Project 1: Data Verification and Discrepancy Detection

Project Definition: The objective of this project is to develop an auto Al solution that compares the Information from invoices shared by customer with the data that is present in database (here excel file) and provide the discrepancies in the data.

You are provided with few PDF invoices from customers and an excel file (master data). Develop a solution that compares the information from invoices shared of each item in a PDF invoice file with the corresponding information in an Excel file.

The solution should identify any discrepancies in total prices between the two files and generate an output Excel file containing the identified discrepancies.

Sample Input:

1. Excel file containing data such as:

| Item | Quantity | Unit Price | Total |

|---------|---------------|----------------|----------|

| Apple | 3 | 5.71 | 17.13 |

2. PDF Invoice file containing information such as:

From: Fresh Grocery Store

Address: 123 Market St, Foodville, TX 75001

Phone: (123) 456-7890

Email: contact@freshgrocery.com

To: Alex Smith

Address: 456 Residential Rd, Homestead, FL 33030

| Item | Quantity | Unit Price | Total |

| Apples | 3 | 6.71 | 20.13 |

**Sample Output:**

Excel file containing any discrepancies found:

| Item |Total price in Invoice | Total price in master data | Discrepancies |

| Apple | 5.71 | 6.71 | 1 I

Instructions:

1. Use python programming language and required libraries for reading and processing Excel and PDF files.

2. Extract the required information.

3. Compare the extracted data from the PDF file with the corresponding data in the Excel file.

4. Identify any discrepancies or price differences for each item.

5. Store the discrepancies in a new Excel file with appropriate headers (Item, Field, Excel Value, PDF Value, Discrepancies).