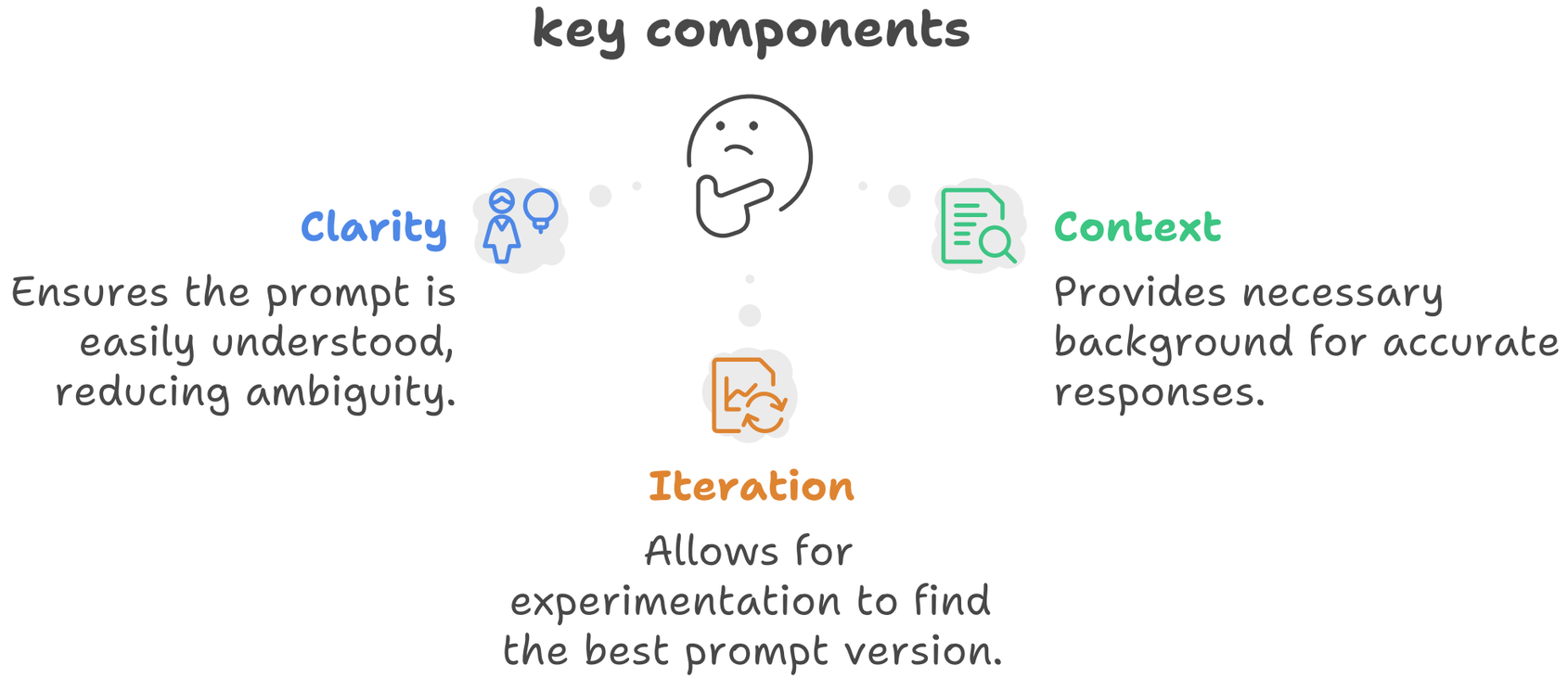
Prompt Engineering Guide for Data Analyst

What is Prompt Engineering?

