📋 **Data & Document Completeness Checker – CSV & PDF Validation Tool**

## Project Overview

Build a **Streamlit web application** to automatically check the completeness of client data and documents. Upload a **CSV file**, a **Rules DOCX**, and multiple **PDFs**, and the app will:

* Validate required **data fields** in the CSV.
* Check for the presence of **required documents** in PDFs.
* Highlight **missing fields and documents** per client.
* Provide **summary statistics** and a **per-client breakdown**.
* Visualize common missing fields for quick analysis.

This tool is ideal for **financial analysts, compliance teams, and data managers** to ensure that all required information and documents are complete.

## Installation Guide

### 1. Prerequisites

* Python 3.8+ installed.
* Terminal/Command Prompt access.

Check Python version:

python --version

### 2. Download Source Code

* Download the project folder (e.g., checker.zip).
* Extract the folder to your desired location.

### 3. Set Up Virtual Environment (Recommended)

cd path/to/project-folder  
python -m venv venv  
# Windows  
venv\Scripts\activate  
# Mac/Linux  
source venv/bin/activate

### 4. Install Required Libraries

pip install --upgrade pip  
pip install streamlit pandas python-docx PyMuPDF

### 5. Run the App

streamlit run app.py

* Open the Local URL shown in terminal (e.g., http://localhost:8501) in your browser.

### 6. Using the App

1. Upload a **CSV file** with client data.
2. Upload a **Rules DOCX** specifying required fields and documents.
3. Upload one or more **PDFs**.
4. View:
   * Summary metrics (total, complete, incomplete cases)
   * Per-client validation results
   * Common missing fields

### 7. Stop the App

* Press **Ctrl + C** in terminal to stop the Streamlit server.

### 8. Optional Tips

* Windows users: If PyMuPDF fails, run:

pip install --upgrade pip setuptools wheel  
pip install PyMuPDF

* Virtual environments keep dependencies isolated from system Python.