Behind the numbers: exploring the sales story of a supermarket

Sales data is a powerful tool that holds tremendous potential for supermarkets to discover growth opportunities. It's not just about understanding revenue; by examining sales numbers, supermarkets can uncover emerging trends, adjust their product selection, fine-tune pricing strategies, and improve the overall shopping experience for customers. In today's competitive market, it is crucial for supermarkets to grasp and make the most of sales data in order to flourish.

In this project, I looked at a dataset from a supermarket and tried to find answers to a few questions:

- Popular days and hours for ordering: I figured out the days and times when most customers place their orders. This helps the supermarket know the busy periods and plan accordingly.
- Popular and frequently ordered and reordered products: I checked which products are in high demand and which ones customers tend to reorder. This helps the supermarket understand what customers like and what they prefer to buy repeatedly.
- Number of products added to the cart: I looked at how many products customers usually add to their shopping carts. This gives an idea of how much they buy in one go.
- Basic customer analysis: I briefly analysed customer information like age and gender to see if there are any patterns in their buying habits.

By exploring these aspects of the dataset, I gained insights into popular products, shopping trends and customer behaviour and preferences. This information can help the supermarket make better decisions to improve customer satisfaction and optimise their operations.

Data source and ingestion

The datasets were taken from Kaggle: (https://www.kaggle.com/datasets/hunter0007/ecommerce-dataset-for-predictive-marketing-2023 and https://www.kaggle.com /datasets /sindraanthony9985/marketing-data-for-a-supermarket-in-united-states).

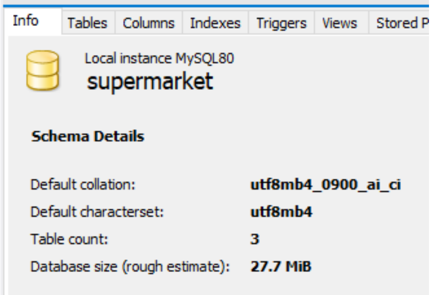
I made some modifications to the original datasets. I split the "ECommerce_ consumer_behaviour.csv" file into two separate files: "Products" with product_id, product_name, department_id and department_name columns and "Transactions" with the following columns: transaction_id, customer_id, order_number, order_day_of_week, order_hour_of_day, days_since_previous_order, product_id, add_to_cart_order and reordered. From the other dataset, I only kept the "Supermarket_CustomerMembers.csv" file and renamed it as "Customers".

Regarding data cleanliness, the overall quality was fairly good. However, I did make some adjustments. I replaced the values for transaction_id, product_id and customer_id. Additionally, the days_since_prior_order column had multiple empty rows. To ensure smooth data loading and avoid errors, I filled those empty rows with "null".

Creating the database, tables and loading the data

Once the dataset was prepared, the next step was to load it into MySQL. My approach was to create a database named "supermarket" and then for each CSV file I have created a corresponding table.

Here is an overview of the schema and tables:



Local instance MySQL80

supermarket

Schema Details

Default collation:	**utf8mb4_0900_ai_ci**
Default characterset:	**utf8mb4**
Table count:	**3**
Database size (rough estimate):	**27.7 MiB**

Fig. 1 Database schema

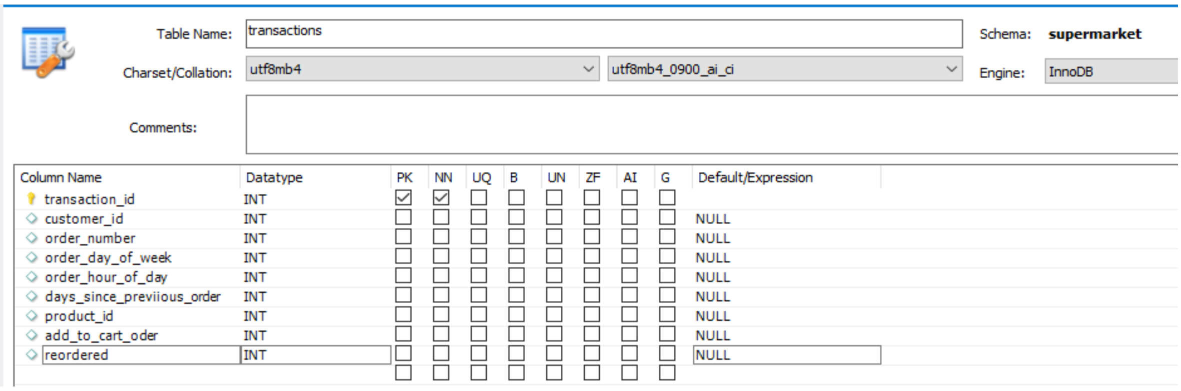


Table Name:	transactions	Schema:	**supermarket**
Charset/Collation:	utf8mb4 / utf8mb4_0900_ai_ci	Engine:	InnoDB
Comments:			

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
transaction_id	INT	☑	☑	□	□	□	□	□	□	
customer_id	INT	□	□	□	□	□	□	□	□	NULL
order_number	INT	□	□	□	□	□	□	□	□	NULL
order_day_of_week	INT	□	□	□	□	□	□	□	□	NULL
order_hour_of_day	INT	□	□	□	□	□	□	□	□	NULL
days_since_previious_order	INT	□	□	□	□	□	□	□	□	NULL
product_id	INT	□	□	□	□	□	□	□	□	NULL
add_to_cart_oder	INT	□	□	□	□	□	□	□	□	NULL
reordered	INT	□	□	□	□	□	□	□	□	NULL
		□	□	□	□	□	□	□	□	

