Retail Data Analysis

SQL SERVER, SSMS

Skills

data cleaning, SQL querying, data analysis, business insights, aggregate functions, subqueries, CTEs.

Description

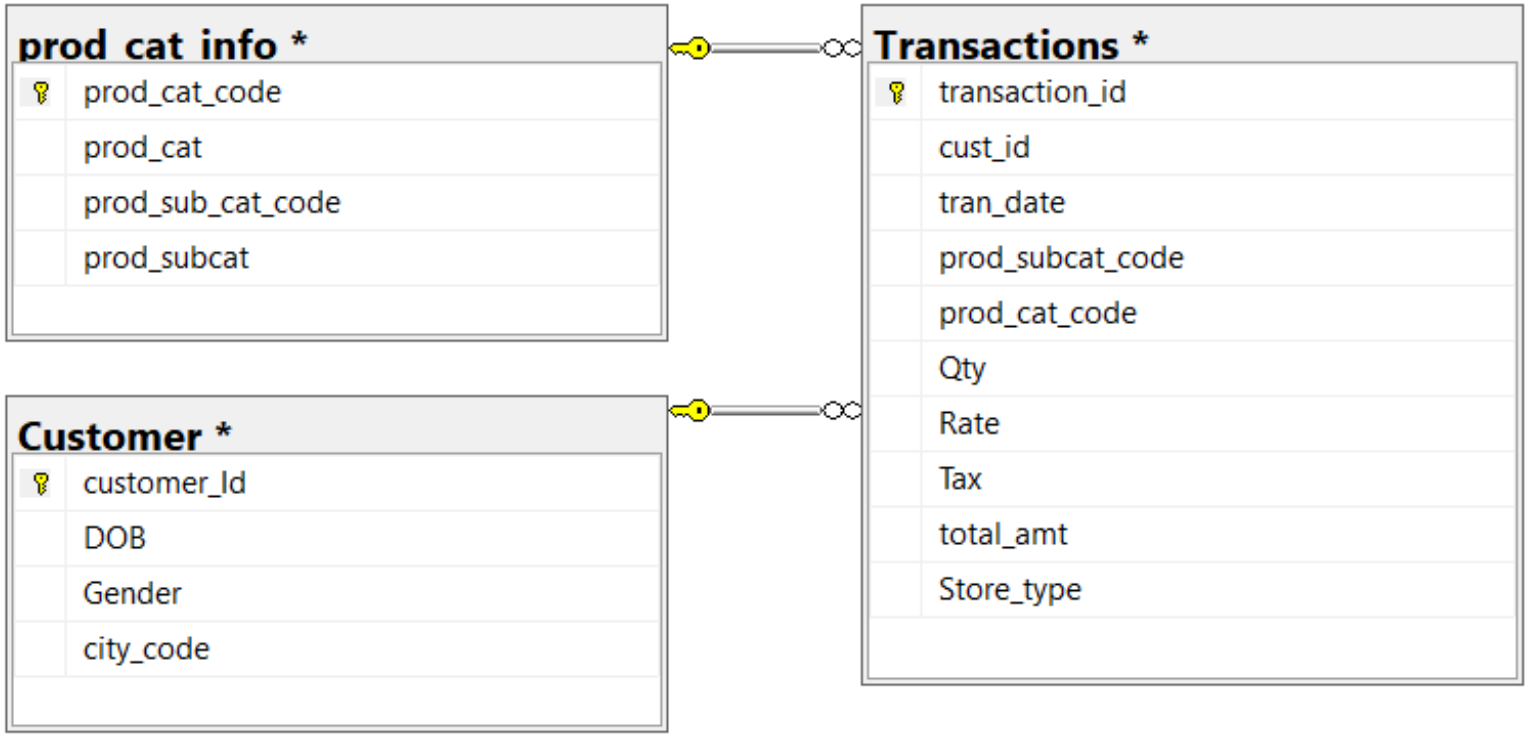
The project involved working with a relational retail database consisting of three tables: customer details, product hierarchy, and transaction data. The transactions include product categories, sales quantities, pricing, tax, and total amounts across different store types. The workflow includes data cleaning (such as date formatting), calculating revenue, analyzing return and sales trends by product sub-category, and identifying high-performing store types.

