Spotify Data Analysis

Natalie Tubia

May 13, 2024

1 Introduction

For the project, I will be looking at the most streamed Spotify songs 2023 dataset to explore various questions regarding music trends. I have also expanded upon my analysis by incorporating data about the top 2000 songs of all time.

The questions I will be evaluating are:

1. Are different song features (beats per minute, instrumentality, etc.) correlated with the number of streams that tracks received in 2023?
2. What are the most popular genres of all time?
   2.1. How does this change over time?
3. Might more streams lend to a higher presence in user Spotify playlists?
4. How does a track's release year affect streams? Do older or newer tracks get streamed more?
5. How does playlist presence compare between streaming services (Spotify, Apple, Deezer)?
6. How do audio feature patterns change over time?

2 Data Summary

The data contains 953 songs, with 24 different columns:

- track_name: 953 songs are in the data
- artist(s)_name: There are 644 artists represented in the data
- artist_count: The data also accounts for tracks with multiple collaborators
- released_year: Songs in the dataset have release dates between 1930 and 2023
- released_month
- released_day
- in_spotify_playlists: How many Spotify playlists the song can be found in
- in_spotify_charts: Presence and rank of the song in Spotify charts
- streams: The number of streams received by each song
- in_apple_playlists: How many Apple Music playlists the song can be found in
- in_apple_charts: Presence and rank of the song in Apple Music charts
- in_deezer_playlists: How many Deezer playlists the song can be found in
- in_deezer_charts: Presence and rank of the song in Deezer charts
- in_shazam_charts: Presence and rank of the song in Shazam charts

- bpm: Beats per minute
- key: Key of the song
- mode: Mode of the song (major/minor)
- danceability_%: Percentage rank of how danceable the song is
- valence_%: Percentage rank of the positivity of the song's content
- energy_%: Percentage energy level of the song
- acousticness_%: Percentage of acoustic sound levels
- instrumentalness_%: Percentage of instrumental content (as opposed to vocal)
- liveness_%: Percentage of live performance content
- speechiness_%: Percentage of spoken lyrics

The "All Time Top 2000s Mega Dataset" includes the same audio features, but uses a "Popularity" metric (between 1-100) instead of streams, and additionally includes song genres.

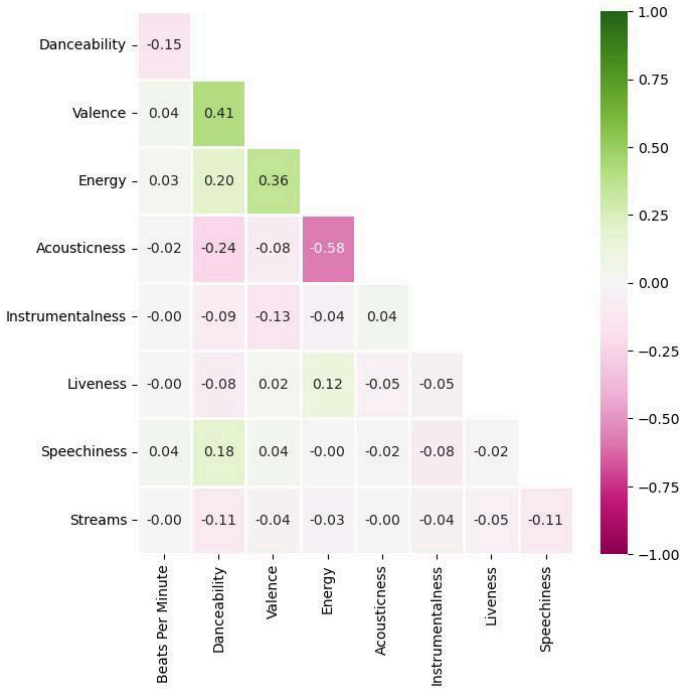

Figure 1: Based on the correlation matrix, we see that the main audio features in the dataset are not correlated with streams. There are some moderate correlations between audio features, including acousticness and energy ($r \approx -0.58$), and valence and danceability ($r \approx 0.41$). These values indicate that higher acousticness is associated with less energetic tracks and that more positive messaging/moods are associated with more danceable songs.