Fig 2. Transactions table

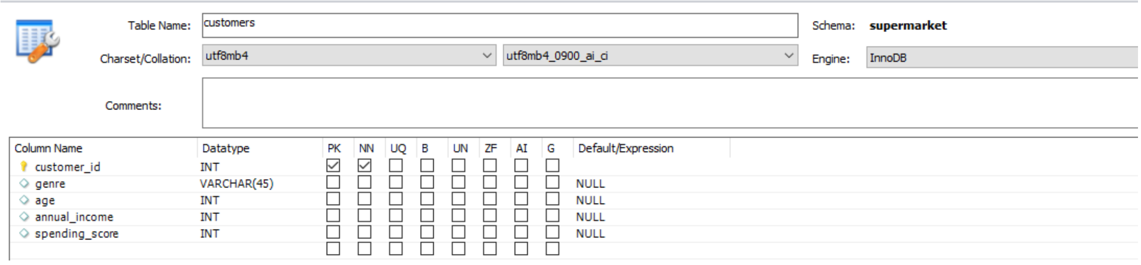


Table Name:	customers	Schema:	**supermarket**
Charset/Collation:	utf8mb4 / utf8mb4_0900_ai_ci	Engine:	InnoDB
Comments:			

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
customer_id	INT	☑	☑	□	□	□	□	□	□	
genre	VARCHAR(45)	□	□	□	□	□	□	□	□	NULL
age	INT	□	□	□	□	□	□	□	□	NULL
annual_income	INT	□	□	□	□	□	□	□	□	NULL
spending_score	INT	□	□	□	□	□	□	□	□	NULL
		□	□	□	□	□	□	□	□	

Fig. 3 Customers table

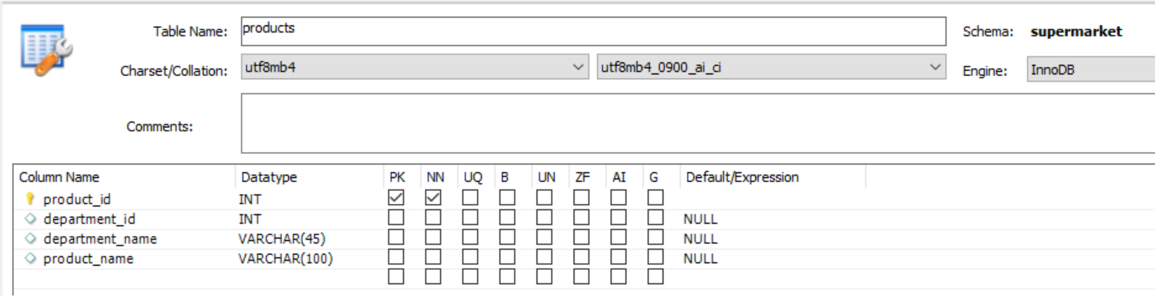
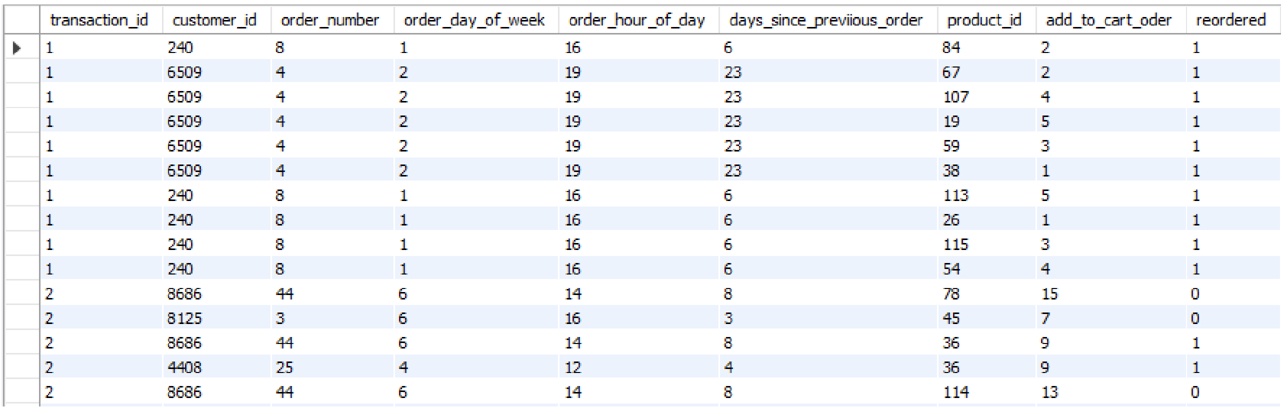


Table Name:	products	Schema:	**supermarket**
Charset/Collation:	utf8mb4 / utf8mb4_0900_ai_ci	Engine:	InnoDB
Comments:			

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
product_id	INT	☑	☑	□	□	□	□	□	□	
department_id	INT	□	□	□	□	□	□	□	□	NULL
department_name	VARCHAR(45)	□	□	□	□	□	□	□	□	NULL
product_name	VARCHAR(100)	□	□	□	□	□	□	□	□	NULL
		□	□	□	□	□	□	□	□	

