**Amjad Altuwayjiri**

**Capstone Project**

**Weekly report #02**

This week, I got involved in the coding part of the project, and methodical work of data transformation. I developed a function that reads PDF files, turning pages into text by specifying the number of page needed, a straightforward yet crucial step. Following this, I crafted a filtering function. Its task was: to remove Arabic texts, split the content into lines, and then remove any lines that were empty or filled with non-text or non-numbers. Clearness and precision in data were the real challenges more than simply writing a code.

The process of grouping these cleaned lines into rows and columns was less about complexity and more about intent. My aim was to align the data with specific categories - year, month, city, and various indexes. This wasn't just a technical exercise; it was a reflection on the essence of data organization. By carefully categorizing this information, I'm not just organizing data; I'm creating a foundation for understanding.

The future plan is to stack the information under their correct category carefully, as they are currently scattered. It is automating the process by following patterns. This isn't merely about sorting; it's a thoughtful process of placing each piece of data where it truly belongs, ensuring that every bit of information is accounted for.

This work is not about the difficulty of the tasks but about their purpose. It’s about setting a stage where data can speak, where numbers and texts transform into insights that can inform decisions, shape understandings, and possibly guide future actions. The aim is to build a dataset not just for the sake of having one but to create a tool that can be used to explore, to question.