

SQL Data Analysis Portfolio Project

■ STEP 1: Create Table

```
CREATE TABLE retail_sales (transactions_id INT, sale_date DATE, sale_time TIME, customer_id INT, gender VARCHAR(5),  
age INT, category VARCHAR(20), quantiy INT, price_per_unit FLOAT, cogs FLOAT, total_sale FLOAT  
);
```

The screenshot shows a database interface with a tree view of tables. Under 'Tables (1)', there is a single entry for 'retail_sales'. Expanding this entry reveals 'Columns (11)' which list all the fields defined in the CREATE TABLE statement.

■ STEP 2: Import CSV Data

```
COPY (  
    transactions_id, sale_date, sale_time, customer_id, gender, age, category, quantiy, price_per_unit, cogs,  
    total_sale  
)  
FROM 'D:\SQL - Retail Sales Analysis_utf.csv'  
DELIMITER ','  
CSV HEADER;
```

The screenshot shows a PostgreSQL client window with multiple tabs. The active tab is titled 'public.retail_sales/sql_p1/postgres@PostgreSQL 17'. The query pane contains the following SQL code:

```
1 SELECT * FROM public.retail_sales  
2 LIMIT 100  
3
```

The results pane displays the data from the 'retail_sales' table, showing 100 rows of sales information. The columns are: transactions_id, sale_date, sale_time, customer_id, gender, age, category, quantiy, price_per_unit, cogs, and total_sale. The data includes various purchase details such as dates ranging from 2022-01-06 to 2022-11-23, and categories like Clothing, Beauty, and Electronics.

■ STEP 3: Initial Exploration

```
SELECT * FROM retail_sales;  
SELECT COUNT(*) FROM retail_sales;
```

-- STEP 4: Null Check & Data Cleaning

```
SELECT *  
FROM 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 quantiy IS NULL  
OR price_per_unit IS NULL  
OR cogs IS NULL  
OR total_sale IS NULL;
```

```
DELETE FROM 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 quantiy IS NULL  
OR price_per_unit IS NULL  
OR cogs IS NULL  
OR total_sale IS NULL;
```

Cleaned Table removing rows having null rows

Data Output Messages Notifications

SQL

	transactions_id	sale_date	sale_time	customer_id	gender
1	180	2022-11-05	10:47:00	117	Male
2	522	2022-07-09	11:00:00	52	Male
3	559	2022-12-12	10:48:00	5	Femal
4	1180	2022-01-06	08:53:00	85	Male
5	1522	2022-11-14	08:35:00	48	Male
6	1559	2022-08-20	07:40:00	49	Femal
7	163	2022-10-31	09:38:00	144	Femal
8	303	2022-04-22	11:09:00	54	Male
9	421	2022-04-08	08:43:00	66	Femal
10	979	2022-05-18	10:18:00	6	Femal
11	1163	2022-05-04	10:52:00	120	Femal
12	1303	2022-03-19	08:59:00	58	Male
13	1421	2022-01-17	07:07:00	59	Femal
14	1979	2022-08-17	11:34:00	102	Femal
15	610	2022-12-18	06:56:00	137	Femal
16	1610	2022-11-23	10:18:00	1	Femal
17	32	2022-07-16	09:11:00	150	Male
18	231	2022-07-09	07:02:00	50	Femal
19	683	2022-03-06	10:22:00	82	Male
20	1032	2022-04-01	08:15:00	1	Male
21	1231	2022-01-29	07:05:00	12	Femal
22	1683	2022-05-04	07:19:00	98	Male
23	367	2022-12-30	07:03:00	123	Femal
24	391	2022-03-25	07:37:00	68	Male
25	1797	2022-04-16	11:47:00	147	Male
26	1001	2022-02-28	07:01:00	77	Male

Total rows: 1987 | Query complete 00:00:00.155

■ STEP 5: Basic Metrics

```
SELECT SUM(quantity * price_per_unit) FROM retail_sales;
SELECT COUNT(DISTINCT customer_id) AS customer_id FROM retail_sales;
SELECT COUNT(DISTINCT category) AS unique_catcategory FROM retail_sales;
```

sum	double precision
908230	
customer_id	bigint
1	155
unique_catcategory	bigint
1	3

