Retail Sales Analysis

By Sudhanshu

Date: 2025.05.06

CONTENTS

01

Introduction

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. This project is ideal for those who are starting their journey in data analysis and want to build a solid foundation in SQL.

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.

02

Database Setup

Database Creation

```
CREATE DATABASE Retail_Sales;
```

Table Creation

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(35), quantity INT, price_per_unit FLOAT, cogs FLOAT, total_sale FLOAT);

03

Data Exploration & Cleaning

Record Count

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

Customer Count

```
SELECT COUNT(DISTINCT customer_id)
FROM retail_sales;
```

Category Count

```
SELECT DISTINCT category
FROM retail_sales;
```

Null Value 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;
```

Data Deletion on Null Value findings

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;

04

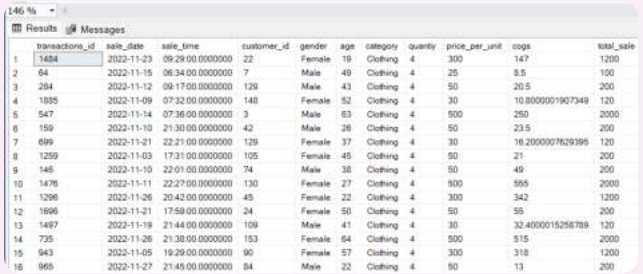
Data Analysis & Findings

Analysis Queries/Sales on a Specific Date

	transactions_id	sale_date	sale_time	customer_id	gender	age	category	quantity	price_per_unit	cogs	total_sale
1	180	2022-11-05	10:47:00.0000000	117	Male	41	Clothing	3	300	129	900
2	240	2022-11-05	11:49:00.0000000	95	Female	23	Beauty	1	300	123	300
3	1256	2022-11-05	09:58:00.0000000	29	Male	23	Clothing	2	500	190	1000
4	1587	2022-11-05	20:06:00.0000000	140	Female	40	Beauty	4	300	105	1200
5	1819	2022-11-05	20:44:00.0000000	83	Female	35	Beauty	2	50	13.5	100
6	943	2022-11-05	19:29:00.0000000	90	Female	57	Clothing	4	300	318	1200
7	1896	2022-11-05	20:19:00.0000000	87	Female	30	Electronics	2	25	30.75	50
8	1137	2022-11-05	22:34:00.0000000	104	Male	46	Beauty	2	500	145	1000
9	856	2022-11-05	17:43:00.0000000	102	Male	54	Electronics	4	30	9.300000190734863	120
10	214	2022-11-05	16:31:00.0000000	53	Male	20	Beauty	2	30	8.100000381469727	60
11	1265	2022-11-05	14:35:00.0000000	86	Male	55	Clothing	3	300	111	900

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

Analysis Queries/Clothing Transactions



146 %

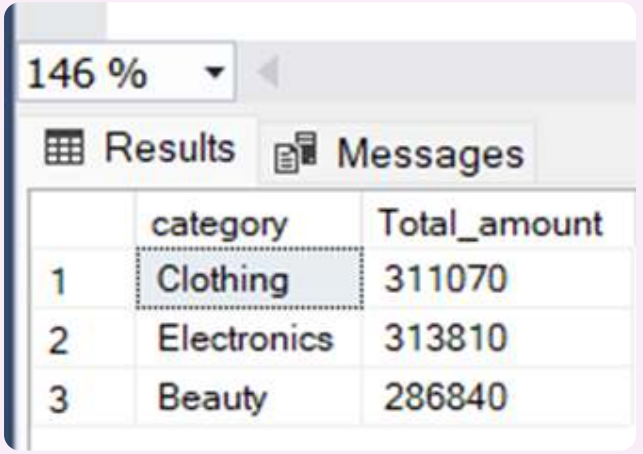
Results | Messages

	transactions_id	sale_date	sale_time	customer_id	gender	age	category	quantiy	price_per_unit	cogs	total_sale
1	1484	2022-11-23	09:29:00.0000000	22	Female	19	Clothing	4	300	147	1200
2	64	2022-11-15	06:34:00.0000000	7	Male	49	Clothing	4	25	8.5	100
3	284	2022-11-12	09:17:00.0000000	129	Male	43	Clothing	4	50	20.5	200
4	1885	2022-11-09	07:32:00.0000000	148	Female	52	Clothing	4	30	10.8000001907349	120
5	547	2022-11-14	07:36:00.0000000	3	Male	63	Clothing	4	500	250	2000
6	159	2022-11-10	21:30:00.0000000	42	Male	26	Clothing	4	50	23.5	200
7	899	2022-11-21	22:21:00.0000000	129	Female	37	Clothing	4	30	16.2000007629395	120
8	1259	2022-11-03	17:31:00.0000000	105	Female	45	Clothing	4	50	21	200
9	148	2022-11-10	22:01:00.0000000	74	Male	38	Clothing	4	50	49	200
10	1476	2022-11-11	22:27:00.0000000	130	Female	27	Clothing	4	500	585	2000
11	1296	2022-11-26	20:42:00.0000000	45	Female	22	Clothing	4	300	342	1200
12	1696	2022-11-21	17:59:00.0000000	24	Female	50	Clothing	4	50	55	200
13	1497	2022-11-19	21:44:00.0000000	109	Male	41	Clothing	4	30	32.4000015258789	120
14	735	2022-11-26	21:38:00.0000000	153	Female	64	Clothing	4	500	515	2000
15	943	2022-11-05	19:29:00.0000000	90	Female	57	Clothing	4	300	318	1200
16	965	2022-11-27	21:45:00.0000000	84	Male	22	Clothing	4	50	13	200