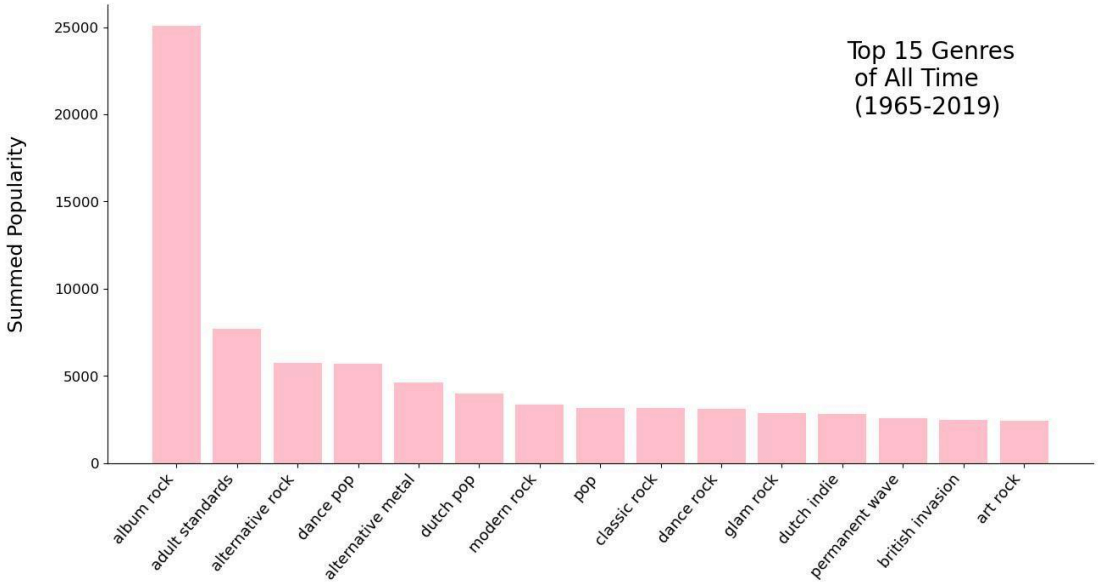

Figure 2: By summing the popularity metric by genre in the top 2000 dataset, we see that the most historically popular genre (as of 2019) is album rock by a wide margin. Behind it are adult standards (swing, jazz, show tunes, “easy listening”), alternative rock, and dance-pop.

