# Retail Sales Analysis SQL Project

## Project Overview

**Project Title**: Retail Sales Analysis

## Objectives

1. **Set up a retail sales database**: Create and populate a retail sales database with the provided sales data.
2. **Data Cleaning**: Identify and remove any records with missing or null values.
3. **Exploratory Data Analysis (EDA)**: Perform basic exploratory data analysis to understand the dataset.
4. **Business Analysis**: Use SQL to answer specific business questions and derive insights from the sales data.

## Project Structure

### 1. Database Setup

* **Database Creation**: The project starts by creating a database named Retail\_sales\_analysis.
* **Table Creation**: A table named retail\_sales is created to store the sales data. The table structure includes columns for transaction ID, sale date, sale time, customer ID, gender, age, product category, quantity sold, price per unit, cost of goods sold (COGS), and total sale amount.

CREATE DATABASE IF NOT EXISTS Retail\_sales\_analysis;

USE Retail\_sales\_analysis;

DROP TABLE IF EXISTS retail\_sales;

CREATE TABLE retail\_sales (

transactions\_id INT PRIMARY KEY,

sale\_date DATE,

sale\_time TIME,

customer\_id INT,

gender VARCHAR(15),

age INT,

category VARCHAR(50),

quantiy INT,

price\_per\_unit FLOAT,

cogs FLOAT,

total\_sale FLOAT

);

### 2. Data Exploration & Cleaning

* **Record Count**: Determine the total number of records in the dataset.
* **Customer Count**: Find out how many unique customers are in the dataset.
* **Category Count**: Identify all unique product categories in the dataset.
* **Null Value Check**: Check for any null values in the dataset and delete records with missing data.

-- Count total rows

SELECT COUNT(\*) AS total\_rows FROM retail\_sales;

-- Find rows with NULLs

SELECT \* FROM retail\_sales

WHERE transactions\_id IS NULL

OR sale\_date IS NULL

OR sale\_time IS NULL

OR gender IS NULL

OR category IS NULL

OR quantiy IS NULL

OR cogs IS NULL

OR total\_sale IS NULL;

-- Delete rows with NULLs

DELETE FROM retail\_sales

WHERE transactions\_id IS NULL

OR sale\_date IS NULL

OR sale\_time IS NULL

OR gender IS NULL

OR category IS NULL

OR quantiy IS NULL

OR cogs IS NULL

OR total\_sale IS NULL;

-- Unique customers

SELECT COUNT(DISTINCT customer\_id) AS unique\_customers FROM retail\_sales;

-- Categories

SELECT DISTINCT category FROM retail\_sales;

### 3. Data Analysis & Findings

The following SQL queries were developed to answer specific business questions:

1. **Sales made on 2022-11-05**

SELECT \*

FROM retail\_sales

WHERE sale\_date = '2022-11-05';

2. **Clothing sales with quantity > 4 in Nov-2022**

SELECT \*

FROM retail\_sales

WHERE category = 'Clothing'

AND DATE\_FORMAT(sale\_date, '%m-%y') = '11-2022'

AND quantiy > 3;

3. **Total sales per category**

SELECT

category,

SUM(total\_sale) AS net\_sale,

COUNT(\*) AS total\_orders

FROM retail\_sales

GROUP BY category;

-- Q4: **Avg age of Beauty customers**

SELECT

ROUND(AVG(age), 2) AS avg\_age

FROM retail\_sales

WHERE category = 'Beauty';

-- Q5: **Transactions > 1000**

SELECT \*

FROM retail\_sales

WHERE total\_sale > 1000;

-- Q6: **Transactions by gender & category**

SELECT

category,

gender,

COUNT(\*) AS total\_trans

FROM retail\_sales

GROUP BY category, gender

ORDER BY category;

-- Q7: **Best-selling month each year (by avg sale)**

SELECT year, month, avg\_sale

FROM (

SELECT

YEAR(sale\_date) AS year,

MONTH(sale\_date) AS month,

AVG(total\_sale) AS avg\_sale,

RANK() OVER (PARTITION BY YEAR(sale\_date) ORDER BY AVG(total\_sale) DESC) AS rnk

FROM retail\_sales

GROUP BY YEAR(sale\_date), MONTH(sale\_date)

) t

WHERE rnk = 1;

-- Q8: **Top 5 customers by sales**

SELECT

customer\_id,

SUM(total\_sale) AS total\_sales

FROM retail\_sales

GROUP BY customer\_id

ORDER BY total\_sales DESC

LIMIT 5;

-- Q9: **Unique customers per category**

SELECT

category,

COUNT(DISTINCT customer\_id) AS cnt\_unique\_customers

FROM retail\_sales

GROUP BY category;

-- Q10: **Orders by shift**

SELECT shift, COUNT(\*) AS total\_orders

FROM (

SELECT \*,

CASE

WHEN HOUR(sale\_time) < 12 THEN 'Morning'

WHEN HOUR(sale\_time) BETWEEN 12 AND 17 THEN 'Afternoon'

ELSE 'Evening'

END AS shift

FROM retail\_sales

) t

GROUP BY shift;

## Findings

* **Customer Demographics**: The dataset includes customers from various age groups, with sales distributed across different categories such as Clothing and Beauty.
* **High-Value Transactions**: Several transactions had a total sale amount greater than 1000, indicating premium purchases.
* **Sales Trends**: Monthly analysis shows variations in sales, helping identify peak seasons.
* **Customer Insights**: The analysis identifies the top-spending customers and the most popular product categories.

## Reports

* **Sales Summary**: A detailed report summarizing total sales, customer demographics, and category performance.
* **Trend Analysis**: Insights into sales trends across different months and shifts.
* **Customer Insights**: Reports on top customers and unique customer counts per category.

## Conclusion

This project serves database setup, data cleaning, exploratory data analysis, and business-driven SQL queries. The findings from this project can help drive business decisions by understanding sales patterns, customer behavior, and product performance.