Project Report: Retail Sales Analysis

Project Title:

Retail Sales Analysis

Problem Statement:

To analyze retail sales data to gain insights into customer behavior, product performance, and sales trends using SQL while addressing specific business questions to guide decision-making.

Objectives:

1. Set up a retail sales database: Create and populate a retail sales database.
2. Data Cleaning: Identify and remove records with missing or null values.
3. Exploratory Data Analysis (EDA): Perform basic analysis to understand the dataset.
4. Business Analysis: Address specific business questions using SQL queries to derive actionable insights.

Tools and Technologies:

1. **Database Management System:** MySQL
2. **Programming Language:** SQL
3. **Database Design Tools:** SQL client (e.g., pgAdmin)

Description:

This project uses a retail sales dataset to build SQL skills, focusing on data cleaning, exploration, and business-driven analysis. The dataset includes transaction details, customer demographics, product categories, and sales metrics.

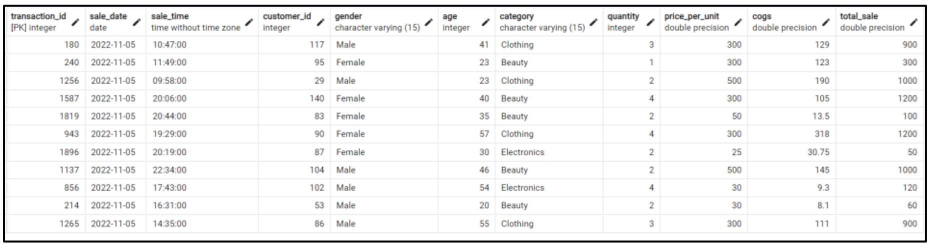
Methodology:

1. **Database Setup:**
   - Created a database named retail_sales.
   - Created a table retail_sales with columns for transaction ID, sale details, customer demographics, product category, and sales metrics.
2. **Data Cleaning and Exploration:**
   - Checked for missing/null values and removed records with incomplete data.
   - Counted total records, unique customers, and product categories.
3. **Business Analysis:**
   Addressed the following business questions using SQL:

Business Questions and Queries:

1. **Retrieve all columns for sales made on '2022-11-05':**

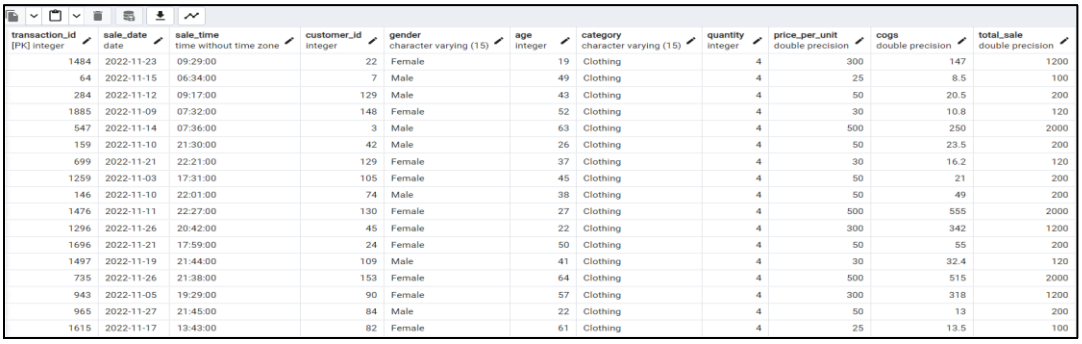
```
SELECT * FROM retail_sales WHERE sale_date = '2022-11-05';
```



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

2. **Retrieve all transactions where the category is 'Clothing' and the quantity sold is more than 4 in November 2022:**

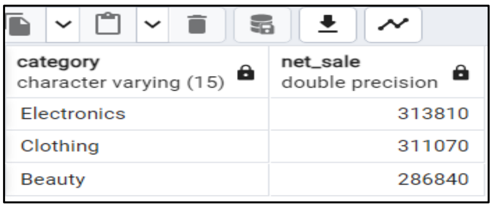
```
SELECT *

FROM retail_sales

WHERE category = 'Clothing'

  AND TO_CHAR(sale_date, 'YYYY-MM') = '2022-11'

  AND quantity > 4;
```