DATA AVAILABILITY

A retail store would like to understand customer behavior using their point of sale data (POS). The data set we will be using for this exercise comprises of three tables:

1. Customer : Customer demographics
2. Transactions : Customer transaction details
3. prod_cat_info : Product category and sub category information

The following diagram explains the relationship between tables-



DATA PREPARATION AND UNDERSTANDING

1. What is the total number of rows in each of the three tables in the database?

```
SELECT 'Customer' AS TABLE_NAME,COUNT(*) AS TOTAL_ROWS FROM Customer
UNION ALL
SELECT 'Prod_cat_info' AS TABLE_NAME,COUNT(*) AS TOTAL_ROWS FROM prod_cat_info
UNION ALL
SELECT 'Transactions' AS TABLE_NAME,COUNT(*) AS TOTAL_ROWS FROM Transactions
```

	TABLE_NAME	TOTAL_ROWS
1	Customer	5647
2	Prod_cat_info	23
3	Transactions	23053

2. What is the total number of transactions that have a return?

```
SELECT COUNT(*) AS TOTAL_TRANSACTIONS
FROM TRANSACTIONS AS T
WHERE T.Qty<0
```

	TOTAL_TRANSACTIONS
1	2177

3. As you would have noticed, the dates provided across the datasets are not in a correct format. AS first step, please convert the date variables into valid date formats before preceeding ahead.

```
SELECT FORMAT(CONVERT(DATE,DOB,103),'dd-MM-yyyy') FROM Customer
SELECT FORMAT(CONVERT(DATE,tran_date,103),'dd-MM-yyyy') FROM Transactions
```

	(No column name)
1	02-01-1970
2	07-01-1970
3	08-01-1970
4	10-01-1970
5	11-01-1970
6	15-01-1970
7	15-01-1970
8	16-01-1970

Q4. What is the time range of the transaction data available for analysis? Show the output in number of days, months, years simultaneously in different columns.

```
SELECT MIN(T.tran_date) AS START_DATE, MAX(T.tran_date)END_DATE,
DATEDIFF(DAY,MIN(T.tran_date),MAX(T.tran_date)) AS TOTAL_DAYS,
DATEDIFF(MONTH,MIN(T.tran_date),MAX(T.tran_date)) AS TOTAL_MONTHS,
DATEDIFF(YEAR,MIN(T.tran_date),MAX(T.tran_date)) AS TOTAL_YEARS
FROM Transactions AS T
```

	START_DATE	END_DATE	TOTAL_DAYS	TOTAL_MONTHS	TOTAL_YEARS
1	2011-01-25	2014-02-28	1130	37	3

Q5. Which product category does the sub-category "DIY" belongs to?

```
SELECT P.prod_cat
FROM prod_cat_info AS P
WHERE P.prod_subcat = 'DIY'
```

	prod_cat
1	Books

DATA ANALYSIS

1. Which channel is most frequently used for transactions?

```
SELECT TOP 1 T.Store_type
FROM Transactions AS T
GROUP BY T.Store_type
ORDER BY COUNT(*) DESC
```

	Store_type
1	e-Shop

2. What is the count of male and female customers in the database?

```
SELECT C.Gender,COUNT(*) AS TOTAL_CUSTOMERS
FROM Customer AS C
WHERE C.GENDER IS NOT NULL
GROUP BY C.Gender
```

	Gender	TOTAL_CUSTOMERS
1	F	2753
2	M	2892

3. From which city do we have the maximum number of customers and how many?

```
SELECT TOP 1 C.city_code, COUNT(*) AS TOTAL_CUSTOMERS
FROM Customer AS C
GROUP BY C.city_code
ORDER BY TOTAL_CUSTOMERS DESC;
```

	city_code	TOTAL_CUSTOMERS
1	3	595

4. How many sub-categories are there under the Books category?

```
SELECT P.prod_cat,count(P.prod_subcat) AS NO_OF_SUB_CATEGORIES
FROM prod_cat_info AS P
WHERE P.prod_cat = 'Books'
GROUP BY P.prod_cat
```

	prod_cat	NO_OF_SUB_CATEGORIES
	Books	6

5. What is the maximum quantity of products ever ordered?

```
SELECT (MAX(T.Qty)) AS MAXIMUM_QTY
FROM Transactions AS T
```

