RELIANCE SMART STORE SALES ANALYSIS USING SQL

Find DUPLICATE (Data cleaning)

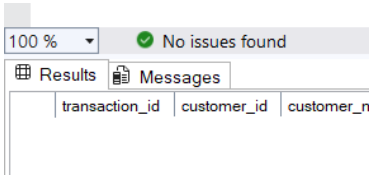
```
WITH CTE AS (
SELECT transaction_id,
ROW_NUMBER() OVER (PARTITION BY transaction_id ORDER BY transaction_id) AS
ROW_NUM
FROM Sales)
SELECT * FROM CTE WHERE ROW_NUM >  1
```

Results | Messages

	transaction_id	ROW_NUM
1	TXN240646	2
2	TXN342128	2
3	TXN855235	2
4	TXN981773	2

Removal of duplicate (Data cleaning)

```
WITH CTE AS (
SELECT  *,
ROW_NUMBER() OVER (PARTITION BY transaction_id ORDER BY transaction_id) AS
ROW_NUM
FROM Sales)
DELETE FROM CTE WHERE ROW_NUM = 2
```



Null Value Checking (Data cleaning)

```
SELECT * FROM
sales_store WHERE
transaction_id IS NULL or   customer_id IS NULL  or customer_name IS
NULL  or product_id IS NULL  or
product_name IS NULL or product_category IS NULL or quantiy IS NULL
or prce IS NULL or
payment_mode IS NULL or purchase_date IS NULL  or time_of_purchase IS NULL
or
status IS NULL
```

Results | Messages

	transaction_id	customer_id	customer_name	customer_age	gender	product_id	product_name	product_category	quantiy	prce	payment_mode	purchase_date	time_of_purchase	status
1	NULL	NULL	NULL	NULL	NULL	4524.00	T-Shirt	Clothing	4	14788	NULL	NULL	NULL	NULL
2	TXN432798	CUST1003	NULL	NULL	NULL	5717.00	Dining Table	Furniture	3	2346	EMI	2023-04-30	05:46:48.0000000	cancelled
3	TXN977900	NULL	Ehsaan Ram	25	M	8212.00	Milk	Groceries	3	12342	Cash	2023-08-13	07:46:18.0000000	returned
4	TXN985663	NULL	Damini Raju	49	Female	3367.00	Notebook	Books	5	8130	EMI	2023-01-24	01:40:38.0000000	cancelled

Updating Null Values Customer id (Data cleaning)

```
select * from sales WHERE customer_name = 'Ehsaan Ram'
UPDATE sales SET customer_id = 'CUST9494' WHERE transaction_id =
'TXN9779000'
select * from sales WHERE customer_name = 'Ehsaan Ram'
```

Results | Messages

	transaction_id	customer_id	customer_name	customer_age	gender	product_i
1	TXN553735	CUST9494	Ehsaan Ram	25	M	3115.00
2	TXN977900	NULL	Ehsaan Ram	25	M	8212.00
3	TXN495746	CUST9494	Ehsaan Ram	25	M	7385.00
4	TXN586594	CUST9494	Ehsaan Ram	25	M	4479.00

	transaction_id	customer_id	customer_name	customer_age	gender	product_i
1	TXN553735	CUST9494	Ehsaan Ram	25	M	3115.00
2	TXN977900	CUST9494	Ehsaan Ram	25	M	8212.00
3	TXN495746	CUST9494	Ehsaan Ram	25	M	7385.00
4	TXN586594	CUST9494	Ehsaan Ram	25	M	4479.00

Business insights we are going to answer using SQL query

1) Top 5 Most Delivered products by Quanitity

```
SELECT TOP 5
product_name,
SUM(quantiy) AS total_quantity
FROM Salee
WHERE status = 'delivered'
GROUP BY product_name
ORDER BY total_quantity DESC
```

100 % | No issues found

Results | Messages

	product_name	total_quantity
1	Wardrobe	70
2	Vegetables	69
3	Sofa	66
4	Dining Table	65
5	Fruits	60

2) Which Product are Mostly Cancelled?

```
SELECT TOP 5
product_name,
COUNT(*) AS total_times_cancelled
FROM Salee
   WHERE status = 'cancelled'
   GROUP BY product_name
   ORDER BY total_times_cancelled  DESC
```

