Beverley Appiatse
Professor Williamson
DS 2002
19 March 2025

First Data Project Reflection: Agriculture Data Processor

One of the main challenges I had when implementing this Agricultural Data Processor was the discrepancies in the data formats from the CSV files and USDA API. I realized that the two CSV files had inconsistent data which required me to instill additional cleaning steps to standardize the data and for it to merge. I found the integration process of the two different data types to be difficult because it required a lot of time to ensure that the data was properly formatted. I also found debugging to be a little difficult, as many problems arose from the differences in column structures and data types.

Despite these challenges, some aspects of the project were easier than anticipated. For example, I found the implementation of error handling to be easier than expected once I identified common issues such as missing values or API failures. Additionally creating the visualization was easier than expected. I used Matplotlib and Seaborn to create the visualizations, and it was relatively straightforward. I had not anticipated the merging of the datasets to be as difficult as it was. It was difficult to make sure that the datasets actually merged and that the merged dataset accurately represented information from both sources. There were many differences in the formatting, column names, and units of measurement in the datasets that required a lot of attention to detail and standardization.

A utility like the Agriculture Data Processor could be useful for future data projects where multiple sources of data need to be integrated and analyzed. Many situations in the real world will require multiple datasets from different origins that will also need to be preprocessed.

This process can ultimately lead to more accurate analyses and will make the dataset more accessible for future initiatives.