

Here is a step-by-step guide to get your Python script running.

### Step 1: Install the Required Library

This script depends on a library called `PyPDF2` . You'll need to install it using `pip` , which is Python's package manager.

1. Open your terminal or command prompt.
2. Type the following command and press Enter:

```
pip install PyPDF2
```

(If you have multiple Python versions, you might need to use `pip3 install PyPDF2` )

### Step 2: Save the Python Code

1. Save the Python code I provided in a file.
2. Name the file something memorable, like `extract_pdf_text.py` .
3. Place this file in a folder where you can easily find it.

### Step 3: Add Your PDF File

1. Find the PDF file you want to extract text from.
2. **Copy** this PDF file into the **exact same folder** where you just saved your `extract_pdf_text.py` script.
3. Rename the PDF if you like (e.g., `my_report.pdf` ).

### Step 4: Edit the Python Script

1. Open the `extract_pdf_text.py` script in any text editor (like Notepad, VS Code, Sublime Text, etc.).
2. Find this line near the bottom of the script:

```
pdf_file_name = 'your_file.pdf'
```

3. Change `'your_file.pdf'` to the **exact name of your PDF file**, including the `.pdf` extension. For example:

```
pdf_file_name = 'my_report.pdf'
```

4. Save the change to the `.py` file.

### Step 5: Run the Script

1. Go back to your terminal or command prompt.
2. Use the `cd` (change directory) command to navigate to the folder where you saved your script and PDF. For example:

- `cd C:\Users\YourName\Documents\PDF_Project` (Windows)
- `cd /home/YourName/Documents/PDF_Project` (Linux/Mac)

3. Once you are in the correct directory, type the following command and press Enter:

```
python extract_pdf_text.py
```

(Again, you might need `python3` instead of `python` depending on your setup: `python3 extract_pdf_text.py` )

## What to Expect

The script will run in your terminal.

1. It will first print `Attempting to extract text from: my_report.pdf...`
2. Then, it will print the total number of pages it found.
3. Finally, it will print all the text it successfully extracted, followed by a message that it saved the text to a new `.txt` file.

**Important Note:** PDF text extraction is not always perfect. If the PDF is a scan (just images of text) or has a very complex layout, the script might not be able to "read" the text, or the output might look jumbled. This script works best on PDFs where you can normally highlight the text with your mouse.