- Variables
    - Track_name, artist(s)_name, artist_count, released_year, released_month, released_day, in_spotify_playlists, in_spotify_charts, streams, bpm, key, mode (major/minor)
      These variables could be used to address questions related to song popularity, artist collaboration trends, streaming trends over time, and more within the context of Spotify streaming data.
- Describe what the variables mean
    - Tells the person using the data what the name of the song is, the artist who sings on the song, how many artists there are, and when it was released including month, day, and year. It also includes how many Spotify playlists it is in and how many streams a song has received. Lastly, it includes the beats per minute or BPM for short, the key it is in, and if it is major or minor.

Throughout the past two weeks, I began looking at Spotify data for 2023 to view the most streamed songs. I looked at a bunch of metrics such as track name, artist name, release year, BPM, number of streams, and more. Throughout my research, I curated a plethora of graphs such as histograms and scatter plots to better conceptualize the story in which the data is trying to get across. The histograms were used to create a visual between variables such as streams vs. BPM. Hypothesis testing and regression analysis were conducted to explore relationships between variables and to test different hypotheses. If I had more time, I would have loved to explore time series analysis and examine streaming trends over a period of time, especially for the latter half of the year, around October through December. Late October through the last week of January is known as the “Holiday Freeze”, this is important because Christmas music is played a lot during this time period, so it would have been interesting to see how popular songs are especially throughout that time. Genre analysis would have also been interesting to look into as well. Within the music scene there has been talk about how pop music is dying, looking at the stats to see what genre of music is most popular would definitely help with that argument; which brings me to my next point of missing variables. I would have loved to go over genre and how popular a specific artist was during that time period as every artist has its wave of up and down within their musical career and their musical peak. One of the biggest assumptions made during the time of analysis was assuming there was a linear relationship between variables, especially in my regression analysis. Some challenges faced were how to handle invalid data or missing data. A prime example of this was that when importing the gathered data, the data was not in UTF-8, which caused major issues. This dataset was over 500 lines, so I was unable to go through each line and fix the mistakes hand-by-hand, instead, I had to figure out ways to work around it. Another challenge was finding the resources and remembering all of the things we learned over the past 12 weeks, not only in this course but in previous courses as well. In this course, we did not use R, the programming language, I had to pull up information on how to construct scatterplots and other graphs from another course and tie it in with this course. It was rather fascinating seeing the courses blend, doing so I feel I was able to show mastery over the skills I have learned before.