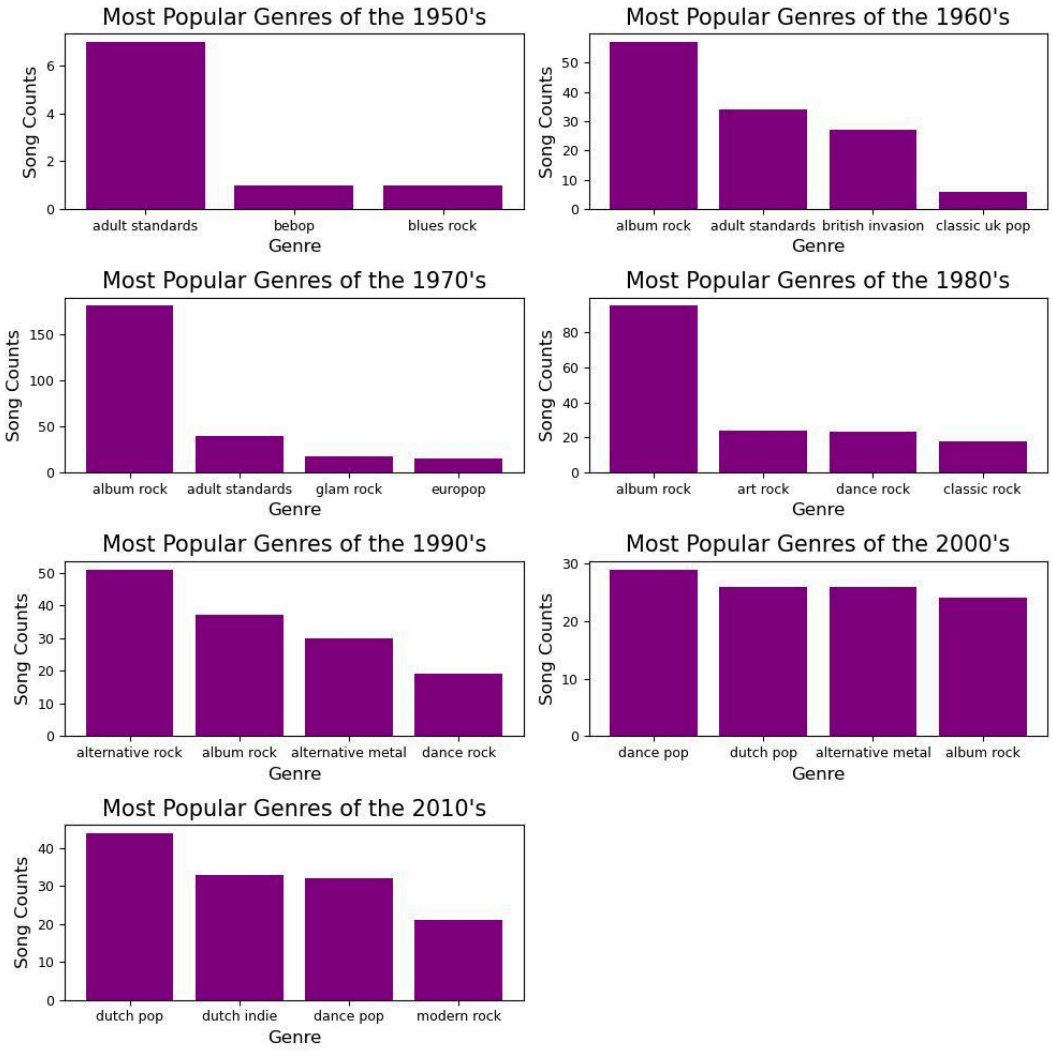

Figure 3: When breaking down genre trends by decade, we see that album rock holds the most significant spot until the 1990s when alternative rock surpasses it. By the 2000s pop music begins to take over, and we see the first instance of significance for indie (independent) music in the 2010s. From the 1960s on, album rock's subcategories (glam, art, dance, classic, modern) also remain consistently prevalent throughout the years.

