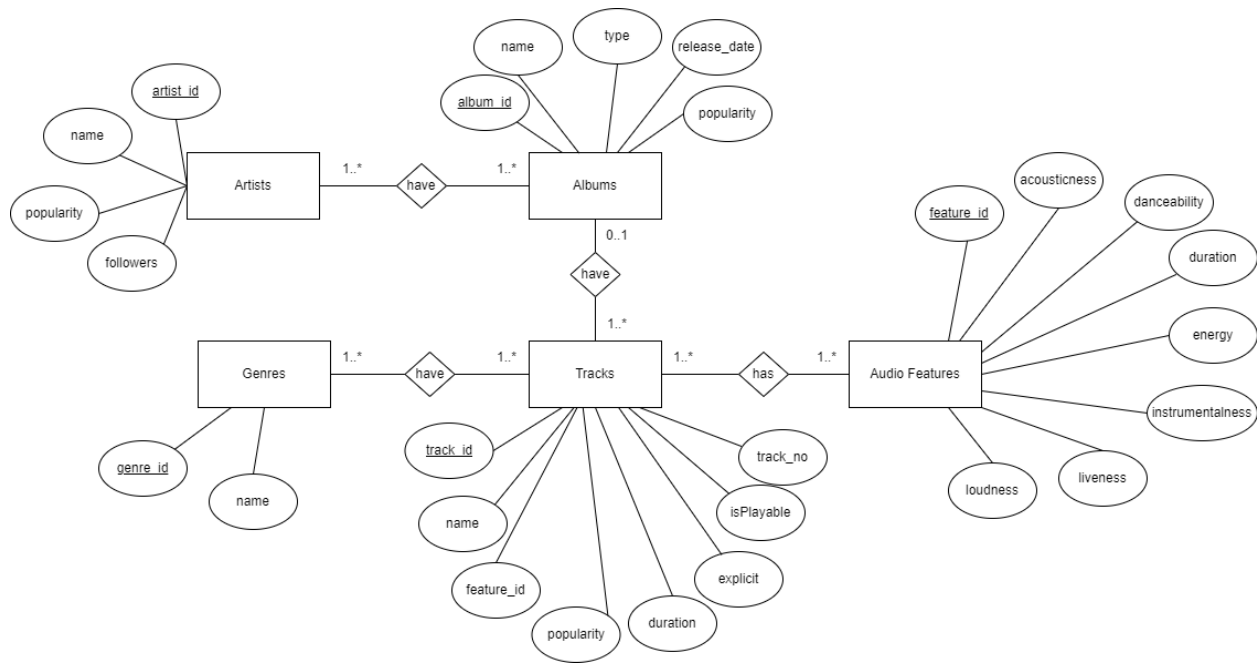The dataset includes the following attributes:

1. **duration_ms**: The length of the song in milliseconds.
2. **popularity**: The popularity score of the song on a scale of 0-100.
3. **acousticness**: A measure of the acousticness of the song, ranging from 0.0 to 1.0.
4. **danceability**: A measure of how danceable a song is, ranging from 0.0 to 1.0.
5. **energy**: A measure of the energy of a song, ranging from 0.0 to 1.0.
6. **instrumentalness**: A measure of the instrumentalness of the song, ranging from 0.0 to 1.0.
7. **key**: The key the song is in (0-11).
8. **liveness**: A measure of the liveness of the song, ranging from 0.0 to 1.0.
9. **loudness**: The overall loudness of the song in decibels (dB).
10. **mode**: The modality (major or minor) of the song (0 = minor, 1 = major).
11. **speechiness**: A measure of the level of spoken words in a song, ranging from 0.0 to 1.0.
12. **tempo**: The tempo of the song in beats per minute (BPM).
13. **time_signature**: The time signature of the song.
14. **valence**: A measure of the positiveness of the song, ranging from 0.0 to 1.0.
15. **artist_name**: The name of the artist who created the song.
16. **track_name**: The name of the song.
17. **track_id**: The unique identifier for the song in Spotify's database.
18. **genres**: Genres to which different tracks belong.