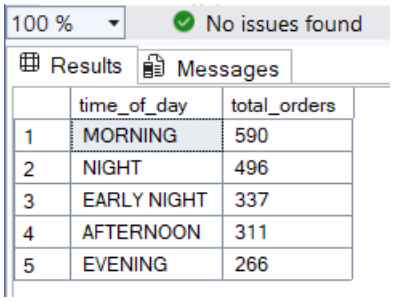
100 % | No issues found

Results | Messages

	product_name	total_times_cancelled
1	Comics	24
2	Sweater	23
3	Chair	21
4	Vegetables	21
5	Smartphone	20

3) Time Of Highest Purchases

```
SELECT time_slot AS time_of_day,
 COUNT(*) AS total_orders
FROM (SELECT CASE
WHEN DATEPART(HOUR, time_of_purchase) BETWEEN 0 AND 5THEN 'NIGHT'
WHEN DATEPART(HOUR, time_of_purchase) BETWEEN 5 AND 12 THEN 'MORNING'
WHEN DATEPART(HOUR, time_of_purchase) BETWEEN 12 AND 16 THEN 'AFTERNOON'
WHEN DATEPART(HOUR, time_of_purchase) BETWEEN 16 AND 19 THEN 'EVENING'
WHEN DATEPART(HOUR, time_of_purchase) BETWEEN 19 AND 23 THEN 'EARLY NIGHT'
END AS time_slot
FROM Salee
) AS sub
GROUP BY time_slot
ORDER BY total_orders DESC
```



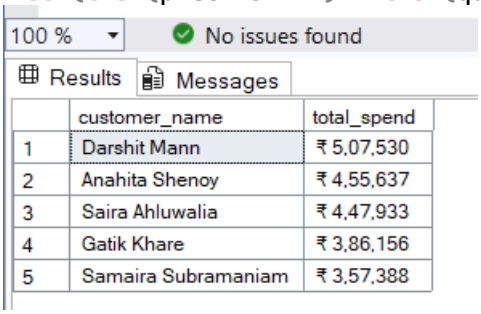
100 % No issues found

Results Messages

	time_of_day	total_orders
1	MORNING	590
2	NIGHT	496
3	EARLY NIGHT	337
4	AFTERNOON	311
5	EVENING	266

4) Top Highest Spending Customers

```
SELECT TOP 5
 customer_name,
 FORMAT(SUM(CAST(prce AS INT) * CAST(quantiy AS INT)), 'C0', 'en-IN') AS
total_spend
FROM salee
GROUP BY customer_name
ORDER BY SUM(CAST(prce AS INT) * CAST(quantiy AS INT)) DESC;
```



100 % No issues found

Results Messages

	customer_name	total_spend
1	Darshit Mann	₹ 5,07,530
2	Anahita Shenoy	₹ 4,55,637
3	Saira Ahluwalia	₹ 4,47,933
4	Gatik Khare	₹ 3,86,156
5	Samaira Subramaniam	₹ 3,57,388

5) WHICH CATEGORY GENERATE HIGHEST REVENUE

```
SELECT TOP 5
product_category,
FORMAT(SUM(CAST(prce AS INT) * CAST(quantiy AS INT)), 'C0', 'en-IN') AS
Revenue
FROM  Salee
GROUP BY product_category
ORDER BY  SUM(CAST(prce AS INT) * CAST(quantiy AS INT)) DESC;
```

100 % No issues found

Results Messages

	product_category	Revenue
1	Accessories	₹ 1,03,65,306
2	Clothing	₹ 1,01,95,727
3	Books	₹ 99,12,929
4	Furniture	₹ 96,59,478
5	Electronics	₹ 95,04,028

6) Return And Cancellation Rate Product per Category

```
SELECT
    product_category,
    FORMAT(COUNT(CASE WHEN status = 'cancelled' THEN 1 END) * 100.0 /
COUNT(*), 'N3', 'en-IN') + ' %' AS cancelled_percentage
FROM  Salee
GROUP BY product_category
ORDER BY
COUNT(CASE WHEN status = 'cancelled' THEN 1 END) * 100.0 / COUNT(*) DESC;
```

