**Project: A Century of the Danceability of Music**

**Project Team**: Velveeta

**Project Team Members**: Haley Peterson, Emily Wuenstel, Collin Page, Drew Miller

**Thesis**: The “danceability” of music is correlated to measurable qualities of the music. Using data from 1921-2021, there are both positive and negative correlations to measurable attributes of a song that impact its “danceability”.

**Project Overview & Methodology**: Based on Spotify data for songs from 1921-2021, we investigated music’s “danceability” (see Definitions, *below*) over the past century. Incorporating Python, Pandas, Matplotlib, and APIs, our project: 1) plots the Danceability and qualities of songs over time to identify whether temporal trends exist; 2) evaluates qualities of music to test their correlative impact on a song’s Danceability; 3) compares song’s individual qualities to one another, independently of danceability; and 4) evaluates music by genres to identify geographical origins/genre types’ Danceability.

**Initial Data Analysis**: Initial analyses showed the data to be inconclusive based on our starting project parameters. Data for the years 1921-1944 contributed to outliers that skewed our analyses and results due to limited comparative volume of songs, and data for the year 2021 are incomplete; we concluded that revising of our data set to be inclusive of songs from the period 1945-2020 provides 76 years of data that are more reliable and serve as a better representation of songs per year and decade. The revised data largely eliminated extraneous and incomplete points from our analysis, validating our hypothesis.

**Research Questions and Results**:

* Over the past century, how has Danceability changed by over time?

We observed dramatic changes over time in plotting line graphs of song qualities with Danceability, our main metric. Through interpretation of these line graphs, there is a highly positive trend of increased Danceability over time. Our observations across several metrics across the entire time period of our initial data set made clear that was a high amount of volatility within the first 25 years’ worth of data; this created doubt in the reliability of this subset of our data range, resulting in our revisions to our data set.

* What musical qualities have the biggest impact on a song’s Danceability?

Analysis of songs by year from 1945-2020 revealed that Acousticness is correlated the strongest with a song’s Danceability. However, they are negatively correlated so as a song’s Acousticness decreases the song’s Danceability is expected to increase. Danceability revealed to also have strong positive correlations to Energy and Loudness. Making these attributes the two other key indicators that affect a song’s “danceableness”.

* Of the quality with the highest positive relationship with Danceability, with what other quality(ies) did it have the closest statistical relationship?

With data aggregated by decade for a manageable number of plots for visual representation, we employed regression analyses for songs’ qualities that showed the highest positive or negative correlation to danceability. After running these regressions via scatter plots, we confirmed through our statistical analyses what could only be inferred in comparing each quality to Danceability: each of the qualities with the highest and lowest statistically significant contribution to Danceability, as judged by the p and r values of the regression equation, were also highly correlated to one another in a similar fashion. Acousticness, which had the greatest negative effect on Danceability, showed strongly negative associations with Energy and Loudness, while Energy and Loudness themselves proved highly corroborative.

* Exclusive of the United States, are there countries with higher levels of Danceability, and which geographically-specific genres are the most “danceable”?

An analysis of country-specific genres with aggregated music from 1921–2020 demonstrates that the genres with the highest danceability scores are not necessarily what would be generally considered as a “popular” music genre. For instance, British Children’s Music, while highly danceable, is clearly not recognized globally as a “popular style” of music amongst all countries analyzed. This shows that Danceability is not indicative of popularity, regardless of geography, be it on a regional or country level.

**Conclusion:** Our analysis provides data-driven evidence that, over the past 76 years:

*1) Across decades, there has been a steady increase of Danceability since the Post-War years, with a decline of Acousticness and rise in both Energy and Loudness;*

*2) As indicated by a song’s qualities, Acousticness has the highest correlation to a song’s Danceability, followed closely by Energy and Loudness;*

*3) Among qualities, Acousticness and Energy had the highest correlation to one another, although strongly negative, while the next closest correlation of qualities was a strong positive relationship between Energy and Loudness; and*

*4) There are geographic influences, as indicated by genre that impact a song’s Danceability.*

**Definitions** (Qualities of Music):

1. **Acousticness**: “A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.”
2. **Danceability**: “Describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.”
3. **Liveliness**: “Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.”
4. **Loudness**: “The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typical range between -60 and 0 db.”
5. **Instrumentalness**: “Predicts whether a track contains no vocals. “Ooh” and “aah” sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly “vocal”. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.”
6. **Speechiness**: “Detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g., talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.”
7. **Tempo**: “The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.”
8. **Valence**: “A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g., happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g., sad, depressed, angry).”