

Week 2 : Report

SQL Analysis & RFM-Based Customer Insights

Objective

- ✓ The objective of Week 2 was :
- to perform **structured data analysis using SQL** to extract meaningful business insights from retail sales data.
- The focus was on sales trends, customer behavior, and churn analysis using the **RFM** (**R**ecency, **F**requency, **M**onetary) framework.

Dataset & Tools Used

- **Database:** PostgreSQL
- **Tables:** Sales, Customer, Shopping Mall
- **Techniques:** SQL querying, aggregation, window functions, RFM segmentation

	customer_id [PK] character varying (20)	gender character varying (10)	age integer	payment_method character varying (20)	age_group text	price numeric (10,2)
1	C150594	Female	60	Cash	46-60	1200.32
2	C556552	Female	51	Cash	46-60	1200.32
3	C375325	Female	53	Cash	46-60	1200.32
4	C163444	Female	57	Cash	46-60	1200.32
5	C247049	Female	54	Cash	46-60	1200.32
6	C180375	Female	55	Cash	46-60	1200.32
7	C111251	Female	50	Cash	46-60	1200.32
8	C194063	Female	52	Cash	46-60	1200.32
9	C101897	Female	60	Cash	46-60	1200.32
10	C323574	Female	58	Cash	46-60	1200.32
11	C997987	Female	59	Cash	46-60	1200.32
12	C143595	Female	57	Cash	46-60	1200.32
13	C265780	Female	59	Cash	46-60	1200.32
14	C176271	Female	46	Cash	46-60	1200.32
15	C836040	Female	53	Cash	46-60	1200.32
16	C162539	Female	51	Cash	46-60	1200.32

Total rows: 99457 Query complete 00:00:01.148

Data Output Messages Notifications

SQL Showing rows: 1 to 10 Page No: 1

	shopping_mall [PK] character varying (100)	construction_year integer	area_sqm integer	location character varying (50)	store_count integer
1	South Coast Plaza	1967	250000	Costa Mesa	270
2	Westfield Valley Fair	1986	220000	Santa Clara	230
3	The Grove	2002	56000	Los Angeles	140
4	Westfield Century City	1964	133000	Los Angeles	200
5	Beverly Center	1982	111000	Los Angeles	160
6	Fashion Valley	1977	161000	San Diego	180
7	Stanford Shopping Center	1956	120000	Palo Alto	140
8	Glendale Galleria	1976	145000	Glendale	190
9	Irvine Spectrum Center	1995	120000	Irvine	130
10	Del Amo Fashion Center	1961	232000	Torrance	220

	invoice_no [PK] character varying (20)	customer_id character varying (20)	category character varying (50)	quantity integer	invoice_date date	price numeric (10,2)	shopping_mall character varying (100)
1	I293112	C176086	Clothing	2	2021-01-13	600.16	Del Amo Fashion Center
2	I249223	C188064	Clothing	2	2021-01-31	600.16	Del Amo Fashion Center
3	I406982	C283221	Clothing	2	2021-01-23	600.16	Del Amo Fashion Center
4	I255495	C279912	Clothing	2	2021-01-13	600.16	Del Amo Fashion Center
5	I159994	C531218	Clothing	2	2021-01-06	600.16	Del Amo Fashion Center
6	I161020	C842177	Clothing	2	2021-01-06	600.16	Del Amo Fashion Center
7	I100464	C129826	Clothing	2	2021-01-05	600.16	Del Amo Fashion Center
8	I614497	C116094	Clothing	2	2021-01-13	600.16	Del Amo Fashion Center
9	I230867	C932744	Clothing	2	2021-01-02	600.16	Del Amo Fashion Center
10	I293236	C180724	Clothing	2	2021-01-06	600.16	Del Amo Fashion Center
11	I124433	C573415	Clothing	2	2021-01-07	600.16	Del Amo Fashion Center
12	I301466	C325355	Clothing	2	2021-01-14	600.16	Del Amo Fashion Center
13	I284799	C160431	Clothing	2	2021-01-03	600.16	Del Amo Fashion Center
14	I258935	C252745	Clothing	2	2021-01-18	600.16	Del Amo Fashion Center
15	I299049	C249614	Clothing	2	2021-01-04	600.16	Del Amo Fashion Center

