ALTERSENSE.

PDF Data Extraction to Excel

Objective

You are required to write a Python program that extracts structured data from purchase order PDF files and saves the results into an Excel sheet. The program should be reusable for multiple PDFs with the same structure but varying countries.

Details

1. **Input**:
    - A purchase order PDF file.
    - Each PDF may contain data for multiple countries.
    - The structure of the PDF will remain consistent across files.

2. **Data Fields to Extract**:
    - Order No
    - Country
    - Product Description
    - Season
    - Type of Construction
    - No. of Pieces
    - Sales Mode
    - Time of Delivery
    - Invoice Average Price

3. **Output**:
    - A single Excel file containing all extracted data.
    - Each record should be stored as a row.
    - Columns should include the static fields and additional columns for **Time of Delivery** and **Invoice Average Price** per country.
    - Ensure the output format is clean, tabular, and ready for analysis.

4. **Requirements**:
    - Use **Python**.
    - You may use any Python PDF parsing libraries (`pdfplumber`, `PyPDF2`, `camelot`, `tabula`, etc.).
    - Use `openpyxl` or `pandas` for Excel export.
    - Code must be modular, well-documented, and handle edge cases (e.g., missing fields, variable number of countries).
    - Ensure the script can process **multiple PDFs** in the same run.

5. **Deliverables**:
    - A GitHub repository containing:
        - Source code.
        - A README.md explaining:
            - Setup instructions.
            - Libraries used.
            - How to run the script.
            - Example input PDF and output Excel.
    - Well-structured, readable Python code following best practices (PEP8 compliance).
6. **Evaluation Criteria**:
    - Correctness and completeness of data extraction.
    - Code readability, modularity, and documentation.
    - Reusability for future PDFs with the same structure.
    - Proper Excel formatting.
7. **Submission Deadline:**
    - **All completed tasks are to be submitted by Sunday, 28th September 2025, at 11:00 AM.**