	MAXIMUM_QTY
1	5

6. What is the net total revenue generated in categories Electronics and Books

```
SELECT T.prod_cat_code,SUM(T.total_amt) AS TOTAL_REVENUE
FROM Transactions AS T
WHERE T.prod_cat_code IN (SELECT DISTINCT p.prod_cat_code
      FROM prod_cat_info AS P
      WHERE P.prod_cat IN ('Electronics','Books'))
GROUP BY T.prod_cat_code
```

	prod_cat_code	TOTAL_REVENUE
1	3	10722463.6401215
2	5	12822694.0382767

7. How many customers have >10 transactions with us, excluding returns?

```
SELECT COUNT(*) AS TOTAL_CUSTOMERS
FROM(SELECT T.cust_id,COUNT(T.transaction_id) AS TOTAL_TRANSACTIONS
  FROM Transactions AS T
  WHERE T.Qty > 0
  GROUP BY T.cust_id
  HAVING COUNT(T.transaction_id) > 10
) AS A
```

	TOTAL_CUSTOMERS
1	6

8. What is the combined revenue earned from the Electronics and Clothing categories, from Flagship Stores

```
SELECT SUM(t.total_amt) AS TOTAL_REVENUE
FROM Transactions AS T
WHERE T.prod_cat_code IN(SELECT DISTINCT p.prod_cat_code
      FROM prod_cat_info AS P
      WHERE P.prod_cat IN ('Electronics','Clothing'))
      AND T.Store_type = 'Flagship store'
```

	TOTAL_REVENUE
1	3409559.27131653

9. What is the total revenue generated from Male customers in Electronics category? Output should display total revenue by prod sub-cat.

```
SELECT P.prod_subcat, SUM(T.total_amt) AS TOTAL_REVENUE
FROM Transactions AS T
INNER JOIN prod_cat_info AS P
ON T.prod_cat_code = P.prod_cat_code
AND T.prod_subcat_code = P.prod_sub_cat_code
WHERE T.cust_id IN (SELECT C.customer_Id FROM Customer AS C
    WHERE C.Gender = 'M')
AND P.prod_cat = 'Electronics'
GROUP BY P.prod_subcat
```

	prod_subcat	TOTAL_REVENUE
1	Personal Appliances	1107593.43469238
2	Mobiles	1192413.23558807
3	Computers	1091417.34134674
4	Audio and video	1138983.16963196
5	Cameras	1172702.24649048

10. What is the percentage of sales and returns by product sub category; display only top 5 sub categories in terms of sales.

```
SELECT TOP 5 P.prod_subcat AS SUB_CATEGORY,
 (SUM(T.total_amt)*100/(SELECT SUM(total_amt) FROM Transactions)) AS
PERCENTAGE_SALES,
 (SUM(T.Qty)*100/(SELECT SUM(Qty) FROM Transactions)) AS PERCENTAGE_RETURNS
FROM Transactions AS T
INNER JOIN prod_cat_info AS P
ON T.prod_cat_code = P.prod_cat_code
AND T.prod_subcat_code = P.prod_sub_cat_code
GROUP BY P.prod_subcat
ORDER BY SUM(T.total_amt) DESC;
```

	SUB_CATEGORY	PERCENTAGE_SALES	PERCENTAGE_RETURNS
1	Mens	12.7056013721471	12
2	Women	12.6949584636058	12
3	Kids	8.76102622483405	8
4	Mobiles	4.62888679806982	4
5	Fiction	4.59502014470774	4

11. For all customers aged between 25 to 35 years find what is the net total revenue generated by these customers in last 30 days of transactions from max transaction date available in the data

```
SELECT T.cust_id, SUM(T.total_amt) AS TOTAL_REVENUE
FROM Transactions AS T
WHERE T.cust_id IN (SELECT C.customer_Id FROM Customer AS C
    WHERE DATEDIFF(YEAR,CONVERT(DATE,C.DOB,103),GETDATE()) BETWEEN 25 AND 35)
    AND (CONVERT(DATE,T.tran_date,103) BETWEEN
     (SELECT DATEADD(DAY,-30,MAX(CONVERT(DATE,T.tran_date,103))) FROM
Transactions AS T)
    AND (SELECT MAX(CONVERT(DATE,T.tran_date,103)) FROM Transactions AS T))
GROUP BY T.cust_id
```

	cust_id	TOTAL_REVENUE
1	267058	876.265014648438
2	267067	676.260009765625
3	267243	0
4	267265	8116.22509765625
5	267360	2234.31005859375
6	267467	2700.6201171875
7	267489	782.340026855469
8	267651	4071.92504882813
9	267656	0
10	267794	967.97998046875
11	267858	5060.89990234375
12	268047	7127.25
13	268050	2274.09008789063