Prompt Engineering is the process of designing and refining input prompts to maximize the effectiveness of AI models in generating accurate and relevant outputs. It is a crucial skill for optimizing AI-driven tasks, enhancing human-AI interactions, and improving data science solutions



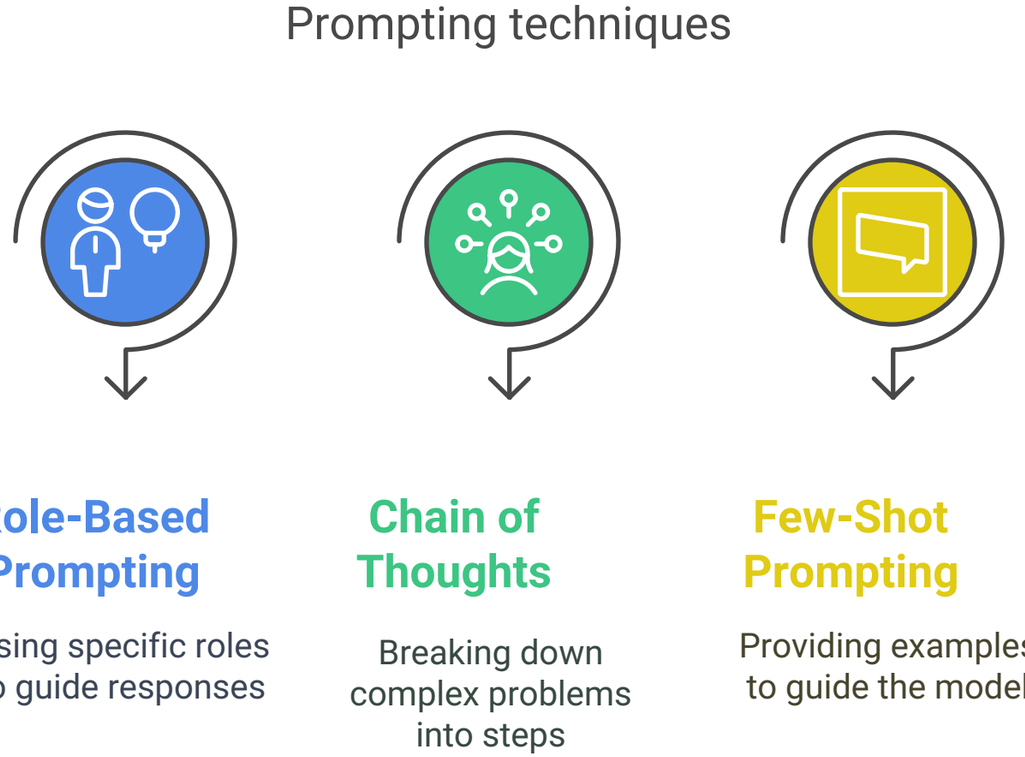
Why It Matters for Data Analysts?

Data analysts often work with complex queries, reports, and insights. Prompt engineering helps:

- Automate repetitive tasks
- Improve clarity and consistency of outputs
- Enhance productivity and decision-making

Techniques Used in This Guide

- Role-Based Prompting
- chain of thoughts
- Few-Shot Prompting (in later slides)



Task 1 – Generate SQL Queries from Business Questions

Before

> Write a SQL query to get total sales per region for 2023.

```
SELECT product_id,
      AVG(sale_amount) AS average_sales
FROM sales
WHERE YEAR(sale_date) = 2025
GROUP BY product_id
ORDER BY average_sales DESC;
```

After

Using Role-Based + Clear Instruction Format

> You are a data analyst. Write a clear and optimized SQL query based on the following request.
>
> **Business Question:** “Find the total sales per region for the year 2023 from the sales_data table.”

OUTPUT:

```
SELECT region, SUM(sales_amount) AS total_sales
FROM sales_data
WHERE YEAR(sale_date) = 2023
```

```
GROUP BY region
ORDER BY total_sales DESC;
```

Task 2 - Summarize insights from raw data or tables

Before

Summarize insights from the following sales data:

Raw Data Sample:

Region	Sales Amount	Sale Date
North	5000	2023-01-10
South	7000	2023-01-15
East	6000	2023-01-20
West	8000	2023-01-25
North	3000	2023-02-10
South	10000	2023-02-15

Output:

- Provide a summary of sales performance across the regions.

Example Output

- Sales are recorded for different regions.
- North region has sales in both months.
- South region shows the highest sales.