Results Messages

	product_category	cancelled_percentage
1	Books	26.205 %
2	Clothing	25.634 %
3	Electronics	24.675 %
4	Accessories	23.547 %
5	Furniture	22.832 %
6	Groceries	22.289 %

	product_category	returned_percentage
1	Accessories	31.498 %
2	Books	25.602 %
3	Clothing	24.789 %
4	Groceries	23.494 %
5	Furniture	23.410 %
6	Electronics	20.779 %

7) Most Preffered Payment Method

```
SELECT
payment_mode,
COUNT(*) as total_payments
FROM SALEE
GROUP BY payment_mode  ORDER BY total_payments DESC
```

100 % No issues found

Results Messages

	payment_mode	total_payments
1	Credit Card	648
2	EMI	350
3	Debit Card	344
4	Cash	332
5	UPI	326

8) Age Purchasing Power

```
    SELECT
           CASE
           WHEN customer_age BETWEEN 18 AND 25 THEN '18-25'
                  WHEN customer_age BETWEEN 25 AND 35 THEN '25-35'
                  WHEN customer_age BETWEEN 35 AND 50 THEN '35-50'
                  WHEN customer_age > 50 THEN '50+'
    END AS Age,
    FORMAT(SUM(CAST(prce AS INT) * CAST(quantiy AS INT)),'C0','en-IN') AS
total_purchase
           FROM SALEE
           GROUP BY CASE
                  WHEN customer_age BETWEEN 18 AND 25 THEN '18-25'
                  WHEN customer_age BETWEEN 25 AND 35 THEN '25-35'
                  WHEN customer_age BETWEEN 35 AND 50 THEN '35-50'
                  WHEN customer_age > 50 THEN '50+'
    END
    ORDER BY SUM(CAST(prce AS INT) * CAST(quantiy AS INT))DESC
```

Results Messages

	Age	total_purchase
1	35-50	₹ 1,94,60,276
2	50+	₹ 1,43,86,538
3	25-35	₹ 1,36,96,027
4	18-25	₹ 1,15,58,780

9) Sales Trend

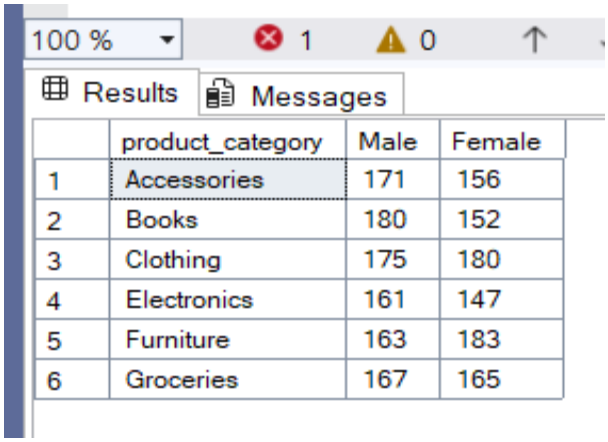
```
    SELECT
    YEAR(purchase_date)AS years,
    MONTH(purchase_date)AS MonthS,
    FORMAT(SUM(CAST(prce AS INT) * CAST(quantiy AS INT)),'C0','en-IN') AS
    total_purchase,
    SUM(quantiy) AS total_Quantity
    FROM Salee
           GROUP BY YEAR(purchase_date) ,MONTH(purchase_date) ORDER BY MonthS
```

Results Messages

	years	MonthS	total_purchase	total_Quantity
1	2023	1	₹ 46,28,608	478
2	2024	1	₹ 3,39,442	31
3	2023	2	₹ 46,98,929	529
4	2023	3	₹ 52,41,364	471
5	2023	4	₹ 49,89,315	505
6	2023	5	₹ 39,02,263	418
7	2023	6	₹ 41,00,112	478
8	2023	7	₹ 51,29,904	577
9	2023	8	₹ 47,88,207	497
10	2023	9	₹ 50,37,847	512
11	2023	10	₹ 58,86,414	547
12	2023	11	₹ 51,09,229	523
13	2023	12	₹ 52,49,987	521

10) Product Category Preffered by Genders

```
SELECT * FROM (
 SELECT gender,product_category
 FROM Salee ) AS source_table
 PIVOT(COUNT(gender)
             FOR gender IN ([Male],[Female])
             ) AS pivot_table
ORDER BY product_category
```



	product_category	Male	Female
1	Accessories	171	156
2	Books	180	152
3	Clothing	175	180
4	Electronics	161	147
5	Furniture	163	183
6	Groceries	167	165