Excel File Transformation Without Delimiter

Project Description

This project was initiated based on a problem found on Upwork. The client provided a PNG image of an Excel file containing various types of information within a single column, without any delimiters. The task was to extract and structure five specific types of data: full name, address, zip code, city, and email. However, some cells also contained additional information such as phone numbers.

Solution

To address this problem, the following steps were undertaken:

1. **Data Format Analysis**: Reviewed the format of the provided data.

2. **Dataset Generation**: Generated 50 rows of sample data using ChatGPT to simulate the problem.

3. **Data Import and Transformation**: Imported the dataset into Power Query and performed the necessary transformations.

4. **Custom Function Implementation**: Developed a custom function to accurately extract city and email information, which were the most challenging to separate.

Limitations

1. **Synthetic Data**: The dataset was generated using ChatGPT, resulting in city names predominantly from Germany and Austria.

2. **Error Rate**: The automated process had an error rate of approximately 10%.

3. **Manual Adjustments**: Some manual transformations were required to achieve the desired accuracy.

Conclusion

Despite the challenges and limitations, the data was successfully transformed and structured to meet the client's requirements. Further refinements could reduce the error rate and improve automation.