SQL Case Study-Basic | Solutions

Approaches used/Concepts covered:

- Aggregate Functions
- Alternate Queries
- SET Operators
- JOINS- Basic & Advance

Business Context: A retail store would like to understand customer behaviour using their point-of-sale data (POS).

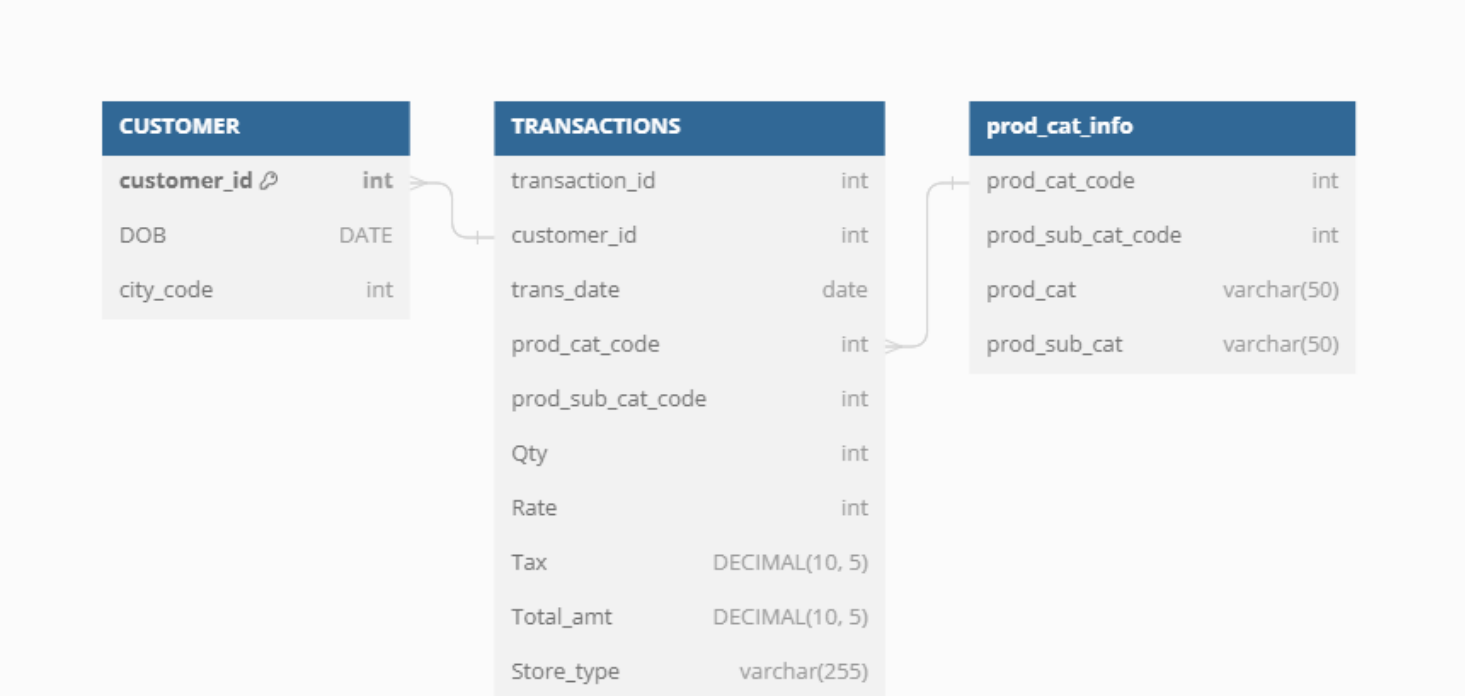
Data Availability: Data is available in three tables-

CUSTOMER: Customer Demographics.

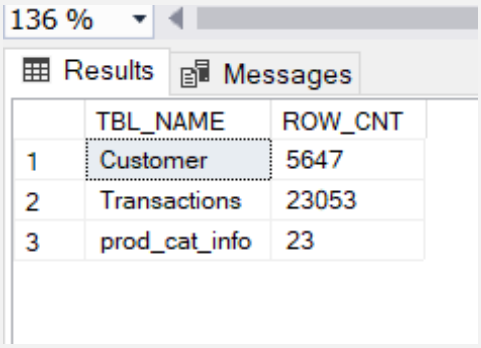
TRANSACTIONS: Customer transactions details.

Product Category Info: Category and sub-category of the products.

