

# Sample Document

This is a sample document created for testing the PDF Rip pipeline.

## Introduction

This document contains various types of content to test the PDF processing capabilities:

- Text content with different formatting
- Lists and bullet points
- Tables with structured data
- Mathematical equations (simulated)
- Contact information for PII testing

The goal is to verify that the PDF Rip pipeline can accurately extract and convert all these elements to

## Contact Information

Email: test@example.com

Phone: (555) 123-4567

Address: 123 Test Street, Test City, TC 12345

## Sample Table

Name	Age	Department	Salary
John Doe	30	Engineering	\$75,000
Jane Smith	28	Marketing	\$65,000
Bob Johnson	35	Sales	\$80,000
Alice Brown	32	HR	\$70,000

## Mathematical Content

Sample equation:  $E = mc^2$

Another equation:  $\int f(x) dx = F(x) + C$

Summation:  $\sum_{i=1}^n i = n(n+1)/2$

## Code Examples

```
def hello_world():
    print("Hello, World!")

class Example:
    def __init__(self):
        self.name = "test"

    def get_name(self):
        return self.name
```

## Conclusion

This document demonstrates various content types that the PDF Rip pipeline should be able to process.

1. Structured text with headings
2. Lists and bullet points
3. Tables with data
4. Mathematical notation
5. Code blocks
6. Contact information (for PII testing)

The pipeline should convert all of this content to clean, well-formatted Markdown while preserving the original structure.