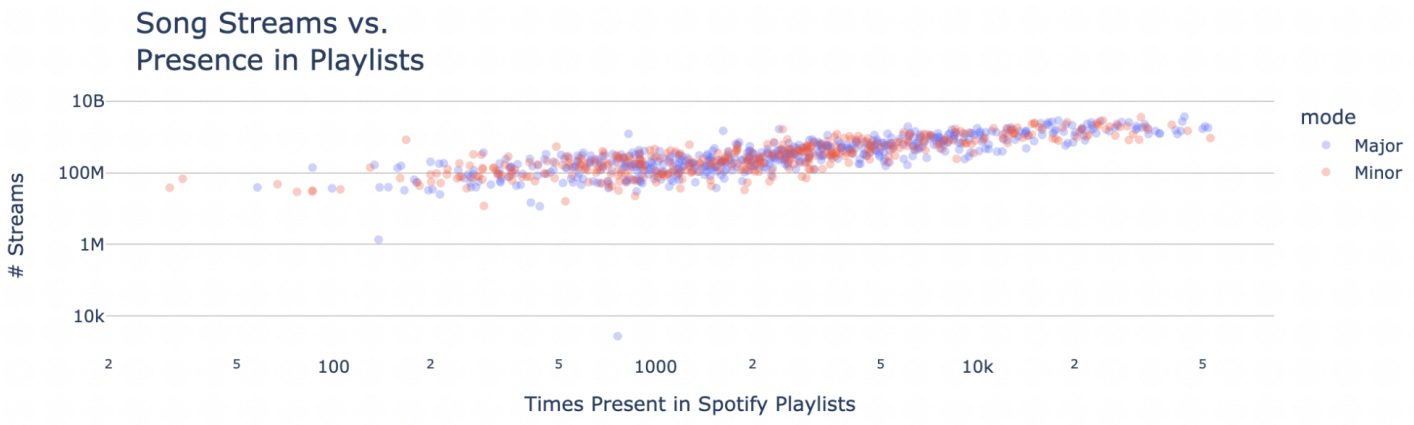

Figure 4: From the scatterplot, we see a very moderate positive linear relationship between the number of streams and the number of times a song is present in a Spotify playlist. The plot also splits the tracks into a mode of major or minor, but there does not appear to be any significant pattern between mode and streams/playlist presence. All in all, the scatterplot indicates an association between streams and playlist presence in that more streams may be connected to higher playlist presence.

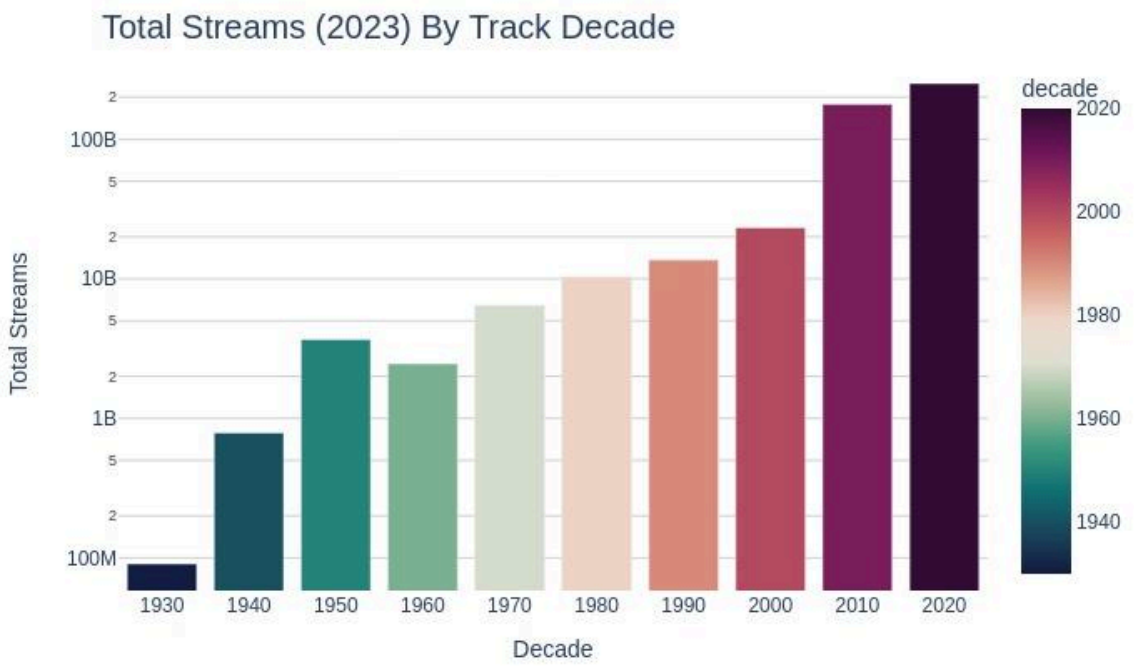

Figure 5: This bar plot shows that, for 2023, newer songs received substantially more streams, as opposed to tracks that have been around for longer. Songs released between 2010-2023 received the largest quantity of streams this past year, with a stark drop in streams for songs released before 2010. Another interesting insight the plot reveals is that there were more streams for tracks released in 1950 than in 1960, likely due to the lasting popularity of certain prolific artists, namely Elvis Presley, especially after the release of two biopics (*Elvis*, 2022) and (*Priscilla*, 2023).