ER Diagram:



```sql
/*DATA PREPERATION AND UNDERSTANDING*/
/*1. What is the total number of rows in each of the 3 tables in the database? */

select 'Customer' as [TBL_NAME], count(*) as [ROW_CNT] from Customer
UNION ALL
select 'Transactions' as [TBL_NAME], count(*) as [ROW_CNT] from Transactions
UNION ALL
select 'prod_cat_info' as [TBL_NAME], count(*) as [ROW_CNT] from prod_cat_info;
```

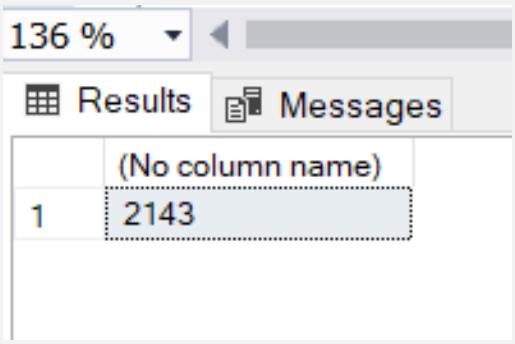


136 %

Results | Messages

	TBL_NAME	ROW_CNT
1	Customer	5647
2	Transactions	23053
3	prod_cat_info	23

```sql
/*2. What is the total number of transactions that have a return? */

select count(*) from Transactions where Qty like '%-%' AND total_amt like '%-%'
```



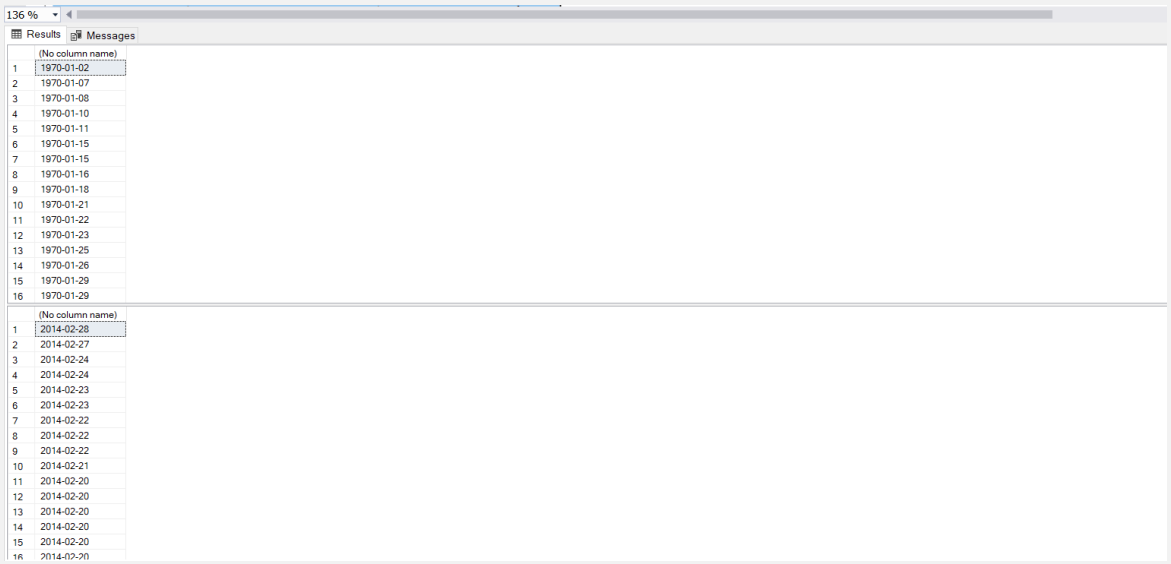
136 %

Results | Messages

	(No column name)
1	2143

```sql
/*3. As you would have noticed, the dates provided across the datasets are not in a
correct format. As first steps, pls convert the date variables into valid date
formats
before proceeding ahead. */

select CONVERT(DATE, DOB, 105) from Customer;
select CONVERT(DATE, tran_date, 105) from Transactions;
```

-



	(No column name)
1	1970-01-02
2	1970-01-07
3	1970-01-08
4	1970-01-10
5	1970-01-11
6	1970-01-15
7	1970-01-15
8	1970-01-16
9	1970-01-18
10	1970-01-21
11	1970-01-22
12	1970-01-23
13	1970-01-25
14	1970-01-26
15	1970-01-29
16	1970-01-29

