Q1. How can Python be utilized to manage .pdf files within the realm of data handling? Could you detail a step-by-step procedure or algorithm for addressing this issue?

For data handling purposes there are several libraries within Python which can handle PDF files such as **PyPDF2, PDFMiner** and **ReportLab**. Using these libraries, Python enables you to interact with PDFs, extract relevant data, and integrate them into your data handling workflows seamlessly. For example, Read PDFs, Write PDFs, Merge or Split PDFs, Edit PDFs and convert to other formats

General Step-by-step algorithm to read the PDF file content.

1. Import Library and Open PDF
2. Create PDF Reader and Get Page Count
3. Loop through Pages and Extract Text
4. Print Page Content
5. Finish Processing

Detailed Step-by-step algorithm to read the PDF file content.

1. Import the **PyPDF2** library into your script.
2. Open the target PDF file in binary reading mode using **open(pdf_file_name, 'rb').**
3. Create a PdfReader object from the opened file using **PyPDF2.PdfReader(pdf_file)**.
4. Determine the total number of pages in the PDF using **len(pdf_reader.pages)**. This stores the value in the variable num_pages.
5. Start a loop that iterates through each page index from 0 to num_pages - 1 using **for page_num in range(num_pages).**
6. Inside the loop, access the current page object using **page = pdf_reader.pages[page_num].**
7. Extract the text content from the current page using the **extract_text()** method of the page object. Store the extracted text in the variable **text**.
8. Within the loop, format the output using string formatting. Prepend the current page number (page_num + 1) with "Page " and append a colon.
9. Print the formatted page number followed by the extracted text from the variable text.
10. Append a newline character ("\n") at the end to separate page content.
11. Once the loop iterates through all pages, the script execution finishes.

Here's a small program that how to read the content of the pdf file using **PyPDF2** library.

```
import PyPDF2

with open('Assignment7.pdf', 'rb') as pdf_file:
    pdf_reader = PyPDF2.PdfReader(pdf_file)

    num_pages = len(pdf_reader.pages)

    for page_num in range(num_pages):
        page = pdf_reader.pages[page_num]
        text = page.extract_text()
        print(f"Page {page_num + 1}:\n{text}\n")
```