12. Which product category has seen the max value of returns in the last 3 months of transactions?

```
 SELECT TOP 1 T.prod_cat_code, P.prod_cat,SUM(T.total_amt) AS
TOTAL_RETURN_VALUE
 FROM Transactions AS T
 INNER JOIN prod_cat_info AS P
 ON T.prod_cat_code = P.prod_cat_code AND T.prod_subcat_code =
P.prod_sub_cat_code
 WHERE T.QTY < 0
 AND CONVERT(DATE,T.tran_date,103) BETWEEN
 (SELECT DATEADD(MONTH,-3,MAX(CONVERT(DATE,T.tran_date,103))) FROM
Transactions AS T)
 AND (SELECT MAX(CONVERT(DATE,T.tran_date,103)) FROM Transactions AS T)
 GROUP BY T.prod_cat_code, P.prod_cat
 ORDER BY TOTAL_RETURN_VALUE DESC
```

	prod_cat_code	prod_cat	TOTAL_RETURN_VALUE
1	4	Bags	-28657.0701293945

13. Which store-type sells the maximum products; by value of sales amount and by quantity sold?

```
SELECT TOP 1 T.Store_type, SUM(T.total_amt) AS SALES_AMT, SUM(T.Qty) AS TOTAL_QTY
FROM Transactions AS T
GROUP BY T.Store_type
ORDER BY SALES_AMT DESC, TOTAL_QTY DESC
```

	Store_type	SALES_AMT	TOTAL_QTY
1	e-Shop	19824816.0530701	22763

14. What are the categories for which average revenue is above overall average.

```
SELECT T.prod_cat_code, P.prod_cat ,AVG(T.total_amt) AS AVG_REVENUE
FROM Transactions AS T
INNER JOIN prod_cat_info AS P
ON T.prod_cat_code = P.prod_cat_code AND T.prod_subcat_code = P.prod_sub_cat_code
GROUP BY T.prod_cat_code, P.prod_cat
HAVING AVG(T.total_amt) > (SELECT AVG(T.total_amt) FROM Transactions AS T)
```

	prod_cat_code	prod_cat	AVG_REVENUE
1	5	Books	2112.81826302137
2	1	Clothing	2111.87077417631
3	3	Electronics	2189.15141692966

15. Find the average and total revenue by each sub-category for the categories which are among top 5 categories in terms of quantity sold.

```
SELECT T.prod_subcat_code,P.prod_subcat ,AVG(T.total_amt) AS AVG_REVENUE,
SUM(T.total_amt) AS TOTAL_REVENUE
 FROM Transactions AS T
 INNER JOIN prod_cat_info AS P
 ON T.prod_cat_code = P.prod_cat_code AND T.prod_subcat_code =
P.prod_sub_cat_code
 WHERE T.prod_cat_code IN ( SELECT TOP 5 T.prod_cat_code FROM Transactions
AS T
      GROUP BY T.prod_cat_code
      ORDER BY SUM(T.Qty) DESC)
GROUP BY T.prod_subcat_code,P.prod_subcat
```

	prod_subcat_code	prod_subcat	AVG_REVENUE	TOTAL_REVENUE
1	12	Academic	2125.48521033586	2055344.19839478
2	10	Audio and video	2247.96000075941	2140057.92072296
3	11	Bath	2059.84961563215	2107226.15679169
4	9	Cameras	2165.87853154991	2133390.35357666
5	11	Children	2136.66750498601	2211450.86766052
6	3	Comics	2037.68001891616	2100848.09950256
7	5	Computers	2181.74983576892	2090116.34266663
8	6	DIY	2108.37240019703	2085180.30379486
9	7	Fiction	2140.22078679963	2232250.28063202
10	2	Furnishing	2084.00695634859	2098595.00504303
11	3	Kids	2136.59914373483	2110959.95401001
12	4	Kids	2125.99262194449	2145126.55554199
13	10	Kitchen	2008.95819427284	2083289.64746094
14	1	Mens	2112.06439070489	1989564.65604401
15	4	Mens	2128.26314327709	2058030.45954895
16	4	Mobiles	2181.0888684825	2248702.62340546
17	10	Non-Fiction	2129.10387280073	2137620.28829193
18	8	Personal Appli...	2170.98395035983	2110196.39974976
19	12	Tools	2024.37144695792	2149882.47666931
20	1	Women	2071.78813731635	2082147.07800293
21	3	Women	1989.58307732153	2085083.06503296