	(No column name)
1	2014-02-28
2	2014-02-27
3	2014-02-24
4	2014-02-24
5	2014-02-23
6	2014-02-23
7	2014-02-22
8	2014-02-22
9	2014-02-22
10	2014-02-21
11	2014-02-20
12	2014-02-20
13	2014-02-20
14	2014-02-20
15	2014-02-20

-

```
/*4. What is the time range of the transaction data available for analysis? Show the
output in number of days, months and years simultaneously in different columns.
*/

select DATEDIFF(DAY, (MIN(TRAN_DATE)), (MAX(TRAN_DATE))) as [DAYS],
DATEDIFF(MONTH, (MIN(TRAN_DATE)), (MAX(TRAN_DATE))) as [MONTHS],
DATEDIFF(YEAR, (MIN(TRAN_DATE)), (MAX(TRAN_DATE))) as [YEAR]
from Transactions;
```

```
/*5. Which product category does the sub-category “DIY” belong to? */

select distinct prod_cat, prod_subcat from prod_cat_info where prod_subcat = 'DIY';
```

Results Messages

	prod_cat	prod_subcat
1	Books	DIY

```sql
/*DATA ANALYSIS*/

/*1. Which channel is most frequently used for transactions?*/

select TOP 1 Store_type, count(Store_type)  as [Ttl_Tran] from Transactions
group by Store_type
order by Ttl_Tran DESC;
```

Results | Messages

	Store_type	Ttl_Tran
1	e-Shop	9311

```sql
/*2. What is the count of Male and Female customers in the database? */

select GENDER, count(customer_Id) as [Ttl_Cus]
from Customer where GENDER IS NOT NULL group by GENDER;
```

Results | Messages

	GENDER	Ttl_Cus
1	F	2753
2	M	2892

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

select TOP 1 city_code, count(customer_Id) as [Ttl_Cus]
from Customer
group by city_code
order by Ttl_Cus desc;
```

Results | Messages

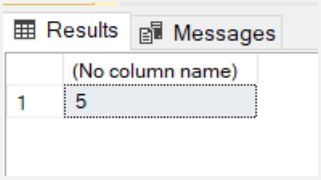
	city_code	Ttl_Cus
1	3	595

```sql
/*4. How many sub-categories are there under the Books category? */

select prod_cat, COUNT(prod_subcat)
from prod_cat_info where prod_cat = 'Books'
group by prod_cat;
```

Results | Messages

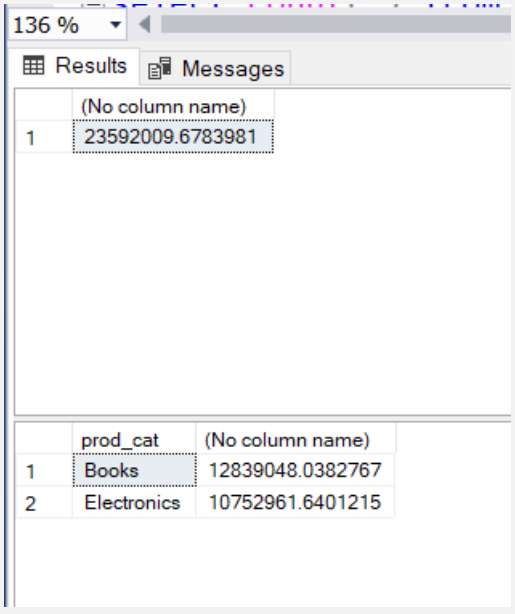
	prod_cat	(No column name)
1	Books	6

```sql
/*5. What is the maximum quantity of products ever ordered? */

select MAX(QTY) from Transactions where QTY not like '-%';
```



	(No column name)
1	5

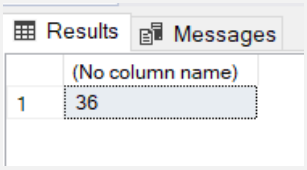
```sql
/*6. What is the net total revenue generated in categories Electronics and Books?
*/

select SUM(total_amt)
from Transactions A
INNER JOIN prod_cat_info B on A.prod_cat_code = B.prod_cat_code AND
A.prod_subcat_code = B.prod_sub_cat_code
where B.prod_cat IN ('Electronics', 'Books')

