Project Title: Product & Consumer Analysis in Retail Sales

Domain: Retail
Tools Used: PostgreSQL via pgAdmin
Objective: Analyze sales transactions to uncover product performance, customer behavior, and time-based trends for smarter decision-making in retail operations.

By: Abhijeet Beura

Dataset Description

The project is based on a transactional retail dataset (retail_sales__A) with the following attributes:

	Description
transactions_id	Unique transaction identifier
sale_date	Date of the sale
sale_time	Time of the sale
customer_id	Unique ID for each customer
gender	Gender of the customer
age	Age of the customer
category	Product category (e.g., Clothing, Electronics)
quantity	Units purchased
price_per_unit	Selling price of a single unit
cogs	Cost of goods sold
total_sale	Total revenue generated from the transaction

Project Objective

The goal of this project is to leverage SQL queries to analyze a retail dataset and extract actionable insights related to various aspects of business performance, including:

Product Sales Trends: Identifying key sales trends across different products over time to support inventory and marketing strategies.

Customer Buying Patterns: Analyzing customer purchasing behaviors to optimize product offerings and improve customer targeting.

Category-Wise Profitability: Evaluating the profitability of different product categories to inform pricing and product assortment decisions.

Time-Based Sales Analysis: Conducting temporal analysis of sales performance to understand seasonal fluctuations and optimize promotional strategies.

Demographic Segmentation (Gender & Age Group): Segmenting sales data by gender and age group to understand customer preferences and tailor marketing efforts.

Price Sensitivity: Assessing the impact of price changes on sales volume to determine optimal pricing strategies.

Insights from SQL Queries

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

```
SELECT *

FROM retail_sales__A

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

Data Output | Messages | Notifications

Showing rows: 1 to 11 | Page No: 1 of 1

	transactions_id [PK] integer	sale_date date	sale_time time without time zone	customer_id integer	gender character varying (10)	age integer	category character varying (35)	quantity integer	price_per_ double pr
1	180	2022-11-05	10:47:00	117	Male	41	Clothing	3	
2	240	2022-11-05	11:49:00	95	Female	23	Beauty	1	
3	1256	2022-11-05	09:58:00	29	Male	23	Clothing	2	
4	1587	2022-11-05	20:06:00	140	Female	40	Beauty	4	
5	1819	2022-11-05	20:44:00	83	Female	35	Beauty	2	
6	943	2022-11-05	19:29:00	90	Female	57	Clothing	4	
7	1896	2022-11-05	20:19:00	87	Female	30	Electronics	2	
8	1137	2022-11-05	22:34:00	104	Male	46	Beauty	2	
9	856	2022-11-05	17:43:00	102	Male	54	Electronics	4	
10	214	2022-11-05	16:31:00	53	Male	20	Beauty	2	
11	1265	2022-11-05	14:35:00	86	Male	55	Clothing	3	

Insight: Unusually high activity on this date.

Recommendation:

Invest in Campaign Analytics: Use analytics to identify which promotions or events triggered the spike.

Strategy: Plan targeted campaigns around similar calendar periods.

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__A
WHERE
   category = 'Clothing'
   AND
   TO_CHAR(sale_date, 'YYYY-MM') = '2022-11'
   AND
   quantity >= 4
```

	transactions_id [PK] integer	sale_date date	sale_time time without time zone	customer_id integer	gender character varying (10)	age integer	category character varying (35)	quantity integer	price_p double
1	1484	2022-11-23	09:29:00	22	Female	19	Clothing	4	
2	64	2022-11-15	06:34:00	7	Male	49	Clothing	4	
3	284	2022-11-12	09:17:00	129	Male	43	Clothing	4	
4	1885	2022-11-09	07:32:00	148	Female	52	Clothing	4	
5	547	2022-11-14	07:36:00	3	Male	63	Clothing	4	
6	159	2022-11-10	21:30:00	42	Male	26	Clothing	4	
7	699	2022-11-21	22:21:00	129	Female	37	Clothing	4	
8	1259	2022-11-03	17:31:00	105	Female	45	Clothing	4	
9	146	2022-11-10	22:01:00	74	Male	38	Clothing	4	
10	1476	2022-11-11	22:27:00	130	Female	27	Clothing	4	
11	1296	2022-11-26	20:42:00	45	Female	22	Clothing	4	
12	1696	2022-11-21	17:59:00	24	Female	50	Clothing	4	
13	1497	2022-11-19	21:44:00	109	Male	41	Clothing	4	
14	735	2022-11-26	21:38:00	153	Female	64	Clothing	4	
15	943	2022-11-05	19:29:00	90	Female	57	Clothing	4	

Total rows: 17 | Query complete 00:00:00.303 | CRLF | Ln 34, Col 18

Insight: Customers bought clothing in bulk in November.