After Using Few-Shot Prompting

You are a data analyst. Based on the following raw sales data, provide a summary of insights regarding overall sales, top-performing regions, and any potential trends.

Here is the data:

Raw Data Sample:

Region	Sales Amount	Sale Date
North	5000	2023-01-10
South	7000	2023-01-15
East	6000	2023-01-20
West	8000	2023-01-25
North	3000	2023-02-10
South	10000	2023-02-15

Few-Shot Example Data 1:

Region	Sales Amount	Sale Date
North	5000	2023-01-10
South	7000	2023-01-15
West	8000	2023-01-20

Insight Example 1:

- The **North region** showed a moderate performance with 5,000 in sales.
- The **South region** had the second-highest performance with 7,000 in sales.

Few-Shot Example Data 2:

Region	Sales Amount	Sale Date
East	6000	2023-01-20
North	3000	2023-02-10
South	10000	2023-02-15

- **Insight Example 2:**
    - The **South region** performed significantly well in February with 10,000 in sales.
    - The **East region** had consistent sales of 6,000, without any major fluctuation.
    - The **North region** saw a drop in February, with only 3,000 in sales.

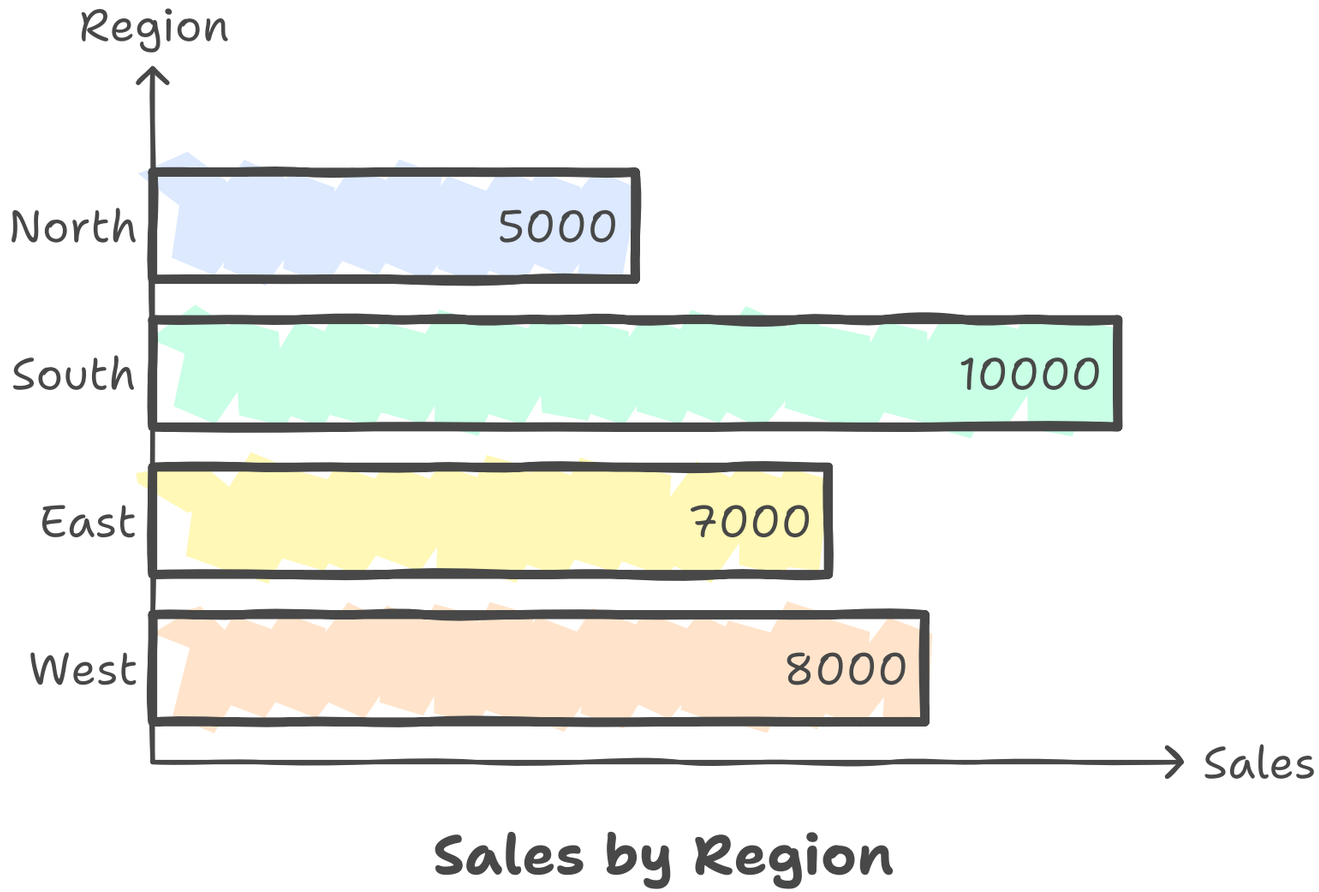
Output:

- **North Region**: Total sales = 8,000. Sales dropped to 3,000 in February.
- **South Region**: Total sales = 17,000. Highest sales recorded in February (10,000).
- **East Region**: Total sales = 6,000. Consistent sales without large fluctuations.
- **West Region**: Total sales = 8,000. High performance in January (8,000).

Task 3: Explain Charts/Graphs in Plain English

Before:

Explain this chart.



Output:

- North is 5000
- South is 10000

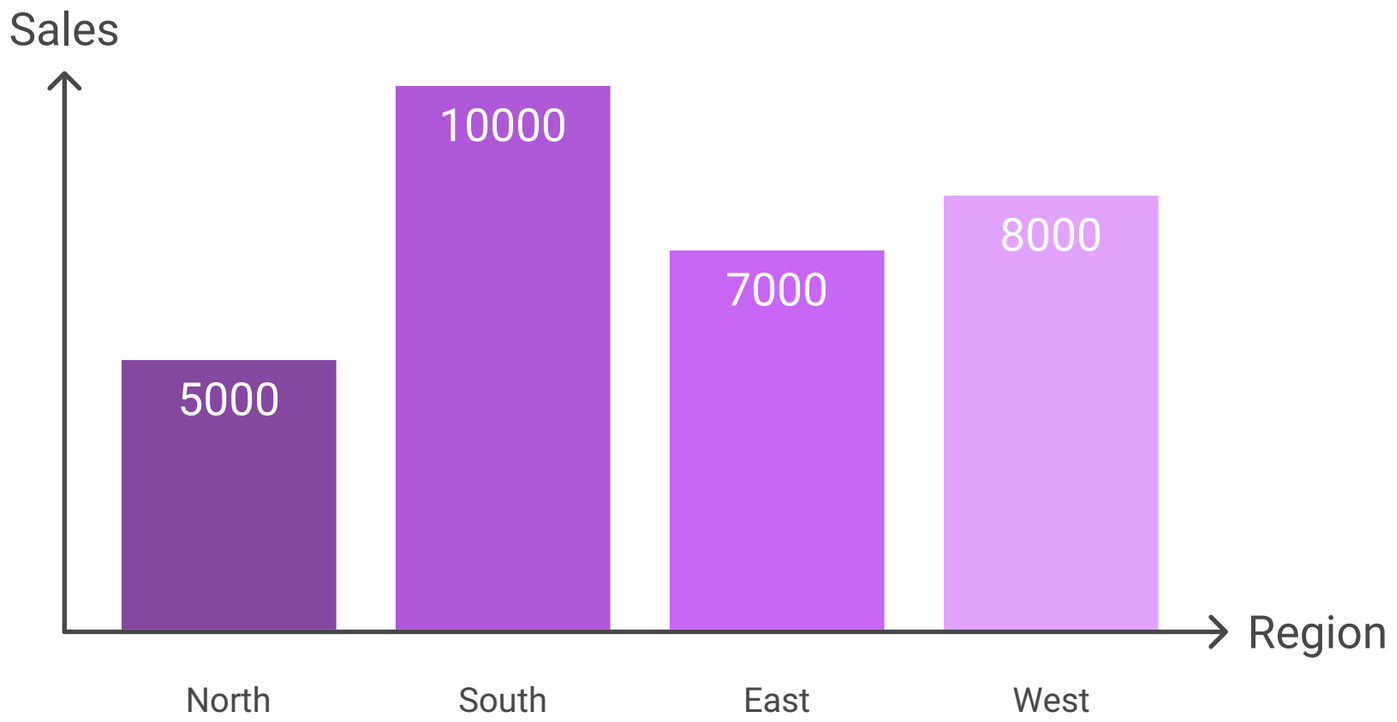
- West is 8000

After (Role-Based + Clear Instruction Format)

You are a data analyst. Explain the insights from the following bar chart in **simple, plain English** as if you're speaking to a non-technical stakeholder.

Sales by region in 2023:

- North: 5000
- South: 10000
- East: 7000
- West: 8000



Sales by Region in 2023

Output:

The South region had the highest sales with 10,000 units, making it the top performer. The North had the lowest sales at 5,000 units. East and West regions performed moderately, with 7,000 and 8,000 units respectively. This suggests that sales strategies are working best in the South and may need improvement in the North."

Task 4: Draft Analytical Reports or Executive Summaries

Before:

Write a report on sales data.

Output:

Sales data from Q1 shows North: 20K, South: 35K, East: 25K. The trend seems to be upward.

After : Role-Based + Clear Instruction