transaction_id [PK] integer	sale_date date	sale_time time without time zone	customer_id integer	gender character varying (15)	age integer	category character varying (15)	quantity integer	price_per_unit double precision	cogs double precision	total_sale double precision
1484	2022-11-23	09:29:00	22	Female	19	Clothing	4	300	147	1200
64	2022-11-15	06:34:00	7	Male	49	Clothing	4	25	8.5	100
284	2022-11-12	09:17:00	129	Male	43	Clothing	4	50	20.5	200
1885	2022-11-09	07:32:00	148	Female	52	Clothing	4	30	10.8	120
547	2022-11-14	07:36:00	3	Male	63	Clothing	4	500	250	2000
159	2022-11-10	21:30:00	42	Male	26	Clothing	4	50	23.5	200
699	2022-11-21	22:21:00	129	Female	37	Clothing	4	30	16.2	120
1259	2022-11-03	17:31:00	105	Female	45	Clothing	4	50	21	200
146	2022-11-10	22:01:00	74	Male	38	Clothing	4	50	49	200
1476	2022-11-11	22:27:00	130	Female	27	Clothing	4	500	555	2000
1296	2022-11-26	20:42:00	45	Female	22	Clothing	4	300	342	1200
1696	2022-11-21	17:59:00	24	Female	50	Clothing	4	50	55	200
1497	2022-11-19	21:44:00	109	Male	41	Clothing	4	30	32.4	120
735	2022-11-26	21:38:00	153	Female	64	Clothing	4	500	515	2000
943	2022-11-05	19:29:00	90	Female	57	Clothing	4	300	318	1200
965	2022-11-27	21:45:00	84	Male	22	Clothing	4	50	13	200
1615	2022-11-17	13:43:00	82	Female	61	Clothing	4	25	13.5	100

3. **Calculate the total sales (total_sale) for each category:**

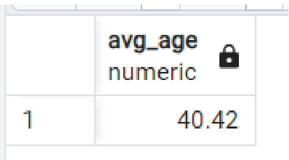
```
SELECT category, SUM(total_sale) AS net_sale, COUNT(*) AS total_orders

FROM retail_sales

GROUP BY category;
```



category character varying (15)	net_sale double precision
Electronics	313810
Clothing	311070
Beauty	286840

4. **Find the average age of customers who purchased items from the 'Beauty' category:**

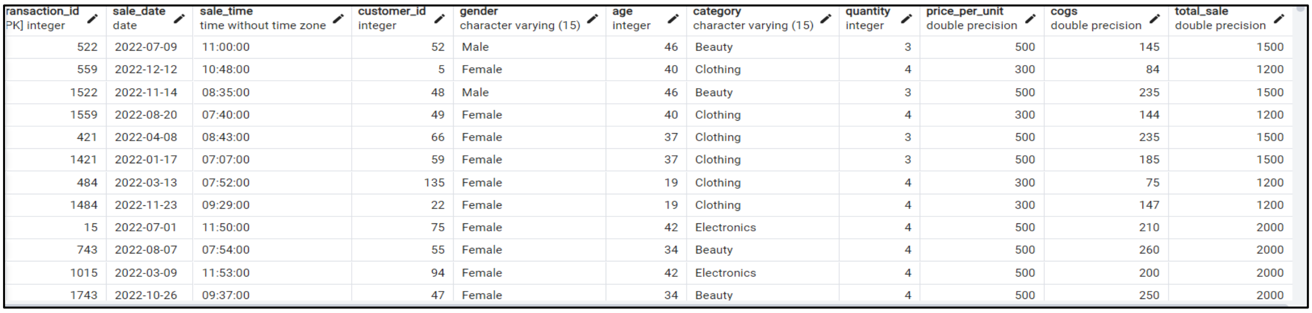
```
SELECT ROUND(AVG(age), 2) AS avg_age

FROM retail_sales

WHERE category = 'Beauty';
```