■ STEP 6: Business Questions & Analysis

-- Q1: Sales on a Specific Date

```
SELECT *
```

```
FROM retail_sales
```

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

	transactions_id	sale_date	sale_time	customer_id	gender	age	category	quantity	price_per_unit	cogs	total_sale
	integer	date	time without time zone	integer	character varying (15)	integer	character varying (20)	integer	double precision	double precision	double precision
1	180	2022-11-05	10:47:00	117	Male	41	Clothing	3	300	129	900
2	240	2022-11-05	11:49:00	95	Female	23	Beauty	1	300	123	300
3	1256	2022-11-05	09:58:00	29	Male	23	Clothing	2	500	190	1000
4	1587	2022-11-05	20:06:00	140	Female	40	Beauty	4	300	105	1200
5	1819	2022-11-05	20:44:00	83	Female	35	Beauty	2	50	13.5	100
6	943	2022-11-05	19:29:00	90	Female	57	Clothing	4	300	318	1200
7	1896	2022-11-05	20:19:00	87	Female	30	Electronics	2	25	30.75	50
8	1137	2022-11-05	22:34:00	104	Male	46	Beauty	2	500	145	1000
9	856	2022-11-05	17:43:00	102	Male	54	Electronics	4	30	9.3	120
10	214	2022-11-05	16:31:00	53	Male	20	Beauty	2	30	8.1	60
11	1265	2022-11-05	14:35:00	86	Male	55	Clothing	3	300	111	900

-- Q2: Clothing Sales >10 Units in Nov-2022

```
SELECT *
```

```
FROM retail_sales
```

```
WHERE category = 'Clothing'
```

```
AND TO_CHAR(sale_date, 'yyyy-mm') = '2022-11'
```

```
AND quantity = 4;
```

	transactions_id	sale_date	sale_time	customer_id	gender	age	category	quantity	price_per_unit	cogs	total_sale
	integer	date	time without time zone	integer	character varying (15)	integer	character varying (20)	integer	double precision	double precision	double precision
1	1484	2022-11-23	09:29:00	22	Female	19	Clothing	4	300	147	1200
2	64	2022-11-15	06:34:00	7	Male	49	Clothing	4	25	8.5	100
3	284	2022-11-12	09:17:00	129	Male	43	Clothing	4	50	20.5	200
4	1885	2022-11-09	07:32:00	148	Female	52	Clothing	4	30	10.8	120
5	547	2022-11-14	07:36:00	3	Male	63	Clothing	4	500	250	2000
6	159	2022-11-10	21:30:00	42	Male	26	Clothing	4	50	23.5	200
7	699	2022-11-21	22:21:00	129	Female	37	Clothing	4	30	16.2	120
8	1259	2022-11-03	17:31:00	105	Female	45	Clothing	4	50	21	200
9	146	2022-11-10	22:01:00	74	Male	38	Clothing	4	50	49	200
10	1476	2022-11-11	22:27:00	130	Female	27	Clothing	4	500	555	2000
11	1296	2022-11-26	20:42:00	45	Female	22	Clothing	4	300	342	1200
12	1604	2022-11-21	17:50:00	94	Female	50	Clothing	4	50	55	200

-- Q3: Total Sales by Category

SELECT

category,

SUM(total_sale) AS Total_Sales,

COUNT(*) AS Total_Orders

FROM retail_sales

GROUP BY category

ORDER BY SUM(total_sale) DESC;

	category character varying (20)	total_sales double precision	total_orders bigint
1	Electronics	311445	678
2	Clothing	309995	698
3	Beauty	286790	611

-- Q4: Average Age of Beauty Buyers

SELECT CAST(AVG(age) AS DECIMAL(10,2)) AS avg_age

FROM retail_sales

WHERE category = 'Beauty';

	avg_age numeric (10,2)
1	40.42

-- Q5: High-Value Transactions

SELECT *

FROM retail_sales

WHERE total_sale > 1000;

