

Data Visualizations and Observations

Sarah Schwartz

September 29, 2022

Data Background

Artist Name	Main Genre	Number of Followers	Popularity
Wild Child	austin americana	144740	52
Tessa Violet	alt z	647740	57
Halsey	dance pop	18402472	83
OK Go	alternative rock	483708	56
Regina Spektor	indie pop	1294864	62
Modest Mouse	alternative rock	1310339	63
The Killers	alternative rock	6655364	77
the GazetteE	j-metal	288014	48
Dolly Parton	classic country pop	1900254	71
Ed Sheeran	pop	102584457	91
The Vaccines	indie rock	717553	59
Bloc Party	alternative dance	1036944	60
Simple Plan	canadian pop punk	3215030	68
Creedence Clearwater Revival	classic rock	5287266	77
Miike Snow	alternative dance	612803	55
Adele	british soul	42433872	84
Grace VanderWaal	post-teen pop	1299911	56
Fall Out Boy	emo	9548528	78
blink-182	pop punk	7095758	76
Taylor Swift	pop	58988141	94

The data in the following visualizations was collected from Spotify as a part of a project for Vanguard's Data Engineer Apprenticeship application process. The artists in Table 1 were those whose information was entered into the database and then queried and visualized. A description of all track features used during the analysis can be found on slide 7.

Table 1. Table 1 contains all of the artists whose albums and songs were used during this project.