	avg_age
	numeric
1	40.42

5. Find all transactions where the total_sale is greater than 1000:

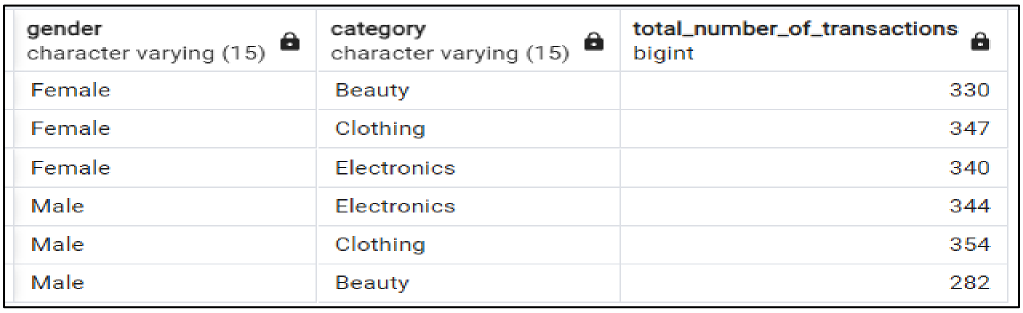
```
SELECT *

FROM retail_sales

WHERE total_sale > 1000;
```



ransaction_id	sale_date	sale_time	customer_id	gender	age	category	quantity	price_per_unit	cogs	total_sale
PK] integer	date	time without time zone	integer	character varying (15)	integer	character varying (15)	integer	double precision	double precision	double precision
522	2022-07-09	11:00:00	52	Male	46	Beauty	3	500	145	1500
559	2022-12-12	10:48:00	5	Female	40	Clothing	4	300	84	1200
1522	2022-11-14	08:35:00	48	Male	46	Beauty	3	500	235	1500
1559	2022-08-20	07:40:00	49	Female	40	Clothing	4	300	144	1200
421	2022-04-08	08:43:00	66	Female	37	Clothing	3	500	235	1500
1421	2022-01-17	07:07:00	59	Female	37	Clothing	3	500	185	1500
484	2022-03-13	07:52:00	135	Female	19	Clothing	4	300	75	1200
1484	2022-11-23	09:29:00	22	Female	19	Clothing	4	300	147	1200
15	2022-07-01	11:50:00	75	Female	42	Electronics	4	500	210	2000
743	2022-08-07	07:54:00	55	Female	34	Beauty	4	500	260	2000
1015	2022-03-09	11:53:00	94	Female	42	Electronics	4	500	200	2000
1743	2022-10-26	09:37:00	47	Female	34	Beauty	4	500	250	2000

6. Find the total number of transactions (transaction_id) made by each gender in each category:

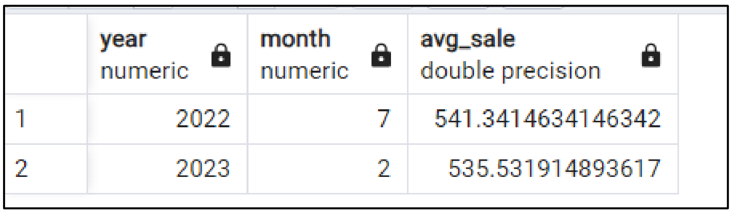
```
SELECT category, gender, COUNT(*) AS total_trans

FROM retail_sales

GROUP BY category, gender

ORDER BY category;
```



gender	category	total_number_of_transactions
character varying (15)	character varying (15)	bigint
Female	Beauty	330
Female	Clothing	347
Female	Electronics	340
Male	Electronics	344
Male	Clothing	354
Male	Beauty	282

7. Calculate the average sale for each month and find the best-selling month in each year:

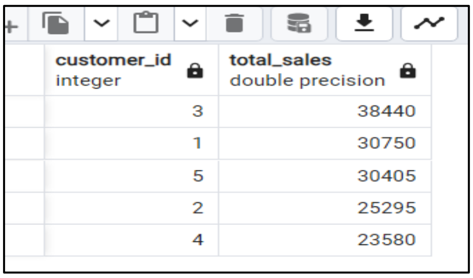
```
SELECT year, month, avg_sale

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 year, month

) AS t1

WHERE rank = 1;
```



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

8. Find the top 5 customers based on the highest total sales:

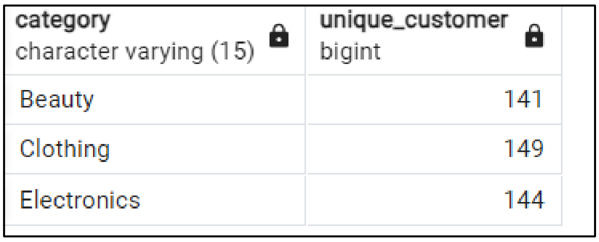
```
SELECT customer_id, SUM(total_sale) AS total_sales

FROM retail_sales

GROUP BY customer_id

ORDER BY total_sales DESC

LIMIT 5;
```



