**Project Overview**

**Project Title**: Retail Sales Analysis  
**Database**: retail\_sales

This project is designed to demonstrate SQL skills and techniques typically used by data analysts to explore, clean, and analyze retail sales data. The project involves setting up a retail sales database, performing exploratory data analysis (EDA), and answering specific business questions through SQL queries.

**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 p1\_retail\_db.
* **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 TABLE retail\_sales(

transactions\_id INT primary key,

sale\_date date,

sale\_time time,

customer\_id FLOAT,

gender VARCHAR(50),

age INT,

category VARCHAR(50),

quantiy FLOAT,

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.

SELECT COUNT(\*) FROM retail\_sales;

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

SELECT DISTINCT category FROM retail\_sales;

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**

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 retail\_sales

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

**2.Write a SQL query to retrieve all transactions where the category is**

**'Clothing' and the quantity sold is more than 4 in the month of Nov-2022:**

select \* from retail\_sales

where category = 'Clothing' and quantiy >=4 and to\_char(sale\_date,'YYYY-MM') ='2022-11' ;

**3.Write a SQL query to calculate the total sales and total count for each category.**

SELECT category , SUM(total\_sale) AS net\_sale,

COUNT(\*) as total\_orders

FROM retail\_sales

GROUP BY category ;

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

select avg(age) as avg\_age

from retail\_sales

where category = 'Beauty';

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

select transactions\_id , total\_sale

from retail\_sales

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 retail\_sales

group by gender,category;

**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

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 rank

FROM retail\_sales

GROUP BY 1, 2

) as t1

WHERE rank = 1

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

select customer\_id , sum(total\_sale) as net\_sale

from retail\_sales

group by customer\_id

order by net\_sale desc

limit 5 ;

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

SELECT category,

COUNT(DISTINCT customer\_id) as cnt\_unique\_cs

FROM retail\_sales

GROUP BY category;

**10.Write a SQL query to create each shift and number of orders**

**(Example Morning <12, Afternoon Between 12 & 17, Evening >17):**

select case

when extract(hour from sale\_time ) < 12 then 'Morning'

when extract (hour from sale\_time) between 12 and 17 then 'evening'

else 'Evening'

End as shift,

count(\*) as total\_order

from retail\_sales

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.