Visualization 1: Heatmap of Track Feature Correlations

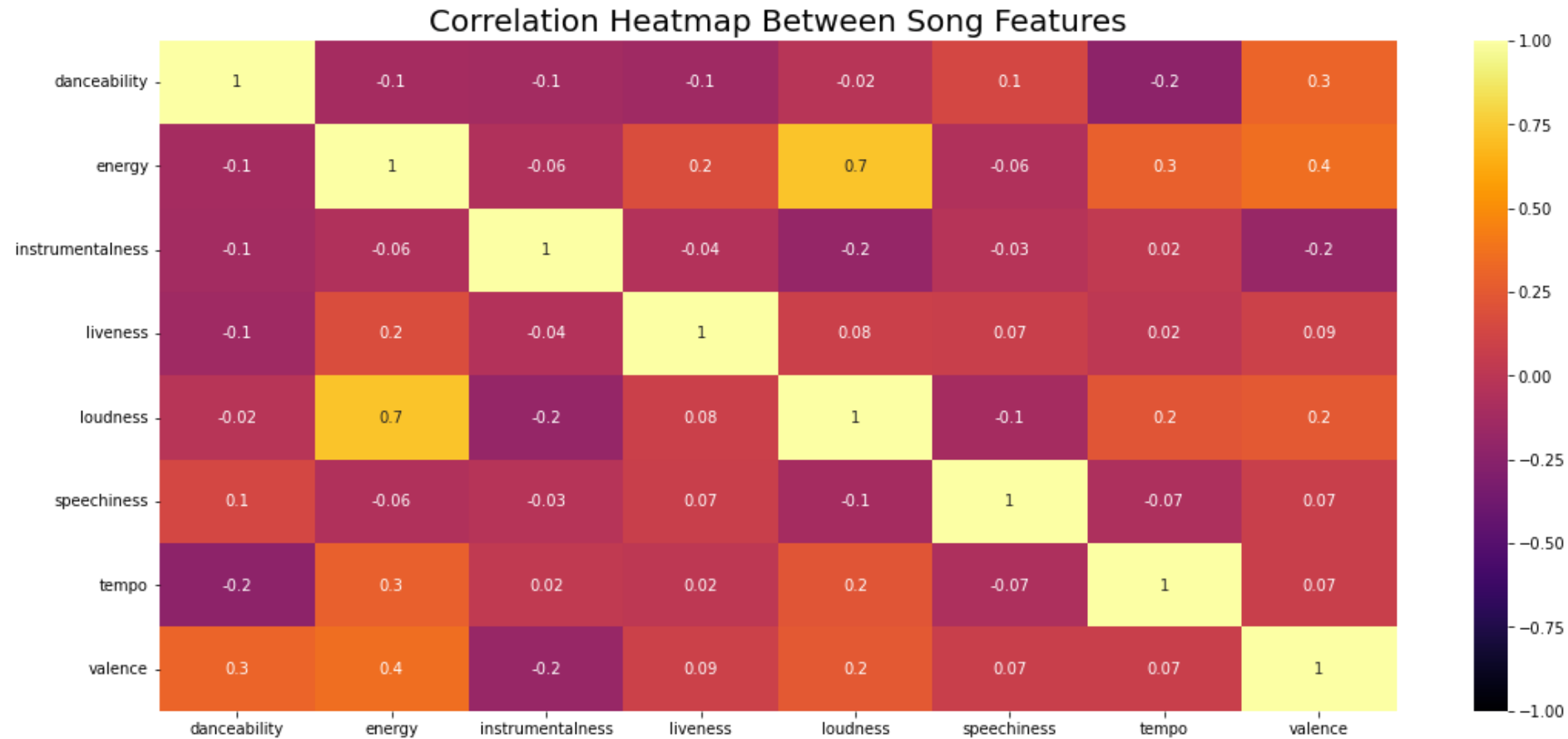


Figure 1. This heatmap shows the correlation between all of the numerical track features. The closer the color is to light yellow and the closer the value it to 1, the more positively correlated the two features are. Conversely, the closer the color is to black and the value is to negative one, the more negatively correlated the two values are. Loudness and energy appear to have a strong positive correlation, with an R-value of 0.7. Most other values have no correlation (close to zero) or are only somewhat correlated. The most interesting observation to me is that danceability and tempo have a slightly negative correlation. Description of all track features can be found on Slide 7.

Visualization 1.5: Scatter Plot of Selected Track Features

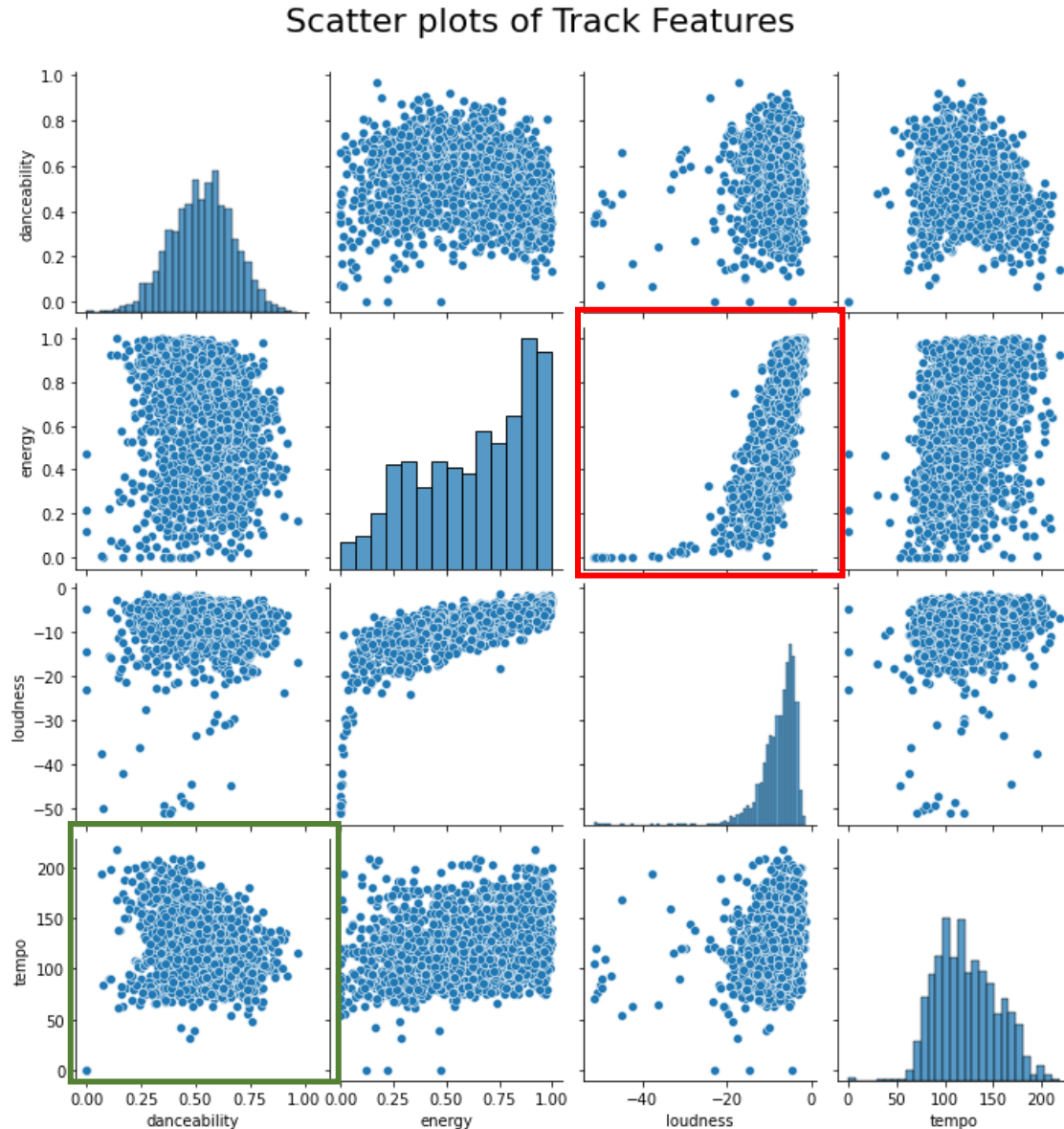


Figure 2. Here are the scatter plots for the same numerical track features from Figure 1. The box in green is the Tempo vs. Danceability which has a negative correlation with an R-value of -0.2 in the previous figure. The red box highlights the Energy vs. Loudness scatterplot data that was shown to have a positive correlation with an R-value of 0.7. Description of all track features can be found on Slide 7.

