Customer Insights

SQL Challenge

Sannidhya Das

@Steel Data
@Matthew Steel

Introduction

I'm Sannidhya Das , the Customer Insights Analyst at 'The General Store.' Today, we're diving deep into the data to extract pivotal insights about our customers. Our focus is to equip the marketing team with actionable information that directly impacts decision-making.

The objective? Uncover key details: Who are our most valuable customers? What products do they prefer? When and how do they shop? Through SQL analysis, we're dissecting tables to reveal patterns and trends that will guide marketing strategies.

This presentation is all about data-driven decision-making.

By the end, we aim to arm the marketing team with insights that enhance customer engagement,satisfaction, and loyal

Let's get straight into the numbers and extract the information that will steer 'The General Store' towards even greater succe

Thank you, and let's dive into the data.

Schema Diagram

baskets

order_id	product_id
1	1
1	2
1	5
2	4
3	3
4	2
4	1
5	3
5	5
6	4
6	3
6	1
7	2
7	1
8	3
8	3

products

product_id	category	price
1	food	5.99
2	sports	12.49
3	vitamins	6.99
4	food	0.89
5	vitamins	15.99

country

country_id	country_name	head_office
1	UK	London
2	USA	New York
3	China	Beijing

customers

customer_id	first_shop	age	rewards	can_email
1	2022-03-20	23	yes	no
2	2022-03-25	26	no	no
3	2022-04-06	32	no	no
4	2022-04-13	25	yes	yes
5	2022-04-22	49	yes	yes
6	2022-06-18	28	yes	no
7	2022-06-30	36	no	no
8	2022-07-04	37	yes	yes

orders

order_id	customer_id	date_shop	sales_channel	country_id
1	1	2023-01-16	retail	1
2	4	2023-01-20	retail	1
3	2	2023-01-25	retail	2
4	3	2023-01-25	online	1
5	1	2023-01-28	retail	3
6	5	2023-02-02	online	1
7	6	2023-02-05	retail	1
8	3	2023-02-11	online	3

Questions

Answer the following questions

Then write a LinkedIn post saying what you have learnt or enjoyed

Make sure to tag @Steel Data and @Matthew Steel

1. What are the names of all the countries in the country table?
2. What is the total number of customers in the customers table?
3. What is the average age of customers who can receive marketing emails (can_email is set to 'yes')?
4. How many orders were made by customers aged 30 or older?
5. What is the total revenue generated by each product category?
6. What is the average price of products in the 'food' category?
7. How many orders were made in each sales channel (sales_channel column) in the orders table?
8. What is the date of the latest order made by a customer who can receive marketing emails?
9. What is the name of the country with the highest number of orders?
10. What is the average age of customers who made orders in the 'vitamins' product category?

Q1. What are the names of all the countries in the country table?

```
SELECT country_name FROM country;
```

Output Messages Notifications

country_name character varying (50)
UK
USA
China

Q2. What is the total number of customers in the customers table?

```
SELECT COUNT(*) FROM customers;
```