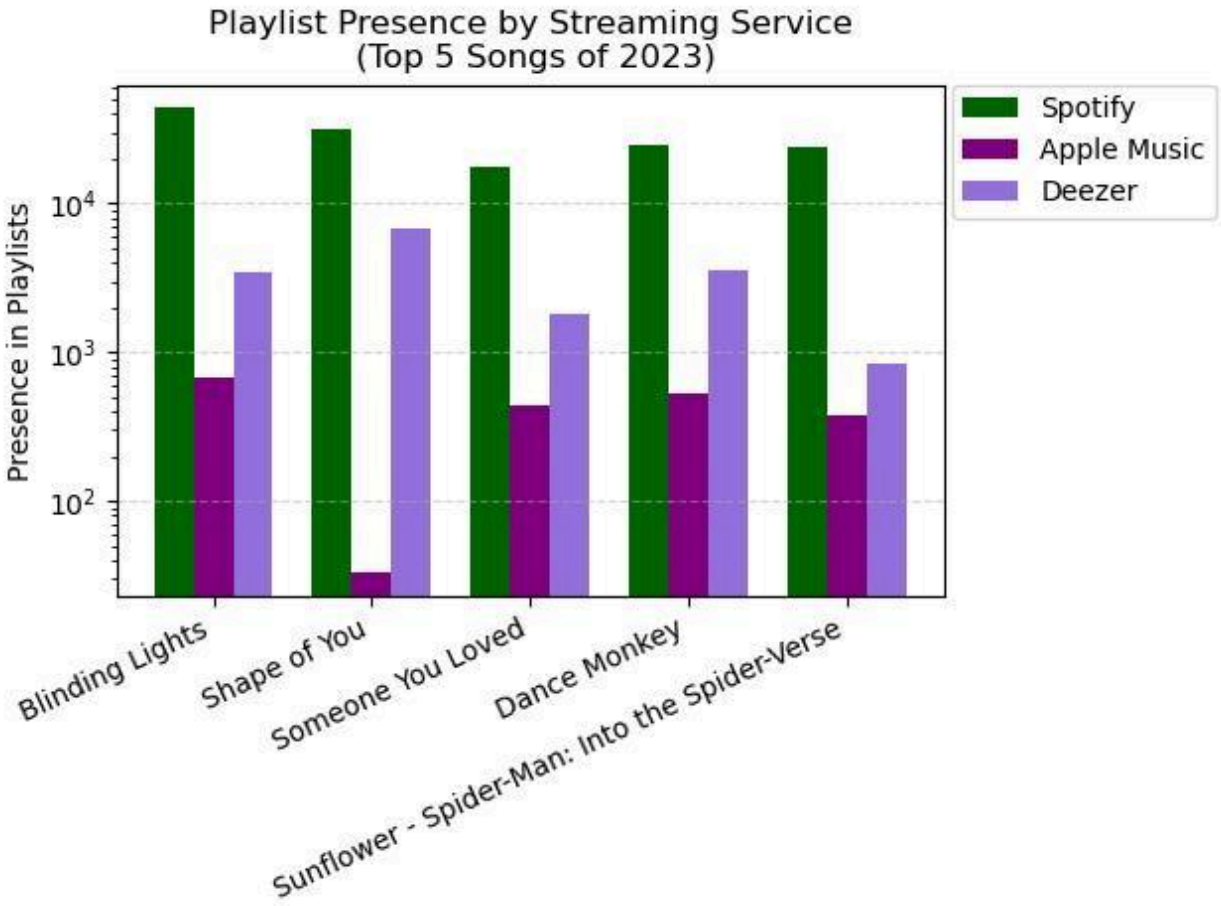

Figure 6: The above grouped bar plot reveals that songs, the top 5 songs of 2023 specifically, most often get placed in Spotify playlists. One more unexpected factor seen in the plot is that Deezer playlists surpass Apple Music playlists for each song, and even come decently close to Spotify for "Shape of You" and "Dance Monkey." This might be due to Deezer's extensive playlist curation features that are nearly comparable to Spotify's, including Deezer Flow, an algorithm that creates a personalized mix of songs based on user preferences, and Shaker, a collaborative playlist feature that creates mixes of friends' top songs from any service.

