Attributes:

1. Plays
    - Each row contains the number of plays that the song has at its marked date, this can be flattened into a singular play or used to find the total number of plays at that time. Its numeric and must be positive, but has no upper bound. (Though I think it would be hard to listen to the same song one million times.)
2. Skips
    - Each row contains the number of times a song was skipped up to its marked date, this, like plays, can be flattened into a singular skip or used to show the total skips at any given time. It has the same limits as plays.
3. Song Title
    - Contains the title of the song, Not numeric, categorical, or ordinal. It just just a string. Songs can share the same name and thus are differentiated by a primary key.
4. Song Length
    - The length in minutes and seconds of the duration of a song. This has a natural ordering from shortest to longest (or the reverse). It is important to note that all plays must be listened to for the ENTIRE duration of the song.
5. Artist
    - The artist who preformed the song. This attribute serves as a higher level of category of songs. (Above album, below genre)
6. Album
    - The album that the song comes off of. This is the smallest category of songs. (Song -> Album -> Artist -> Genre -> Library)
7. Genre
    - The genre that the song belongs to. This data is categorical but does not have an intentional order. There are currently 38 categories of genre, but most categories will be relatively small.
8. DateTime
    - The date and time of day that the play or skip occurred. Since this is a measure of time it is naturally ordered.
9. Release Date
    - The date that the song originally came out.
10. Date Added
    - The date that the song was added to my music library. This can be important for finding the overall density of song plays.
11. Scrobbles
    - The amount of times the song has been listened to by those contributing to the Last.fm song database.

Sources:

Since these data sets don't have an exact source, I will add a link to a github project containing all of the methods I used to gather my sources.

- Song Stats: for tracking the time of plays and skips
    - https://apps.apple.com/us/app/song-stats-for-apple-music/id1450329823
- My Itunes library for some song attributes not captured by the tracking application:
    - https://www.apple.com/itunes/
- Last.fm for getting total scrobbles for each song
    - https://www.last.fm/home

Github (NOT A SOURCE): https://github.com/Trevor-Johannessen/CSE-332-HW-1-Resources

Data Points:
~10,200, this number will increase as I update the database with more plays.

Datasets Used:
ITunes, SongStats, Last.fm

Why is my data interesting?

I think my data is really interesting because not only is it relevant to me, but is dynamic. I think the possibility of finding trends and reconstructing what seem like completely separate patterns to be fascinating. I think the data presents many opportunities for different types of conclusions. With patterns we can be fairly certain what is happening by referencing it with events that happen at the same time, but I am also curious about the outliers and what can be said about them. Do outliers tend to be longer or shorter, continuous or with pauses? I think the dataset allows for a lot of extremely interesting questions that serve as a really cool exercise in data analysis.

Hypotheses:

1. My daily routine should be discoverable by looking at trends in data. My class schedule should be apparent by looking for consistent times I listen to music while walking to and from class. The data currently covers three semesters so given the dataset it should be reasonable to find the start and end times of each class.

2. As each semester progresses the music I listen to should change as circumstances of the semester change. I expect to see slower, more somber music during times of stress such as midterms and finals.

3. I would like to find the rate I listen to newly added songs. Mainly does it matter more when a song is released, or when it is added to my library. I expect that newly released music would have more dense periods of plays than old music recently added to my library.

4. I believe that as the day progresses songs may be faster or more energetic as I wake up and begin to tolerate the energy more. Early morning songs may also reflect genres which have slower songs.

5. How unique my listening habits are compared to the users of the last.fm database. Are there songs that I listen to relatively frequently that do not appear nearly as often in the database? If I was to take the ratio of my listens to the whole, what songs present the largest value.

* A little append to say that I am working on adding BPM to the dataset but it is more difficult than I thought since there is no reliable source of truth for a song's BPM. I worked with the Last.FM and Spotify API but none had that data so I am now trying one more web API, this however requires me to streamline some of the data which is difficult. The dataset should still be complete even without BPM.