

This project was a great learning experience in working with real-world data, especially integrating a local music CSV file with data pulled from the Spotify API. While we expected some challenges, a few aspects were easier than we anticipated, while others were more difficult than we thought.

One of the biggest challenges was working with the Spotify API. What specifically made it harder than expected was trying to match the tracks from our CSV file with API responses because of small differences in the song titles and artist names. The challenge we faced was that a ton of the songs didn't return exact matches, so we had to figure out how to handle missing data. We tackled these issues by adding error handling to remove inconsistencies between the data. It was only difficult because we had to define the line between finding exact data matches and settling for no information, which took a lot of time and discussion between us.

One part of the project that was easier than expected was retrieving track metadata from the Spotify API. Setting up authentication for the generated API keys was easy thanks to other work from our class. Once the authentication was set up, making requests using the `spotipy` library was straightforward, and the API responses were well-structured, making it easy to extract relevant information. Additionally, working with the CSV file in `pandas` was simple, since it allowed us for quick data exploration and cleaning. While some aspects of the project were challenging, the overall process of handling and integrating the data was smoother than expected.

Through this project, we strongly believe this kind of ETL pipeline could be really useful for future projects. Automating data retrieval from APIs allows us to work with real-time data rather than static files, which is useful for literally all fields. Learning how to clean and merge data from multiple sources is also a valuable skill for any data-related job. Overall, this project

reinforced key data science concepts and helped us better understand the challenges of working with external data sources. It was helpful to also learn through data that interests us as a group. While dealing with APIs and data inconsistencies was frustrating at times, it was rewarding to see everything come together in the end.