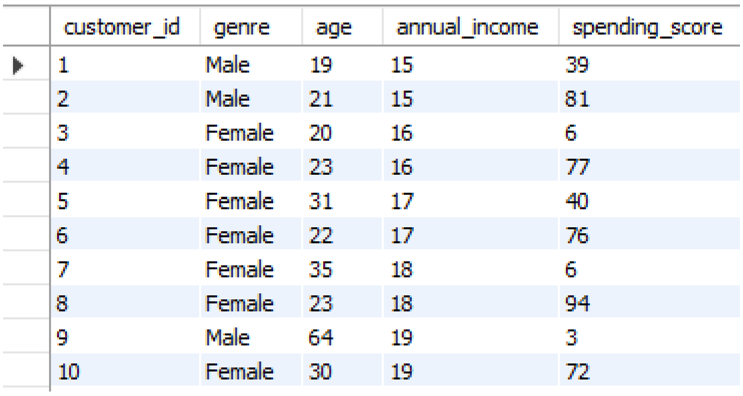
Fig. 4 Products table

Here’s what the tables preview look like:



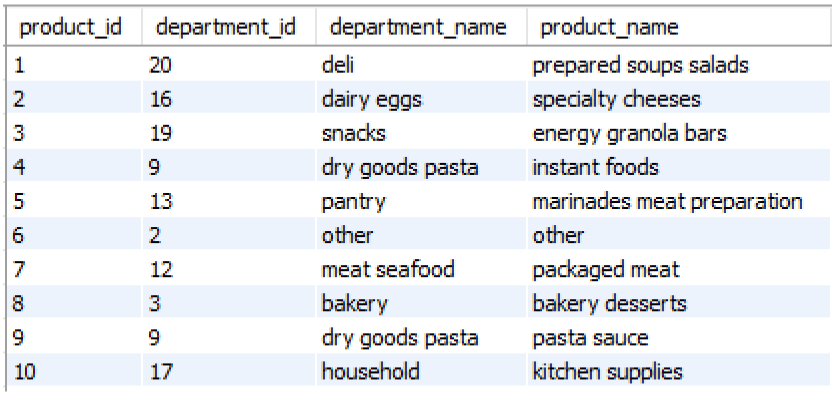
transaction_id	customer_id	order_number	order_day_of_week	order_hour_of_day	days_since_previious_order	product_id	add_to_cart_oder	reordered
1	240	8	1	16	6	84	2	1
1	6509	4	2	19	23	67	2	1
1	6509	4	2	19	23	107	4	1
1	6509	4	2	19	23	19	5	1
1	6509	4	2	19	23	59	3	1
1	6509	4	2	19	23	38	1	1
1	240	8	1	16	6	113	5	1
1	240	8	1	16	6	26	1	1
1	240	8	1	16	6	115	3	1
1	240	8	1	16	6	54	4	1
2	8686	44	6	14	8	78	15	0
2	8125	3	6	16	3	45	7	0
2	8686	44	6	14	8	36	9	1
2	4408	25	4	12	4	36	9	1
2	8686	44	6	14	8	114	13	0

Fig. 5 Transactions table



customer_id	genre	age	annual_income	spending_score
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40
6	Female	22	17	76
7	Female	35	18	6
8	Female	23	18	94
9	Male	64	19	3
10	Female	30	19	72

Fig. 6 Customers table



product_id	department_id	department_name	product_name
1	20	deli	prepared soups salads
2	16	dairy eggs	specialty cheeses
3	19	snacks	energy granola bars
4	9	dry goods pasta	instant foods
5	13	pantry	marinades meat preparation
6	2	other	other
7	12	meat seafood	packaged meat
8	3	bakery	bakery desserts
9	9	dry goods pasta	pasta sauce
10	17	household	kitchen supplies

Fig. 7 Products table

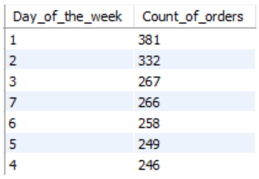
Querying

Once the data was loaded into MySQL, I started querying the tables to derive insights that address the initial analysis questions. I took a step-by-step approach, beginning with determining the most popular days and hours for placing orders.

To find this information, I queried the "transactions" table to determine the days and hours when most orders are placed. By examining this aspect of the dataset, I gained a better understanding of customer ordering patterns. This knowledge could be crucial in determining the optimal times for supermarket operations, staffing and managing inventory effectively.

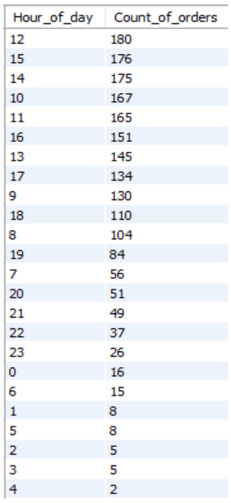
```
# most popular days
SELECT order_day_of_week as Day_of_the_week,
       COUNT(distinct transaction_id) as Count_of_orders
 FROM transactions
 GROUP BY order_day_of_week
 ORDER BY COUNT(distinct transaction_id) desc;

#most popular hours
SELECT order_hour_of_day as Hour_of_day,
            COUNT(distinct transaction_id) as Count_of_orders
FROM transactions
GROUP BY order_hour_of_day
ORDER BY COUNT(distinct transaction_id) desc;
```