Output Messages Notifications

count bigint
8

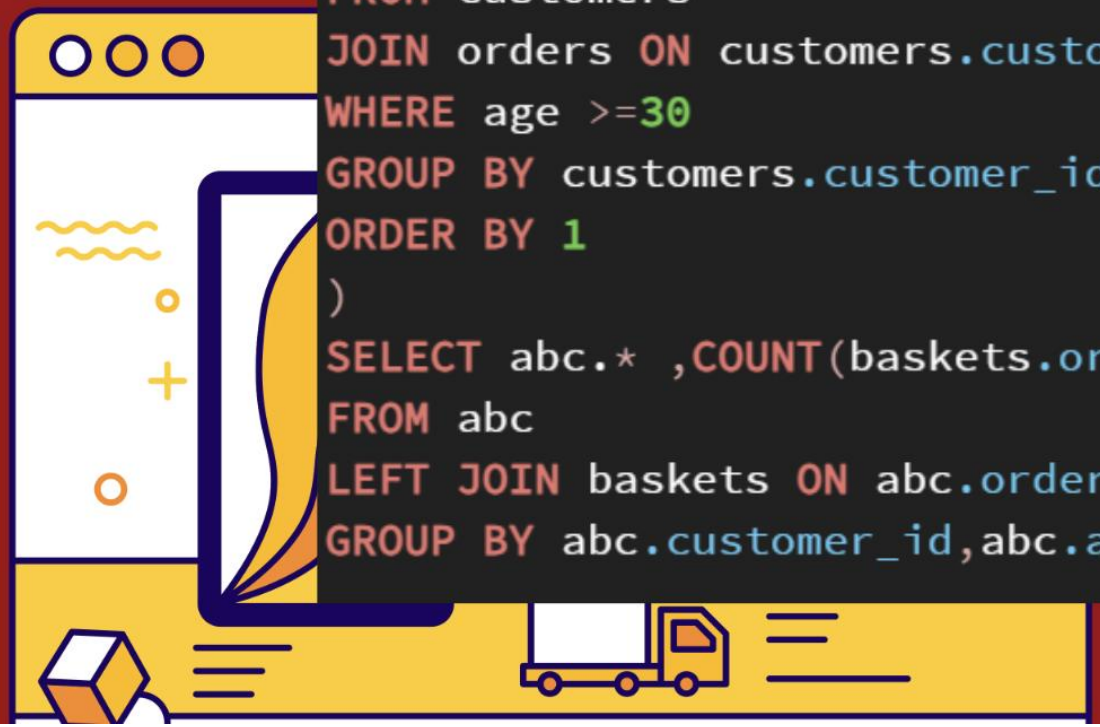
Q3. What is the average age of customers who can receive marketing emails (can_email is set to 'yes')?

```sql
SELECT ROUND(AVG(age),2) AS avg_age
FROM customers
WHERE can_email = 'yes';
```

Output Messages Notifications

avg_age numeric
37.00

Q4. How many orders were made by customers aged 30 or older?



```sql
WITH abc AS (

SELECT customers.customer_id,age,order_id
FROM customers
JOIN orders ON customers.customer_id=orders.customer_id
WHERE age >=30
GROUP BY customers.customer_id,orders.order_id
ORDER BY 1
)
SELECT abc.* ,COUNT(baskets.order_id) AS total_orders
FROM abc
LEFT JOIN baskets ON abc.order_id=baskets.order_id
GROUP BY abc.customer_id,abc.age,abc.order_id;
```

customer_id integer	age integer	order_id integer	total_orders bigint
5	49	6	3
3	32	8	2
3	32	4	2

Q5. What is the total revenue generated by each product category?


```sql
SELECT products.product_id,category,
COUNT(baskets.product_id) AS total_orders,
ROUND(COUNT(baskets.product_id)*price,2) AS total_revenue
FROM products
JOIN baskets ON products.product_id=baskets.product_id
GROUP BY products.product_id
ORDER BY category
```

Output | Messages | Notifications

product_id [PK] integer	category character varying (50)	total_orders bigint	total_revenue numeric
4	food	2	1.78
1	food	4	23.96
2	sports	3	37.47
3	vitamins	5	34.95
5	vitamins	2	31.98

Note : But in the previous output we have two types of foods and vitamins and for getting our total revenue we have to add them also

```sql
WITH abcd AS (

SELECT category,
COUNT(baskets.product_id) AS total_orders,
ROUND(COUNT(baskets.product_id)*price,2) AS total_revenue
FROM products
JOIN baskets ON products.product_id=baskets.product_id
GROUP BY products.category,products.price
ORDER BY category
)
SELECT category,SUM(total_orders) AS total_orders,
SUM(total_revenue) AS total_revenue
FROM abcd
GROUP BY abcd.category
```

category character varying (50)	total_orders numeric	total_revenue numeric
food	6	25.74
sports	3	37.47
vitamins	7	66.93

And this is our final table what we want .....

