DATA ANALYSIS PROJECT – SQL

Retrieve all columns from the Sales Analysis dataset in ascending order by sale date.

| III | Results 🗐 Mess | sages | | | | | | | | | |
|------------|-----------------|------------|-----------|-------------|--------|-----|-------------|----------|----------------|--------|------------|
| | transactions_id | sale_date | sale_time | customer_id | gender | age | category | quantity | price_per_unit | cogs | total_sale |
| 1 | 1269 | 2022-01-01 | 08:09:00 | 71 | Male | 25 | Clothing | 4 | 500.00 | 145.00 | 2000.00 |
| 2 | 1366 | 2022-01-01 | 11:34:00 | 25 | Male | 57 | Clothing | 2 | 50.00 | 23.50 | 100.00 |
| 3 | 958 | 2022-01-01 | 21:14:00 | 116 | Male | 62 | Electronics | 2 | 25.00 | 11.00 | 50.00 |
| ı | 420 | 2022-01-02 | 10:53:00 | 28 | Female | 22 | Clothing | 4 | 500.00 | 200.00 | 2000.00 |
| 5 | 189 | 2022-01-02 | 09:44:00 | 143 | Male | 63 | Beauty | 1 | 50.00 | 14.50 | 50.00 |

-----DATA CLEANING-----

Modify the cogs column in the dataset to display two decimal places.

```
ALTER TABLE Sales_Analysis ALTER COLUMN cogs DECIMAL(10, 2);
```

Modify the price per unit column in the dataset to display two decimal places.

```
ALTER TABLE Sales_Analysis ALTER COLUMN price_per_unit DECIMAL(10, 2);
```

Modify the total sale column in the dataset to display two decimal places.

```
ALTER TABLE Sales_Analysis ALTER COLUMN total_sale DECIMAL(10, 2);
```

Modify the sale time column in the dataset to display hh:mm:ss format.

```
ALTER TABLE Sales_Analysis ALTER COLUMN sale_time TIME(0);
```

Check if there are any NULL values in the dataset.

```
SELECT * FROM Sales_Analysis
WHERE transactions_id IS NULL
OR 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
OR total_sale IS NULL;
```

Delete rows with NULL values from the dataset.

```
DELETE FROM Sales_Analysis
WHERE transactions_id IS NULL
OR 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
OR total_sale IS NULL;
```

Rename the column 'quantity' to 'quantity'.

```
EXEC sp_rename 'Sales_Analysis.quantiy', 'quantity', 'COLUMN';
```

-----DATA ANALYSIS-----

Display the total number of transactions.

```
SELECT COUNT(*) AS 'Total Transactions' FROM Sales_Analysis;

Results Messages

Total Transactions
1 2000
```

Retrieve unique values from the category column.

```
SELECT DISTINCT category AS 'Category' FROM Sales_Analysis;

Results Messages

Category

Category

Category

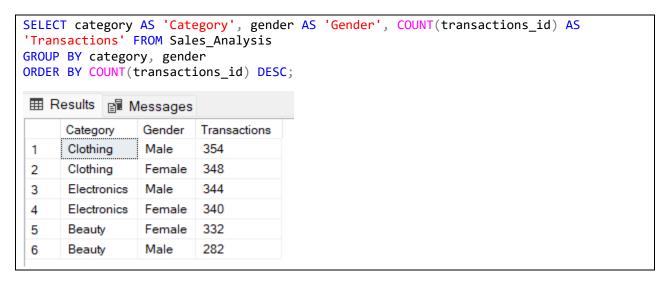
Electronics

Beauty
```

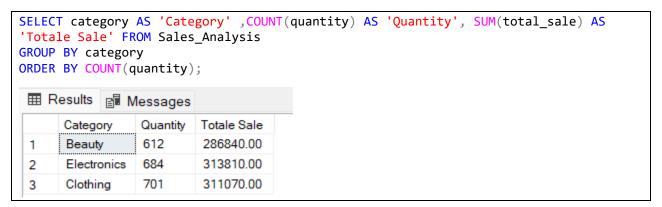
Retrieve the 'Total quantity' and 'Total sale' from the dataset.

```
SELECT COUNT(quantity) AS 'Total Quantity', SUM(total_sale) AS 'Total Sale' FROM Sales_Analysis;
```

Retrieve 'Category', 'Gender' and 'the number of transactions', grouped by 'Category' and 'Gender'



Retrieve 'Category', 'Quantity' and 'the sum of total sales', grouped by 'Category'.



Retrieve the 'Year' and the count of 'Transactions' from the dataset.