Day_of_the_week	Count_of_orders
1	381
2	332
3	267
7	266
6	258
5	249
4	246

Fig. 8 Most popular days



Hour_of_day	Count_of_orders
12	180
15	176
14	175
10	167
11	165
16	151
13	145
17	134
9	130
18	110
8	104
19	84
7	56
20	51
21	49
22	37
23	26
0	16
6	15
1	8
5	8
2	5
3	5
4	2

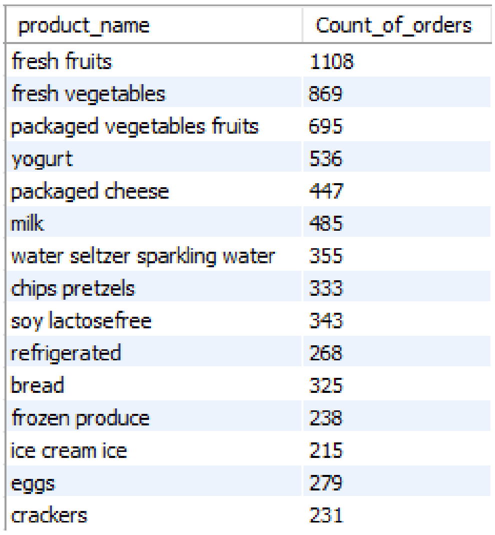
Fig.9 Most popular hours

The analysis revealed that customers prefer to do their shopping mainly on Monday, Tuesday and Wednesday and the peak hours are 12 and 15. This information provides valuable insights into customer behaviour, helping to optimise supermarket operations and staffing during those peak times.

Moving forward, I delved into understanding the most in-demand products. To achieve this, I performed a query to identify the top 15 most frequently ordered and reordered products. To accomplish this, I joined two tables: "transactions" and "products". This analysis could assist in making informed decisions about product inventory, marketing strategies and ensuring the availability of popular items.

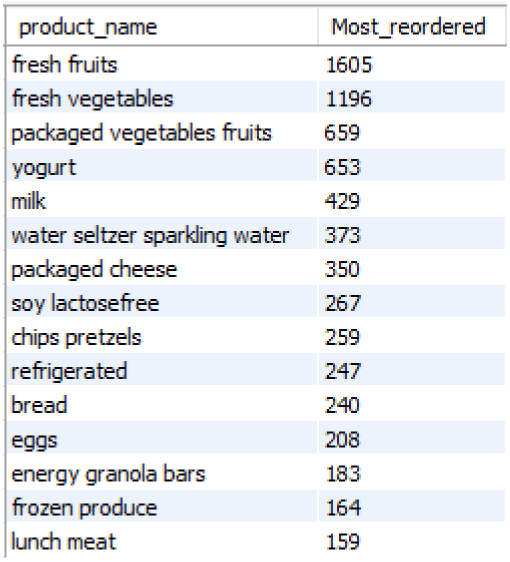
```
# top 15 most sold products
SELECT product_name,
       COUNT(distinct transaction_id) as Count_of_orders
FROM transactions as t
LEFT JOIN products as p ON t.product_id = p.product_id
GROUP BY product_name
ORDER BY COUNT(transaction_id) desc
LIMIT 15;

# top 15 most reordered products
SELECT product_name,
      SUM(reordered) as Most_reordered
FROM transactions as t
LEFT JOIN products as p ON t.product_id = p.product_id
GROUP BY product_name
ORDER BY SUM(reordered) desc
LIMIT 15;
```



product_name	Count_of_orders
fresh fruits	1108
fresh vegetables	869
packaged vegetables fruits	695
yogurt	536
packaged cheese	447
milk	485
water seltzer sparkling water	355
chips pretzels	333
soy lactosefree	343
refrigerated	268
bread	325
frozen produce	238
ice cream ice	215
eggs	279
crackers	231

Fig. 10 Top 15 bought products



product_name	Most_reordered
fresh fruits	1605
fresh vegetables	1196
packaged vegetables fruits	659
yogurt	653
milk	429
water seltzer sparkling water	373
packaged cheese	350
soy lactosefree	267
chips pretzels	259
refrigerated	247
bread	240
eggs	208
energy granola bars	183
frozen produce	164
lunch meat	159

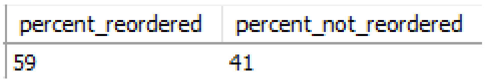
Fig. 11 Top 15 reordered products

As it can be seen, the top 15 most sold products were also the most reordered, except crackers. This observation indicates a strong correlation between sales and reorder patterns for most products. Customers consistently show a preference for purchasing and repurchasing these items, emphasising their popularity and demand.

To gain a deeper understanding of the ordering behaviour, I investigated the distinction between reordered and not reordered products as a percentage and then, specifically categorised by department.

