DS Project 1 Reflection: Christina Barton, Corinna Keum, Maria Vitória Klein

Our project utilizes data from a CSV file describing egg prices over the last 50 years and an API housing milk prices over the same time frame. Our original plan was to scrape the milk data from the original website, but we encountered too many errors to determine if it was possible to get the data through scraping. Eventually, we learned that we needed to use an API to obtain the data. Fortunately, the milk data was available on a website providing free APIs, allowing us to easily proceed with the steps to retrieve the data. However, working with APIs was a new concept, and we faced a learning curve in figuring out how to register for the API key, find the series_ID, and correctly use the base_URL.

One of the significant challenges we encountered during implementation was not knowing when errors would occur, making it difficult to troubleshoot and debug the process. Additionally, figuring out how to authenticate and make requests to the API was initially confusing, especially when handling potential errors or incorrect requests. Errors occurred if the API key was incorrect, the series_ID was invalid, or the URL was not properly formatted.

Although we had some trouble, many aspects were easier than expected. Combining and cleaning up our data was simpler than we had anticipated, likely due to the fact that our data was fairly simple and collected from reliable and complete sources. This meant there were fewer missing values to account for, making the cleaning process straightforward. Renaming columns and merging the data into a complete set was also relatively simple, especially with the help of error messages that allowed us to quickly identify and fix any formatting or compatibility issues.

This ETL process we created could be very useful for future data projects that involve retrieving and analyzing data from APIs or CSV files. While our project was relatively simple, going through the steps of converting file formats, cleaning data, and implementing error messages gave us valuable experience that can be applied to more complex analysis tasks. Making sure our data is compatible across different formats and platforms is crucial, and the error-checking mechanisms we developed can be adapted to improve data processing in future projects. The skills we gained from this project will be especially helpful when working with larger and more complicated datasets that require careful handling and analysis.