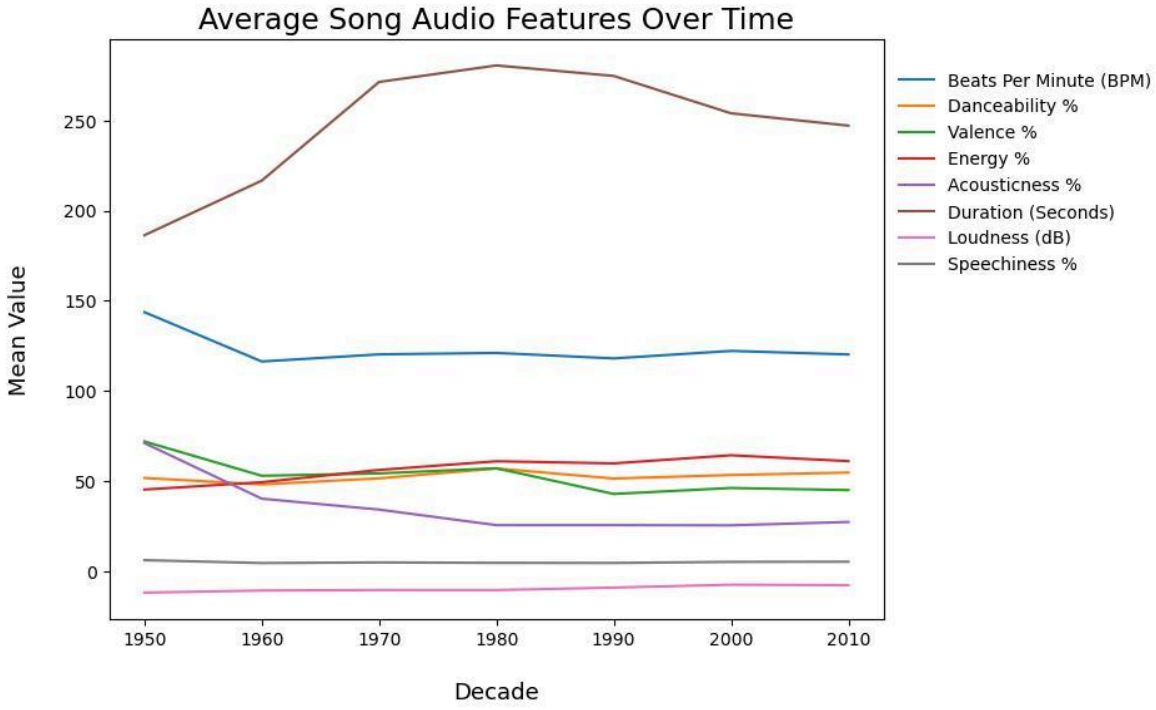

Figure 7: Using a multivariate time series, we see that various audio features have changed over six decades. We see that there have been significant changes in duration, acousticness, valence, and BPM. However, because many of these features are measured on varying scales, I took a closer look at each individual variable to gain some more insight into potential trends.