Visualization 2: Violin Plots of Selected Track Features

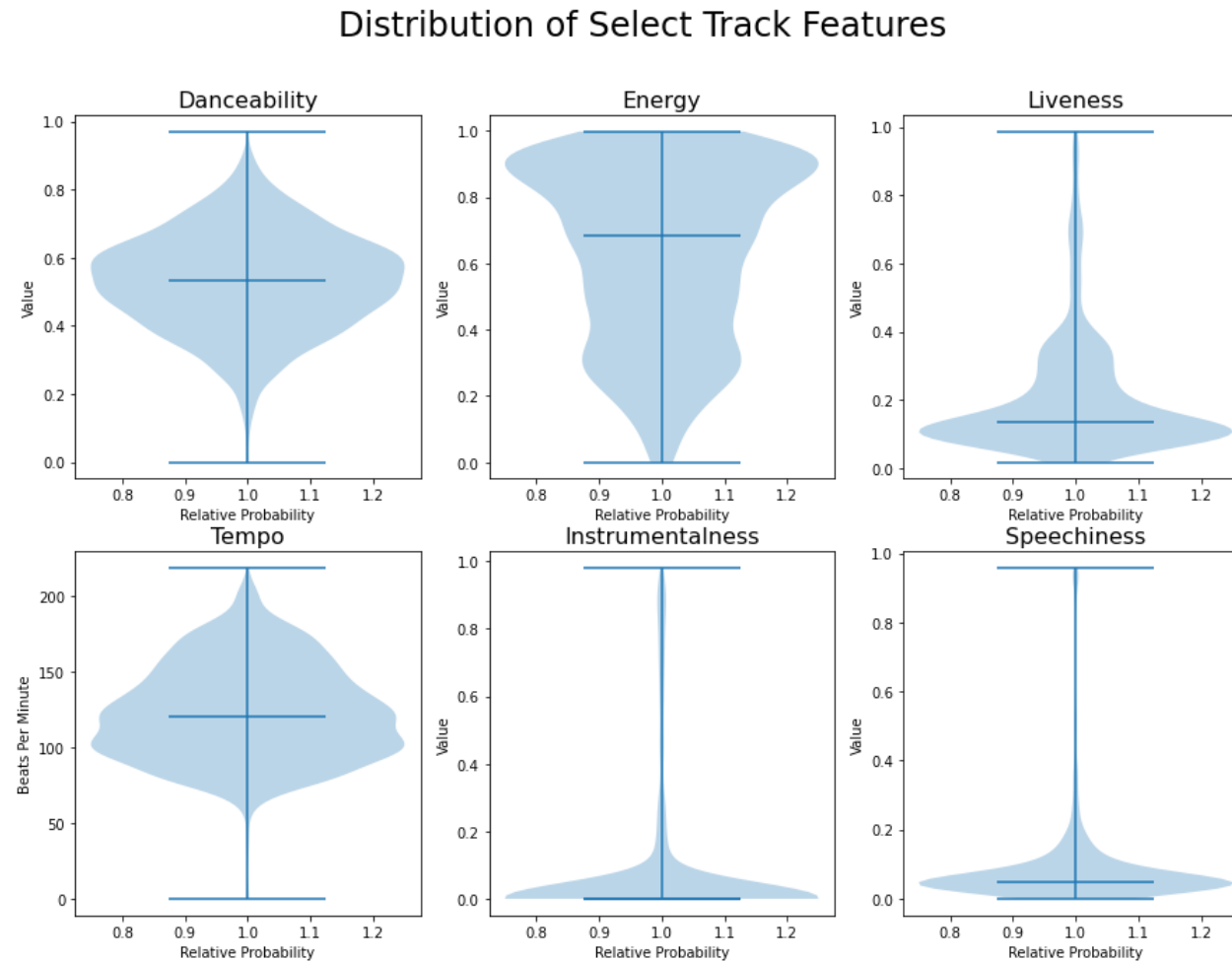


Figure 3. Above are the violin plots for Danceability, Energy, Liveness, Tempo, Instrumentalness, and Speechiness. For each feature, the minimum, maximum and mean values are all represented as bars, while the overall distribution is shown by the light shading. Danceability is the most similar to a normal distribution. I find the distribution of the Instrumentalness to be interesting since it appears that the mean and minimum values are very similar, but there is a small group of highly instrumental tracks.