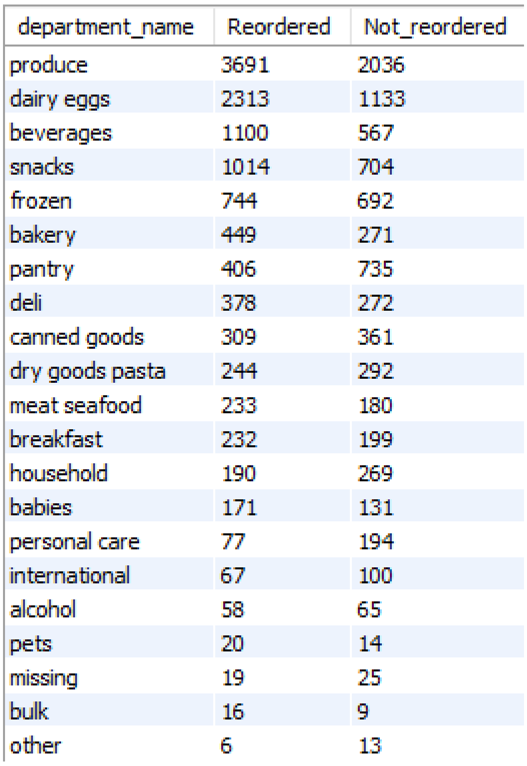
```
    # percent reordered vs not reordered
    SELECT ROUND(((COUNT(CASE WHEN reordered = 1 THEN 1 END))/(SELECT
COUNT(reordered) FROM transactions))*100) as percent_reordered,
              ROUND(((COUNT(CASE WHEN reordered = 0 THEN 1 END))/(SELECT
      COUNT(reordered) FROM transactions))*100) as percent_not_reordered
    FROM transactions;

    # reordered vs not reordered by department
    SELECT department_name,
              COUNT(CASE WHEN reordered = 1 THEN 1 END) as Reordered,
              COUNT(CASE WHEN reordered = 0 THEN 1 END) as Not_reordered
    FROM transactions as t
    LEFT JOIN products as p ON t.product_id = p.product_id
    GROUP BY department_name
    ORDER BY reordered desc;
```



percent_reordered	percent_not_reordered
59	41

Fig. 12 Percentage of reordered and not reordered



department_name	Reordered	Not_reordered
produce	3691	2036
dairy eggs	2313	1133
beverages	1100	567
snacks	1014	704
frozen	744	692
bakery	449	271
pantry	406	735
deli	378	272
canned goods	309	361
dry goods pasta	244	292
meat seafood	233	180
breakfast	232	199
household	190	269
babies	171	131
personal care	77	194
international	67	100
alcohol	58	65
pets	20	14
missing	19	25
bulk	16	9
other	6	13

Fig. 13 Reordered status by departments

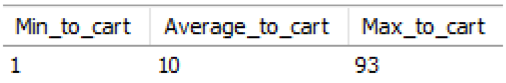
By looking at the above output it is evident that the majority of customers tend to reorder products from departments such as produce, dairy eggs, beverages and snacks. These departments showed a higher frequency of reordered products compared to not ordered ones. It's interesting to observe that in 9 out of the 21 departments in the supermarket, there were more products that were not ordered compared to those that were ordered.

Understanding the departments with a higher reorder rate could provide valuable insights, as it can highlight the areas where customers have established loyalty and repeat purchase behaviour, enabling the supermarket to optimise its inventory, promotional strategies and customer engagement efforts accordingly.

Another important aspect to look at was the number of products added to cart, specifically the min, average and max. To examine this, I have created a subquery that calculates the count of products for each transaction. From this subquery, I have extracted the minimum, average and maximum values to gain insights into customers' purchasing habits.

```
# min, avg and max number of products to cart
SELECT MIN(prod_count) as Min_to_cart,
            ROUND(AVG(prod_count)) as Average_to_cart,
             MAX(prod_count) as Max_to_cart
FROM
             (SELECT transaction_id, COUNT(product_id) as Prod_count
              FROM transactions
              GROUP BY transaction_id) as Count_to_cart;
```

And the output is:



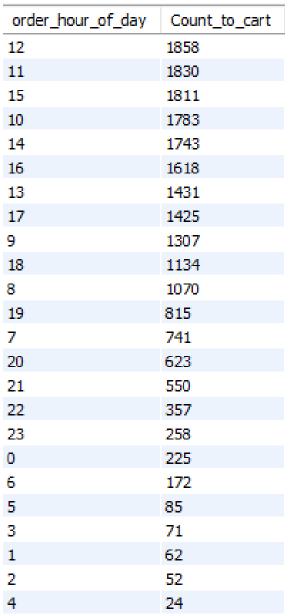
Min_to_cart	Average_to_cart	Max_to_cart
1	10	93

Fig. 14 Min, average and max products to cart

The minimum value indicated that some customers only added one product to their cart, while the average value revealed that most customers added around 10 products per order. On the other hand, the maximum value showed that some customers added as many as 93 products in a single order. This data point identified customers who engaged in larger shopping sprees or may be more likely to make bulk purchases. This statistic showed the general purchasing behaviour of customers and helped in inventory planning, resource allocation and promotional strategies.