Q6. What is the average price of products in the 'food' category?

```
SELECT category, ROUND(AVG(price),2) AS avg_price
FROM products
WHERE category = 'food'
GROUP BY category;
```

Output Messages Notifications

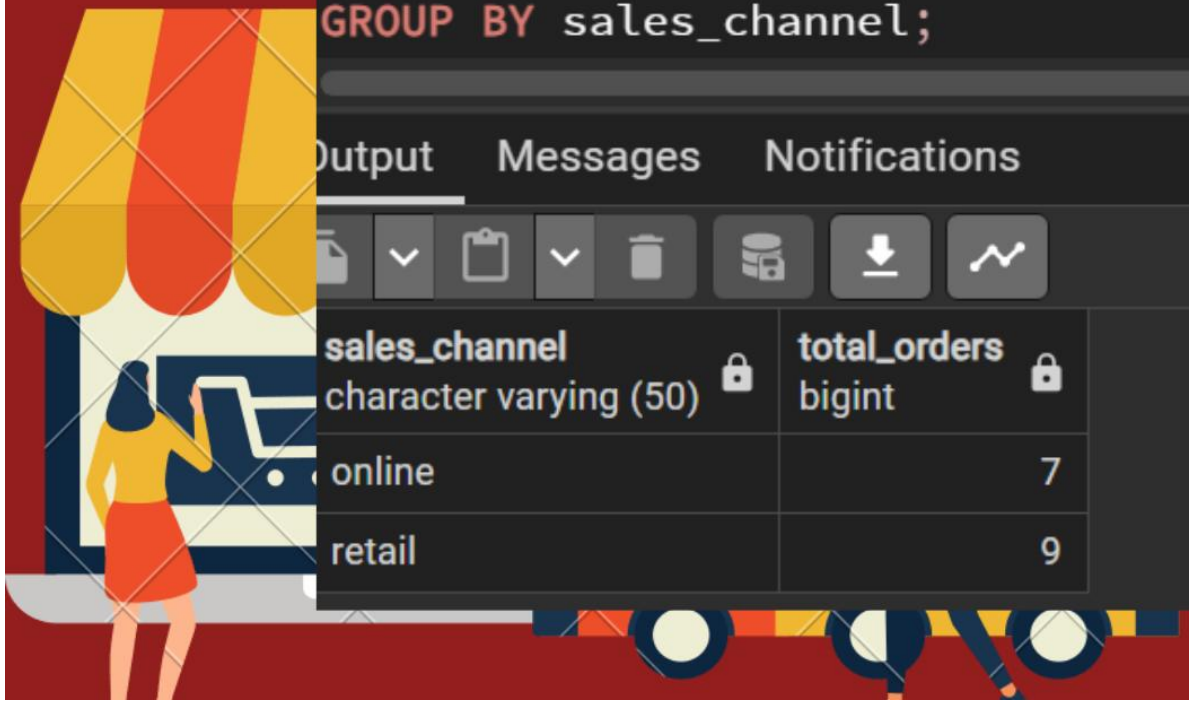
category character varying (50)	avg_price numeric
food	3.44

Q7. How many orders were made in each sales channel (sales_channel column) in the orders table?

```sql
SELECT sales_channel,COUNT(baskets.order_id) AS total_orders
FROM orders
JOIN baskets ON orders.order_id=baskets.order_id
GROUP BY sales_channel;
```

