**Amjad Altuwayjiri**

**Capstone Project**

**Weekly report #05**

This week, I resolved a major issue that had been bothering me. The organization of numbers within the dataset wasn’t aligning as neatly into the columns as I had anticipated. After a careful review, I made the necessary adjustments to ensure that each number found its proper place, bringing much-needed order to the chaos.

In my quest for clarity and relevance, I also made the decision to delete the third and fourth rows from each page of the PDF data. These rows represented the difference in change from one month to the next and the month prior a detail that, while interesting, isn't essential for our current cleaning objectives. It's after the pure data, the raw numbers that speak to the heart of the information we're trying to understand. Anything extra can be calculated through thorough analysis later on.

With the cities listed in order within the dataset, I then did a loop to automate their organization. This step, though technical, was straightforward compared to the challenge that followed which was ensuring the accuracy of the now organized data. My method was hands-on and somewhat traditional manual checks through a random selection of rows post-pipeline, cross-referencing back to the original PDF files. This approach, while labor-intensive, was crucial for identifying and eliminating mistakes, and ensuring the integrity of the dataset.

This exercise was a reminder that even seemingly straightforward tasks can become unexpectedly complex. What I initially thought would be a quick fix turned into a deep dive into the data, underscoring the importance of flexibility and patience in this process.

Looking forward, my plan for next week is to continue refining the dataset by looping through the months and adding the year. It is essential that this step is taken in order to construct a complete and consistent dataset that reflects the data's temporal dimension accurately. Even though the path forward is not always straightforward, I'm moving closer to an organized and fully realized dataset with each step.