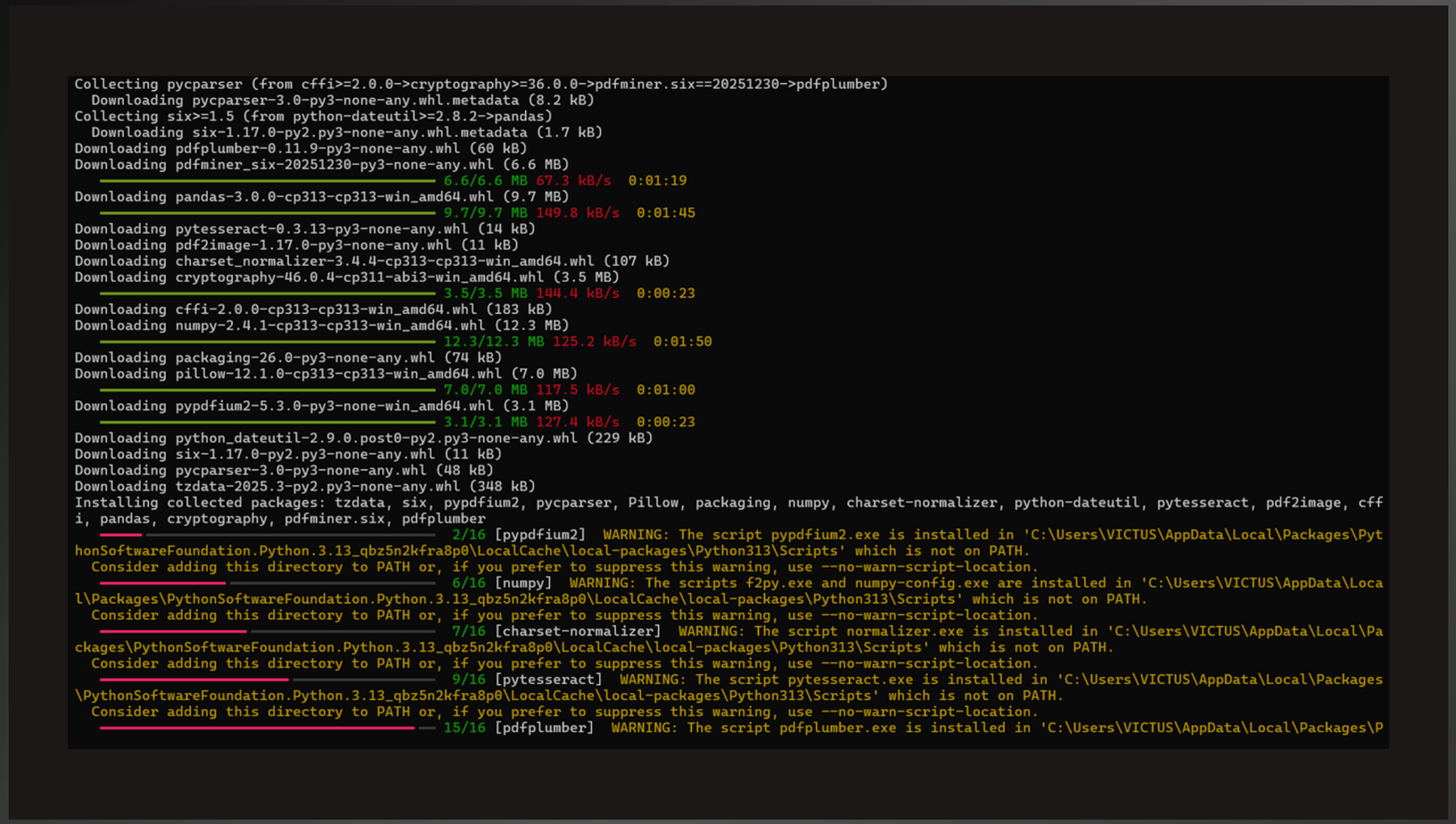
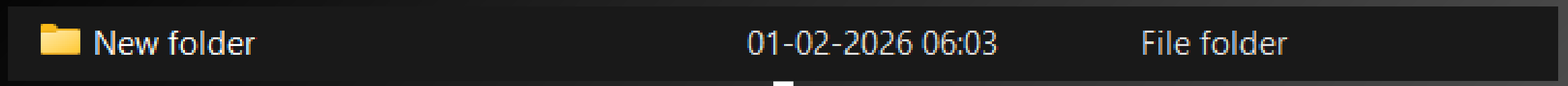
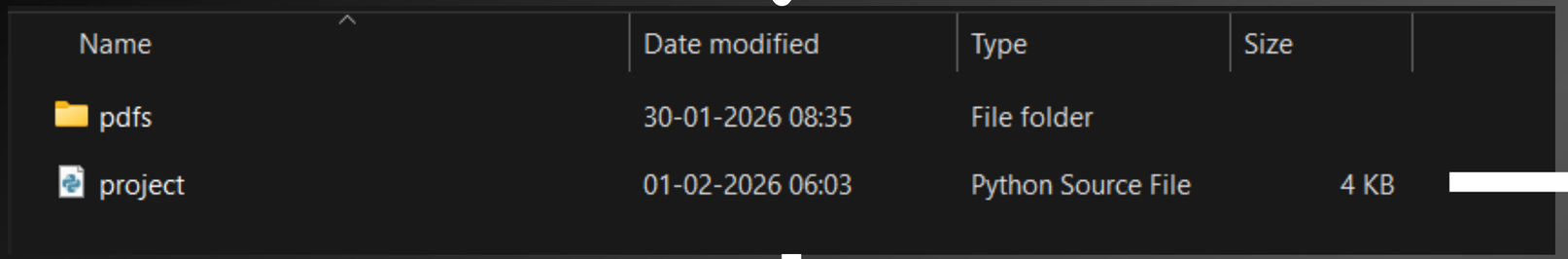
PROJECT : P111 - SCRIPTS AND WEBSITE

