 **PyMuPDF (fitz):**

PyMuPDF stands out as the most effective library for this task. It performs well in extracting structured text, including headings and subheadings, by leveraging features such as font size, style, and layout analysis. It handles formatting effectively and introduces minimal noise and requiring preprocessing to achieve cleaner results.

 **PDFMiner:**

PDFMiner provides advanced text and layout analysis capabilities, including text coordinates, which can help identify headings. However, it tends to introduce significant noise during text extraction, requiring preprocessing to achieve cleaner results.

 **pdfplumber:**

Pdfplumber struggles with structured text extraction. It often breaks down text into single-character tokens, making it unsuitable for heading and subheading extraction. While it is designed to extract tables, text, and metadata, it does not handle hierarchical content effectively.

 **PyPDF2:**

PyPDF2 is primarily designed for raw text extraction and lacks features for analyzing layout or formatting. While lightweight and straightforward for simple tasks, it is not capable of identifying structured elements like headings, making it the least suitable option for this analysis.suited for simple text extraction tasks but not for heading identification.

**PyPDFium2:**

PyPDFium2 is capable of extracting text efficiently and accurately, with better handling of text layout compared to PyPDF2. However, its primary focus is on rendering PDFs and extracting raw text. While it provides tools for extracting text position and layout, it does not inherently support advanced features like font style or size analysis, which are critical for identifying headings and subheadings.

In conclusion, PyMuPDF is recommended as the primary library for extracting headings and subheadings from PDF documents due to its precision and features that support structured text analysis. PDFMiner can be used as a secondary option with additional preprocessing. Combining these tools with custom heuristics or NLP techniques can further enhance heading detection accuracy.