```
SELECT *
FROM retail_sales
WHERE category = 'Clothing'
AND quantiy >= 4
AND FORMAT(sale_date, 'yyyy-MM') = '2022-11';
```

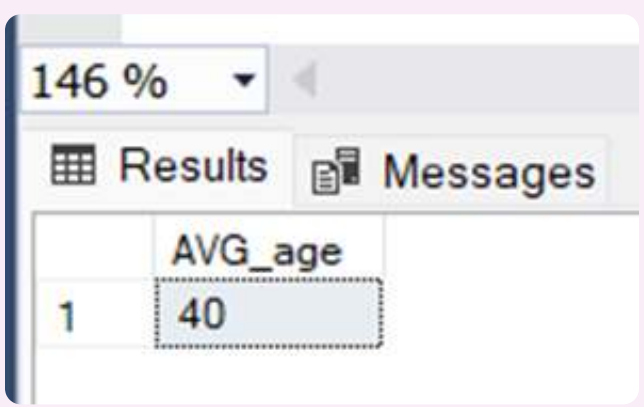
Analysis Queries/Total Sales Per Category

```
Select category, SUM(total_sale) AS Total_amount
from retail_sales
group by category;
```

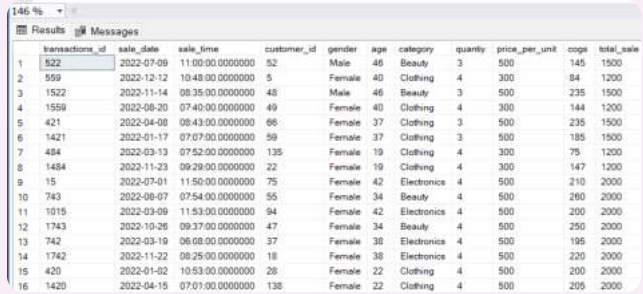


Analysis Queries/Average Age of Customers

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



Analysis Queries/Transactions Over 1000



146 %

Results | Messages

	transactions_id	sale_date	sale_time	customer_id	gender	age	category	quantiy	price_per_unit	cogs	total_sale
1	522	2022-07-09	11:00:00.0000000	52	Male	46	Beauty	3	500	145	1500
2	559	2022-12-12	10:48:00.0000000	5	Female	40	Clothing	4	300	84	1200
3	1522	2022-11-14	08:35:00.0000000	48	Male	46	Beauty	3	500	235	1500
4	1559	2022-08-20	07:40:00.0000000	49	Female	40	Clothing	4	300	144	1200
5	421	2022-04-08	08:43:00.0000000	66	Female	37	Clothing	3	500	235	1500
6	1421	2022-01-17	07:07:00.0000000	59	Female	37	Clothing	3	500	185	1500
7	484	2022-03-13	07:52:00.0000000	135	Female	19	Clothing	4	300	75	1200
8	1484	2022-11-23	09:29:00.0000000	22	Female	19	Clothing	4	300	147	1200
9	15	2022-07-01	11:50:00.0000000	75	Female	42	Electronics	4	500	210	2000
10	743	2022-08-07	07:54:00.0000000	55	Female	34	Beauty	4	500	260	2000
11	1015	2022-03-09	11:53:00.0000000	94	Female	42	Electronics	4	500	200	2000
12	1743	2022-10-26	09:37:00.0000000	47	Female	34	Beauty	4	500	250	2000
13	742	2022-03-19	06:08:00.0000000	37	Female	38	Electronics	4	500	195	2000
14	1742	2022-11-22	08:25:00.0000000	18	Female	38	Electronics	4	500	220	2000
15	420	2022-01-02	10:53:00.0000000	28	Female	22	Clothing	4	500	200	2000
16	1420	2022-04-15	07:01:00.0000000	138	Female	22	Clothing	4	500	205	2000

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

Analysis Queries/Transactions by Gender and Category

