

Project Documentation: PDF to JSON Extractor using Streamlit

1. Project Title

PDF Table Extractor and JSON Converter

2. Objective

The objective of this project is to extract tabular data from PDF files, display it on a web-based interface, and convert it into JSON format.

This process simplifies the handling of structured datasets stored in PDFs, making them machine-readable and usable for further applications like data analysis, visualization, and machine learning.

3. Technologies Used

- **Python** – Core programming language.
 - **Streamlit** – For building the frontend interface.
 - **pdfplumber** – To read and extract tabular data from PDF documents.
 - **pandas** – To structure and clean tabular data into DataFrames.
 - **json** – To convert processed data into JSON format.
-

4. Code Implementation

```
import streamlit as st

import pdfplumber

import pandas as pd

import json

# File uploader

uploaded_file = st.file_uploader("Upload a Dataset PDF", type=["pdf"])

if uploaded_file is not None:

    with pdfplumber.open(uploaded_file) as pdf:

        table = pdf.pages[0].extract_table()

        if table:
```

NAME-KASA SATYA SAMPATH KUMAR

```
# Convert table into dataframe  
df = pd.DataFrame(table[1:], columns=table[0])  
  
# Clean dataframe  
df = df.dropna(axis=1, how="all")  
  
df = df.loc[:, df.columns.notna()]  
  
df = df.loc[:, df.columns.str.strip() != ""]  
  
df = df.loc[:, ~df.columns.duplicated()]  
  
# Display dataframe  
st.write(df)  
  
# Convert to JSON  
json_string = json.dumps(df.to_dict(orient="records"), indent=4)  
st.code(json_string, language="json")  
  
# Save JSON to file  
with open("output.json", "w", encoding="utf-8") as f:  
    f.write(json_string)
```

5. Working Procedure

1. The user uploads a PDF file (dataset.pdf) containing a structured table.
 2. The **first page** of the PDF is processed using the pdfplumber library.
 3. The extracted table is converted into a **pandas DataFrame**.
 4. Data cleaning is performed by:
 - o Dropping empty columns.
 - o Removing blank headers.
 - o Removing duplicate column names.
 5. The clean dataset is displayed in a **Streamlit table**.
 6. The DataFrame is converted into **JSON format**.
 7. The JSON output is displayed on the frontend and saved locally as **output.json**.
-

