**Amjad Altuwayjiri**

**Capstone Project**

**Weekly report #03**

This week, my focus shifted towards defining the features of the data more precisely. I embarked on the task of creating columns—56 in total—to accurately represent each category found within the data. This step was crucial for transforming the raw information into a structured format that’s ready for analysis. The challenge, though, lay in the details by ensuring that each number was correctly placed under its corresponding category.

To achieve precision, I made a bold move by deleting all textual content, both in Arabic and English. This might seem drastic, but it was a calculated decision to eliminate any potential confusion and ensure that numbers landed in their rightful order. Then proceeded to manually add the column names. This approach is time-consuming since it is manual and might be not the most correct way, yet it's essential for laying down a clear, unambiguous foundation for the dataset.

Looking ahead, the plan is to add columns for city, year, and month. Despite the structure of the PDF data, some sort of pattern appears, suggesting a way forward. I’m considering leveraging this pattern by developing a loop that could automate the process of adding these critical contextual details.