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

9. Find the number of unique customers who purchased items from each category:

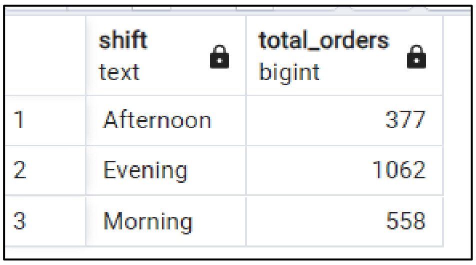
```
SELECT category, COUNT(DISTINCT customer_id) AS cnt_unique_cs

FROM retail_sales

GROUP BY category;
```



category character varying (15)	unique_customer bigint
Beauty	141
Clothing	149
Electronics	144

10. **Create each shift and count the number of orders per shift (Morning <12, Afternoon 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
)
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	558

Findings:

1. **Customer Demographics:**
   Sales span various age groups, with notable purchases in categories like "Clothing" and "Beauty."
2. **High-Value Transactions:**
   Several transactions exceeded $1000, indicating premium purchases.
3. **Sales Trends:**
   The monthly and shift-wise analysis provided insights into customer shopping habits and peak periods.
4. **Customer Insights:**
   Identified top-spending customers and unique buyers per category.

Insights

1. Customer Demographics

- Age-Based Preferences: Customers from various age groups actively purchase items. Notably, categories like "Clothing" and "Beauty" see significant engagement, which suggests these categories are universally appealing.
- Gender Trends: Sales data segmented by gender for each category reveal potential areas for targeted marketing and promotions**.**

---

2. High-Value Transactions

- Premium Buyers: Several transactions exceeded $1,000, indicating the presence of a premium customer base. Products associated with these high-value purchases can be analyzed further to design exclusive offers or personalized experiences.

---

3. Sales Trends

- Seasonality and Peaks: The analysis of monthly average sales and identification of the best-selling month in each year provide insights into seasonality, enabling better inventory and marketing strategies.
- Shift Analysis: Sales are distributed across morning, afternoon, and evening shifts, with peak periods identified. Retailers can allocate resources such as staff or promotional offers accordingly.

---

4. Product Performance

- Category Performance: Categories with the highest net sales (e.g., "Clothing," "Beauty") indicate strong market demand. These categories should remain a priority for inventory management and promotional campaigns.
- Customer Engagement by Category: A count of unique buyers per category highlights customer engagement, helping in identifying underperforming or niche categories.

---

5. Customer Insights

- Top-Spending Customers: The top 5 customers based on total sales represent a valuable segment. Personalizing their experience with loyalty programs or exclusive offers can increase retention and lifetime value.
- Unique Buyers per Category: Tracking the number of distinct buyers in each category helps measure customer penetration and category appeal**.**

---

6. Operational Efficiency

- Order Distribution by Shift: Morning, afternoon, and evening order data indicate when most transactions occur, enabling optimized staffing and operations during peak hours.

Recommendations:

1. **Targeted Marketing Campaigns:**

- Focus marketing efforts on high-performing categories like "Clothing" and "Beauty."
- Design campaigns tailored to specific demographics based on age and gender trends.

2. **Customer Retention Strategies:**

- Implement loyalty programs for high-spending customers.
- Offer targeted discounts or promotions to unique buyers in underperforming categories.

3. **Seasonal and Peak Planning:**

- Align inventory and promotional activities with peak sales months and times of the day.
- Prepare for seasonal variations to maximize sales opportunities.

4. **Operational Improvements:**

- Optimize staffing and logistics during identified peak hours.
- Evaluate shift-based performance to streamline customer service and reduce wait times.

5. **Future Analysis:**

- Integrate promotional data and customer feedback to understand the drivers of high-value transactions and underperforming categories.
- Leverage advanced analytics to predict future sales trends and customer preferences.

Conclusion:

This project provided hands-on experience with SQL for database creation, cleaning, and analysis. The queries addressed business questions that offer actionable insights into customer behavior, sales patterns, and product performance.

Future Scope:

1. Integrate additional datasets (e.g., promotions, feedback) for deeper analysis.
2. Automate reporting using visualization tools like Power BI or Tableau.
3. Apply advanced analytics for predictive sales forecasting.