Comprehensive PDF Test Document

Table of Contents

Introduction

This document contains synthetic data designed to test PDF extraction capabilities. It includes various elements commonly found in PDF documents such as formatted text, tables, images, form fields, and other complex structures.

Text Formatting Examples

Font Variations

This is regular text. **This text is bold.** *This text is italic.* ***This text is bold and italic.*** <u>This text is underlined.</u>

Font Sizes

Normal size text.

Heading 1 (Largest)

Heading 2

Heading 3

Heading 4

Heading 5

Heading 6 (Smallest heading)

Text Colors

When creating the PDF, apply these colors:

- Red text
- Blue text
- Green text
- Yellow text with black background

Special Characters

Math symbols: ∑ π ∫ ∞ √ Δ ≤ ≥ ≠ ∩ ∪ Currency symbols: $ € £ ¥ ₹ ₽ ₩ Arrows: → ← ↑ ↓ ↔ ↕ Miscellaneous: © ® ™ • § ¶ † ‡

Mathematical Equations

Simple equation: $y = mx + b$

Complex equation: $\int_0^\infty e^{-x^2} \, dx = \sqrt{\pi}/2$

Matrix:

| a b c |

| d e f |

| g h i |

Multiple Languages

English: Hello world Spanish: Hola mundo French: Bonjour le monde German: Hallo Welt Chinese: 你好世界 Japanese: こんにちは世界 Arabic: مرحبا بالعالم (right-to-left text) Russian: Привет, мир

Structural Elements

Bulleted List

- Item 1
- Item 2
  - Subitem 2.1
  - Subitem 2.2
- Item 3

Numbered List

1. First item
2. Second item
   1. Subitem 2.1
   2. Subitem 2.2
3. Third item

Block Quote

This is a block quote. It contains text that is set apart from the main content to highlight important information or quotations from other sources.

This is the second paragraph of the block quote.

Footnotes

Here is text with a footnote[^1].

Another sentence with a different footnote[^2].

[^1]: This is the first footnote content. [^2]: This is the second footnote with additional information.

Text in Columns

[When creating the PDF, format this section in two columns]

Column 1 text: Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus lacinia odio vitae vestibulum vestibulum. Cras porttitor metus velit, ut aliquam lorem egestas et. Ut at ullamcorper nulla.

Column 2 text: Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo.

Tables and Data

Simple Table

Name	Age	Occupation
John	28	Engineer
Mary	34	Doctor
Sam	41	Teacher
Lisa	25	Designer

Complex Table with Merged Cells

Category	Q1 Sales	Q2 Sales	Q3 Sales	Q4 Sales	Annual Total
Clothing	$12,500	$15,750	$11,200	$18,300	$57,750
Electronics	$31,200	$28,500	$42,100	$65,700	$167,500
Home & Garden	$8,750	$22,300	$16,500	$9,200	$56,750
Quarterly Total	**$52,450**	**$66,550**	**$69,800**	**$93,200**	**$282,000**

Large Table for Multi-page Testing

ID	First Name	Last Name	Email	Phone	Department	Position	Start Date	Salary	Status
001	John	Smith	john.smith@example.com	555-1234	Marketing	Manager	2018-03-15	$85,000	Active
002	Mary	Johnson	mary.j@example.com	555-2345	Finance	Analyst	2019-06-22	$72,000	Active
003	Robert	Williams	rob.w@example.com	555-3456	IT	Developer	2017-11-08	$91,000	Active
004	Sarah	Brown	sarah.b@example.com	555-4567	HR	Specialist	2020-01-14	$68,000	Active
005	Michael	Jones	mike.j@example.com	555-5678	Sales	Representative	2018-09-30	$76,000	Active
006	Jennifer	Davis	jen.d@example.com	555-6789	Marketing	Coordinator	2021-04-05	$62,000	Active
007	David	Miller	david.m@example.com	555-7890	IT	Administrator	2019-08-17	$88,000	Active
008	Lisa	Wilson	lisa.w@example.com	555-8901	Finance	Controller	2017-05-23	$95,000	Active
009	James	Taylor	james.t@example.com	555-9012	Operations	Supervisor	2020-11-02	$79,000	Active
010	Patricia	Anderson	pat.a@example.com	555-0123	HR	Manager	2018-02-14	$87,000	Active
011	Thomas	White	thomas.w@example.com	555-1235	Sales	Director	2016-10-18	$110,000	Active
012	Barbara	Clark	barb.c@example.com	555-2346	Marketing	Specialist	2020-03-28	$65,000	Active
013	Daniel	Lewis	dan.l@example.com	555-3457	IT	Manager	2018-07-09	$98,000	Active
014	Nancy	Lee	nancy.l@example.com	555-4568	Finance	Analyst	2019-12-01	$71,000	Active
015	Paul	Walker	paul.w@example.com	555-5679	Operations	Coordinator	2021-02-15	$64,000	Active

Table with Numerical Data

Product	Units Sold	Unit Price	Total Revenue	Profit Margin	Profit
Widget A	1,250	$12.99	$16,237.50	35%	$5,683.13
Widget B	875	$24.50	$21,437.50	42%	$9,003.75
Widget C	2,100	$8.75	$18,375.00	28%	$5,145.00
Widget D	432	$65.00	$28,080.00	55%	$15,444.00
Widget E	1,600	$15.25	$24,400.00	31%	$7,564.00
Total	**6,257**	**-**	**$108,530.00**	**39.54%**	**$42,839.88**

Visual Elements

Charts and Graphs

[When creating the PDF, include:

1. A pie chart showing market share distribution
2. A bar chart comparing monthly sales
3. A line graph showing growth trends over time]

Images

[When creating the PDF, include:

1. A photograph
2. A logo
3. A diagram or flowchart]

Diagram

[Create a simple flowchart in your PDF with boxes connected by arrows, showing a process flow]

Form Elements

[When creating your PDF, include these form elements:]

Text Fields

Name: [] **Email: []** Comments: [_____]

Checkboxes

[ ] Option 1 [ ] Option 2 [ ] Option 3

Radio Buttons

( ) Yes ( ) No ( ) Maybe

Dropdown Menu

Select an option: [▼]

- Option 1
- Option 2

- Option 3

Advanced Features

Hyperlinks

[Include these as actual hyperlinks in your PDF]

- [Visit Example Website](#)

- [Internal link to Introduction](#)

- [Email Contact](#)

Sidebar Content

[Format this as a sidebar in your PDF]

Important Information This sidebar contains supplementary information that is related to but separate from the main content. Sidebars are often used to highlight key points or provide additional context.

Watermark

[Add a watermark saying "CONFIDENTIAL" or "DRAFT" diagonally across some pages]

Annotations

[Add comment annotations to several places in your PDF]