Total rows: 99457 Query complete 00:00:00.619 CRLF Ln 38, Col 1

Data Preparation & Cleaning (SQL-Level Handling)

- Imported cleaned datasets from Excel into PostgreSQL.
- Ensured correct **data types** for invoice dates, numeric revenue fields, and IDs.
- Handled **NULL values** by filtering invalid rows during aggregation.

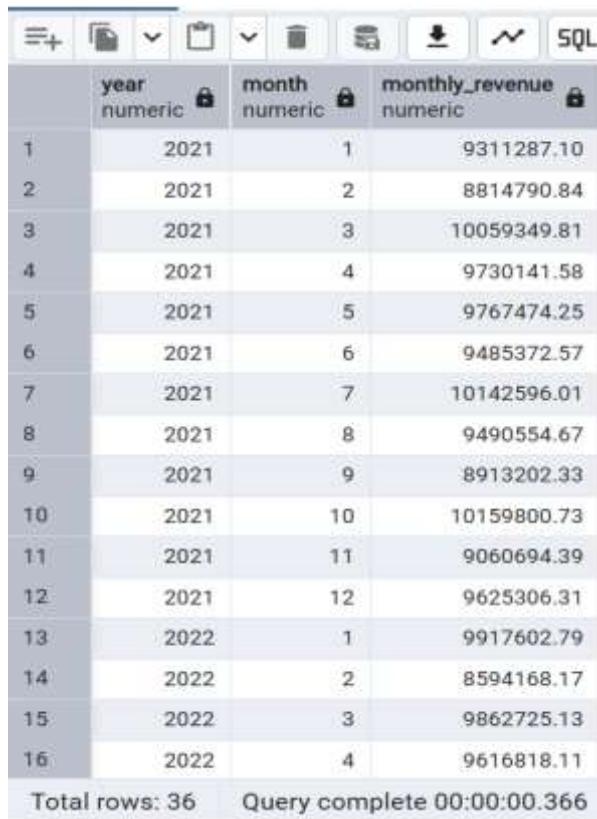
- Standardized **date formats** to enable monthly and yearly analysis.
- Validated data integrity before performing joins and calculations.

Key SQL Analysis Performed

➤ Sales & Revenue Analysis

- Calculated **total revenue** and **total number of orders**.
- Analyzed **monthly revenue trends** to understand growth patterns.
- Identified **highest revenue-generating shopping malls and locations**.
- Evaluated **category-wise revenue contribution**.

