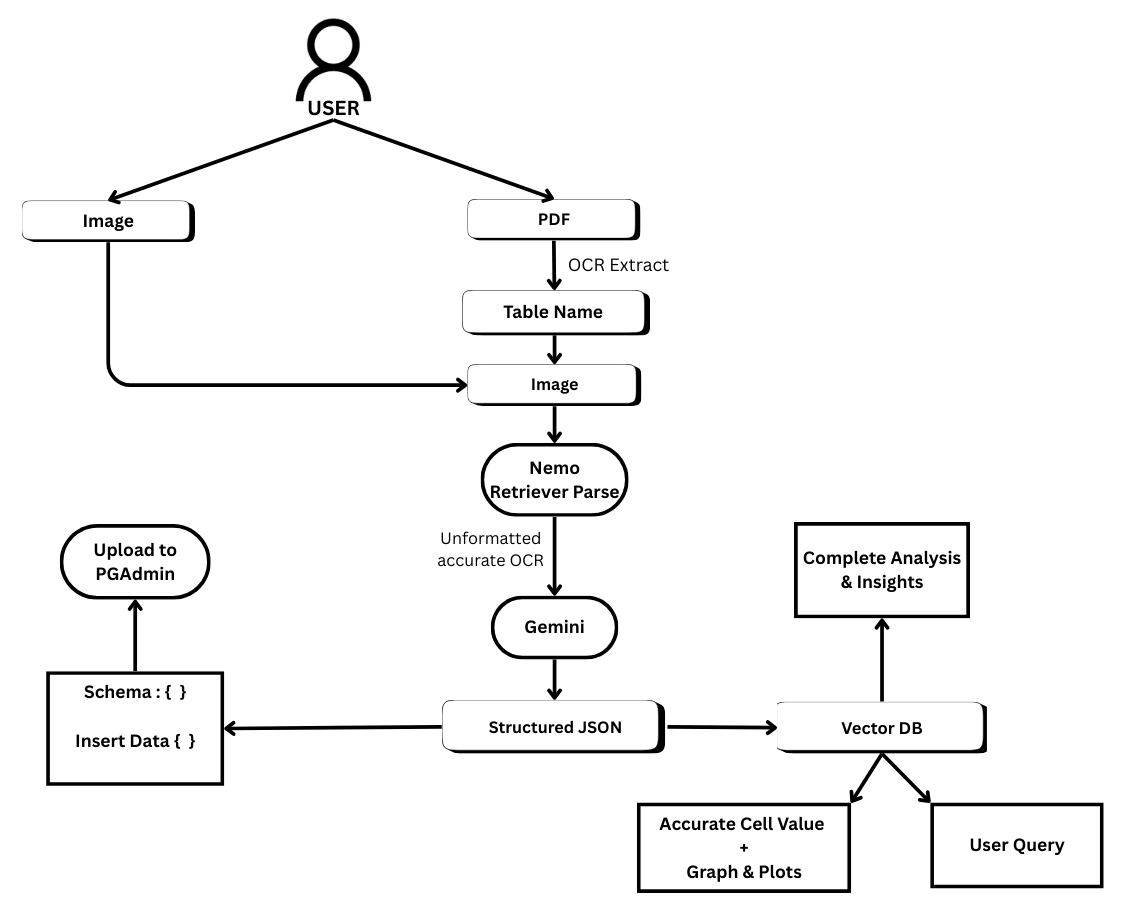
# **LEGACY DATA EXTRACTION**

1. **User Guide**
   1. **HW requirements:** Nil
   2. **SW requirements(with versions):**
      1. Python
      2. Flask
      3. HTML
      4. CSS
      5. JS
   3. **Flow diagram:**



* 1. **Python libraries used:**
     1. faiss
     2. numpy
     3. openai
     4. flask
     5. final\_image
     6. werkzeug.utils
     7. fitz
     8. pytesseract
     9. langchain
     10. PIL
     11. Psychopg2
  2. **Environment details:**
     1. VS Code
  3. **APIs used with cost:**
     1. Openai- $2 per million tokens
     2. Nvidia nim
     3. Gemini
  4. **Test cases along with execution time:**
     1. Scanned pdf of the “National Sample Survey” April-June 1951(60 pages)- ~5 mins
     2. Textual pdf (10 pages)- ~ 1 min
  5. **List of files:**
  6. **Steps to execute:**
     1. Setup the environment(Eg: Python installed)
     2. Setup the environment variables(Add necessary API keys and database credentials inside .env)
     3. Run the flask file
     4. Open the web application
     5. Upload the PDF and query data
  7. **Screenshots of execution:**
  8. **Contact details of team members:**
     1. Name: Kuhan M

E-mail ID: 22d126@psgitech.ac.in

Mobile No: 7845036121

Institute: PSG Institute of Technology and Applied Research

Course: B.Tech. Artificial Intelligence and Data Science

Year of passing: 2026

Contribution of team member: Backend/AI

* + 1. Name: S. Gautham Prasath

E-mail ID: 23d112@psgitech.ac.in

Mobile No:8287072540

Institute: PSG Institute of Technology and Applied Research

Course: B.Tech. Artificial Intelligence and Data Science

Year of passing: 2027

Contribution of team member: Backend/AI

* + 1. Name: Niranjana

E-mail ID: 23d133@psgitech.ac.in

Mobile No: 9047617017

Institute: PSG Institute of Technology and Applied Research

Course: B.Tech. Artificial Intelligence and Data Science

Year of passing: 2027

Contribution of team member: Frontend

* + 1. Name: Risheekesh K.G.

E-mail ID: 23d140@psgitech.ac.in

Mobile No:9384762478

Institute: PSG Institute of Technology and Applied Research

Course: B.Tech. Artificial Intelligence and Data Science

Year of passing: 2027

Contribution of team member: Backend/Frontend

* + 1. Name: Saisree Anusha B

E-mail ID: 23d146@psgitech.ac.in

Mobile No:8825910880

Institute: PSG Institute of Technology and Applied Research

Course: B.Tech. Artificial Intelligence and Data Science

Year of passing: 2027

Contribution of team member: Backend/Database

1. **Source code**