You are a business analyst. Based on the quarterly sales data, write an executive summary for leadership. Highlight key trends, regional performance, and strategic implications.

Output:

In Q1, the South region led sales with $35K, showing a 15% growth over the previous quarter. The North and East followed with $20K and $25K, respectively. The South's performance is attributed to the new promotional campaign. Recommendation: Expand successful strategies from South to other regions.

Task 5: Recommend KPIs or Metrics Based on Business Context

Before:

What are the KPIs for a retail company?

Output:

- Sales
- Customer satisfaction
- Profit

After (Role-Based + Clear Instruction Format)

You are a business intelligence analyst. Based on a retail company's goal to improve customer retention and inventory efficiency, recommend specific KPIs aligned to these objectives.

Output:

Customer Retention KPIs:

- Repeat Purchase Rate
- Customer Lifetime Value
- Churn Rate

Inventory Efficiency KPIs:

- Inventory Turnover Ratio
- Stockout Rate
- Holding Cost per Unit

Task 6: Clean and Format Messy/Unstructured Data

Before:

Fix this messy Excel data.

Output:

Student_ID	Course_Name	Module_Completed	Score	Completion_Date	Time_Spent_Min	Feedback_Rating
S1001	AI Basics	Module 1	80	01-05-2024	30	4
S1002	python	Module 2	90	01-05-2024	45	5
S1002	python	Module 2	90	01-05-2024	45	5
S1003	Data science	Module 1	75	2/5/24	thirty-five	3

After (Few-Shot + Clear Instruction)

You are a data cleaning assistant. Clean the student training dataset.

