

# GEN AI

## UNIT 2

### Document Processing in the Era of GenAI

**Document processing** means working with documents to get important information from them.

#### Examples of Documents

- **Structured:** Forms or tables with a clear layout (e.g., Excel sheets).
- **Semi-Structured:** Documents with some patterns (e.g., invoices, resumes).
- **Unstructured:** Free text with no set format (e.g., contracts, emails).

#### Steps in Simple Terms

1. **Collect:** Gather documents (digital or paper).
2. **Review:** Read and understand them.
3. **Extract:** Pick out the key information.
4. **Save:** Store the data digitally or in files.
5. **Search:** Find the data when needed.

#### Here are the steps for document processing using Generative AI.

1. **Upload Document:** Provide the document (text, image, or PDF) for processing.
2. **Convert Document:** If needed, the AI converts images or PDFs to text using OCR (Optical Character Recognition).
3. **Clean and Organize:** The AI cleans up the text by removing errors or

unnecessary parts.

4. **Analyze Content:** The AI reads and understands the text using Natural Language Processing (NLP).

5. **Generate Output:** The AI creates the desired output, like summaries, translations, or structured data.

6. **Review and Refine:** The AI checks and improves the output for accuracy and clarity.

7. **Save or Share:** The final processed document or data is ready for use or sharing.

**Challenges in traditional document processing and how GenAI can help.**

| Challenge          | Explanation   | GenAI Solution                             | How It Helps                                 |
|--------------------|---|--|--|
| Time-consuming     | Manual processing takes too much time.                            | Automates data extraction.                 | Saves time by completing tasks faster.       |
| Human errors       | Mistakes happen in manual data entry and interpretation.          | Uses AI to interpret documents accurately. | Ensures reliable and error-free outputs.     |
| High costs         | Hiring large teams for processing increases expenses.             | Reduces human involvement.                 | Lowens operational costs.                    |
| Unstructured data  | Irregular formats make it hard to extract useful insights.        | Organizes and standardizes data.           | Makes messy data easy to understand and use. |
| Scalability issues | Manual methods can't handle growing document volumes effectively. | Processes documents in bulk.               | Scales up easily to handle large workloads.  |

# Extracting, refining, and analyzing data using GenAI models.

## 1. Extracting Data




GenAI pulls out useful information from raw data sources like documents, images, or databases.

- How it works:

- Reads text from files, even if the format is messy (like scanned PDFs).
- Finds specific details such as names, dates, totals, or important keywords.

- Example:

From a bill, GenAI can extract the bill number, total amount, and due date.

|   |                 |  |              |               |
|---|-----------------|--|--------------|---------------|
|    |                 | <b>ABC Design</b> 3839 Maxwell Street - Denver, CO 80303<br>Email: info@abcdesigners.com Phone: 303-555-0173<br>Website: abcdesigners.com          |              |               |
| <b>Bill to:</b> John Smith<br>442 Swansea Street,<br>Denver, CO 80303<br>United States  |                 | <b>Invoice:</b> 2021-005<br><b>Invoice Date:</b> 05.21.2021<br><b>Due Date:</b> 06.20.2021   |              |               |
| <b>Description</b>  | <b>Quantity</b> | <b>Unit</b>  | <b>Price</b> | <b>Amount</b> |
| Design project  | 12              | hour   | 50.00        | 600.00        |
| Discount 10%  |                 |  |              | -60.00        |
| Subtotal without Tax  |                 |  |              | 600.00        |
| <b>Total USD</b>  |                 |  |              | <b>540.00</b> |
| Amount Paid   |                 |  |              | 0.00          |
| <b>Amount Due (USD)</b>   |                 |  |              | <b>540.00</b> |
| <b>Terms &amp; Conditions</b><br>Net 30.<br>Please make the payment via the payment link below.<br>10% discount applied for new customer. |                 |  |              |               |
| <br><b>Pay online</b>                                  |                 | Go to: <a href="https://invoice.sumup.com/s/F8wY5gBRk">https://invoice.sumup.com/s/F8wY5gBRk</a><br>Or scan the QR code using your phone's camera. |              |               |
|   |                 | <br>Page 1 of 1 for Invoice #2021-005                         |              |               |

