1. **Why did you choose the dataset you used for your EDA? What about the context of the data interested you? What relationships did you expect to find among your variables of interest?**

The dataset was obtained from Zenoob's Music Information Retrieval collection and was originally compiled using data from Spotify. The Music Genre Dataset includes weekly charts of the 200 most-streamed songs across various countries and territories, as well as a global aggregate chart.

I chose this topic to combine my passion for music with my academic interest in data science. Not only does the dataset study popular music trends, but provides insight into the defining characteristics of hit songs—such as language, danceability, key, and other musical attributes. Thereby, its analysis may uncover what makes a song successful. The dataset's context intrigued me because it captures how audiences around the world engage with music on a weekly basis, with cultural preferences and listening habits across globally. I additionally hoped to gain exposure to music data analytics and explore its potential as a future career path through this project.

This analysis has helped me understand that popularity isn't driven by one musical characteristic alone. Instead, it emerges from a combination of accessible lyrical content, engagement, and overall production appeal. My prior thoughts and biases have been redirected to be based on this study. The explicit nature of a song impacts the markets in which it may be available, but does not directly correlate to a song's popularity. Thereby, both explicit and non-explicit music were evenly represented in Spotify's top hits. The number of artists on a song does not determine whether or not a song is a hit, however, it affects the audience it may reach. With more artists on a song, a larger audience is reached, demonstrating that collaborative music is typically more successful.

Musical structure related to the duration and tempo, a foundation, of a hit. My investigation has revealed that while tempo and duration may have a wide range and vary throughout the Spotify top hits, there is a certain range most common among hits. Thereby, successful songs have a typical structure, reflecting broad trends.

2. **What challenges did you face while completing the assignment, and how did you overcome those challenges? What do you plan to do differently next time as a result? List at least one specific thing you learned about yourself as a data science student as a result.**

One of the primary challenges I faced during data preparation was filtering the dataset to only include the variables I needed. Several variables were interlinked, removing one would sometimes affect the integrity of another, making it difficult to isolate them cleanly. Despite this, I proceeded by focusing on the relevant columns and disregarding the rest in my analysis.

Although the dataset remained larger than intended, this approach ensured that no essential information was lost while keeping the project manageable.

Another challenge I encountered was importing the dataset into my environment for my code to reference. When importing the file in the standard way, the dataset output appeared disorganized, confusing, and lost its structure, most likely due to its size. To overcome this issue, I imported the dataset directly into the environment and wrote my code to pull from it using the "read.delim" function.

Next time I run into an issue, I plan on reaching out for help to both my teachers and online resources. I wasted time troubleshooting myself and retrying without success. This trouble proved to be unnecessary and could have easily been avoided if I had used my resources at the right time. I additionally plan on using these methods as a means of independently troubleshooting, as these challenges have taught me code that I may need to employ in the future.

I have learned that I tend to become frustrated when working alone for too long. I'm more productive when I attend office hours and ask questions. I learn best by running into challenges, sometimes failing to solve them, and then seeking insight from those with more experience.

3. **Describe at least one way your personal background influenced how you analyzed this data or presented your analysis. What might someone else have done differently if analyzing the same dataset?**

My personal background mainly affected the questions I asked. Because I tend to focus on patterns that reflect human behavior and decision-making, I prioritized exploring relationships that revealed those aspects in the data. This focus influenced the variables I compared and the insights I drew. Another analyst, perhaps with a more technical or statistical orientation, might have emphasized different variables.