# Retail Sales Analysis Report using SQL

## 1. Introduction

The retail sales analysis project leverages SQL to derive insights from transactional data. This report outlines the data preprocessing steps, key analysis queries, and their respective findings to understand customer behavior, sales trends, and category performance.

## 2. Data Preprocessing

The dataset used for analysis is accessed from the sql_projects.retail_sales table. The following preprocessing steps were undertaken to ensure data quality:

### Data Quality Check

The dataset was checked for missing values using the following SQL query:

```
SELECT *
FROM sql_projects.retail_sales
WHERE transactions_id IS NULL
  OR 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
  OR total_sale IS NULL;
```

This query identifies rows with null values in critical columns. Any rows with missing data were flagged for further review or removal, ensuring the dataset's completeness and accuracy.

## 3. Data Analysis

### 3.1 Sales on a Specific Date

A query was written to retrieve all transactions made on 2022-11-05:

```
SELECT *
FROM retail_sales
```

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

This analysis identifies sales made on a specific day, useful for tracking performance during events or promotions.

**3.2 High-Quantity Clothing Sales in November 2022**

To find clothing sales with a quantity greater than or equal to 4 during November 2022:

```
SELECT *
FROM retail_sales
WHERE category = 'Clothing'
  AND CHAR(sale_date, 'YYYY-MM') = '2022-11'
  AND quantity >= 4;
```

This query highlights high-demand products within a category during a specific period.

**3.3 Total Sales by Category**

The total sales for each category were calculated:

```
SELECT
    category, SUM(total_sale) AS total_sales
FROM retail_sales
GROUP BY category;
```

This provides an overview of which categories contribute most to revenue.

**3.4 Average Age of Beauty Customers**

The average age of customers purchasing beauty products was calculated:

```
SELECT
    AVG(age)
FROM retail_sales
WHERE category = 'Beauty';
```

Understanding customer demographics aids in targeted marketing efforts.

**3.5 Transactions with High Sales**

To identify transactions where total_sale exceeded 1000:

```
SELECT *
FROM retail_sales
```

```
WHERE total_sale > 1000;
```

This analysis highlights significant sales events.

**3.6 Transactions by Gender and Category**

The number of transactions by gender within each category was calculated:

```
SELECT
    category, gender, COUNT(transactions_id) AS total_transaction
FROM retail_sales ;
GROUP BY category, gender;
```

This helps understand customer preferences and purchasing patterns.

**3.7 Best Selling Month by Year**

To find the best-performing month in each year:

```
SELECT
    year,
    month,
    avg_sale
FROM (
    SELECT
        EXTRACT(YEAR FROM sale_date) AS year,
        EXTRACT(MONTH FROM sale_date) AS month,
        AVG(total_sale) AS avg_sale,
        RANK() OVER (PARTITION BY EXTRACT(YEAR FROM sale_date) ORDER BY AVG(total_sale) DESC) AS ranked
    FROM retail_sales
    GROUP BY month, year
) AS t1
WHERE ranked = 1;
```

This provides insights into seasonal trends.

**3.8 Unique Customers by Category**

The number of unique customers for each category was calculated:

```
SELECT
    category, COUNT(DISTINCT customer_id) AS unique_customer
FROM retail_sales
GROUP BY category;
```

This highlights customer engagement by product category.

### 3.9 Orders by Shift

Transactions were grouped into shifts based on the time of sale:

```
WITH shifting AS (
    SELECT *,
        CASE
            WHEN EXTRACT(HOUR FROM sale_time) < 12 THEN 'Morning'
            WHEN EXTRACT(HOUR FROM sale_time) BETWEEN 12 AND 17 THEN 'Afternoon'
            ELSE 'Evening'
        END AS shift
    FROM retail_sales
)
SELECT
    shift, COUNT(*) AS total_order
FROM shifting
GROUP BY shift;
```

This helps in planning resources and understanding customer behavior during different times of the day.

## 4. Conclusion

The retail sales analysis project showcases the effective use of SQL for:

- Identifying trends and patterns in sales data.
- Understanding customer demographics and preferences.
- Analyzing the performance of product categories and time-based shifts.