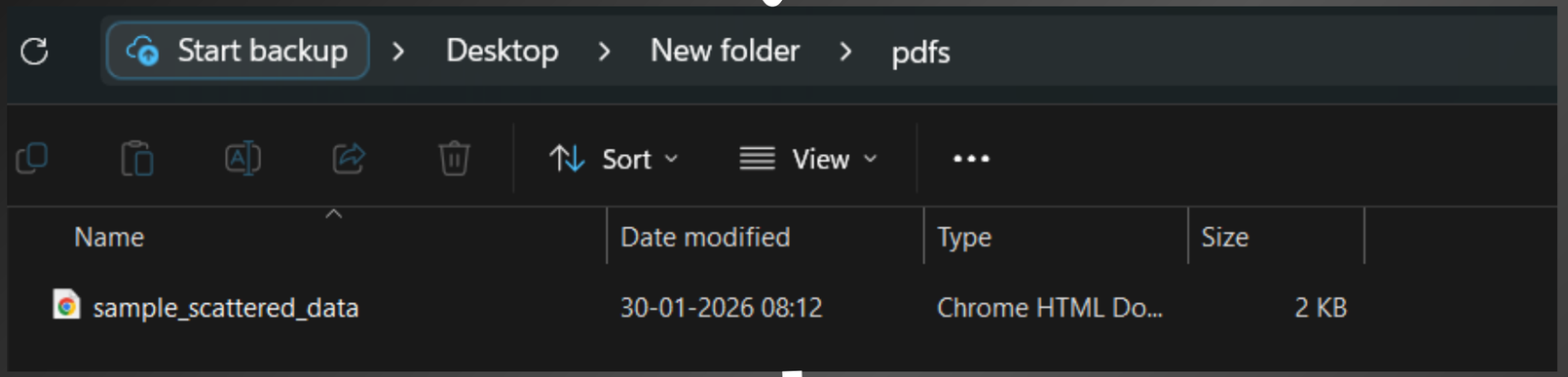


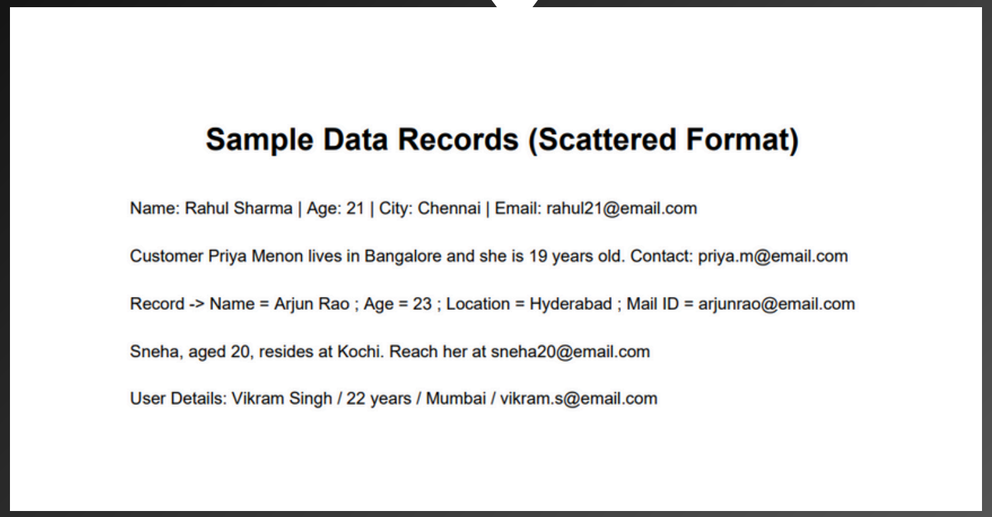
We began by setting up the Python environment and installing required libraries such as pdfplumber for PDF text extraction, pandas for data handling, and re for pattern matching.





