Rix Prakash

DS2002

Data Science Systems

20 October 2024

Data Project 1 Reflection

During this project, one of the main challenges I faced was managing the transformation between different file formats. Since the process of switching the input file into a JSON and SQL are different, I had to create functions that would execute based on conditions. I also had to handle the editing/modification of the files with the data manipulation differently. Adding/removing files from JSON and SQL are different, but fortunately CSV was straightforward to do and was similar to SQL. Specifically with these files, the transformation of complex JSON structures into the relational format proved by the user seemed to be more difficult than I thought. Additionally, managing potential errors that could come into effect like missing columns or formatting issues required me to put exceptions to handle these errors to allow for a smooth execution of the code. Also, I had to make sure the user put in only 2 options that were provided and if they didn't I had to not let the code execute and prompt them to put in a correct choice.

However, despite these challenges, some aspects of the project were easier than I expected. Using the pandas library for example to handle data manipulation turned out to be pretty straightforward for csv files, and similarly with editing SQL files. Also, setting up the user input and flow control in the script was manageable to get the function to continue to use this. With this process, I allowed the user to choose between CSV or SQL output which was easy to ask, however like previously stated, to actually convert it was a little tricky. Finally, the easiest

part was saving the file back to the system or saving it as a sql database. Both of these proved to be simple.

Overall, the utility developed for this project offers value for future data tasks in many different fields as well. Many data projects involve working with diverse data sources, such as the finance industry using data from APIs, CSV files, or and other databases like Bloomberg that offers different types of files. Having a flexible processor like this one can help streamline data processing by automating common tasks like data ingestion, format conversion, and preprocessing. A perfect example of this is with the Warriors NBA team, where they might have collected data from various formats and places. This would help them consolidate it into a specific form they want and can keep on adding through a structured process like this. Eventually, this could be expanded to integrate data from multiple sources into one, or adding features for more advanced data cleaning. Overall, this processor provides a solid foundation for transforming, cleaning, and converting data, which makes this very useful for a wide variety of industries that need to conduct data science projects.