6. Screenshots

ID	Category	Sub-Category	Entity Name	Address	Postcode	Local Authority	Region	Policy Area	Relevant Policy/Legislation	Policy Summary	Status	Funding (GBP Million)	URL
1	Healthcare	Hospital	St Thomas' Hospital	Westminster Bridge Rd, London	SE1 7EH	Lambeth	London	Health & Social Care	Health and Care Act 2022	Aims to improve the integration of health and social care services, patient choice and academic freedom in higher education.	Operational	1500	https://www.guysandstthomas.nhs.uk/
2	Education	University	University of Manchester	Oxford Rd, Manchester	M13 9PL	Manchester	North West	Higher Education	Higher Education (Freedom of Speech) Act 2023	Outlines the government's vision for science and research.	Operational	1200	https://www.manchester.ac.uk/
3	Business	Aviation	Rolls-Royce Holdings plc	62 Buckingham Gate, London	SW1E 6AT	Westminster	London	Industrial Strategy	NHS Research and Development Roadmap	Sets out the government's planning policies for health and social care.	Operational	1000	https://www.gov.uk/gov/coronavirus-covid-19
4	Residential	Housing Estate	Barbican Estate	Six St. London	EC2Y 8QS	City of London	London	Housing & Planning	National Planning Policy Framework	Outlines the government's vision for planning and development.	Completed	N/A	https://www.gov.uk/gov/barbican-planning-policy
5	Healthcare	Hospital	Addenbrooke's Hospital	I Hills Rd, Cambridge	CB2 2QQ	Cambridge	East of England	Health & Social Care	NHS Long Term Plan	Sets out the priorities for the NHS for the next 10 years.	Operational	900	https://www.gov.uk/gov/nhs-long-term-plan
6	Education	College	Birmingham Metropolitan College	Jamaine Road, Birmingham	B4 1PP	Birmingham	West Midlands	Further Education	Skills and Post-16 Education Act 2022	Aims to improve the quality of post-16 education and apprenticeships.	Operational	70	https://www.bmc.ac.uk/
7	Business	Defense	BAE Systems	Stirling Square, London	SW1Y 8HQ	Westminster	London	Defence & Security	National Security Strategy	Sets out the government's approach to protecting the UK's national security interests.	Operational	N/A	https://www.bae-systems.com
8	Residential	New Town	Poundbury	Poundbury, Dorset	DT13 9QE	Dorset	South West	Urban Planning	Leveling Up and Regeneration Act 2023	Outlines the government's approach to providing better infrastructure and opportunities across the UK.	Under Development	N/A	https://www.gov.uk/gov/poundbury
9	Healthcare	Hospital	Queen Elizabeth University Hospital	1045 Great Orme Rd, Glasgow	G51 4TF	Glasgow City	Scotland	Health & Social Care	NHS Recovery Plan (Scotland)	Sets out the Scottish Government's plan for health and social care.	Operational	840	https://www.gov.scot/nhs-recovery-plan
10	Education	University	Cardiff University	Park Pl, Cardiff	CF10 3AT	Cardiff	Wales	Higher Education	Tertiary Education and Research (Wales) Act 2022	Establishes a new framework for tertiary education in Wales.	Operational	500	https://www.gov.wales/tertiary-education-and-research-wales-act-2022
11	Business	Pharmaceuticals	AstraZeneca plc	1 Francis Crick Avenue, Cambridge	CB6 5UA	Cambridge	East of England	Life Sciences	Life Sciences Vision	Sets out a 10-year strategy for the UK's science and innovation base.	Operational	N/A	https://www.gov.uk/gov/astrazeneca-plc
12	Residential	Regeneration	Homelessness Registration	Hume, Manchester	M15	Manchester	North West	Urban Regeneration	The Northern Powerhouse Strategy	Aims to boost economic growth in the North of England through investment in adult social care in rural areas.	Completed	140	https://www.gov.uk/gov/northern-powerhouse-strategy
13	Healthcare	Hospital	Royal Victoria Infirmary	Queen Victoria Rd, Newcastle upon Tyne	NE1 4LP	Newcastle upon Tyne	North East	Health & Social Care	People at the Heart of Care, adult social care reform	Aims to put patients at the heart of the service delivery.	Operational	600	https://www.gov.uk/gov/royal-victoria-infirmary-newcastle-upon-tyne
14	Education	College	Exeter College	Exeter House, London	EX4 4BY	Exeter	South West	Further Education	The UK's Green Industrial Revolution	An action plan for a green industrial revolution.	Operational	N/A	https://www.gov.uk/gov/exeter-college
15	Business	Consumer Goods	GlaxoSmithKline (GSK)	888 Great West Rd, Brentford	TW8 9GS	Brentwood	London	Pharmaceuticals	UK Biopharma Manufacturing Vision	A strategic plan for the UK's biopharma industry.	Under Development	N/A	https://www.gov.uk/gov/gsk
16	Residential	Housing Development	Kilburne Village	Kilburne, London	SE3	Grenfell	London	Housing Development	The London Plan	The spatial development strategy for the Greater London area.	Under Development	1000	https://www.gov.uk/gov/kilburne-village
17	Healthcare	Hospital	Belfast City Hospital	Linen Hall Rd, Belfast	B75 7AB	Belfast	Northern Ireland	Health & Social Care	Health and Wellbeing 2030: Delivering Together	Sets out the strategic direction for health and social care in Northern Ireland.	Operational	180	https://www.gov.uk/gov/belfast-city-hospital
18	Education	Academy	University of Edinburgh	Edith College, South Bridge, Edinburgh	EH8 9PL	Edinburgh	Scotland	Higher Education	The Promise (Scotland)	Aims to improve the lives of children and young people in Scotland.	Operational	1300	https://www.gov.uk/gov/university-of-edinburgh
