1. What is the business problem we are trying to solve

The business wants to identify the top 5 products by total revenue and analyze customer sentiment for those products. This will help understand how well the best-selling products are received by customers.

2. What is the objective of this analysis

The objective is to:

- Calculate total revenue for each product
- Determine the average customer sentiment score for each product
- Compare customer satisfaction with sales performance to find meaningful insights

3. Which datasets are being used

We are using two datasets:

- `sales_data_raw.csv` – contains details of each sale transaction
- `customer_feedback.json` – contains sentiment scores and feedback from customers about products

4. What are the required data fields from each dataset

From `sales_data_raw.csv`:

- `product_id`
- `sale_price`
- `quantity`
- `sale_date`

From `customer_feedback.json`:

- `product_id`
- `customer_id`
- `sentiment_score`
- `review_date`

5. What key metrics will be calculated

- Total Revenue per Product = `sale_price × quantity`
- Average Sentiment Score per Product
- Number of Sales per Product
- Number of Reviews per Product

6. What are the desired insights

- The top 5 revenue-generating products
- The average sentiment score for each of those top 5 products
- Whether higher revenue products also receive more positive feedback
- Which products might need improvement based on negative sentiment despite high sales

7. How will the insights be used

- Targeted marketing for top-performing and well-rated products
- Product development or customer service improvements for products that sell well but receive negative feedback
- Strategic decision-making on inventory, promotions, or pricing