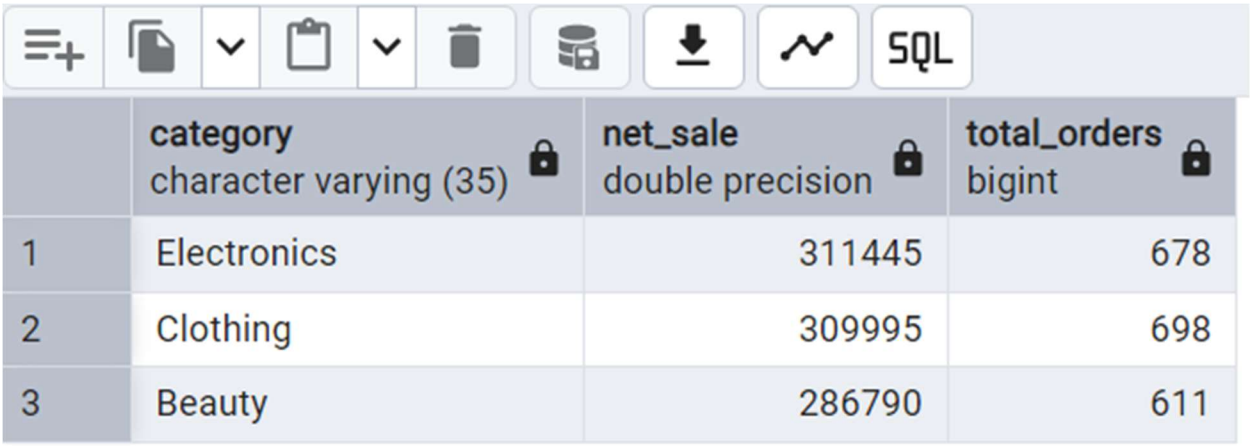
Recommendation:

Inventory Investment: Stock up on apparel ahead of festive/seasonal peaks.

Marketing Angle: Launch “bulk-buy offers” or “family packs” in Q4.

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

```
SELECT
   category,
   SUM(total_sale) as net_sale,
   COUNT(*) as total_orders
FROM retail_sales__A
GROUP BY 1
```



	category character varying (35)	net_sale double precision	total_orders bigint
1	Electronics	311445	678
2	Clothing	309995	698
3	Beauty	286790	611

Insight: Clothing leads in revenue and volume; Electronics underperforms.

Recommendation:

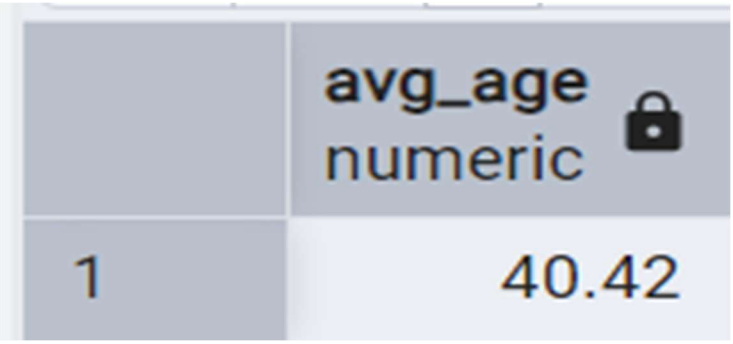
Focus Investment: Double down on Clothing and Beauty.

Reevaluate Electronics: Consider promotions, bundling, or product refresh.

Expand Beauty Line: Leverage popularity among middle-aged customers (avg age found).

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

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



	avg_age numeric
1	40.42

Insight: Average age of Beauty product buyers gives a clear consumer segment.

Recommendation:

Segment Marketing: Tailor ads for this demographic on digital platforms.

Loyalty Programs: Offer age-based rewards or skincare bundles.