19	Business	Automotive	Jaguar Land Rover	Abbey Rd, Whitley, Coventry	<td>Coventry</td> <td>West Midlands</td> <td>Automotive Sector</td> <td>The Race to Zero</td> <td>A strategy for zero-emission road transport.</td> <td>Operational</td> <td>N/A</td> <td>https://www.gov.uk/gov/jaguar-land-rover</td>	Coventry	West Midlands	Automotive Sector	The Race to Zero	A strategy for zero-emission road transport.	Operational	N/A	https://www.gov.uk/gov/jaguar-land-rover
20	Residential	Heritage	Grange Town, Newcastle	Newcastle upon Tyne	NE1	Newcastle upon Tyne	North East	Heritage & Conservation	National Heritage Act 1980	Makes provision for the protection of ancient monuments and buildings.	Completed	40	https://www.gov.uk/gov/national-heritage-act-1980
21	Healthcare	Cancer Centre	The Christie NHS Foundation Trust	Newcastle Rd, Manchester	M20 4BX	Manchester	North West	Cancer Care	NHS Cancer Programme	Aims to improve cancer outcomes and services in England.	Operational	300	https://www.gov.uk/gov/the-christie-nhs-foundation-trust
22	Education	College	Leeds City College	Park Ln, Leeds	LS1 1AA	Leeds	Yorkshire and the Humber	Further Education	The UK's Green Industrial Revolution	Aims to put employers at the heart of the skills system.	Operational	70	https://www.gov.uk/gov/leeds-city-college
23	Business	Pharmaceuticals	GlassSmithDrive (GSD)	888 Great West Rd, Brentford	TW8 9GS	Brentwood	London	Pharmaceuticals	UK Biopharma Manufacturing Vision	Aims to make the UK the best place in the world to do business.	Operational	N/A	https://www.gov.uk/gov/glassmithdrive-gsd
24	Residential	Social Housing	Park Hill flats, Sheffield	Park Hill, Sheffield	S3	Sheffield	Yorkshire and the Humber	Social Housing (Regulation) Act 2023	Aims to improve the regulation of social housing.	Completed	100	https://www.gov.uk/gov/park-hill-flats-sheffield	
25	Healthcare	Children's Hospital	Great Ormond Street Hospital	Great Ormond St, London	WC1N 3JH	London	Children's Health	Children and Young People's Health Policy	Aims to improve the health and well-being of children and young people.	Operational	400	https://www.gov.uk/gov/great-ormond-street-hospital	
26	Education	University	Imperial College London	Exhibition Rd, London	SW7 2AZ	Kensington and Chelsea	London	Science & Technology	UK Innovation Strategy	Makes provision for the protection of ancient monuments and buildings.	Operational	1100	https://www.gov.uk/gov/imperial-college-london
27	Business	Retail	Tesco PLC	Tesco House, Shire Park, Kent	A17 1GA	Watton	Midlands	Retail	The UK's Food Strategy	Sets out the government's vision for a sustainable food system.	Operational	N/A	https://www.gov.uk/gov/tesco-plc
28	Residential	Regeneration	New Brighton, Manchester	Manchester	M1	Manchester	North West	Urban Regeneration	The Estates Gazette Regeneration Report	Proposes changes into the UK's most active regeneration locations.	Completed	300	https://www.gov.uk/gov/new-brighton-manchester
29	Healthcare	Cancer Centre	The Royal Marsden NHS Foundation T 223 Fulham Rd, London	SW3 8JJ	Kensington and Chelsea	London	Cancer Research	Dame Tessa Jowell Brain Cancer Mission	Aims to put patients at the heart of cancer research and care in the UK.	Operational	200	https://www.gov.uk/gov/royal-marsden-nhs-foundation-trust	
30	Education	University	Queen's University Belfast	University Rd, Belfast	B77 1NN	Belfast	Northern Ireland	Higher Education	The Higher Education Strategy for Northern Ireland	Recommends the strategic direction for higher education in Northern Ireland.	Operational	400	https://www.gov.uk/gov/queens-university-belfast
31	Business	Technology	DeepMind	8 Pancras Square, London	NC1 4AG	Camden	London	Artificial Intelligence	National Strategy	Recommends the strategic direction for artificial intelligence in the UK.	Operational	N/A	https://www.gov.uk/gov/deepmind
32	Residential	Garden Village	Elstree Garden City	Elstree, Hertfordshire	DA10 7AA	Dartford and Gravesham	South East	Housing & Planning	Garden Communities Programme	A programme to support the development of new garden towns and villages.	Under Development	200	https://www.gov.uk/gov/elstree-garden-city
33	Healthcare	Mental Health	The Phoenix Hospital	Rushmoor	SW15 5JJ	Wandsworth	London	Mental Health	NHS Mental Health Implementation Plan	Improves mental health services across the UK.	Operational	50	https://www.gov.uk/gov/phoenix-hospital

```
streamlit > 🐄 pdf_text.py > ...
1  import streamlit as st
2  import pdfplumber
3  import pandas as pd
4  import json
5  uploaded_file = st.file_uploader("Upload a Dataset PDF", type=["pdf"])
6  if uploaded_file is not None:
7      with pdfplumber.open(uploaded_file) as pdf:
8          table = pdf.pages[0].extract_table()
9          if table:
10              df = pd.DataFrame(table[1:], columns=table[0])
11              df = df.dropna(axis=1, how="all")
12              df = df.loc[:, df.columns.notna()]
13              df = df.loc[:, df.columns.str.strip() != ""]
14              df = df.loc[:, ~df.columns.duplicated()]
15              st.write(df)
16              json_string = json.dumps(df.to_dict(orient="records"), indent=4)
17              st.code(json_string, language="json")
18              with open("output.json", "w", encoding="utf-8") as f:
19                  f.write(json_string)
```

