DS 2002: Data Science Systems Final Project

Cody Liddle, Tamer El-Aridi, and Katrina Garcia

For this project, we selected our datasets based on alignment with our overall goal: developing a chatbot capable of answering music-related questions and providing song recommendations. We utilized the Spotify API to retrieve live data and supplemented it with a cleaned dataset from Kaggle, which originally contained over 1.6 million songs. Due to GitHub's file size constraints, we refined the dataset by removing duplicates and irrelevant entries, ultimately reducing it to approximately 5.2 MB. This preprocessing made the dataset more manageable and improved overall system performance.

Throughout the process, we encountered several challenges. Setting up API authentication for Spotify proved to be less intuitive than expected, particularly when managing access tokens and securely storing credentials. Additionally, cleaning the dataset was time intensive, and building an ETL pipeline that integrated both static and live data sources required careful planning and testing.

Despite these difficulties, the project provided valuable learning experiences. We strengthened our ability to work with external APIs, design and implement ETL pipelines from scratch, and integrate multiple components into a cohesive system. It was especially rewarding to see that even a relatively simple chatbot could deliver meaningful results when supported by well structured and relevant data.

Given more time, we would have expanded the system's functionality by incorporating Spotify user authentication to enable personalized recommendations. We also hoped to enhance the

chatbot's conversational abilities by implementing memory and more advanced natural language understanding. Additionally, building a user facing frontend and incorporating more filtering options, such as by genre or mood, would have significantly improved the user experience. Finally, more comprehensive error handling would have contributed to a more effective and polished application.