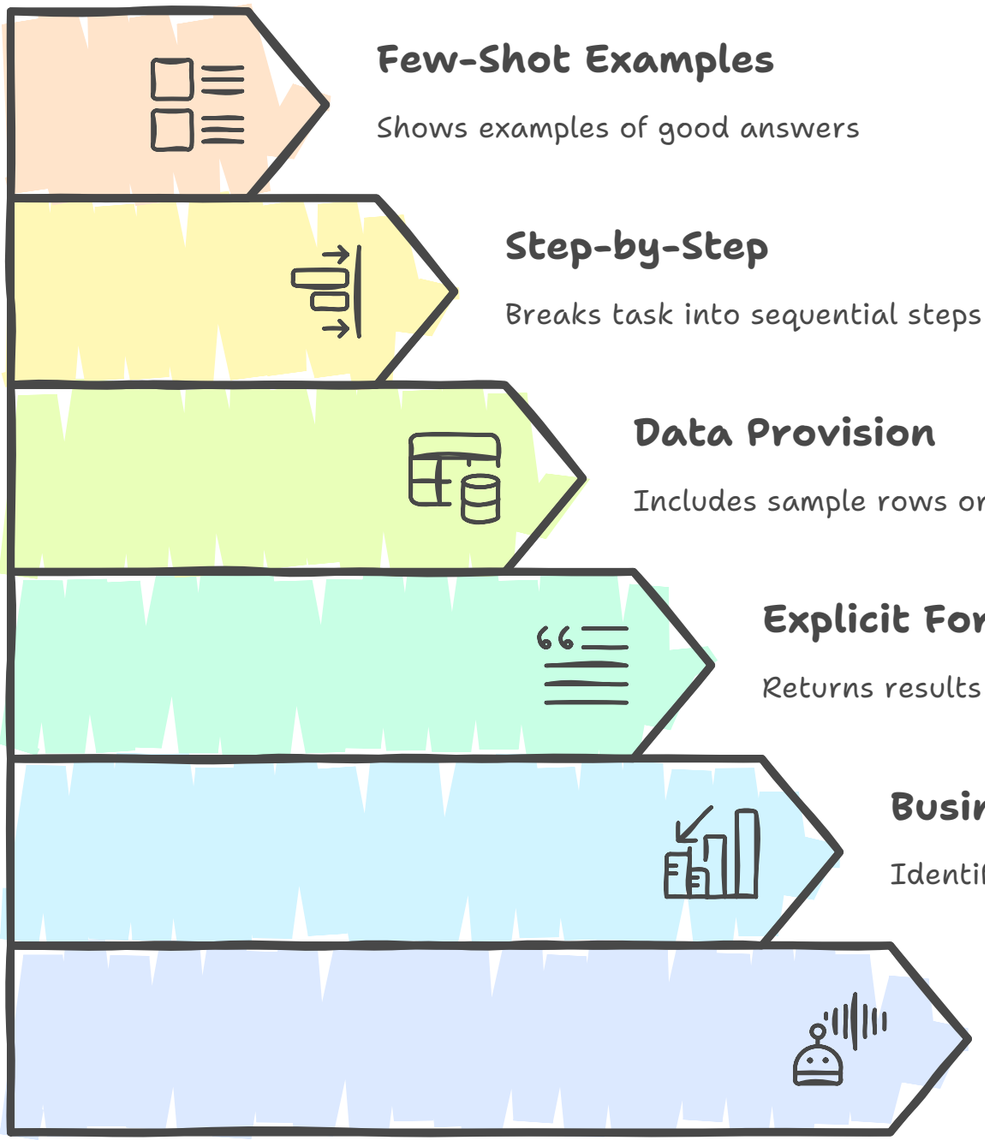
- Capitalize course names
- Format date to **YYYY-MM-DD**
- Ensure scores and time are numeric
- Remove duplicates
- Rename columns to camel case for clarity

output:

StudentID	CourseName	ModuleCompleted	Score	CompletionDate	TimeSpentMinutes	FeedbackRating
S1001	Ai Basics	Module 1	80	2024-05-01	30	4
S1002	Python	Module 2	90	2024-05-01	45	5
S1003	Data Science	Module 1	75	2024-05-02	35	3

Tips:

Do's



Few-Shot Examples

Shows examples of good answers

Step-by-Step

Breaks task into sequential steps

Data Provision

Includes sample rows or schema details

Explicit Formats

Returns results as a structured table

Business Context

Identifies sales trends by region

Clear Role

Guides model's tone and focus

Made with Napkin

Don' ts

Iteration Needed

Revise prompts based on feedback received.

Vague Instructions

Avoid giving vague instructions to AI.

Interpretability

Request a short explanation of each result.

Data Ambiguities

Be specific about data ambiguities always.

Mixed Objectives

Separate multiple objectives into distinct prompts.

Made with Napkin