This project served as a great synthesis of what we learned from this course, particularly with regards to the specific technical implementations related to large language model concepts (embeddings, vector databases, semantic search) but also with lessons throughout the duration of our MSBA curriculum, including but not limited to: proper code development, presentation development, teamwork and communication skills, and professionalism in an end-to-end data science project.

Our team faced a number of challenges throughout this project. For example, we experimented with new and different packages, libraries, and models not explicitly covered in the course - to successfully incorporate these new items in our project in a way that we could showcase our growth, we had to take what we had already learned in the course as background, general context and adapt to new circumstances (and read plenty of documentation). We also were challenged to connect our key findings and takeaways to business impacts. Although we started with some good ideas on the direction that we wanted this project to go and the overarching motivations, we deeply considered the user, the user experience, and how our new app solved user pain points. In some ways, this process drew upon our learnings from other courses in the MSBA curriculum such as product management, but ultimately our group felt that this also showcased our growth as not just pure data scientists but as business analysts who can take real business problems, turn them into data analytics projects, and come out with actionable takeaways and summaries.

As mentioned previously, since our team incorporated previously unseen tools, we naturally were excited by certain components of our work by the end of the project. Notably, our team enjoyed working with the Spotify API and the depth of Spotify data available. Granted, we were unable to work with music audio-related data (and mentioned in our presentation and report that we would like to do so in a hypothetical future iteration of this project), but the audio-related data that we were able to find, in addition to the lyrics data that we had to begin with, were interesting to work with. Our team also liked working with the Llama 2 model for keyword generation. It was fun to be able to try out a different generative model and incorporate it in a useful way. Lastly, our project group enjoyed interacting with the front-end Gradio interface and developing example queries. Especially as all of us are avid users of Spotify, it was interesting to develop our own (basic) app interface, still with the recognizable stylings of Spotify, and experiment with different search terms across both Spotify itself as well as our deployed app and analyze the difference in results.

Ultimately, this project was a success end-to-end. We were pleased with the outputs (code, app, report, and slide deck) and enjoyed working on a project with topics that we were all interested in and passionate about.