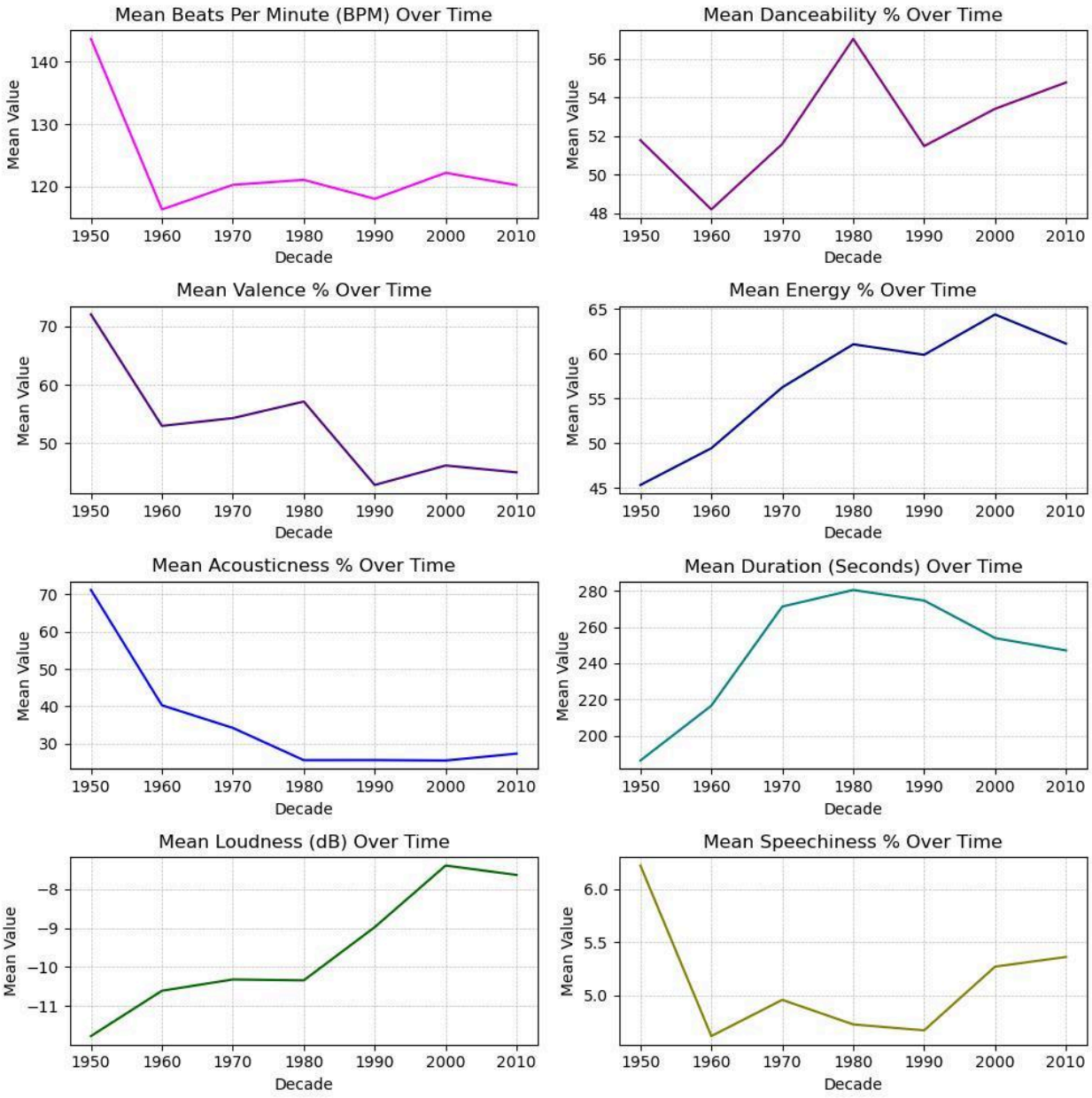

Figure 8: Splitting the variables into each of their own plots, we see that there is some type of substantial shift in all features over the years. Most notably, songs have gotten much longer, with a peak of 280 seconds in the 1980s, leveling off to about 250 most recently. Tracks have also seen roughly a 20% decrease in musical positivity and about a 20% increase in energy since the 1950s. Not surprisingly, danceability had a spike in the 1980s, and acousticness has decreased significantly, meaning more electronically mixed music. Songs have gotten louder and less speechy, and after a ≈25-point decrease after the 1950s, BPM has remained fairly constant.