8. Sample Output

```
[{
    "ID": "1",
    "Category": "Healthcare",
    "Sub-Category": "Hospital",
    "Entity Name": "St Thomas' Hospital",
    "Address": "Westminster Bridge Rd, London",
    "Postcode": "SE1 7EH",
    "Local Authority": "Lambeth",
    "Region": "London",
    "Policy Area": "Health & Social Care",
    "Relevant Policy/Legislation": "Health and Care Act 2022",
    "Policy Summary": "Aims to improve the integration of health and social",
    "Status": "Operational"
    "Funding (GBP Million)": "1500",
    "URL": "https://www.guysandstthomas.nhs.uk/"
}, {
    "ID": "2",
    "Category": "Education",
    "Sub-Category": "University",
    "Entity Name": "University of Manchester",
    "Address": "Oxford Rd, Manchester",
    "Postcode": "M13 9PL",
    "Local Authority": "Manchester",
    "Region": "North West",
    "Policy Area": "Higher Education",
    "Relevant Policy/Legislation": "Higher Education (Freedom of Speech) Act 2023",
    "Policy Summary": "Sets out the government's approach to protecting the UK's freedom of speech and academic freedom in higher education.",
    "Status": "Operational"
    "Funding (GBP Million)": "1200",
    "URL": "https://www.manchester.ac.uk/"
}, {
    "ID": "3",
    "Category": "Business",
    "Sub-Category": "Aviation",
    "Entity Name": "Rolls-Royce Holdings plc",
    "Address": "62 Buckingham Gate, London",
    "Postcode": "SW1E 6AT",
    "Local Authority": "Westminster",
    "Region": "London",
    "Policy Area": "Industrial Strategy",
    "Relevant Policy/Legislation": "NHS Research and Development Roadmap",
    "Policy Summary": "Outlines the government's vision for science and innovation in the NHS and the NHS Research and Development Roadmap.",
    "Status": "Operational"
    "Funding (GBP Million)": "1000",
    "URL": "https://www.gov.uk/gov/nhs-research-and-development-roadmap"
}, {
    "ID": "4",
    "Category": "Residential",
    "Sub-Category": "Housing Estate",
    "Entity Name": "Barbican Estate",
    "Address": "Six St, London",
    "Postcode": "EC2Y 8QS",
    "Local Authority": "City of London",
    "Region": "London",
    "Policy Area": "Housing & Planning",
    "Relevant Policy/Legislation": "National Planning Policy Framework",
    "Policy Summary": "Sets out the government's planning policies for housing and planning in England, including the National Planning Policy Framework (NPPF).",
    "Status": "Operational"
    "Funding (GBP Million)": "900",
    "URL": "https://www.gov.uk/gov/national-planning-policy-framework"
}, {
    "ID": "5",
    "Category": "Healthcare",
    "Sub-Category": "Hospital",
    "Entity Name": "Addenbrooke's Hospital",
    "Address": "I Hills Rd, Cambridge",
    "Postcode": "CB2 2QQ",
    "Local Authority": "Cambridge",
    "Region": "East of England",
    "Policy Area": "Health & Social Care",
    "Relevant Policy/Legislation": "NHS Long Term Plan",
    "Policy Summary": "Sets out the priorities for the NHS for the next 10 years, including the NHS Long Term Plan (LTP).",
    "Status": "Operational"
    "Funding (GBP Million)": "900",
    "URL": "https://www.gov.uk/gov/nhs-long-term-plan"
}, {
    "ID": "6",
    "Category": "Education",
    "Sub-Category": "College",
    "Entity Name": "Exeter College",
    "Address": "Exeter House, London",
    "Postcode": "EX4 4BY",
    "Local Authority": "Exeter",
    "Region": "South West",
    "Policy Area": "Further Education",
    "Relevant Policy/Legislation": "The UK's Green Industrial Revolution",
    "Policy Summary": "An action plan for a green industrial revolution, including the UK's Green Industrial Revolution (GIR) and the Local Skills Improvement Plan (LSIP).",
