**Luci Barnett and Ben Gierhart**  
**Team Project 1 Report**  
**DS160 – Introduction to Data Science**  
**Fall 2024**

For this project, we used the [Taylor Swift Discography (TTPD included) dataset](https://www.kaggle.com/datasets/delfinaoliva/taylor-swift-discography). For context, TTPD stands for *The Tortured Poets Department*, Taylor Swift’s latest album. This is a regularly updated dataset by Kaggle user Delfina Oliva. It was last updated on September 24 and gathers data from Spotify, Youtube, and Wikipedia. There are 28 columns in total. Here is a description of the ones that are not self-explanatory:

* **ID** - unique identifier for each track
* **duration\_ms** - the duration of the track in milliseconds
* **videoclip\_views** - the number of views of the track's videoclip on YouTube
* **spotify\_streams** - the number of streams (listens) the track has on Spotify
* **spotify\_global\_peak** - the highest global ranking the track achieved on Spotify charts
* **track\_number** - the order the song appears on it's album
* **album\_physical\_sales** - the physical sales of the album
* **acousticness** - a measure of how acoustic the track is (0 to 1)
* **danceability** - a measure of how suitable the track is for dancing (0 to 1) based on a combination of musical elements including rhythm stability, tempo and beat
* **energy** - the intensity and activity level of the track (0 to 1). Typically, energetic tracks feel fast, loud, and noisy
* **instrumentalness** - a measure of how much of the track is instrumental (0 to 1). The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content
* **liveness** - a measure of how "live" the track feels, indicating audience presence (0 to 1). Tracks with higher liveness values are more likely to have been performed in a live setting
* **loudness** - the overall loudness of the track in decibels (dB). 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** - a measure of the amount of spoken words in the track (0 to 1). The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value
* **tempo** - the tempo or speed of the track in beats per minute (BPM). Tempo is the speed or pace of a given piece and derives directly from the average beat duration
* **valence** - a measure of how positive or negative the track sounds (0 to 1). Tracks with high valence sound more positive

Our rationale is that Taylor Swift is currently the biggest pop star in the world, and so we wanted to look at this comprehensive dataset to learn more about the discography and career of this once-in-a-generation artist. In order to do so, we explored multiple individual variables using data summary functions and visually with histograms and a pie chart. We also used scatterplots, box plots, line graphs, bar charts, and a heat map while visually exploring the relationships between different variables.

Our statistical analysis revealed several notable relationships within Taylor Swift's discography. We found a strong negative linear relationship between acousticness and energy, indicating that more acoustic tracks tend to have lower energy levels. We also observed a moderate positive correlation (0.48) between valence and energy, meaning that songs with higher valence, which are generally perceived as happier, tend to have more energetic compositions.

In terms of commercial success, we found that danceability is positively correlated with Spotify streams, implying that more danceable tracks are generally more popular. Additionally, energy levels showed a slight positive correlation with album physical sales, suggesting that energetic tracks may contribute to better sales performance. These findings suggest that both musical attributes and perceived emotion play a role in shaping a track’s popularity and commercial impact within Taylor Swift’s body of work.

We also made other interesting findings. Albums categorized as “Country” had the highest number of physical album sales on average, while those categorized as “Indie Folk” had the lowest. However, the opposite is true when it comes to streams on Spotify. It’s also worth pointing out that tracks on albums categorized as “Indie Folk” only make up 17.2% of Taylor Swift’s discography. The standard edition of the original *1989* album is the highest in physical sales at just over 8 million copies. The most streamed track is “Cruel Summer” from *Lover* at almost 2 billion streams. With our pie chart, we can see that combining the “Country” and “Country Pop” album genres still gives 26.5% of the discography, and this group contains the highest performing albums in terms of physical sales.

In conclusion, Taylor Swift is a statistically interesting artist whose discography covers a number of genres and whose career spans multiple decades. She is also an artist who has seen vast changes in technology responsible for significant shifts in consumers’ tastes and access to music.