GROUP_34 (SPOTIFY 12-M SONGS DATASET)

Anushka Churi (ac2224)
Meeti Dixit (md2335)
Neha Kulkarni (nk4349)
Shreya Sakpal (ss7807)

ER DIAGRAM

SPOTIFY 12-M SONGS DATASET ER DIAGRAM

RELATIONAL MODEL

albums (**album_id**, name, type, release_date, popularity)

artists (**artists_id**, name, popularity, followers)

tracks (**track_id**, name, *audio_feature_id*, popularity, duration, explicit, isPlayable, track_no)

genres (**genre_id**, name)

audio_features (**audio_feature_id**, acousticness, danceability, duration, energy, instrumentalness, liveness, loudness)

GROUP_34 (SPOTIFY 12-M SONGS DATASET)

Anushka Churi (ac2224)
Meeti Dixit (md2335)
Neha Kulkarni (nk4349)
Shreya Sakpal (ss7807)

albums_artists (***album_id***, ***artist_id***)

tracks_genre (***track_id***, ***genre_id***)

SQL QUERIES TO LOAD THE TABLES

CREATE TABLE spotify.artists(
    artist_id   varchar NOT NULL,
    name   varchar,
    popularity   int,
    followers   int,
    PRIMARY KEY(artist_id)
);

CREATE TABLE spotify.albums(
    album_id varchar NOT NULL,
    name varchar,
    type varchar,
    release_date int,
    popularity int,
    PRIMARY KEY (album_id)
);

CREATE TABLE spotify.audio_features(
    audio_feature_id varchar NOT NULL,
    acousticness decimal,
    danceability decimal,
    duration integer,
    energy decimal,
    instrumentalness decimal,
    liveness decimal,
    loudness decimal,
    PRIMARY KEY (audio_feature_id)
 );

GROUP_34 (SPOTIFY 12-M SONGS DATASET)

Anushka Churi (ac2224)
Meeti Dixit (md2335)
Neha Kulkarni (nk4349)
Shreya Sakpal (ss7807)

```
CREATE TABLE spotify.tracks(
   track_id varchar NOT NULL,
   name varchar,
   feature_id varchar,
   popularity integer,
   track_num integer,
   duration integer,
   audio_feature_id varchar,
   PRIMARY KEY (track_id)
   FOREIGN KEY (audio_feature_id) references audio_features(audio_feature_id)
);

CREATE TABLE spotify.genres(
   genre_names varchar,
   genre_id integer,
   PRIMARY KEY (genre_id)
);

CREATE TABLE spotify.albums_artists(
   album_id varchar NOT NULL,
   artist_id   varchar NOT NULL,
   PRIMARY KEY (album_id, artist_id)
   FOREIGN KEY (album_id) references albums(album_id)
   FOREIGN KEY (artist_id) references artists(artist_id)
);


CREATE TABLE spotify.tracks_genres(
   track_id varchar NOT NULL,
   genre_id   varchar NOT NULL,
   PRIMARY KEY (track_id, genre_id)
   FOREIGN KEY (track_id) references tracks(track_id)
   FOREIGN KEY (genre_id) references genres(genre_id)
);
```

GROUP_34 (SPOTIFY 12-M SONGS DATASET)

Anushka Churi (ac2224)
Meeti Dixit (md2335)
Neha Kulkarni (nk4349)
Shreya Sakpal (ss7807)

PROGRAM TO LOAD THE TABLES

TABLE ALBUMS
\copy albums from '/Users/shreyasakpal/Desktop/Spotify/album.csv' with (FORMAT csv, null '\N');

TABLE ARTISTS
\copy artists from '/Users/shreyasakpal/Desktop/Spotify/artist.csv' with (FORMAT csv, null '\N');

TABLE TRACKS
\copy tracks from '/Users/shreyasakpal/Desktop/Spotify/tracks.csv' with (FORMAT csv, null '\N');

TABLE AUDIO_FEATURES
\copy audio_features from '/Users/shreyasakpal/Desktop/Spotify/audio_features.csv' with (FORMAT csv, null '\N');

TABLE GENRES
\copy genres from '/Users/shreyasakpal/Desktop/Spotify/genres.csv' with (FORMAT csv, null '\N');