select B.prod_cat, SUM(total_amt)
from Transactions A
INNER JOIN prod_cat_info B on A.prod_cat_code = B.prod_cat_code AND
A.prod_subcat_code = B.prod_sub_cat_code
where B.prod_cat IN ('Electronics', 'Books')
group by prod_cat;
```



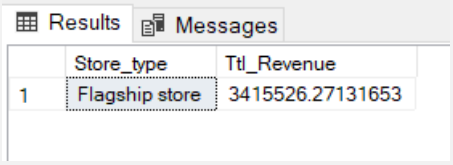
	(No column name)
1	23592009.6783981

	prod_cat	(No column name)
1	Books	12839048.0382767
2	Electronics	10752961.6401215

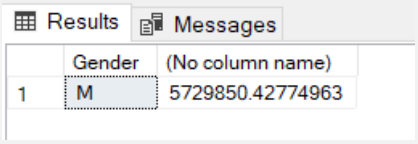
```sql
/*7. How many customers have >10 transactions with us, excluding returns? */

select count(*) from
(select cust_id, count(transaction_id) as [Ttl_Cus] from Transactions
group by cust_id
having count(transaction_id) > 10)E;
```




```sql
/*8. What is the combined revenue earned from the “Electronics” & “Clothing”
categories, from “Flagship stores”?*/

select A.Store_type, SUM(total_amt) as [Ttl_Revenue]
from Transactions A
INNER JOIN prod_cat_info B on A.prod_cat_code = B.prod_cat_code AND
A.prod_subcat_code = B.prod_sub_cat_code
where B.prod_cat IN ('Electronics', 'Clothing') AND A.Store_type in ('Flagship
store')
group by A.Store_type;
```




```sql
/*9. What is the total revenue generated from “Male” customers in “Electronics”
category? Output should display total revenue by prod sub-cat. */

select A.Gender, SUM(B.total_amt)
from Customer A
INNER JOIN Transactions B on A.customer_Id = B.cust_id
INNER JOIN prod_cat_info C on B.prod_cat_code = C.prod_cat_code AND
B.prod_subcat_code = C.prod_sub_cat_code
where A.Gender IN ('M') AND C.prod_cat IN ('Electronics')
group by A.Gender;
```

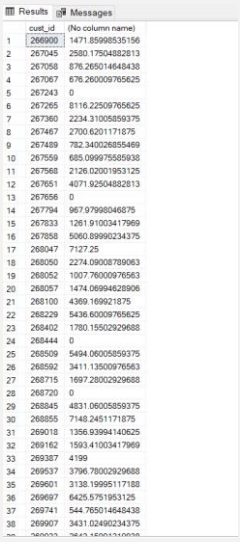


```sql
/*10.What is percentage of sales and returns by product sub category; display only
top
5 sub categories in terms of sales? */

select TOP 5 B.prod_subcat,
(SUM(total_amt)/(select SUM(total_amt) from Transactions)) * 100 as [% SALES],
(COUNT(CASE WHEN Qty < 0 THEN QTY ELSE NULL END)/SUM(QTY)) * 100 as [% RETURNS]
from Transactions A
INNER JOIN prod_cat_info B on A.prod_cat_code = B.prod_cat_code AND
A.prod_subcat_code = B.prod_sub_cat_code
group by B.prod_subcat
order by [% SALES] DESC;
```

Results | Messages

	prod_subcat	% SALES	% RETURNS
1	Women	12.7361945946486	0
2	Mens	12.6937123344749	0
3	Kids	8.76345659535307	0
4	Mobiles	4.61794161655303	0
5	Fiction	4.58460888838087	0

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

select cust_id, SUM(total_amt)
from Transactions
where cust_id IN
                        (
                        select customer_Id
                        from Customer
                        where DATEDIFF(YEAR, DOB, GETDATE()) BETWEEN 25 AND 35
                        )
AND tran_date BETWEEN DATEADD(DAY, -30, (select max(tran_date) from Transactions))
AND (select MAX(tran_date) from Transactions)
group by cust_id;
```