Revenue by month and year

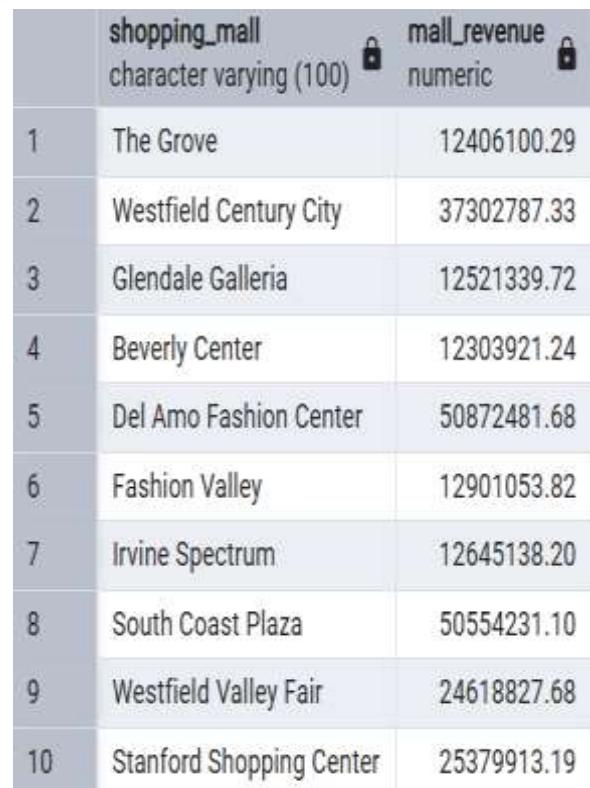


The screenshot shows a database query results interface with a toolbar at the top containing various icons for file operations and SQL navigation. The main area displays a table with three columns: 'year' (numeric), 'month' (numeric), and 'monthly_revenue' (numeric). The data spans from January 2021 to April 2022. The table has 16 rows, with the last two rows being from 2022. The total number of rows is 36, and the query took 0.366 seconds to complete.

	year numeric	month numeric	monthly_revenue numeric
1	2021	1	9311287.10
2	2021	2	8814790.84
3	2021	3	10059349.81
4	2021	4	9730141.58
5	2021	5	9767474.25
6	2021	6	9485372.57
7	2021	7	10142596.01
8	2021	8	9490554.67
9	2021	9	8913202.33
10	2021	10	10159800.73
11	2021	11	9060694.39
12	2021	12	9625306.31
13	2022	1	9917602.79
14	2022	2	8594168.17
15	2022	3	9862725.13
16	2022	4	9616818.11

Total rows: 36 Query complete 00:00:00.366

Revenue by shopping mall



The screenshot shows a database query results interface with a toolbar at the top containing various icons for file operations and SQL navigation. The main area displays a table with two columns: 'shopping_mall' (character varying) and 'mall_revenue' (numeric). The data lists 10 different shopping malls with their respective revenue amounts. The total number of rows is 10, and the query took 0.000 seconds to complete.

	shopping_mall character varying (100)	mall_revenue numeric
1	The Grove	12406100.29
2	Westfield Century City	37302787.33
3	Glendale Galleria	12521339.72
4	Beverly Center	12303921.24
5	Del Amo Fashion Center	50872481.68
6	Fashion Valley	12901053.82
7	Irvine Spectrum	12645138.20
8	South Coast Plaza	50554231.10
9	Westfield Valley Fair	24618827.68
10	Stanford Shopping Center	25379913.19

Categories with revenue > 10M

Data Output Messages Notifications

	category character varying (50)	total_revenue numeric
1	Clothing	113996791.04
2	Shoes	66553451.47
3	Technology	57862350.00

Top categories by Revenue

	category character varying (50)	category_revenue numeric
1	Clothing	113996791.04
2	Shoes	66553451.47
3	Technology	57862350.00
4	Cosmetics	6792862.90
5	Toys	3980426.24
6	Food & Beverage	849535.05
7	Books	834552.90
8	Souvenir	635824.65

Customer Purchase Behavior

- Calculated **Average Order Value (AOV)**.
- Analyzed customer distribution by **gender, age group, and payment method**.
- Observed that most customers made **only a single purchase**, indicating limited repeat behavior.

	avg_order_value numeric		total_revenue numeric		total_orders bigint
1	2528.79	1	251505794.25	1	99457

RFM Analysis (Recency, Frequency, Monetary)

Methodology

- Recency:** Calculated as the number of days since the customer's last purchase, using the maximum invoice date as the reference date.
- Frequency:** Count of distinct invoices per customer.
- Monetary:** Total revenue contributed by each customer.
- Used SQL window functions (**NTILE**) to score customers from 1–5 for R, F, and M.