I also wanted to determine the time of day when most products are added to the cart. I have created the below query to answer the question.

```
SELECT order_hour_of_day, COUNT(product_id) as Count_to_cart
FROM transactions
GROUP BY order_hour_of_day
ORDER BY COUNT(product_id) desc;
```



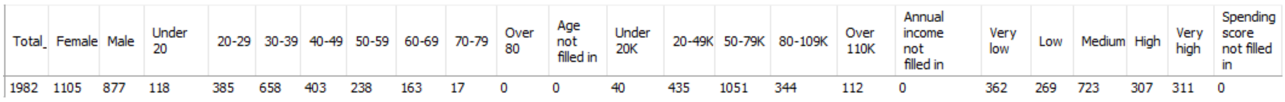
order_hour_of_day	Count_to_cart
12	1858
11	1830
15	1811
10	1783
14	1743
16	1618
13	1431
17	1425
9	1307
18	1134
8	1070
19	815
7	741
20	623
21	550
22	357
23	258
0	225
6	172
5	85
3	71
1	62
2	52
4	24

Fig. 15 Peak hours for most products added to cart

With this I identified the peak hours during which customers tend to add more items to their carts, specifically 12, 11 and 15 hours. These findings align with our prior understanding of peak shopping hours, further confirming the significance of these time periods. This information could be valuable in optimising staffing, inventory management and overall operational efficiency during those busy periods.

Next, I wanted to gain a better understanding of customers and their characteristics. To achieve this, I have created a query that categorises customers based on their gender, age, income (thousands) and spending score (1-100). By running this query, I gained insights into the diverse traits of the customer base.

```
# number of customer by gender, age and income
SELECT COUNT(*) as Total_customers,
      COUNT(CASE WHEN genre = 'Female' THEN 1 END) as Female,
      COUNT(CASE WHEN genre = 'Male' THEN 1 END) as Male,
      SUM(IF(age<20,1,0)) as 'Under 20',
      SUM(IF(age BETWEEN 20 and 29,1,0)) as '20-29',
      SUM(IF(age BETWEEN 30 and 39,1,0)) as '30-39',
      SUM(IF(age BETWEEN 40 and 49,1,0)) as '40-49',
      SUM(IF(age BETWEEN 50 and 59,1,0)) as '50-59',
      SUM(IF(age BETWEEN 60 and 69,1,0)) as '60-69',
      SUM(IF(age BETWEEN 70 and 79,1,0)) as '70-79',
      SUM(IF(age > 80,1,0)) as 'Over 80',
      SUM(IF(age is null,1,0)) as 'Age not filled in',
      SUM(IF(annual_income < 20,1,0)) as 'Under 20K',
      SUM(IF(annual_income BETWEEN 20 and 49,1,0)) as '20-49K',
      SUM(IF(annual_income BETWEEN 50 and 79,1,0)) as '50-79K',
      SUM(IF(annual_income BETWEEN 80 and 109,1,0)) as '80-109K',
      SUM(IF(annual_income > 110,1,0)) as 'Over 110K',
      SUM(IF(annual_income is null,1,0)) as 'Annual income not filled in',
      SUM(IF(spending_score <= 20,1,0)) as 'Very low',
      SUM(IF(spending_score BETWEEN 21 and 40,1,0)) as 'Low',
      SUM(IF(spending_score BETWEEN 41 and 60,1,0)) as 'Medium',
      SUM(IF(spending_score BETWEEN 61 and 80,1,0)) as 'High',
      SUM(IF(spending_score > 81,1,0)) as 'Very high',
      SUM(IF(spending_score is null,1,0)) as 'Spending score not filled in'
FROM customers;
```



Total_	Female	Male	Under 20	20-29	30-39	40-49	50-59	60-69	70-79	Over 80	Age not filled in	Under 20K	20-49K	50-79K	80-109K	Over 110K	Annual income not filled in	Very low	Low	Medium	High	Very high	Spending score not filled in
1982	1105	877	118	385	658	403	238	163	17	0	0	40	435	1051	344	112	0	362	269	723	307	311	0

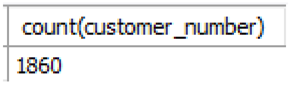
Fig. 16 Customers overview

Categorising customers by gender provided an understanding of the distribution between male and female customers, allowing the supermarket to tailor marketing campaigns and promotions accordingly. Analysing age groups helped identify different generational segments, enabling for customised products and services to meet the specific needs and preferences of each age group. The split by income level provided insights into the purchasing power and potential spending capacity of different customer segments. Lastly, analysing the spending score helped identify high-value customers and target them with personalised offers and rewards.

For a more comprehensive analysis, I wanted to determine the number of customers who have placed a single order and the number of customers who have placed three or more orders. This provides insights into customer loyalty and engagement with the supermarket.

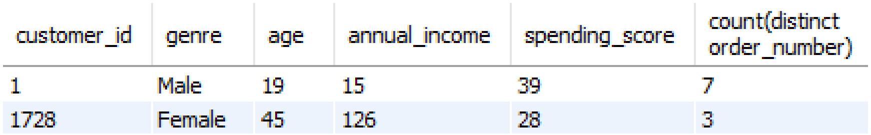
```
# number of customers that ordered once
SELECT COUNT(customer_number) as ordered_once
FROM
                    (SELECT customer_id as customer_number
                    FROMtransactions
                    GROUP BY customer_id
                    HAVING COUNT(distinct order_number) = 1) as order_count;

# profile of customers who placed 3 or more orders
 SELECT t.customer_id, genre, age, annual_income, spending_score,
             COUNT(DISTINCT order_number)
FROM transactions as t
LEFT JOIN customers as c ON t.customer_id = c.customer_id
GROUP BY t.customer_id
HAVING COUNT(DISTINCT order_number) >= 3
ORDER BY COUNT(DISTINCT order_number) desc;
```



