Retail Sales — SQL Project Problem Statement

Background

Novus Retail Pvt. Ltd., a mid-sized multi-category retail company (fictional), has provided its complete **2022 transaction-level sales dataset** for analysis. The dataset includes fields such as Transaction_Id, Sale_Date, Sale_Time, Customer_Id, Gender, Age, Category, Quantity, Price_Per_Unit, Cogs, and Total_Sale. However, the raw dataset may contain **duplicates, null values, inconsistent data types, and formatting issues** that must be addressed before any meaningful analysis can be performed.

To ensure accuracy and reliability of insights, the project requires extensive **data cleaning, data type corrections, duplicate handling, and null-treatment** in SQL before moving into exploration and business analysis.

Objective / Problem Statement

The goal of this project is to **clean, transform, and analyse Novus Retail's 2022 sales dataset** to uncover insights about revenue performance, customer demographics, high-value segments, product/category trends, and operational patterns.
These insights will help Novus Retail:

- Optimise inventory planning,

- Identify high-performing and underperforming categories,

- Strengthen customer-based targeting strategies, and

- Improve overall sales and operational decision-making.

Key Business Questions

1. **Total Revenue**

   o What is the total revenue generated in 2022?

2. **Revenue By Category & Product Performance**

   o Which product categories contribute the most to sales and revenue?

   o Which categories show weak performance?

3. **Monthly & Seasonal Trends**

   o How do sales vary across months?

   o Which months show peak or low demand?

4. **Customer Behaviour & Segmentation**

o   Who are the top customers based on total spending?

o   What are the spending patterns of different customer groups?

5. **Demographic Insights**

o   How do age and gender groups contribute to overall sales?

o   Which demographics prefer certain product categories?

6. **High-Value Transactions**

o   How many orders fall under high-value transactions (e.g., Total_Sale > 1000)?

7. **Order Patterns by Time of Day**

o   During which shift (morning/afternoon/evening) do most sales occur?

Stakeholders & Expected Outputs

Stakeholders

- **Sales & Marketing Team**
  Needs insights on customer demographics, top spenders, and category demand to design targeted campaigns.

- **Inventory & Supply Chain Team**
  Requires monthly/category performance and product demand trends to optimise stock levels and reduce shortages or overstock.

- **Finance Department**
  Needs accurate revenue reporting, category-level contribution, and high-value transaction tracking.

- **Senior Management / Strategy Team**
  Looks for actionable insights to improve business decisions and revenue growth.

Expected Outputs

- A **clean, processed SQL dataset** with duplicates removed, nulls handled, and corrected data types.

- SQL-based summary metrics:

  o   Total Revenue

  o   Average Order Value

  o   Unique Customers

  o   Order Count

- Category-wise and monthly revenue breakdowns.

- Top customer list and demographic spending insights.

- Shift-based sales analysis.

- A list of high-value transactions.

- A clear analytical foundation for building a future dashboard.