## 2. Refining Data

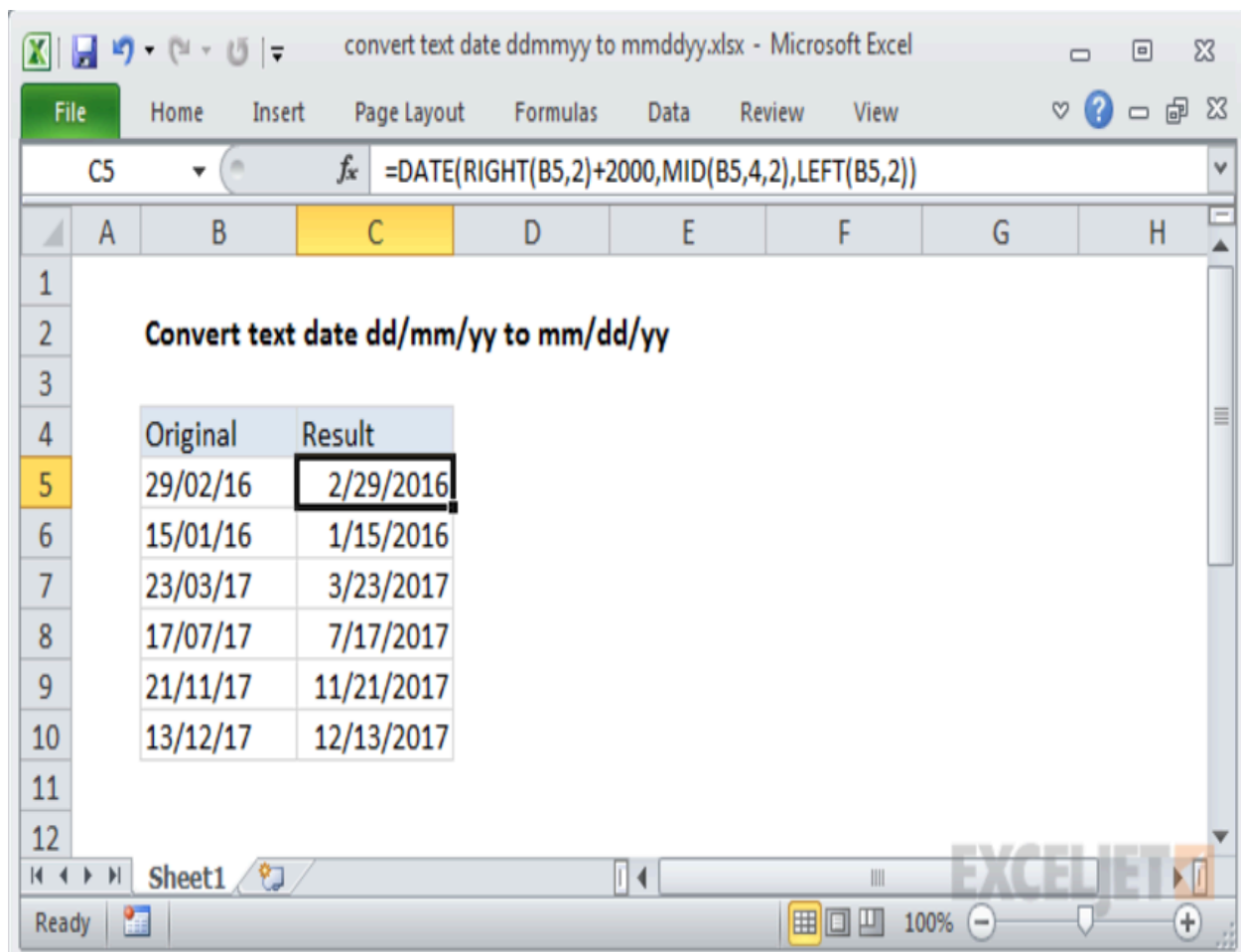
Refining means cleaning and organizing the extracted data to make it neat and ready to use.