count(customer_number)
1860

Fig. 17 Number of customers with one order

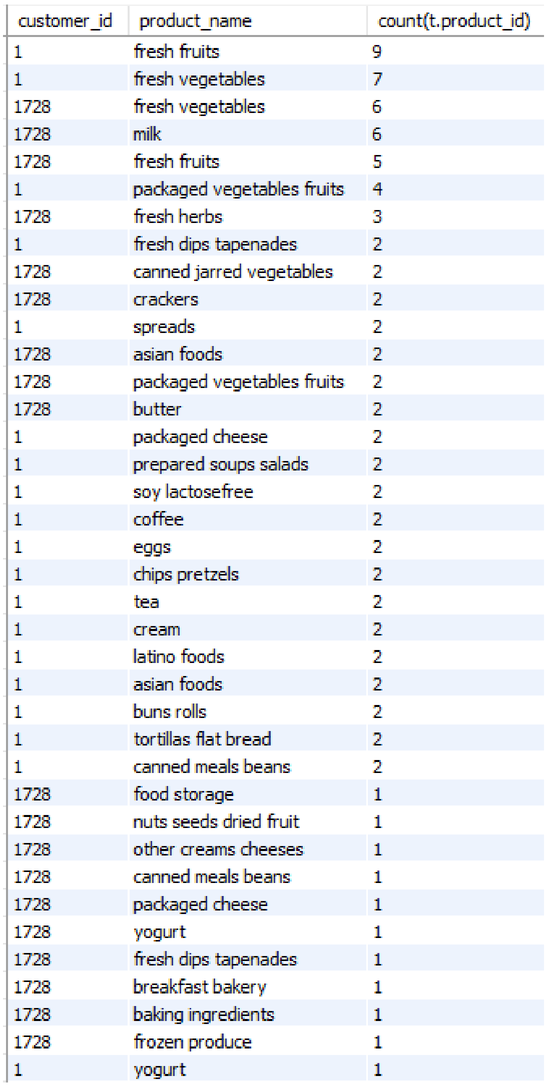


customer_id	genre	age	annual_income	spending_score	count(distinct order_number)
1	Male	19	15	39	7
1728	Female	45	126	28	3

Fig. 18 Customers that have placed 3 or more orders

To determine the specific products that customer 1 and customer 1728 have ordered, I have queried the dataset to retrieve their respective order details.

```
SELECT customer_id, product_name, COUNT(t.product_id)
FROM transactions as t
LEFT JOIN products as p ON t.product_id = p.product_id
GROUP BY customer_id, product_name
HAVING customer_id = 1 or customer_id = 1728
ORDER BY COUNT(product_id) desc;
```



customer_id	product_name	count(t.product_id)
1	fresh fruits	9
1	fresh vegetables	7
1728	fresh vegetables	6
1728	milk	6
1728	fresh fruits	5
1	packaged vegetables fruits	4
1728	fresh herbs	3
1	fresh dips tapenades	2
1728	canned jarred vegetables	2
1728	crackers	2
1	spreads	2
1728	asian foods	2
1728	packaged vegetables fruits	2
1728	butter	2
1	packaged cheese	2
1	prepared soups salads	2
1	soy lactosefree	2
1	coffee	2
1	eggs	2
1	chips pretzels	2
1	tea	2
1	cream	2
1	latino foods	2
1	asian foods	2
1	buns rolls	2
1	tortillas flat bread	2
1	canned meals beans	2
1728	food storage	1
1728	nuts seeds dried fruit	1
1728	other creams cheeses	1
1728	canned meals beans	1
1728	packaged cheese	1
1728	yogurt	1
1728	fresh dips tapenades	1
1728	breakfast bakery	1
1728	baking ingredients	1
1728	frozen produce	1
1	yogurt	1

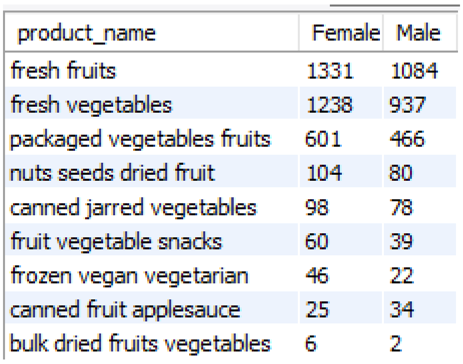
Fig. 19 Products bought by most frequent customers

Upon examining the purchasing patterns of the most frequent customers, it becomes evident that they primarily purchased fruits, vegetables, and milk. This finding aligns with the previous identification of the most popular products among customers. It suggests a correlation between the preferences of the frequent customers and the overall demand for these items.

One last aspect I wanted to explore was the distribution of customers' buying preferences, focusing on fruits and vegetables, as well as meat and alcohol, based on gender. To conduct this analysis, I have created the below queries:

