

PROMPT 1: SETTING THE CONVERSATION

For this one conversation, respond with pure, concise answers. No unnecessary enthusiasm, no preambles, no filler phrases, and no 'Sure!' or 'Of course!' at the beginning. Just answer the question or complete the task directly, with clarity and precision. Assume the user wants maximum efficiency and minimal emotional padding. Stay strictly on-topic.

We are working on the data provided by an FMCG client. They have stored their entire orders data in an excel (.xlsx) file. We have to perform data preprocessing activities such as normalization, data enrichment, modelling etc. before business analysis.

You are a Data engineer with 10+ years of experience.

We will assign you various tasks related to data preprocessing one-by-one. You have to provide solutions for each task.

Return your solutions in a step-by-step guide with clear heading and bullet points.

Use a simple and professional tone, which a non-technical professional can understand.

Focus on practical solutions we can implement. In case, we require you to perform any task directly, we will let you know. Reply with "OKAY" if you understand.

PROMPT 2: DATA NORMALIZATION STEP

{

"Task": "Generate an action plan for data normalization",

"Instructions": ["Review and understand the raw data structure", "Generate a detailed step-by-step normalization plan", "Provide example normalized table structures (schemas only, without sample data)"],

"Raw_Data_Fields": ["order_id", "dt_order", "segment", "category", "sub_category", "product", "mrp_inr", "cost_per_unit", "store_name", "country", "state", "city", "quantity", "discount"]

}

PROMPT 3: BRAND NAME CLEANING

{

"Task": "Clean 'product' column to extract brand names",

"Instructions": ["Identify and extract only the full brand name from each product entry", "If sub-brands are present, return only the parent brand (e.g., Schwarzkopf Color Ultime, Schwarzkopf Professional Igora, Schwarzkopf Live Colour → Schwarzkopf)"],

"Output_Format": ["uncleaned_product_name", "cleaned_brand_name"],

"Data": [<inserted raw data>]

}

PROMPT 4: CALENDAR TABLE

```
{  
  "Task": "Generate a calendar table",  
  "Instructions": [  
    "Write Power Query M code to create a calendar table",  
    "Extract the minimum and maximum order date from fact_order[dt_order]",  
    "Generate a continuous list of dates between the min and max",  
    "Ensure the column datatype is set to 'date'"  
  ],  
  "Output": "M Code"  
}
```

PROMPT 5: ADMINISTRATIVE ZONES

5A (Google AI Mode):

“What are different administrative regions of India such as North, South etc. Which government body decides the segmentation?”

5B (ChatGPT):

```
{  
  "Task": "Transform unstructured to structured data",  
  "Instructions": [  
    "Review the provided information",  
    "Identify Zone names and their respective States/UTs",  
    "Return the result in a structured table format"  
  ],  
  "Output_Format": ["Zone_Name", "State/UT"],  
  "Information": [<inserted unstructured data>  
}
```