- **How it works:**

- Fixes errors like wrong spellings or typos.
- Standardizes data formats (e.g., turning all dates into **YYYY-MM-DD**).
- Removes duplicates or unnecessary parts.

- **Example:**

If you have dates written differently, like “19/11/2024” and “November 19, 2024,” GenAI changes them all to **2024-11-19**.



The screenshot shows a Microsoft Excel window titled "convert text date ddmmyy to mmddyy.xlsx". The formula bar displays the formula `=DATE(RIGHT(B5,2)+2000,MID(B5,4,2),LEFT(B5,2))` in cell C5. The worksheet contains a table titled "Convert text date dd/mm/yy to mm/dd/yy" with two columns: "Original" and "Result". The table lists several dates in dd/mm/yy format and their corresponding mm/dd/yyyy results. The date 29/02/16 is highlighted in the original image.

| Original | Result     |
|----------|------------|
| 29/02/16 | 2/29/2016  |
| 15/01/16 | 1/15/2016  |
| 23/03/17 | 3/23/2017  |
| 17/07/17 | 7/17/2017  |
| 21/11/17 | 11/21/2017 |
| 13/12/17 | 12/13/2017 |

### 3. Analyzing Data

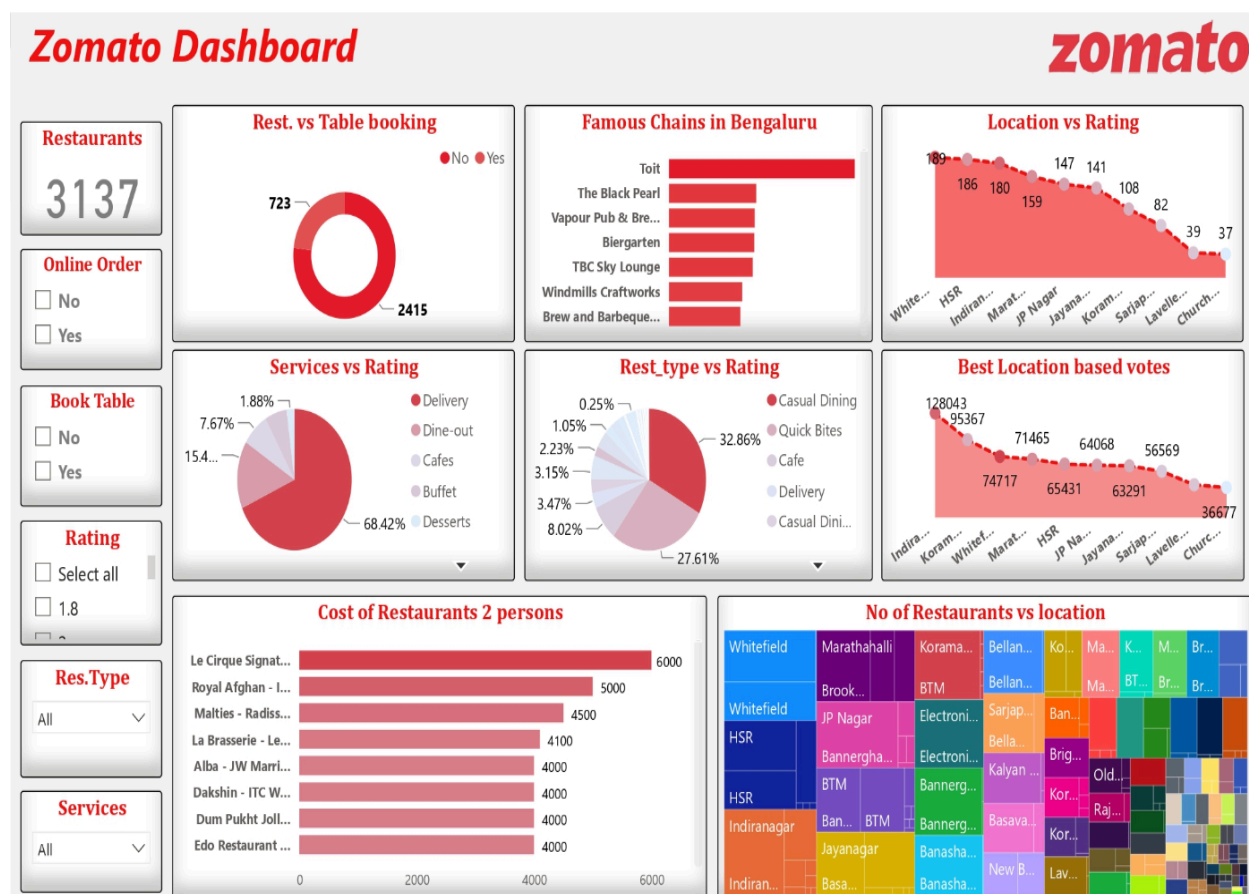
Analyzing means using the clean data to find patterns, trends, or insights.

