


PROJECT : P111 - SCRIPTS AND WEBSITE

```
Collecting pycparser (from cffi>=2.0.0->cryptography>=36.0.0->pdfminer.six==20251230->pdfplumber)
  Downloading pycparser-3.0-py3-none-any.whl.metadata (8.2 kB)
Collecting six>=1.5 (from python-dateutil>=2.8.2->pandas)
  Downloading six-1.17.0-py2.py3-none-any.whl.metadata (1.7 kB)
Downloading pdfplumber-0.11.9-py3-none-any.whl (60 kB)
Downloading pdfminer_six-20251230-py3-none-any.whl (6.6 MB)
   6.6/6.6 MB 67.3 kB/s 0:01:19
Downloading pandas-3.0.0-cp313-cp313-win_amd64.whl (9.7 MB)
   9.7/9.7 MB 149.8 kB/s 0:01:45
Downloading pytesseract-0.3.13-py3-none-any.whl (14 kB)
Downloading pdf2image-1.17.0-py3-none-any.whl (11 kB)
Downloading charset_normalizer-3.4.4-cp313-cp313-win_amd64.whl (107 kB)
Downloading cryptography-46.0.4-cp311-abi3-win_amd64.whl (3.5 MB)
   3.5/3.5 MB 144.4 kB/s 0:00:23
Downloading cffi-2.0.0-cp313-cp313-win_amd64.whl (183 kB)
Downloading numpy-2.4.1-cp313-cp313-win_amd64.whl (12.3 MB)
   12.3/12.3 MB 125.2 kB/s 0:01:50
Downloading packaging-26.0-py3-none-any.whl (74 kB)
Downloading pillow-12.1.0-cp313-cp313-win_amd64.whl (7.0 MB)
   7.0/7.0 MB 117.5 kB/s 0:01:00
Downloading pypdfium2-5.3.0-py3-none-win_amd64.whl (3.1 MB)
   3.1/3.1 MB 127.4 kB/s 0:00:23
Downloading python_dateutil-2.9.0.post0-py2.py3-none-any.whl (229 kB)
Downloading six-1.17.0-py2.py3-none-any.whl (11 kB)
Downloading pycparser-3.0-py3-none-any.whl (48 kB)
Downloading tzdata-2025.3-py2.py3-none-any.whl (348 kB)
Installing collected packages: tzdata, six, pypdfium2, pycparser, Pillow, packaging, numpy, charset-normalizer, python-dateutil, pytesseract, pdf2image, cffi, pandas, cryptography, pdfminer.six, pdfplumber
   2/16 [pypdfium2] WARNING: The script pypdfium2.exe is installed in 'C:\Users\VICTUS\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.13_qbz5n2kfra8p0\LocalCache\local-packages\Python313\Scripts' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
   6/16 [numpy] WARNING: The scripts f2py.exe and numpy-config.exe are installed in 'C:\Users\VICTUS\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.13_qbz5n2kfra8p0\LocalCache\local-packages\Python313\Scripts' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
   7/16 [charset-normalizer] WARNING: The script normalizer.exe is installed in 'C:\Users\VICTUS\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.13_qbz5n2kfra8p0\LocalCache\local-packages\Python313\Scripts' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
   9/16 [pytesseract] WARNING: The script pytesseract.exe is installed in 'C:\Users\VICTUS\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.13_qbz5n2kfra8p0\LocalCache\local-packages\Python313\Scripts' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
  15/16 [pdfplumber] WARNING: The script pdfplumber.exe is installed in 'C:\Users\VICTUS\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.13_qbz5n2kfra8p0\LocalCache\local-packages\Python313\Scripts' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress this warning, use --no-warn-script-location.
```



We began by setting up the Python environment and installing required libraries such as pdfplumber for PDF text extraction, pandas for data handling, and re for pattern matching.

 New folder

01-02-2026 06:03

File folder












Name	Date modified	Type	Size	
 pdfs	30-01-2026 08:35	File folder		
 project	01-02-2026 06:03	Python Source File	4 KB	




THIS IS THE CODE WHICH WE HAVE WRITTEN FOR AUTOMATION OF THE PDFS THAT ARE PRESENT IN THE FOLDER:pdfs



  Start backup > Desktop > New folder > pdfs

      Sort >  View > ...

Name	Date modified	Type	Size	
 sample_scattered_data	30-01-2026 08:12	Chrome HTML Do...	2 KB	



Sample Data Records (Scattered Format)

Name: Rahul Sharma | Age: 21 | City: Chennai | Email: rahul21@email.com

Customer Priya Menon lives in Bangalore and she is 19 years old. Contact: priya.m@email.com

Record -> Name = Arjun Rao ; Age = 23 ; Location = Hyderabad ; Mail ID = arjunrao@email.com

Sneha, aged 20, resides at Kochi. Reach her at sneha20@email.com

User Details: Vikram Singh / 22 years / Mumbai / vikram.s@email.com

THIS IS A SAMPLE
SCATTERED PDF THAT
WE MADE TO CHECK
OUR CODE

```
Found 1 PDF file(s). Processing...  
Processing: sample_scattered_data.pdf  
Extracted 5 record(s)  
Extraction complete.  
Output saved as: final_output.csv  
PS C:\Users\VICTUS\Desktop\New folder>
```

THIS WAS RUN IN VS
CODE

Name	Date modified	Type	Size
pdfs	30-01-2026 08:35	File folder	
final_output	01-02-2026 06:24	Microsoft Excel Co...	1 KB
project	01-02-2026 06:03	Python Source File	4 KB

THE OUTPUT CSV FILE
AUTOMATICALLY
ADDED INTO THE
FOLDER

THE CSV OUTPUT

H4								
	A	B	C	D	E	F	G	
1	Name	Age	City	Email				
2	Rahul Sharma	21	Chennai	rahul21@email.com				
3	Sneha	20	Kochi	sneha20@email.com				
4	Vikram Singh	22	Mumbai	vikram.s@email.com				
5	Arjun Rao	23	Hyderabad	arjunrao@email.com				
6	Priya Menon	19	Bangalore	priya.m@email.com				
7								
8								
9								
10								

< >

final_output

+