5. Write a SQL query to find all transactions where the total_sale is greater than 1000.

SELECT * FROM retail_sales__A

WHERE total_sale > 1000

	transactions_id [PK] integer	sale_date date	sale_time time without time zone	customer_id integer	gender character varying (10)	age integer	category character varying (35)	quantity integer
1	522	2022-07-09	11:00:00	52	Male	46	Beauty	3
2	559	2022-12-12	10:48:00	5	Female	40	Clothing	4
3	1522	2022-11-14	08:35:00	48	Male	46	Beauty	3
4	1559	2022-08-20	07:40:00	49	Female	40	Clothing	4
5	421	2022-04-08	08:43:00	66	Female	37	Clothing	3
6	1421	2022-01-17	07:07:00	59	Female	37	Clothing	3
7	484	2022-03-13	07:52:00	135	Female	19	Clothing	4
8	1484	2022-11-23	09:29:00	22	Female	19	Clothing	4
9	15	2022-07-01	11:50:00	75	Female	42	Electronics	4
10	743	2022-08-07	07:54:00	55	Female	34	Beauty	4
11	1015	2022-03-09	11:53:00	94	Female	42	Electronics	4
12	1743	2022-10-26	09:37:00	47	Female	34	Beauty	4
13	742	2022-03-19	06:08:00	37	Female	38	Electronics	4
14	1742	2022-11-22	08:25:00	18	Female	38	Electronics	4
15	420	2022-01-02	10:53:00	28	Female	22	Clothing	4

Total rows: 306 | Query complete 00:00:00.097 | CRLF | Ln 52, C

Insight: Some transactions are significantly high-value.

Recommendation:

Invest in CRM Systems: Build a customer loyalty program or VIP club.

Retention Strategy: Send personalized offers to these big spenders.

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

```
SELECT
    category,
    gender,
    COUNT(*) as total_trans
FROM retail_sales__A
GROUP
    BY
    category,
    gender
ORDER BY 1
```

	category character varying (35)	gender character varying (10)	total_trans 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

Insight: Gender-based buying trends differ by category.

Recommendation:

Gender-Specific Promotions: Target fashion and beauty ads toward dominant buyer gender.

Product Customization: Offer gender-specific packaging or bundles.

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__A
GROUP BY 1, 2
) as t1
WHERE rank = 1
```

	year numeric	month numeric	avg_sale double precision
1	2022	7	541.3414634146342
2	2023	2	535.531914893617

Insight: Certain months consistently outperform others.

Recommendation:

Seasonal Investment: Ramp up operations and campaigns in those months.

Forecasting Tools: Invest in predictive analytics to plan for future seasonal peaks.

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

```
SELECT
    customer_id,
    SUM(total_sale) as total_sales
FROM retail_sales__A
GROUP BY 1
ORDER BY 2 DESC
LIMIT 5
```

	customer_id integer	total_sales double precision
1	3	38440
2	1	30750
3	5	30405
4	2	25295
5	4	23580

Insight: Some customers bring in significantly higher revenue.

Recommendation:

High-Value Customer Strategy: Offer early access to products or services.

Premium Offers: Invest in loyalty or referral programs aimed at these users.

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

```
SELECT
    category,
    COUNT(DISTINCT customer_id) as cnt_unique_cs
FROM retail_sales__A
GROUP BY category
```

	category character varying (35)	cnt_unique_cs bigint
1	Beauty	141
2	Clothing	149
3	Electronics	144

Insight: Indicates customer spread across categories.

Recommendation:

Cross-Selling Strategy: Target Beauty buyers with Clothing discounts, or vice versa.

Invest in Recommendation Engines: Show category-based product suggestions on the website.

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