- **How it works:**

- Summarizes long documents into key points.
- Finds trends, like sales going up or down over months.
- Detects emotions in text (e.g., if feedback is happy or unhappy).
- Answers questions based on data (e.g., “What is the total sales for October?”).

- **Example:**

From customer reviews, GenAI can highlight the most common complaints or praise.



## **Case study:**

**Auto summarization, content generation, and parsing complex documents with GenAI.**

### **Introduction:**

**XYZ Corporation, a leading financial services company, was struggling with manual document processing, which was leading to inefficiencies, errors, and a significant amount of time spent on document processing. They had a large volume of documents to process every day, including financial reports, customer feedback, and contracts.**

### **Solution:**

**To address these challenges, XYZ Corporation turned to GenAI, a cutting-edge document processing technology. GenAI implemented its advanced natural language processing capabilities to automate document summarization, classification, and extraction. This enabled XYZ Corporation to quickly and accurately process large volumes of documents, extract relevant information, and make data-driven decisions.**

## **Results:**

**The implementation of GenAI led to significant benefits for XYZ Corporation, including:**

- Reduced manual processing time by 80%: GenAI automated document processing, freeing up staff to focus on higher-value tasks.**
- Improved document accuracy by 95%: GenAI's advanced algorithms ensured high accuracy rates, reducing errors and minimizing risk.**
- Enhanced decision-making capabilities by 90%: GenAI provided insights and analysis, enabling data-driven decisions.**
- Increased customer satisfaction by 25%: GenAI improved response times and accuracy, leading to higher customer satisfaction.**
- Reduced language barriers and improved global communication: GenAI translated documents, breaking down language barriers and enabling global communication.**

## Conclusion:

**The implementation of GenAI transformed XYZ Corporation's document processing operations, improving efficiency, accuracy, and decision-making. With GenAI, organizations can revolutionize their document processing workflows and achieve significant benefits. This case study demonstrates the power of GenAI in document processing and its potential to transform the financial services industry.**

### Step 1: Auto Summarization

First, GenAI summarizes long documents into short, easy points.

GenAI can help by summarizing each vacation package. For example, instead of reading the entire brochure, GenAI could summarize it like this:

- **Destination:** Paris, France
- **Duration:** 5 days, 4 nights
- **Key Highlights:** Eiffel Tower, Louvre Museum, Seine River Cruise
- **Cost:** \$1,500 per person

Now, the travel planner can quickly understand each package without reading through all the details!



