Executive Summary

This report explores what makes a song enjoyable by analyzing Spotify listening data. The goal was to uncover the key features of songs that influence listening preferences and determine whether publicly available data can predict which songs people enjoy. Using audio features such as track popularity, energy, danceability, loudness, and acousticness, the analysis applied both simple statistical models and more advanced machine learning techniques.

The results showed that certain song characteristics are linked to enjoyment. Songs that are popular, energetic, and easy to dance to were more likely to be enjoyed, while quieter or highly acoustic songs were less favored. However, despite identifying these trends, the models' ability to predict enjoyment was limited, with accuracy hovering around 55-57%. This means that, while some patterns could be found, the publicly available data alone is not enough to fully explain why someone enjoys a particular song.

The modest accuracy of these models suggests that Spotify's ability to deliver highly personalized recommendations depends on far more detailed and specific data than what is publicly shared. Beyond basic song features, Spotify likely uses extensive behavioral data, such as when users listen, their mood, habits, and patterns over time. It is also possible that third-party data and other forms of personal identification help refine their algorithms, giving the platform a deeper understanding of individual preferences.

This raises important questions about transparency and data privacy. Users benefit from accurate, personalized music recommendations, but they often do not realize how much data is being collected and analyzed to achieve this. Platforms like Spotify operate sophisticated systems that rely on hidden, proprietary data, which gives them a significant advantage in understanding user behavior but also leaves users in the dark about how their information is used.

In conclusion, while this report highlights some factors that influence song enjoyment, it also underscores the limits of publicly available data. The findings point to the hidden power of proprietary data held by streaming services, which allows them to create far more accurate and tailored recommendations. Future work could explore richer datasets that include listening behavior, time-based trends, and contextual factors to better understand what truly drives music preferences.