    "Status": "Operational"
    "Funding (GBP Million)": "600",
    "URL": "https://www.gov.uk/gov/exeter-college"
}, {
    "ID": "7",
    "Category": "Business",
    "Sub-Category": "Manufacturing Goods",
    "Entity Name": "GlassSmithDrive (GSD)",
    "Address": "888 Great West Rd, Brentford",
    "Postcode": "TW8 9GS",
    "Local Authority": "Brentwood",
    "Region": "London",
    "Policy Area": "Pharmaceuticals",
    "Relevant Policy/Legislation": "UK Biopharma Manufacturing Vision",
    "Policy Summary": "An action plan for the UK's biopharma industry, including the UK Biopharma Manufacturing Vision (BMV) and the Local Skills Improvement Plan (LSIP).",
    "Status": "Under Development"
    "Funding (GBP Million)": "N/A",
    "URL": "https://www.gov.uk/gov/glassmithdrive-gsd"
}, {
    "ID": "8",
    "Category": "Residential",
    "Sub-Category": "Housing Development",
    "Entity Name": "Kilburne Village",
    "Address": "Kilburne, London",
    "Postcode": "SE3",
    "Local Authority": "Grenfell",
    "Region": "London",
    "Policy Area": "Housing Development",
    "Relevant Policy/Legislation": "The London Plan",
    "Policy Summary": "The spatial development strategy for the Greater London area, including the London Plan (LP).",
    "Status": "Under Development"
    "Funding (GBP Million)": "N/A",
    "URL": "https://www.gov.uk/gov/kilburne-village"
}, {
    "ID": "9",
    "Category": "Healthcare",
    "Sub-Category": "Hospital",
    "Entity Name": "Belfast City Hospital",
    "Address": "Linen Hall Rd, Belfast",
    "Postcode": "B75 7AB",
    "Local Authority": "Belfast",
    "Region": "Northern Ireland",
    "Policy Area": "Health & Social Care",
    "Relevant Policy/Legislation": "Health and Wellbeing 2030: Delivering Together",
    "Policy Summary": "Sets out the strategic direction for health and social care in Northern Ireland, including the Health and Wellbeing 2030: Delivering Together (HWT) and the Northern Ireland Health and Social Care Strategy (NIHSS).",
    "Status": "Operational"
    "Funding (GBP Million)": "180",
    "URL": "https://www.gov.uk/gov/belfast-city-hospital"
}, {
    "ID": "10",
    "Category": "Education",
    "Sub-Category": "Academy",
    "Entity Name": "University of Edinburgh",
    "Address": "Edith College, South Bridge, Edinburgh",
    "Postcode": "EH8 9PL",
    "Local Authority": "Edinburgh",
    "Region": "Scotland",
    "Policy Area": "Higher Education",
    "Relevant Policy/Legislation": "The Promise (Scotland)",
    "Policy Summary": "Aims to improve the health and well-being of children and young people in Scotland, including the Promise (Scotland) and the National Health and Care Strategy (NHCS).",
    "Status": "Operational"
    "Funding (GBP Million)": "1300",
    "URL": "https://www.gov.uk/gov/university-of-edinburgh"
}, {
    "ID": "11",
    "Category": "Business",
    "Sub-Category": "Automotive",
    "Entity Name": "Jaguar Land Rover",
    "Address": "Abbey Rd, Whitley, Coventry",
    "Postcode": "CV4 4LF",
    "Local Authority": "Coventry",
    "Region": "West Midlands",
    "Policy Area": "Automotive Sector",
    "Relevant Policy/Legislation": "The Race to Zero",
    "Policy Summary": "A strategy for zero-emission road transport, including the Race to Zero and the National Transport Strategy (NTS).",
    "Status": "Operational"
    "Funding (GBP Million)": "400",
    "URL": "https://www.gov.uk/gov/jaguar-land-rover"
}, {
    "ID": "12",
    "Category": "Residential",
    "Sub-Category": "Heritage",
    "Entity Name": "Grange Town, Newcastle",
    "Address": "Newcastle upon Tyne",
    "Postcode": "NE1",
    "Local Authority": "Newcastle upon Tyne",
