

Match Branded Food Items to Common Food IDs

Project Description

This project was initiated based on a job found on Upwork. The job was to match two rows of information and return matching values. I used Python's pandas, and regular expression libraries to achieve the desired result.

Solution

To address this problem, the following steps were undertaken among many others:

- 1. Dataset Generation:** Generated 50 rows of sample data for a dataset and 20 rows of sample data for another dataset using ChatGPT to simulate the problem.
- 2. Data Import and Transformation:** Imported the two datasets from an Excel file into two pandas data frame.
- 3. Data Format Analysis:** Reviewed the provided data.
- 4. Function to Match Rows:** Created functions for matching rows of data.
- 5. Save the Matched Rows Data:** Saved the matched data into a CSV file.

Limitations

- 1. Synthetic Data:** As the two datasets were generated using artificial intelligence, the data was synthetic.
- 2. Small Datasets:** It's two very small datasets to only validate the logics.

Files

- 1. TestDataset.xlsx:** Raw file.
- 2. Match_Rows.ipynb:** Python cleaning script.
- 3. branded_foods_with_common_ids.csv:** Matched CSV file generated by the Python script as a form of output.
- 4. Task Description.pdf:** Details job description.
- 5. Project Description.pdf:** Documenting the approach to come an solution.