CLIENT REQUIREMENTS DOCUMENT

Project Title: Retail Sales Performance Analysis
Client: Retail Business Management
Prepared For: Data Analysis Team

PROJECT DESCRIPTION:

The **Retail Sales Performance Analysis Project** focuses on examining the sales transactions of a retail business using structured data, including customer demographics, product categories, transaction timestamps, quantities, and pricing. The goal is to identify key sales trends, top-performing product categories, customer purchasing behavior, and high-value transactions. Additionally, the project provides insights into sales performance by time of day and customer segments, enabling data-driven decision-making to optimize inventory, marketing strategies, and overall business profitability.

DATA COLUMNS & DESCRIPTIONS

1. **transactions_id**
    - A unique identifier for each sales transaction.
    - **Purpose**: To uniquely identify and reference each sale.
2. **sale_date**
    - The calendar date on which the sale occurred.
    - **Purpose**: To perform time-based analysis (daily, monthly, yearly trends).
3. **sale_time**
    - The time of day when the sale occurred.
    - **Purpose**: To analyze purchasing patterns across different times of the day.
4. **customer_id**
    - A unique identifier for the customer making the purchase.
    - **Purpose**: To track customer behavior, frequency, and lifetime value.
5. **gender**
    - The gender of the customer (e.g., Male, Female).
    - **Purpose**: To analyze purchasing trends across different demographic segments.

6. **age**
    - The age of the customer at the time of purchase.
    - **Purpose**: To understand the preferences of different age groups.
7. **category**
    - The product category of the item sold (e.g., Clothing, Beauty, Electronics).
    - **Purpose**: To analyze performance and sales volume across different product types.
8. **quantity**
    - The number of units sold in a single transaction.
    - **Purpose**: To analyze sales volume and bulk purchasing trends.
9. **price_per_unit**
    - The selling price for one unit of the product.
    - **Purpose**: To calculate revenue and analyze pricing strategies.
10. **cogs (Cost of Goods Sold)**
    - The direct cost attributable to the production of the goods sold.
    - **Purpose**: To calculate gross profit and margin analysis.
11. **total_sale**
    - The total value of the transaction (quantity * price_per_unit).
    - **Purpose**: To measure revenue generation and identify high-value sales.

ANALYSIS REQUIREMENTS & TASKS

The analysis must be performed using SQL queries on the provided database (`p1_retail_db`). The final delivery should include the results for the following specific tasks:

1. **Daily Sales Extraction**
    - Retrieve the complete details of all sales transactions that occurred on a specific date (`2022-11-05`).
2. **Filtered Transaction Analysis**
    - Identify all transactions from November 2022 where the product category was 'Clothing' and the quantity sold was more than 4 units.
3. **Category Performance Summary**
    - Calculate the total sales revenue (`total_sale`) and the total number of orders for each product category.
4. **Customer Demographic Analysis**
    - Determine the average age of customers who purchased items from the 'Beauty' category.

5. **High-Value Transaction Identification**
    - Find all transactions where the total sale value exceeded $1000.
6. **Sales Distribution by Demographic and Category**
    - Calculate the total number of transactions segmented by both customer gender and product category.
7. **Monthly Sales Trend & Peak Period Identification**
    - Calculate the average sale value for each month. Furthermore, identify the best-selling month (highest average sale) for each calendar year within the dataset.
8. **Top Customer Identification**
    - Identify the top 5 customers based on their cumulative total sales value across all their transactions.
9. **Customer Reach per Category**
    - For each product category, find the number of unique customers who made a purchase.
10. **Operational Analysis by Shift**
    - Categorize each transaction into a time shift based on the time of sale (e.g., Morning, Afternoon, Evening) and calculate the total number of orders processed during each shift.

KEY POINTS FOR DEVELOPING THE QUERY

- Ensure the dataset is clean by handling any missing or NULL values before performing analysis.
- Use appropriate SQL functions for date and time extraction (e.g., `EXTRACT`, `TO_CHAR`) for time-based analysis.
- Queries should be optimized for clarity and performance.
- Results should be presented in a clear and consumable format for business stakeholders.