# Retail Sales Analysis SQL Project

## Objectives

- **Set up a retail sales database:** Create and populate a database with retail sales data.
- **Data Cleaning:** Identify and remove records with missing or null values.
- **Exploratory Data Analysis (EDA):** Perform basic EDA to understand the dataset.
- **Business Analysis:** Use SQL queries to answer business questions and derive insights.

---

## Project Structure

### 1. Database Setup

- **Database Creation:** Create a database named `sales_retail`.

**Table Creation:** A table named `retail_sales` stores the sales data with the following schema:

```
CREATE DATABASE sales_retail;

CREATE TABLE retail_sales
(
    transactions_id INT PRIMARY KEY,
    sale_date DATE,
    sale_time TIME,
    customer_id INT,
    gender VARCHAR(10),
    age INT,
    category VARCHAR(35),
    quantity INT,
    price_per_unit FLOAT,
    cogs FLOAT,
    total_sale FLOAT
```

- );

### 2. Data Exploration & Cleaning

- **Record Count:** Determine the total number of records.
  SELECT COUNT(*) FROM retail_sales;
- **Customer Count:** Identify unique customers.
  SELECT COUNT(DISTINCT customer_id) FROM retail_sales;
- **Category Count:** Find unique product categories.
  SELECT DISTINCT category FROM retail_sales;

**Null Value Check:** Locate and remove records with null values.

```
SELECT * FROM retail_sales
WHERE
   sale_date IS NULL OR sale_time IS NULL OR customer_id IS NULL OR
   gender IS NULL OR age IS NULL OR category IS NULL OR
   quantity IS NULL OR price_per_unit IS NULL OR cogs IS NULL;

DELETE FROM retail_sales
WHERE
   sale_date IS NULL OR sale_time IS NULL OR customer_id IS NULL OR
   gender IS NULL OR age IS NULL OR category IS NULL OR
```

- quantity IS NULL OR price_per_unit IS NULL OR cogs IS NULL;

## 3. Data Analysis & Findings

**Example Queries:**

**Sales on a specific date:**

```
SELECT *
FROM retail_sales
```

- WHERE sale_date = '2022-11-05';

**High-quantity sales in November 2022 for Clothing:**

```
SELECT *
FROM retail_sales
WHERE category = 'Clothing'
  AND TO_CHAR(sale_date, 'YYYY-MM') = '2022-11'
```

- AND quantity >= 4;

**Total sales by category:**

```
SELECT category, SUM(total_sale) AS net_sale, COUNT(*) AS total_orders
FROM retail_sales
```

- GROUP BY category;

---

# Findings

- **Customer Demographics:** Analysis reveals customer age distributions and purchasing patterns across categories.
- **High-Value Transactions:** Premium purchases (total sale > 1000) were identified.
- **Sales Trends:** Monthly trends highlight peak seasons and popular categories.
- **Customer Insights:** Top-spending customers and unique category purchasers were identified.

---

# How to Use

1. **Clone the Repository:**
   git clone <repository_url>
2. **Set Up the Database:** Execute the SQL file `retail_sales.sql` in your PostgreSQL environment to create and populate the database.
3. **Run the Queries:** Use the provided queries in the project documentation to perform analysis.
4. **Explore and Modify:** Feel free to extend the queries or create your own to explore additional aspects of the dataset.

---

# Conclusion

This project provides a comprehensive introduction to SQL for data analysis. By completing it, you will gain hands-on experience in:

- Database creation and management
- Data cleaning and exploratory analysis
- Writing business-driven SQL queries to derive insights

---

### Reports

- **Sales Summary:** Overview of total sales, customer demographics, and category performance.
- **Trend Analysis:** Insights into monthly sales patterns.
- **Customer Insights:** Top customers and unique category purchaser