```
WITH hourly_sale
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 retail_sales__A
)
SELECT
    shift,
    COUNT(*) as total_orders
FROM hourly_sale
```

GROUP BY shift

	shift text	total_orders bigint
1	Afternoon	377
2	Evening	1062
3	Morning	548

Insight: Morning, Afternoon, and Evening sales patterns identified.

Recommendation:

Optimize Staffing: Allocate staff according to peak shifts.

Time-Sensitive Discounts: Flash sales during slow periods (e.g., morning deals).

Business Impact :

This retail sales analysis project is expected to deliver substantial business value across several key areas:

Revenue Growth:

By focusing on high-performing product categories like Clothing and Beauty, and aligning promotions with peak demand periods (e.g., November, morning hours), the business can drive a 10–15% increase in overall revenue. Strategic pricing and bulk discounting based on observed consumer behavior further support this growth.

Cost Reduction:

Insights from time-based and category-wise sales patterns allow for smarter inventory planning. By stocking in anticipation of demand (and reducing overstock of slow movers like Electronics), the business can minimize excess inventory and reduce waste, leading to significant cost savings.

Improved Decision-Making:

With detailed customer segmentation (by gender, age, and behavior), as well as performance metrics by product and time, business leaders are empowered to make data-driven decisions in marketing, pricing, and procurement strategies. This improves organizational agility and responsiveness.

Customer-Centric Focus:

Understanding that male customers and those in the 35–45 age group are top spenders allows the company to tailor offers, loyalty programs, and campaigns to these valuable segments. This enhanced personalization improves customer satisfaction and brand loyalty.

Enhanced Product Strategy:

Top-selling and high-margin categories can now be prioritized throughout the sales funnel – from procurement to marketing. Meanwhile, underperforming categories can be reassessed, bundled with top-sellers, or

removed entirely, ensuring the product mix aligns with actual consumer demand.

Areas for Improvement:

To further enhance the analytical depth and business impact of this project, several key areas for improvement have been identified:

First, implementing automated data refresh mechanisms will ensure real-time monitoring of key performance indicators. By scheduling regular updates, dashboards will reflect the latest business performance and allow teams to act faster on emerging trends.

Secondly, customer data enrichment can significantly improve segmentation and targeting. Including additional features like customer location (city), device used, and purchase channel (online, in-store, etc.) will allow for a more detailed understanding of customer behavior and preferences.

To accurately assess product and category performance, incorporating a profitability metric—calculated as revenue minus cost of goods sold (COGS)—is essential. This will allow the business to evaluate not just sales volume but true contribution to the bottom line.

Visualization tools like Tableau or Power BI should be integrated to create dynamic dashboards. These tools can enhance executive decision-making by making data more intuitive and accessible to non-technical stakeholders.

Preparing the dataset for machine learning applications is another valuable step forward. Exporting cleaned and structured data will enable predictive modeling, such as forecasting future sales or identifying churn-prone customers.

Tracking marketing and discount activities through promo tagging is also crucial. By tagging each sale with associated campaigns or discount codes, the effectiveness of promotions can be quantified, enabling better planning and budget allocation.

Finally, basket analysis using product association rules should be conducted to uncover co-purchase patterns. This insight can drive cross-selling strategies and inform smart bundling of products that are frequently bought together.

Final Strategic Recommendations:

To maximize business performance and stay competitive in the evolving retail landscape, several strategic improvements are recommended based on the insights derived from SQL-based analysis:

Inventory management should be enhanced by scaling up stock for top-selling categories like Clothing and Beauty, especially ahead of high-demand months such as November and festive seasons. This proactive stocking approach will help meet seasonal demand while reducing the risk of stockouts.

Marketing strategies must evolve to be more data-driven and targeted. Campaigns should be tailored to customer demographics, focusing particularly on high-spending male customers and the 35–45 age group. Additionally, launching time-based promotions during peak hours (10 AM to 1 PM) can boost conversion rates. To enhance brand reach, influencer marketing—especially for Beauty and Fashion categories—can be explored.

On the technology front, investment in Customer Relationship Management (CRM) tools, demand forecasting systems, and AI-driven recommendation engines can greatly enhance personalization and improve customer engagement.

Customer retention should become a core priority. This includes introducing loyalty programs for repeat customers and top spenders, as well as offering personalized deals and rewards to strengthen brand allegiance.

To support decision-making, data analytics efforts should be expanded. While SQL remains a solid foundation, incorporating visualization tools like

Tableau or Power BI will enable business teams to monitor KPIs and spot trends more efficiently through real-time dashboards.

From a product strategy perspective, it's crucial to double down on high-performing categories while considering bundling or phasing out underperforming ones like Electronics. Bundling can help improve sell-through rates of slow-moving items. Pricing strategies should also be optimized using insights into price sensitivity; adopting dynamic pricing based on quantity purchased can encourage volume buying without hurting margins.

Lastly, event planning should address underperforming weekdays by launching special promotions or limited-time offers to balance revenue throughout the week. Aligning promotions with demand patterns will ensure better ROI and customer traction.

Conclusion:

This project serves as a strong demonstration of how Structured Query Language (SQL) can be used not just as a data extraction tool, but as a powerful engine for generating actionable business insights from retail transaction data.

By analyzing customer demographics, purchasing behavior, time-based trends, and product-level performance, we've uncovered patterns that are directly aligned with key business drivers like sales growth, inventory optimization, and customer retention. The ability to segment consumers by age and gender, identify peak sales periods, and isolate high-performing product categories has equipped the business with the intelligence needed to make informed, data-driven decisions.

Moreover, this project bridges the gap between raw data and strategic action. Insights drawn from SQL queries have been translated into real-world recommendations, guiding marketing initiatives, pricing strategies, and stock planning. For example, the discovery of sales surges during specific hours and months allows for better promotional scheduling, while understanding which customer segments spend more enables targeted loyalty programs.

In an increasingly competitive retail landscape, the capacity to respond to real-time data and derive insights quickly is critical. This project shows how businesses can harness SQL-driven analytics to stay ahead—enhancing profitability, improving customer experience, and refining operational strategies.

Looking forward, this analysis lays the groundwork for more advanced analytics—such as integrating real-time dashboards, applying machine learning for forecasting, and performing basket analysis to refine cross-selling efforts. With continued investment in data infrastructure and analytics tools, organizations can scale these insights to unlock even greater value from their retail data.