```
Select Count(*) as total_transactions, gender, category
from retail_sales
group by category, gender
order by total_transactions;
```

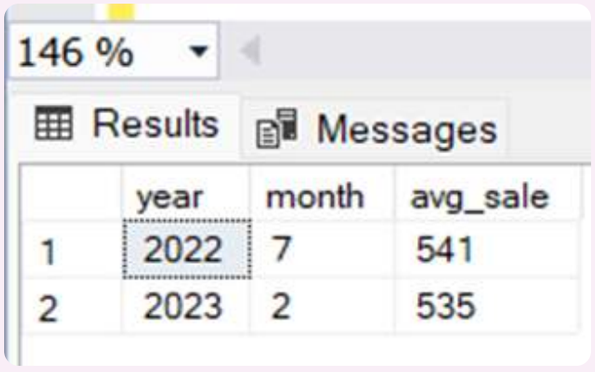
146 %

Results | Messages

	total_transactions	gender	category
1	282	Male	Beauty
2	330	Female	Beauty
3	340	Female	Electronics
4	344	Male	Electronics
5	347	Female	Clothing
6	354	Male	Clothing

Analysis Queries/Average Sales by Month

```
SELECT
   year,  month, avg_sale
FROM
(
SELECT
YEAR(sale_date) AS year, MONTH(sale_date) AS month,
AVG(total_sale) AS avg_sale,
RANK() OVER (
PARTITION BY YEAR(sale_date)
ORDER BY AVG(total_sale) DESC
)
AS rank
FROM retail_sales
GROUP BY YEAR(sale_date), MONTH(sale_date)
) AS t1
WHERE rank = 1;
```



	year	month	avg_sale
1	2022	7	541
2	2023	2	535

Analysis Queries/Top 5 Customers

```
select top 5 customer_id, sum(total_sale) as Total_amount
from Retail_Sales
order by Total_amount
group by customer_id;
```

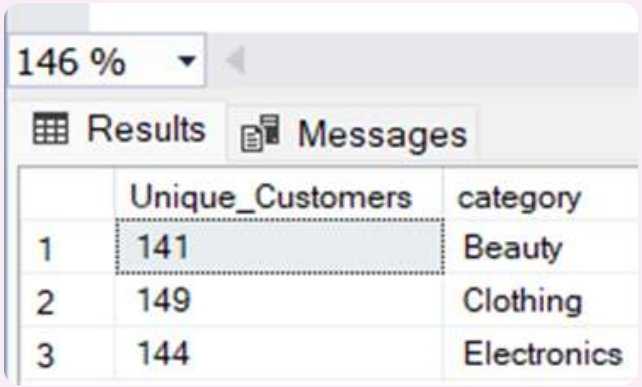
146 %

Results | Messages

	customer_id	Total_amount
1	123	200
2	118	640
3	150	675
4	23	845
5	149	1090

Analysis Queries/Unique Customers per Category

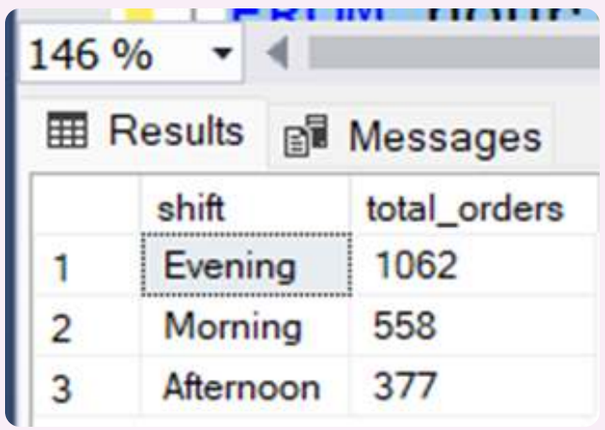
```
SELECT category, COUNT(DISTINCT customer_id) as unique_customer
FROM retail_sales
GROUP BY category;
```



	Unique_Customers	category
1	141	Beauty
2	149	Clothing
3	144	Electronics

Analysis Queries/Orders by Shift