    "Region": "North East",
    "Policy Area": "Heritage & Conservation",
    "Relevant Policy/Legislation": "National Heritage Act 1980",
    "Policy Summary": "Sets out the government's plan to make the UK a global leader in heritage protection, including the National Heritage Act 1980 and the National Strategy for Heritage (NSH).",
    "Status": "Completed"
    "Funding (GBP Million)": "40",
    "URL": "https://www.gov.uk/gov/grange-town-newcastle"
}, {
    "ID": "13",
    "Category": "Healthcare",
    "Sub-Category": "Cancer Centre",
    "Entity Name": "The Christie NHS Foundation Trust",
    "Address": "Newcastle Rd, Manchester",
    "Postcode": "M20 4BX",
    "Local Authority": "Manchester",
    "Region": "North West",
    "Policy Area": "Cancer Care",
    "Relevant Policy/Legislation": "NHS Cancer Programme",
    "Policy Summary": "Aims to improve cancer outcomes and services in England, including the NHS Cancer Programme (NCP) and the National Cancer Strategy (NCS).",
    "Status": "Operational"
    "Funding (GBP Million)": "300",
    "URL": "https://www.gov.uk/gov/the-christie-nhs-foundation-trust"
}, {
    "ID": "14",
    "Category": "Education",
    "Sub-Category": "College",
    "Entity Name": "Leeds City College",
    "Address": "Park Ln, Leeds",
    "Postcode": "LS1 1AA",
    "Local Authority": "Leeds",
    "Region": "Yorkshire and the Humber",
    "Policy Area": "Further Education",
    "Relevant Policy/Legislation": "The UK's Green Industrial Revolution",
    "Policy Summary": "Aims to put employers at the heart of the skills system, including the UK's Green Industrial Revolution (GIR) and the Local Skills Improvement Plan (LSIP).",
    "Status": "Operational"
    "Funding (GBP Million)": "70",
    "URL": "https://www.gov.uk/gov/leeds-city-college"
}, {
    "ID": "15",
    "Category": "Business",
    "Sub-Category": "Manufacturing Goods",
    "Entity Name": "GlassSmithDrive (GSD)",
    "Address": "888 Great West Rd, Brentford",
    "Postcode": "TW8 9GS",
    "Local Authority": "Brentwood",
    "Region": "London",
    "Policy Area": "Pharmaceuticals",
    "Relevant Policy/Legislation": "UK Biopharma Manufacturing Vision",
    "Policy Summary": "An action plan for the UK's biopharma industry, including the UK Biopharma Manufacturing Vision (BMV) and the Local Skills Improvement Plan (LSIP).",
    "Status": "Under Development"
    "Funding (GBP Million)": "N/A",
    "URL": "https://www.gov.uk/gov/glassmithdrive-gsd"
}, {
    "ID": "16",
    "Category": "Residential",
    "Sub-Category": "Housing Development",
    "Entity Name": "Kilburne Village",
    "Address": "Kilburne, London",
    "Postcode": "SE3",
    "Local Authority": "Grenfell",
    "Region": "London",
    "Policy Area": "Housing Development",
    "Relevant Policy/Legislation": "The London Plan",
    "Policy Summary": "The spatial development strategy for the Greater London area, including the London Plan (LP).",
    "Status": "Under Development"
    "Funding (GBP Million)": "N/A",
    "URL": "https://www.gov.uk/gov/kilburne-village"
}, {
    "ID": "17",
    "Category": "Healthcare",
    "Sub-Category": "Hospital",
    "Entity Name": "Belfast City Hospital",
    "Address": "Linen Hall Rd, Belfast",
    "Postcode": "B75 7AB",
    "Local Authority": "Belfast",
    "Region": "Northern Ireland",
    "Policy Area": "Health & Social Care",
    "Relevant Policy/Legislation": "Health and Wellbeing 2030: Delivering Together",
    "Policy Summary": "Sets out the strategic direction for health and social care in Northern Ireland, including the Health and Wellbeing 2030: Delivering Together (HWT) and the Northern Ireland Health and Social Care Strategy (NIHSS).",