	transactions_id integer	sale_date date	sale_time time without time zone	customer_id integer	gender character varying (15)	age integer	category character varying (20)	quantity integer	price_per_unit double precision	cogs double precision	total_sale double precision
1	522	2022-07-09	11:00:00	52	Male	46	Beauty	3	500	145	1500
2	559	2022-12-12	10:48:00	5	Female	40	Clothing	4	300	84	1200
3	1522	2022-11-14	08:35:00	48	Male	46	Beauty	3	500	235	1500
4	1559	2022-08-20	07:40:00	49	Female	40	Clothing	4	300	144	1200
5	421	2022-04-08	08:43:00	66	Female	37	Clothing	3	500	235	1500
6	1421	2022-01-17	07:07:00	59	Female	37	Clothing	3	500	185	1500
7	484	2022-03-13	07:52:00	135	Female	19	Clothing	4	300	75	1200
8	1484	2022-11-23	09:29:00	22	Female	19	Clothing	4	300	147	1200
9	15	2022-07-01	11:50:00	75	Female	42	Electronics	4	500	210	2000
10	743	2022-08-07	07:54:00	55	Female	34	Beauty	4	500	260	2000
11	1015	2022-03-09	11:53:00	94	Female	42	Electronics	4	500	200	2000
12	1743	2022-10-26	09:37:00	47	Female	34	Beauty	4	500	250	2000
13	742	2022-03-19	06:08:00	37	Female	38	Electronics	4	500	195	2000
14	1742	2022-11-22	08:25:00	18	Female	38	Electronics	4	500	220	2000
15	420	2022-01-02	10:53:00	28	Female	22	Clothing	4	500	200	2000
16	1420	2022-04-15	07:01:00	138	Female	22	Clothing	4	500	205	2000
17	592	2022-12-26	09:15:00	77	Female	46	Beauty	4	500	275	2000
18	1592	2022-03-16	09:08:00	81	Female	46	Beauty	4	500	155	2000
19	720	2022-04-08	08:50:00	116	Female	56	Beauty	3	500	235	1500
20	1720	2022-10-10	08:51:00	28	Female	56	Beauty	3	500	190	1500
21	269	2022-09-19	11:31:00	87	Male	25	Clothing	4	500	250	2000
22	320	2022-04-20	08:35:00	57	Female	28	Electronics	4	300	159	1200
23	673	2022-07-04	10:14:00	18	Female	43	Clothing	3	500	270	1500
24	1269	2022-01-01	08:09:00	71	Male	25	Clothing	4	500	145	2000

-- Q6: Transactions by Gender & Category

SELECT

```
category,  
gender,  
COUNT(*) AS no_of_Transctions
```

FROM retail_sales

GROUP BY gender, category

ORDER BY category;

	category character varying (20) 	gender character varying (15) 	no_of_transctions bigint 
1	Beauty	Female	330
2	Beauty	Male	281
3	Clothing	Female	347
4	Clothing	Male	351
5	Electronics	Male	343
6	Electronics	Female	335

-- Q7: Best Selling Month by Year

WITH CTE_Best_Selling_Month AS (

SELECT

```
TO_CHAR(sale_date, 'YYYY') AS sales_year,  
TO_CHAR(sale_date, 'Mon') AS sale_month,  
CAST(AVG(total_sale) AS DECIMAL(10,2)) AS Average_Sales,  
RANK() OVER (  
    PARTITION BY TO_CHAR(sale_date, 'YYYY')  
    ORDER BY AVG(total_sale) DESC  
) AS Best_Selling_Month_rnk
```

FROM retail_sales

GROUP BY 1, 2)

SELECT *

FROM CTE_Best_Selling_Month

WHERE Best_Selling_Month_rnk = 1;

	sales_year text 	sale_month text 	average_sales numeric (10,2) 	best_selling_month_rnk bigint 
1	2022	Jul	541.34	1
2	2023	Feb	535.53	1

-- Q8: Top 5 Customers by Sales

WITH cte_top_5_cust AS (

SELECT

customer_id,

SUM(total_sale) AS sales_total,

RANK() OVER (ORDER BY SUM(total_sale) DESC) AS rnk

FROM retail_sales

GROUP BY customer_id

)

SELECT

customer_id,

sales_total,

rnk

FROM cte_top_5_cust

WHERE rnk <= 5;

	customer_id	sales_total	rnk
1	3	38440	1
2	1	30750	2
3	5	30405	3
4	2	25295	4
5	4	23580	5

-- Q9: Unique Customers per Category

SELECT

category,

COUNT(DISTINCT customer_id) AS unique_customers

FROM retail_sales

GROUP BY category;

	category	unique_customers
1	Beauty	141
2	Clothing	149
3	Electronics	144

-- Q10: Shift-Wise Order Distribution

```
WITH CTE_Shift_Wise_Orders AS (
    SELECT
        transactions_id,
        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 Working_shift
    FROM retail_sales
)
SELECT
    Working_shift,
    COUNT(transactions_id) AS no_of_orders
FROM CTE_Shift_Wise_Orders
GROUP BY Working_shift;
```

	working_shift	no_of_orders
1	Afternoon	377
2	Evening	1062
3	Morning	548

■ STEP 7: Insights & Conclusions

Insights:

1. Clothing and Beauty categories generate the highest revenue.
2. Afternoon shift records the most transactions, indicating peak shopping hours.
3. Top 5 customers contribute disproportionately to total sales—ideal for loyalty targeting.
4. November shows strong sales performance, especially in Clothing.
5. Beauty category attracts younger customers on average.
6. Gender-based purchasing patterns vary significantly across categories.

Conclusions:

1. The business can optimize staffing during afternoon hours to handle peak demand.
2. Marketing efforts should focus on high-value customers and top-selling months.
3. Product segmentation by age and gender can improve personalization.
4. Further analysis with time-series and visual dashboards (Power BI or Python) can enhance decision-making.