```
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;
```



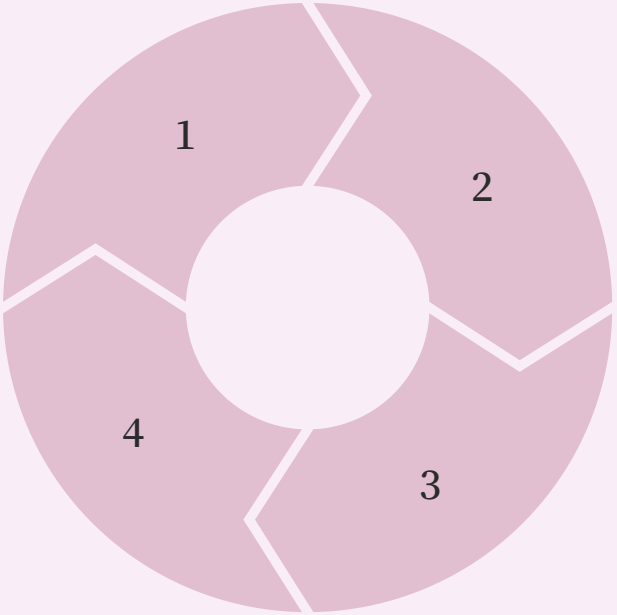
05

Findings

Findings

Customer Demographics: The dataset includes customers from various age groups, with sales distributed across different categories such as Clothing and Beauty.

Customer Insights: The analysis identifies the top-spending customers and the most popular product categories.



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.

06

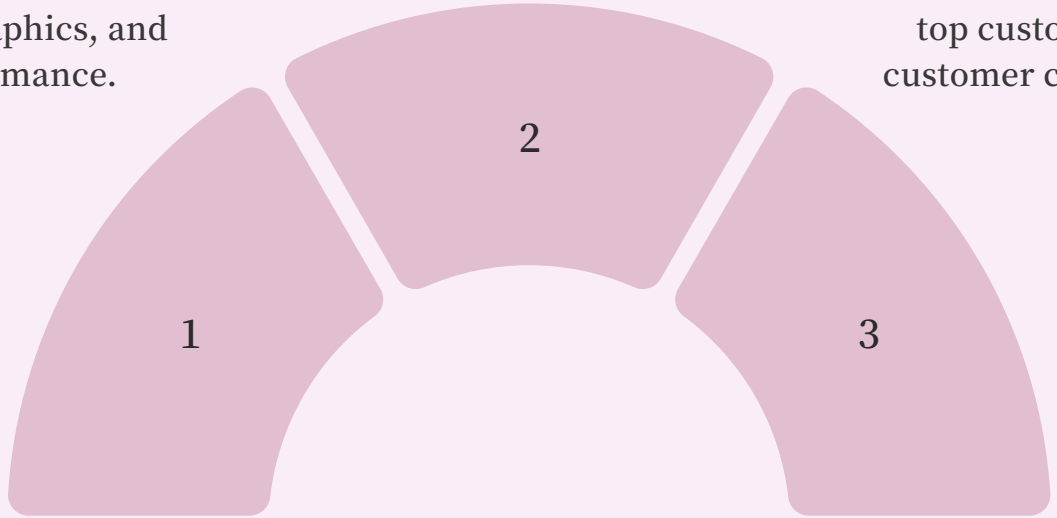
Reports

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.



07

Conclusion

Conclusion

This project serves as a comprehensive introduction to SQL for data analysts, covering database setup, data cleaning, exploratory data analysis, and business-driven SQL queries. The findings from this project can help drive business decisions by understanding sales patterns, customer behavior, and product performance.


Thank You