Visualization 3: Average Danceability of Songs From 1999 to 2022

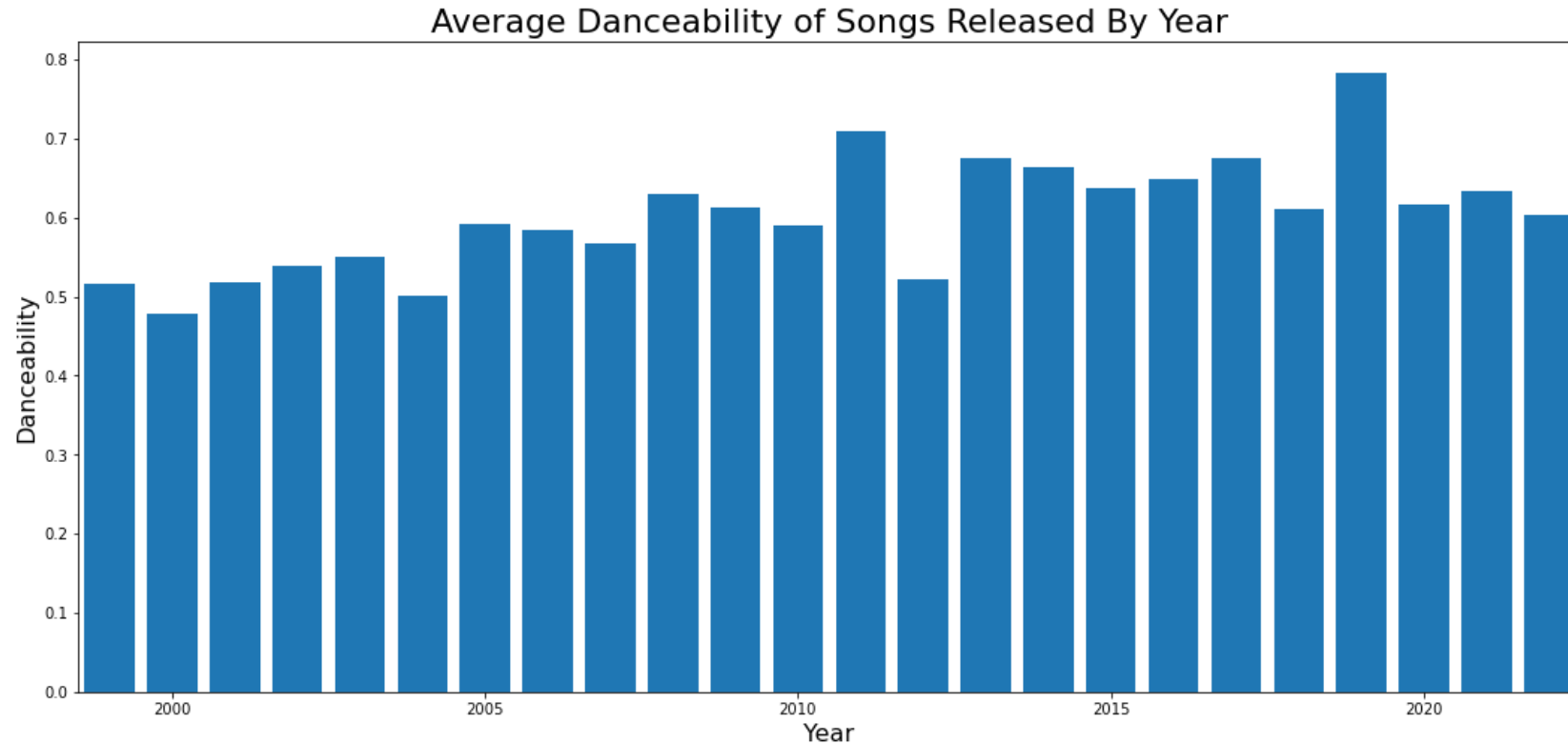


Figure 4. This last chart looks simple, but don't be deceived! There might be an influence from the Covid-19 pandemic in the data. Above is the average danceability of songs released each year. The Danceability for 2019 looks to be much higher than that of the following year. In fact, The average Danceability for 2019 and 2020 are 0.623873 ± 0.174231 and 0.560098 ± 0.12276 with sample sizes of 92 and 102 tracks respectively. The p-value between these two values is 0.034 meaning there is only a 3.4% chance that they are consistent with each other. This can be considered a statistically significant different. However, the values for 2018 and 2020 are statistically consistent, so it is possible that 2019 was just a great year for dancing music.

Track Feature Definitions

Feature Name	Description ¹
Danceability	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.
Energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale.
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.
Liveness	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.
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 typically range between -60 and 0 db.
Speechiness	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.
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.
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).

Note 1: Descriptions as described by Spotify (<https://developer.spotify.com/documentation/web-api/reference/#/operations/get-several-audio-features>)