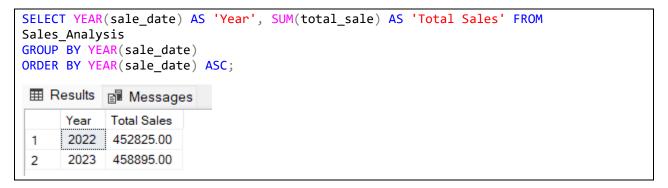
```
SELECT YEAR(sale_date) AS 'Year', COUNT(transactions_id) AS 'Transactions' FROM Sales_Analysis
GROUP BY YEAR(sale_date)
ORDER BY YEAR(sale_date);

Results Messages

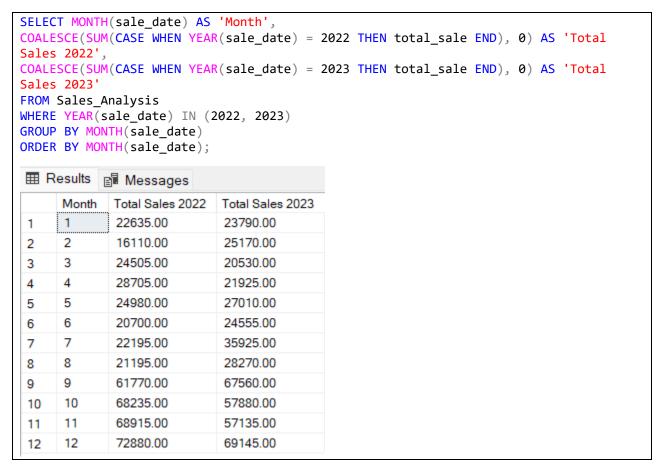
Year Transactions
1 2022 979
2 2023 1021
```

Retrieve the total number of distinct customers.

Retrieve 'Year' and the sum of 'Total sales', grouped by 'Year'.



Compare the total sales between 2022 and 2023 for each month.

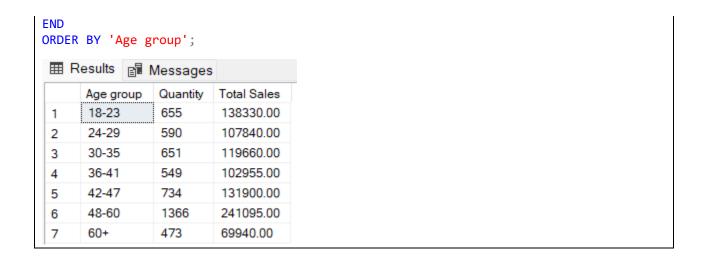


Compare the average sales between 2022 and 2023 for each month.

| COALI DECII COALI DECII FROM WHERI GROUI | H(sale_c ESCE(CA: MAL(10, ESCE(CA: MAL(10, Sales_/ E YEAR(: P BY MOI R BY MOI | 2)), 0.00) AS | EN YEAR(sale_d 'Avg Sales 20 EN YEAR(sale_d 'Avg Sales 20 (2022, 2023) | date) = 2023 THEN total_sale END) AS |
|--|---|----------------|--|--------------------------------------|
| | Month | Avg Sales 2022 | Avg Sales 2023 | |
| 1 | 1 | 397.11 | 396.50 | |
| 2 | 2 | 366.14 | 535.53 | |
| 3 | 3 | 521.38 | 394.81 | |
| 4 | 4 | 486.53 | 466.49 | |
| 5 | 5 | 480.38 | 450.17 | |
| 6 | 6 | 481.40 | 438.48 | |
| 7 | 7 | 541.34 | 427.68 | |
| 8 | 8 | 385.36 | 495.96 | |
| 9 | 9 | 478.84 | 462.74 | |
| 10 | 10 | 467.36 | 399.17 | |
| 11 | 11 | 472.02 | 453.45 | |
| 12 | 12 | 464.20 | 490.39 | |

Retrieve 'Age group', 'Quantity' and 'Total sales', grouped by 'Age group'.

```
SELECT
CASE
      WHEN Age BETWEEN 18 AND 23 THEN '18-23'
      WHEN Age BETWEEN 24 AND 29 THEN '24-29'
      WHEN Age BETWEEN 30 AND 35 THEN '30-35'
      WHEN Age BETWEEN 36 AND 41 THEN '36-41'
      WHEN Age BETWEEN 42 AND 47 THEN '42-47'
      WHEN Age BETWEEN 48 AND 60 THEN '48-60'
ELSE '60+'
END AS 'Age group', SUM(quantity) AS 'Quantity', SUM(total_sale) AS 'Total Sales'
FROM Sales_Analysis
GROUP BY
CASE
      WHEN Age BETWEEN 18 AND 23 THEN '18-23'
      WHEN Age BETWEEN 24 AND 29 THEN '24-29'
      WHEN Age BETWEEN 30 AND 35 THEN '30-35'
      WHEN Age BETWEEN 36 AND 41 THEN '36-41'
      WHEN Age BETWEEN 42 AND 47 THEN '42-47'
      WHEN Age BETWEEN 48 AND 60 THEN '48-60'
ELSE '60+'
```



Retrieve the 'Time period' and the number of 'Transactions', grouped by 'Time period'.

```
SELECT
 CASE
   WHEN DATEPART(HOUR, sale time) < 12 THEN 'Morning'
   WHEN DATEPART(HOUR, sale_time) BETWEEN 12 AND 17 THEN 'Afternoon'
   WHEN DATEPART (HOUR, sale time) BETWEEN 18 AND 21 THEN 'Evening'
   ELSE 'Night'
 END AS 'Time Period', COUNT(transactions_id) AS 'Transactions'
FROM Sales Analysis
GROUP BY
CASE
   WHEN DATEPART(HOUR, sale_time) < 12 THEN 'Morning'
   WHEN DATEPART(HOUR, sale_time) BETWEEN 12 AND 17 THEN 'Afternoon'
   WHEN DATEPART (HOUR, sale time) BETWEEN 18 AND 21 THEN 'Evening'
   ELSE 'Night'
 END
ORDER BY 'Transactions' DESC;
Time Period
                 Transactions
      Evening
                 855
 1
 2
      Morning
                 561
                 377
 3
      Afternoon
      Night
                 207
 4
```