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





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

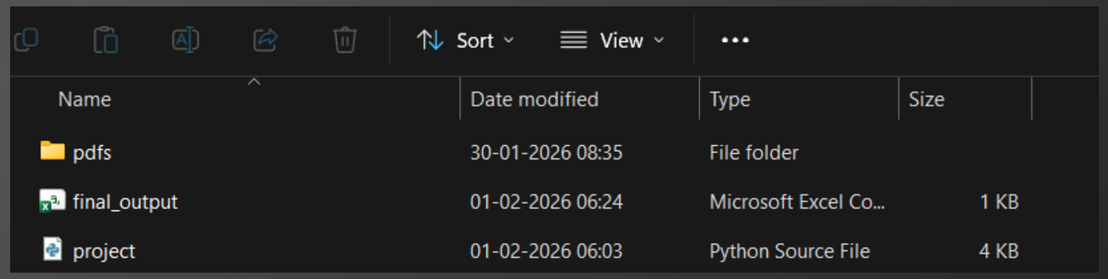
Sneha, aged 20, resides at Kochi. Reach her at sneha20@email.com

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

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

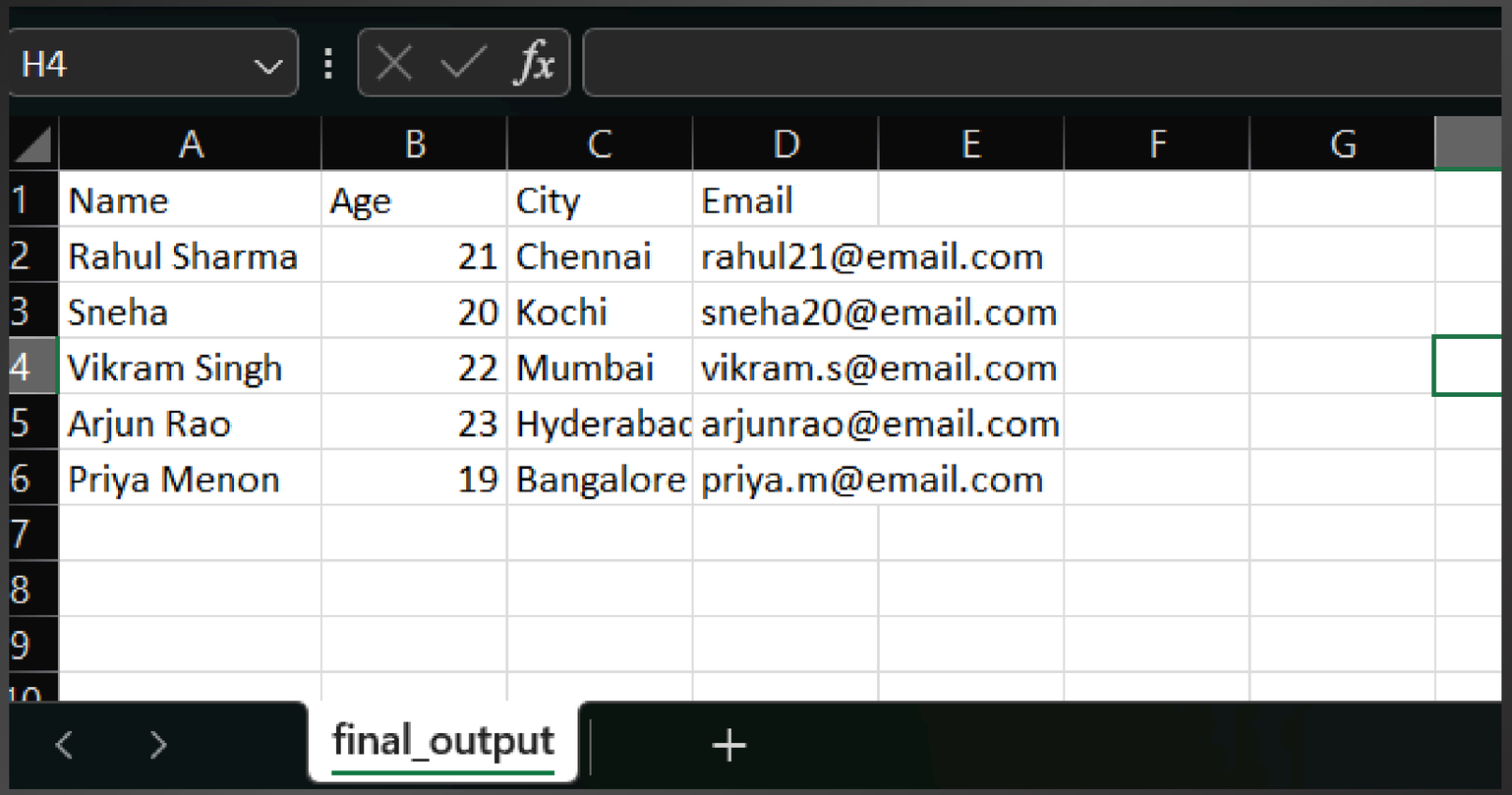
```
Found 1 PDF file(s). Processing...
Processing: sample_scattered_data.pdf
Extracted 5 record(s)
Extraction complete.
Output saved as: final_output.csv
PS C:\Users\VICTUS\Desktop\New folder>
```

THIS WAS RUN IN VS CODE



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	Hyderabac	arjunrao@email.com			
6	Priya Menon	19	Bangalore	priya.m@email.com			
7							
8							
9							

final_output