Calculate R,F,M per customer

	customer_id character varying (20)	recency integer	frequency bigint	monetary numeric	Reference date
					max date
1	C100004	765	1	7502.00	
2	C100005	303	1	2400.68	
3	C100006	718	1	322.56	
4	C100012	868	1	130.75	
5	C100019	889	1	35.84	
6	C100025	1030	1	143.36	
7	C100028	766	1	15.15	
8	C100030	444	1	4801.28	
9	C100034	936	1	1200.32	
10	C100041	982	1	2700.72	
11	C100042	639	1	650.56	
12	C100045	456	1	5.23	
13	C100066	691	1	322.56	
14	C100067	943	1	600.17	
15	C100078	718	1	60.60	
16	C100088	972	1	5.23	

Total rows: 99457 Query complete 00:00:02.278

RFM Scoring using NTILE

	customer_id character varying (20)	recency integer	frequency bigint	monetary numeric	r_score integer	f_score integer	m_score integer
1	C163422	29	1	9602.72	1	1	1
2	C153995	29	1	4801.28	1	2	2
3	C323960	29	1	9602.72	1	1	1
4	C303453	29	1	573.44	1	3	3
5	C332230	29	1	2700.72	1	2	2
6	C106686	29	1	9602.72	1	1	1
7	C303609	29	1	9602.72	1	1	1
8	C248631	29	1	2400.68	1	2	2
9	C148206	29	1	9602.72	1	1	1
10	C198196	29	1	4200.00	1	2	2
11	C208029	29	1	7502.00	1	1	1
12	C238184	29	1	7502.00	1	1	1
13	C566943	29	1	7502.00	1	1	1
14	C176228	29	1	2700.72	1	2	2
15	C132663	29	1	4801.28	1	2	2
16	C118998	29	1	1016.50	1	3	3

Total rows: 99457 Query complete 00:00:02.713

Customer Segmentation

	customer_id character varying (20)	r integer	f integer	m integer	customer_segment text
1	C163422	1	1	1	Lost Customers
2	C153995	1	2	2	Need Attention
3	C323960	1	1	1	Lost Customers
4	C303453	1	3	3	At Risk
5	C332230	1	2	2	Need Attention
6	C106686	1	1	1	Lost Customers
7	C303609	1	1	1	Lost Customers
8	C248631	1	2	2	Need Attention
9	C148206	1	1	1	Lost Customers
10	C198196	1	2	2	Need Attention
11	C208029	1	1	1	Lost Customers
12	C238184	1	1	1	Lost Customers
13	C566943	1	1	1	Lost Customers
14	C176228	1	2	2	Need Attention
15	C132663	1	2	2	Need Attention
16	C118998	1	3	3	At Risk

Total rows: 99457 Query complete 00:00:03.006

Key Finding from RFM Analysis

- Majority of customers fall into:
 - Lost Customers
 - At Risk
 - Need Attention
- No Champions or Loyal Customers were identified due to:
 - Very low purchase frequency (mostly one-time buyers)
 - Long gaps since last purchase (high recency values)
- Indicates high customer churn across multiple California malls.

Business Insights

- Revenue growth is driven mainly by new or one-time customers.
- Customer retention is weak despite multiple mall locations.
- Certain malls and categories contribute disproportionately to total revenue.
- Lack of repeat purchases presents a long-term business risk.

Improvement & Recommendation Strategies

Customer Retention

- Introduce loyalty programs to incentivize repeat purchases.
- Offer discounts or rewards on second and third purchases.

Churn Reduction

- Target **At Risk** and **Need Attention** customers with personalized offers.
- Re-engagement campaigns via email or SMS.

Business Growth

- Focus marketing efforts on high-performing categories and malls.
- Convert high-spend one-time buyers into repeat customers.

Data Strategy

- Track customer lifecycle and repeat behavior continuously.
- Use RFM analysis regularly for marketing and CRM decisions.

Conclusion

The SQL and RFM analysis revealed a strong revenue base but weak customer retention. The absence of loyal customer segments highlights a major opportunity for improving long-term growth through targeted retention and engagement strategies.