Output | Messages | Notifications

sales_channel character varying (50)	total_orders bigint
online	7
retail	9

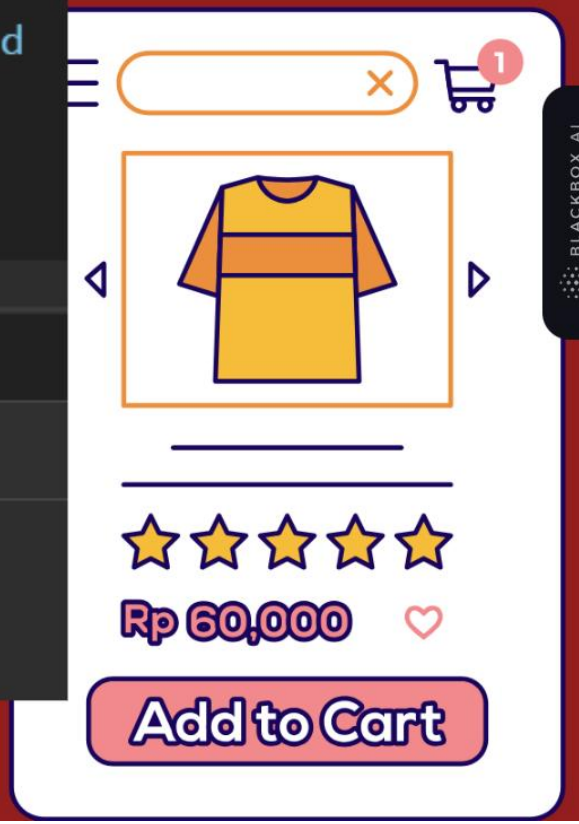


Q8. What is the date of the latest order made by a customer who can receive marketing emails?

```sql
SELECT customers.customer_id,date_shop,customers.can_email
FROM orders
JOIN customers ON customers.customer_id=orders.customer_id
WHERE can_email='yes'
ORDER BY 2 DESC
LIMIT 1;
```

Output | Messages | Notifications

customer_id integer	date_shop date	can_email character varying (50)
5	2023-02-02	yes

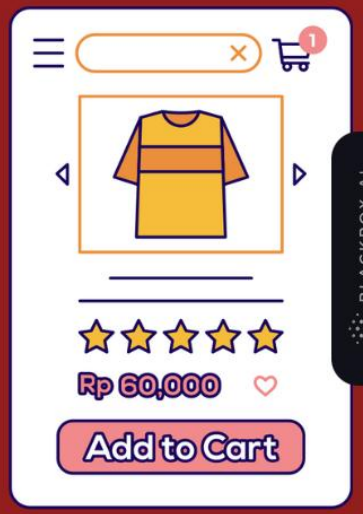


We can answer it another way.....

```sql
SELECT MAX(date_shop) AS latest_order_date
FROM orders
RIGHT JOIN customers ON customers.customer_id=orders.customer_id
WHERE can_email='yes'
```

Output | Messages | Notifications

latest_order_date date
2023-02-02



Q9. What is the name of the country with the highest number of orders?

```sql
SELECT country.country_id,country_name,
COUNT(baskets.order_id) AS total_num_of_orders
FROM country
JOIN orders ON country.country_id=orders.country_id
JOIN baskets ON orders.order_id=baskets.order_id
GROUP BY country.country_id
ORDER BY 3 DESC
LIMIT 1;
```

Output | Messages | Notifications

country_id [PK] integer	country_name character varying (50)	total_num_of_orders bigint
1	UK	11

Q10. What is the average age of customers who made orders in the'vitamins' product category?

```sql
WITH abcde AS(

SELECT DISTINCT(customers.customer_id),age,products.category
FROM customers
JOIN orders ON customers.customer_id=orders.customer_id
JOIN baskets ON orders.order_id=baskets.order_id
JOIN products ON baskets.product_id=products.product_id
WHERE category='vitamins'
)
SELECT ROUND(AVG(age)) AS average_age,category FROM abcde
GROUP BY abcde.category;
```

Output | Messages | Notifications

average_age numeric	category character varying (50)
33	vitamins

Thank You!

Don’t hesitate to follow me

Sannidhya Das