

# SQL Project: Retail Sales Analysis

---

 Created by: Md Nadim ( MS Banking and Financial Analytics - JMI )

 Database: dbbb

 Date: April 2025

## Database Setup

```
CREATE DATABASE dbbb;
USE dbbb;
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(100),
    quantity INT,
    price_per_unit FLOAT,
    cogs FLOAT,
    total_sale FLOAT
);
```

## Basic Data Checks

```
-- Record Count
SELECT COUNT(*) FROM retail_sales;
```

Output:

COUNT(*)
1987

```
-- Unique Customers  
SELECT COUNT(DISTINCT customer_id) AS unique_cust FROM retail_sales;
```

Output:

unique_cust
155

```
-- Unique Categories  
SELECT COUNT(DISTINCT category) AS unique_category FROM retail_sales;
```

Output:

unique_category
3

```
-- Null Check  
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;
```

```
-- Deleting Nulls  
SET SQL_SAFE_UPDATES = 1;  
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;
```

## Data Analysis Queries

### Q1: Sales made on '2022-11-05'

```
SELECT * FROM retail_sales  
WHERE sale_date='2022-11-05';
```

Output:

	transactions_id	sale_date	sale_time	customer_id	gender	age	category	quantity	price_per_unit	cogs	total_sale
1137	2022-11-05	22:34:00	104	Male	46	Beauty	2	500	145	1000	
943	2022-11-05	19:29:00	90	Female	57	Clothing	4	300	318	1200	
856	2022-11-05	17:43:00	102	Male	54	Electronics	4	30	9.3	120	
240	2022-11-05	11:49:00	95	Female	23	Beauty	1	300	123	300	
214	2022-11-05	16:31:00	53	Male	20	Beauty	2	30	8.1	60	
180	2022-11-05	10:47:00	117	Male	41	Clothing	3	300	129	900	
*	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

### Q2: Clothing category, quantity > 4 in November 2022

```
SELECT * FROM retail_sales
WHERE category ='clothing' AND DATE_FORMAT(sale_date, '%Y-%m')='2022-11'
AND quantity >= 4;
```

Output:

	transactions_id	sale_date	sale_time	customer_id	gender	age	category	quantity	price_per_unit	cogs	total_sale
▶	64	2022-11-15	06:34:00	7	Male	49	Clothing	4	25	8.5	100
	146	2022-11-10	22:01:00	74	Male	38	Clothing	4	50	49	200
	159	2022-11-10	21:30:00	42	Male	26	Clothing	4	50	23.5	200
	284	2022-11-12	09:17:00	129	Male	43	Clothing	4	50	20.5	200
	547	2022-11-14	07:36:00	3	Male	63	Clothing	4	500	250	2000
	699	2022-11-21	22:21:00	129	Female	37	Clothing	4	30	16.2	120
	735	2022-11-26	21:38:00	153	Female	64	Clothing	4	500	515	2000

### Q3: Total sales by category

```
SELECT category, SUM(total_sale) as net_sale, COUNT(*) AS total_orders
FROM retail_sales
GROUP BY category;
```

Output:

	category	net_sale	total_orders
▶	Beauty	286790	611
	Clothing	309995	698
	Electronics	311445	678

#### Q4: Average age of customers in 'Beauty' category

```
SELECT ROUND(AVG(age), 2) as avg_age  
FROM retail_sales  
WHERE category='Beauty';
```

Output:

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	avg_age			
▶	40.42			

#### Q5: Transactions with total\_sale > 1000

```
SELECT * FROM retail_sales  
WHERE total_sale > 1000;
```

Output:

Result Grid												Filter Rows:	Edit:	Export/Import:	Wrap Cell Content:
	transactions_id	sale_date	sale_time	customer_id	gender	age	category	quantity	price_per_unit	cogs	total_sale				
▶	13	2023-02-08	17:43:00	106	Male	22	Electronics	3	500	245	1500				
	15	2022-07-01	11:50:00	75	Female	42	Electronics	4	500	210	2000				
	16	2022-06-25	10:33:00	82	Male	19	Clothing	3	500	180	1500				
	31	2023-12-31	17:47:00	3	Male	44	Electronics	4	300	129	1200				
	46	2022-11-08	17:50:00	54	Female	20	Electronics	4	300	84	1200				
	47	2022-10-22	17:22:00	96	Female	40	Beauty	3	500	600	1500				
	54	2022-10-20	10:17:00	142	Female	38	Electronics	3	500	200	1500				

#### Q6: Total transactions by gender in each category

```
SELECT category, gender, COUNT(*) AS total_trans  
FROM retail_sales  
GROUP BY category, gender  
ORDER BY category;
```

Output:

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	category	gender	total_trans		
▶	Beauty	Female	330		
	Beauty	Male	281		
	Clothing	Female	347		
	Clothing	Male	351		
	Electronics	Female	335		
	Electronics	Male	343		

### Q7: Average sale for each month and best-selling month

```
SELECT  
    EXTRACT(YEAR FROM sale_date) AS year,  
    EXTRACT(MONTH FROM sale_date) AS month,  
    AVG(total_sale) AS avg_sale  
FROM retail_sales  
GROUP BY year, month;
```

Output:

	year	month	avg_sale
▶	2022	12	460.7692307692308
	2022	6	481.3953488372093
	2023	8	495.96491228070175
	2023	9	462.73972602739724
	2023	11	453.45238095238096
	2023	7	427.67857142857144
	2022	10	467.13793103448273

### Q8: Top 5 customers by total sales

```
SELECT customer_id, SUM(total_sale) AS total_sale  
FROM retail_sales  
GROUP BY customer_id  
ORDER BY total_sale DESC  
LIMIT 5;
```

Output:

	customer_id	total_sale
▶	3	38440
	1	30750
	5	30405
	2	25295
	4	23580

### Q9: Unique customers per category

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

Output:

Result Grid | Filter Rows: Export: Wrap Cell Content:

	category	uniq_cust
▶	Beauty	141
	Clothing	149
	Electronics	144

### Q10: Shifts and number of orders

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

Output:

Result Grid | Filter Rows: Export: Wrap Cell Content:

	Shift	num_of_orders
▶	evening	1062
	morning	548
	afternoon	377

### Summary & Findings

#### 1. Customer Demographics:

People of all age groups are shopping — from young adults to seniors, buying clothing and beauty products.

#### 2. High-Value Transactions:

Several transactions crossed ₹1000, indicating a few high-spending customers.

#### 3. Sales Trends:

Sales vary across months, helping to identify peak shopping seasons.

#### 4. Customer Insight:

The data reveals the most frequent and highest-spending customers, as well as the most popular product categories.

#### 5. Reporting:

- Sales Summary: Shows total sales, popular categories, and top customers.
- Trend Analysis: Helps identify monthly patterns and plan promotions accordingly.