    "Status": "Operational"
    "Funding (GBP Million)": "180",
    "URL": "https://www.gov.uk/gov/belfast-city-hospital"
}, {
    "ID": "18",
    "Category": "Education",
    "Sub-Category": "University",
    "Entity Name": "Queen's University Belfast",
    "Address": "University Rd, Belfast",
    "Postcode": "BT7 1NN",
    "Local Authority": "Belfast",
    "Region": "Northern Ireland",
    "Policy Area": "Higher Education",
    "Relevant Policy/Legislation": "The Higher Education Strategy for Northern Ireland",
    "Policy Summary": "Recommends the strategic direction for higher education in Northern Ireland, including the Higher Education Strategy for Northern Ireland (HESNI) and the Northern Ireland Strategy for Higher Education (NISHE).",
    "Status": "Operational"
    "Funding (GBP Million)": "400",
    "URL": "https://www.gov.uk/gov/queens-university-belfast"
}, {
    "ID": "19",
    "Category": "Business",
    "Sub-Category": "Technology",
    "Entity Name": "DeepMind",
    "Address": "8 Pancras Square, London",
    "Postcode": "NC1 4AG",
    "Local Authority": "Camden",
    "Region": "London",
    "Policy Area": "Artificial Intelligence",
    "Relevant Policy/Legislation": "National Strategy",
    "Policy Summary": "Recommends the strategic direction for artificial intelligence in the UK, including the National Strategy for Artificial Intelligence (NSAI) and the National Strategy for Data (NSD).",
    "Status": "Operational"
    "Funding (GBP Million)": "N/A",
    "URL": "https://www.gov.uk/gov/deepmind"
}, {
    "ID": "20",
    "Category": "Residential",
    "Sub-Category": "Garden Village",
    "Entity Name": "Elstree Garden City",
    "Address": "Elstree, Hertfordshire",
    "Postcode": "DA10 7AA",
    "Local Authority": "Dartford and Gravesham",
    "Region": "South East",
    "Policy Area": "Housing & Planning",
    "Relevant Policy/Legislation": "Garden Communities Programme",
    "Policy Summary": "A programme to support the development of new garden towns and villages, including the Garden Communities Programme (GCP) and the National Strategy for Garden Communities (NSGC).",
    "Status": "Under Development"
    "Funding (GBP Million)": "200",
    "URL": "https://www.gov.uk/gov/elstree-garden-city"
}, {
    "ID": "21",
    "Category": "Healthcare",
    "Sub-Category": "Mental Health",
    "Entity Name": "The Phoenix Hospital",
    "Address": "Rushmoor",
    "Postcode": "SW15 5JJ",
    "Local Authority": "Wandsworth",
    "Region": "London",
    "Policy Area": "Mental Health",
    "Relevant Policy/Legislation": "NHS Mental Health Implementation Plan",
    "Policy Summary": "Improves mental health services across the UK, including the NHS Mental Health Implementation Plan (MHIP) and the National Strategy for Mental Health (NSMH).",
    "Status": "Operational"
    "Funding (GBP Million)": "50",
    "URL": "https://www.gov.uk/gov/phoenix-hospital"
}]
```

8. Applications

- Extracting structured data from **reports, invoices, research papers, or government datasets.**
 - Converting datasets into **machine-readable JSON** for APIs and web apps.
 - Data preprocessing for **machine learning and AI projects.**
 - Creating digital pipelines for **policy analysis, healthcare, education, and business datasets.**
-

9. Conclusion

This project demonstrates how PDF data can be converted into structured formats for further analysis. Using **Streamlit for UI**, **pdfplumber for extraction**, and **pandas for cleaning**, the system successfully transforms static tables into **dynamic JSON files**, making data more accessible and reusable for modern applications.