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

select TOP 1 B.prod_cat, SUM(total_amt) as [Ttl_RTN]
from Transactions A
INNER JOIN prod_cat_info B on A.prod_cat_code = B.prod_cat_code AND
A.prod_subcat_code = B.prod_sub_cat_code
WHERE tran_date BETWEEN DATEADD(MONTH, -3, (select max(tran_date) from
Transactions)) AND (select MAX(tran_date) from Transactions)
AND total_amt LIKE '-%'
group by B.prod_cat
order by Ttl_RTN DESC
```

Results Messages

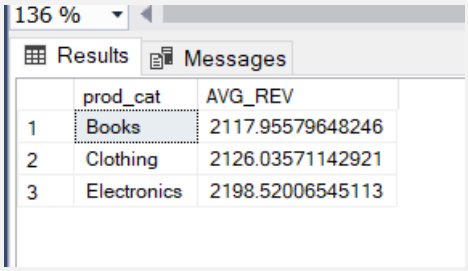
	prod_cat	Ttl_RTN
1	Bags	-28657.0701293945

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

--Solution 1: Using TOP Clause
select TOP 1 Store_type, SUM(total_amt) as [Ttl_SALES], SUM(Qty) as [QTY]
from Transactions
group by Store_type
order by Ttl_SALES DESC;
```

Results Messages

	Store_type	Ttl_SALES	QTY
1	e-Shop	19861723.0530701	22763

```sql
--Solution 2: Using HAVING CLAUSE with ALL Keyword
select Store_type, SUM(total_amt) as [Ttl_SALES], SUM(Qty) as [QTY]
from Transactions
group by Store_type
HAVING SUM(TOTAL_AMT) >=ALL (SELECT SUM(TOTAL_AMT) FROM Transactions GROUP BY
STORE_TYPE)
AND SUM(QTY) >=ALL (SELECT SUM(QTY) FROM Transactions GROUP BY STORE_TYPE)
order by Ttl_SALES DESC;


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

select B.prod_cat, AVG(total_amt) as [AVG_REV]
from Transactions A
INNER JOIN prod_cat_info B on A.prod_cat_code = B.prod_cat_code AND
A.prod_subcat_code = B.prod_sub_cat_code
group by B.prod_cat
HAVING AVG(total_amt) > (select AVG(total_amt) from Transactions)
```



136 %

Results | Messages

	prod_cat	AVG_REV
1	Books	2117.95579648246
2	Clothing	2126.03571142921
3	Electronics	2198.52006545113

```
/*15. Find the average and total revenue by each subcategory for the categories
which
are among top 5 categories in terms of quantity sold. */

select B.prod_cat, B.prod_subcat, AVG(total_amt) as [AVG_REV], SUM(total_amt) as
[Ttl_REV], SUM(Qty)
from Transactions A
INNER JOIN prod_cat_info B on A.prod_cat_code = B.prod_cat_code AND
A.prod_subcat_code = B.prod_sub_cat_code
where prod_cat IN
                        (
                        select top 5 B.prod_cat
                        from Transactions A
                        INNER JOIN prod_cat_info B on A.prod_cat_code =
B.prod_cat_code AND A.prod_subcat_code = B.prod_sub_cat_code
                        group by B.prod_cat
                        order by SUM(Qty) DESC
                        )
group by B.prod_cat, B.prod_subcat;
```

Results | Messages

	prod_cat	prod_subcat	AVG_REV	Ttl_REV	(No column name)
1	Books	Academic	2125.48521033586	2055344.19839478	2298
2	Electronics	Audio and video	2256.64728497154	2143814.92072296	2486
3	Home and kitchen	Bath	2059.84961563215	2107226.15679169	2478
4	Electronics	Cameras	2178.15397108511	2141125.35357666	2413
5	Books	Children	2143.22230914944	2216091.86766052	2487
6	Books	Comics	2047.43935811717	2106815.09950256	2448
7	Electronics	Computers	2202.95851743371	2106028.34266663	2395
8	Books	DIY	2115.33161478709	2087832.30379486	2405
9	Books	Fiction	2142.48683361998	2232471.28063202	2573
10	Home and kitchen	Furnishing	2091.57058155371	2104120.00504303	2435
11	Clothing	Kids	2142.34645796354	2114495.95401001	2407
12	Footwear	Kids	2135.77535272023	2152861.55554199	2527
13	Home and kitchen	Kitchen	2018.63408845352	2087267.64746094	2439
14	Clothing	Mens	2136.4145543985	2063776.45954895	2370
15	Footwear	Mens	2115.71801917535	1990890.65604401	2286
16	Electronics	Mobiles	2181.0888684825	2248702.62340546	2587
17	Books	Non-Fiction	2134.09101524619	2140493.28829193	2458
18	Electronics	Personal Appliances	2176.40617893899	2113290.39974976	2427
19	Home and kitchen	Tools	2030.02872447626	2153860.47666931	2589
20	Clothing	Women	2099.93714086207	2102037.07800293	2397
21	Footwear	Women	2005.44599524685	2095691.06503296	2461