**Retail Sales Analysis**

**Overview**

**Project Title:** Retail Sales Analysis

**Field:** Data Analytics

This is a project which is perfomed to explore, clean and analyze the retail sales data. The project involves creating a database, importing, performing exploratory data analysis (EDA) and answering specific business questions through SQL queries. This Provides basic insight on how queries help in solving the business problems.

**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.

**Details**

**1. Creating Database**

- Database Creation: A database named `retailsales` is created.

- Table Creation: A table named `retailsales` 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 retailsales;

CREATE TABLE retailsales

(

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 in the dataset.

SELECT COUNT(\*) FROM retailsales;

**Customer Count:** Find out how many unique customers are in the dataset.

SELECT COUNT(DISTINCT customer\_id) AS Customer\_Count FROM retailsales;

**Category Count:** Identify all unique product categories in the dataset.

SELECT DISTINCT category AS Categories FROM retailsales;

**Null Value Check:** Check for any null values in the dataset and delete records with missing data.

SELECT \* FROM retailsales

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

quantiy IS NULL

OR

price\_per\_unit IS NULL

OR

cogs IS NULL

OR

total\_sale IS NULL;

DELETE FROM retailsales

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

quantiy IS NULL

OR

price\_per\_unit IS NULL

OR

cogs IS NULL

OR

total\_sale IS NULL;

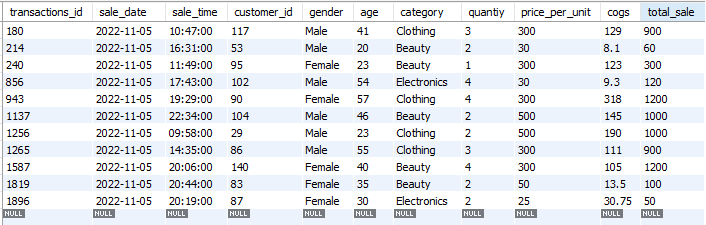
**3. Data Analysis & Findings**

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

**1. Write a SQL query to retrieve all columns for sales made on 2022-11-05**

SELECT \* FROM retailsales

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



**2. Write a SQL query to retrieve all transactions where category is 'clothing' and the quantity sold is more than 3 in month of nov 2022**

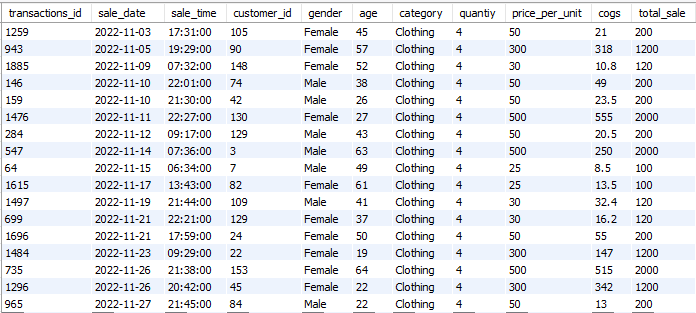
SELECT \* FROM retailsales

WHERE category = 'Clothing'

AND quantiy > 3

AND sale\_date BETWEEN '2022-11-01' AND '2022-11-30'

ORDER BY sale\_date;



**3. Write a SQL query to calculate the total sales (total\_sales) for each category.**

SELECT

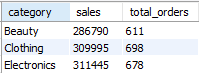
category,

SUM(total\_sale) AS sales,

COUNT(\*) AS total\_orders

FROM retailsales

GROUP BY category;



**4. Write a SQL query to find the average age of customers who puschased items from the 'Beauty' category.**

SELECT AVG(age) FROM retailsales

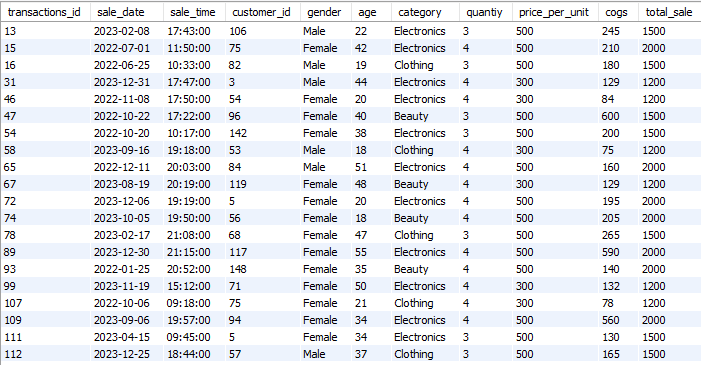
WHERE category = 'Beauty';



**5. Write a SQL query to find all transactions where the total\_sales is greater than 1000.**

SELECT \* FROM retailsales

WHERE total\_sale > 1000;



**6. Write a SQL query to find the total number of transactions (transaction\_id) made by each gender in each category.**

SELECT

gender,

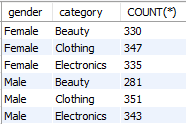
category,

COUNT(\*)

FROM retailsales

GROUP BY gender,category

ORDER BY gender;



**7. Write a SQL query to calculate the average sale for each month. Find out best selling month in each year.**

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 pos

FROM retailsales

GROUP BY year, month

) AS T1

WHERE pos = 1;



**8. Write SQL query to find top 5 customers based on the highest total sales**

SELECT

customer\_id,

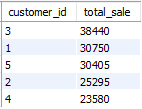
SUM(total\_sale) AS total\_sale

FROM retailsales

GROUP BY customer\_id

ORDER BY total\_sale DESC

LIMIT 5;



**9. Write SQL query to find the number of unique customers who purchased items from each category.**

SELECT

COUNT(DISTINCT customer\_id) AS unique\_cust,

category

FROM retailsales

GROUP BY category

ORDER BY unique\_cust DESC;



**10. Write a SQL query to create each shift and number of orders (Ex. Morning <=12, Afternoon Between 12 and 17, Evening >17)**

WITH hourly\_sales

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 retailsales

)

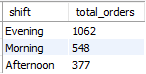
SELECT

shift,

COUNT(\*) AS total\_orders

FROM hourly\_sales

GROUP BY shift;



**4. 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.

**Conclusion**

This is a basic SQL Data Analysis comprising of database setup, data cleaning, exploratory data analytics(EDA) and business-driven SQL queries. Findings from this projects will help solve the business problems and data driven decision making.

**How to use.**

- Download the Folder named 'Retail Sales Analysis'.

- Unzip the file.

- Open the file named 'myfile.sql'.