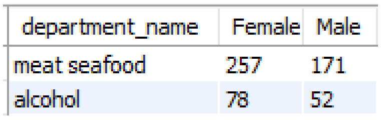
```
# who buys more fruit and veg - men vs women
SELECT product_name, COUNT(CASE WHEN genre = 'Female' THEN 1 END) as
        Female, COUNT(CASE WHEN genre = 'Male' THEN 1 END) as Male
FROM customers as c
RIGHT JOIN transactions as t ON c.customer_id = t.customer_id
LEFT JOIN products as p on t.product_id = p.product_id
GROUP BY product_name
HAVING product_name like '%fruit%' or product_name like '%veg%
ORDER BY Female desc';
```

```
# who buys more meat and alcohol - men vs women
SELECT department_name, COUNT(CASE WHEN genre = 'Female' THEN 1 END)
            as Female, COUNT(CASE WHEN genre = 'Male' THEN 1 END) as Male
FROM customers as c
RIGHT JOIN transactions as t ON c.customer_id = t.customer_id
LEFT JOIN products as p on t.product_id = p.product_id
GROUP BY department_name
HAVING department_name like '%meat%' or department_name like '%alc%';
```



product_name	Female	Male
fresh fruits	1331	1084
fresh vegetables	1238	937
packaged vegetables fruits	601	466
nuts seeds dried fruit	104	80
canned jarred vegetables	98	78
fruit vegetable snacks	60	39
frozen vegan vegetarian	46	22
canned fruit applesauce	25	34
bulk dried fruits vegetables	6	2

Fig. 20 Who buys more fruit and veggies?



department_name	Female	Male
meat seafood	257	171
alcohol	78	52

Fig. 21 What about meat and alcoholic beverages?

Based on the analysis of the purchasing patterns, it is interesting to note that female buyers lead in all the aspects examined, including their preference for fruits and vegetables as well as meat and alcohol. This could be attributed to the fact that a majority of the supermarket's customer base consists of female customers. so it is likely that their buying preferences would reflect in the overall sales patterns.

Visualising insights

The insights obtained from querying the dataset would be better and more easily understood if presented in a visual format. This visual presentation allows for a quicker grasp of the information and makes it more accessible to a wider audience.

To visualise the insights I derived from this dataset I used Tableau Public. Tableau is a great tool for visualising data because it is easy to use and offers powerful features for creating visualisations. I chose Tableau because it is user-friendly, has advanced functions and provides many options for creating engaging visuals.

Most popular days and hours for placing orders

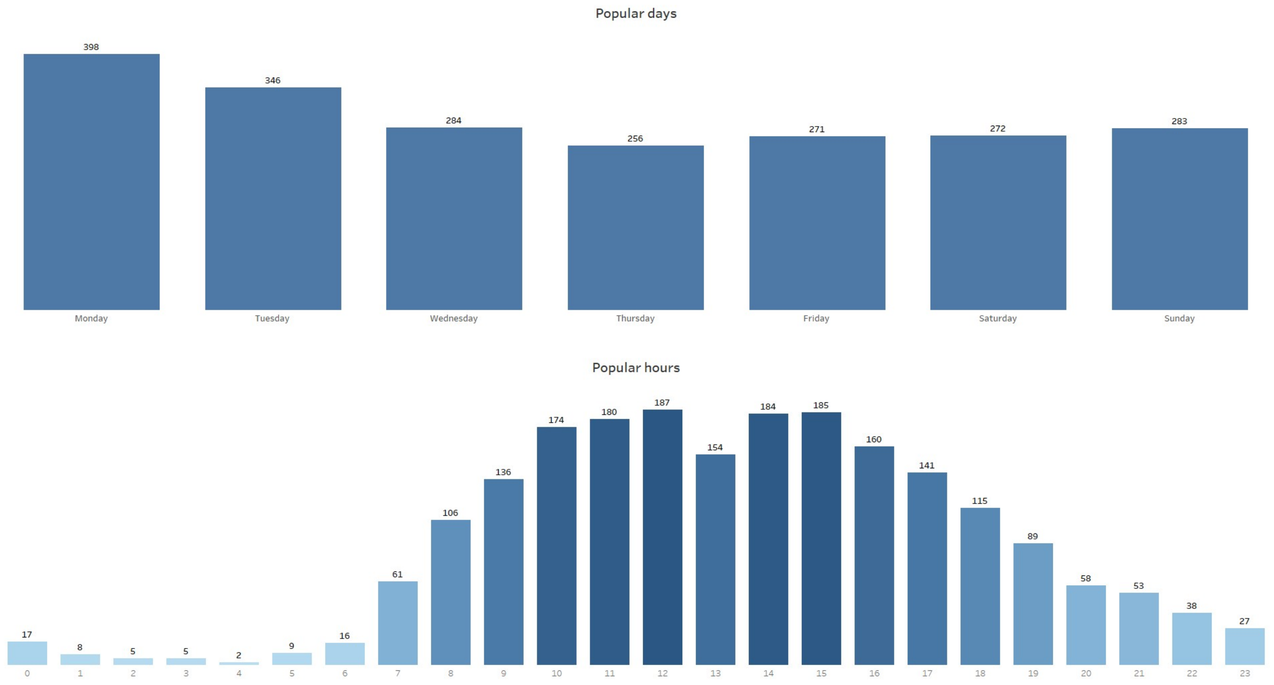


Fig. 22 Most popular days and hours

It can be observed that customers show a preference for shopping primarily on Mondays, Tuesdays and Wednesdays. These days experience higher customer activity compared to other weekdays, with a slight decrease observed on Thursdays. In terms of peak hours, the busiest times for shopping are between 10 AM and 12 PM and 2 PM and 3 PM. These time slots see a significant increase in customer engagement and purchasing.

This information could be valuable for the supermarket as it allows them to plan and allocate resources effectively during these busy periods. By optimising staffing and managing inventory accordingly, the supermarket can ensure a seamless shopping experience and meet customer demands efficiently.

Most in demand and reordered products

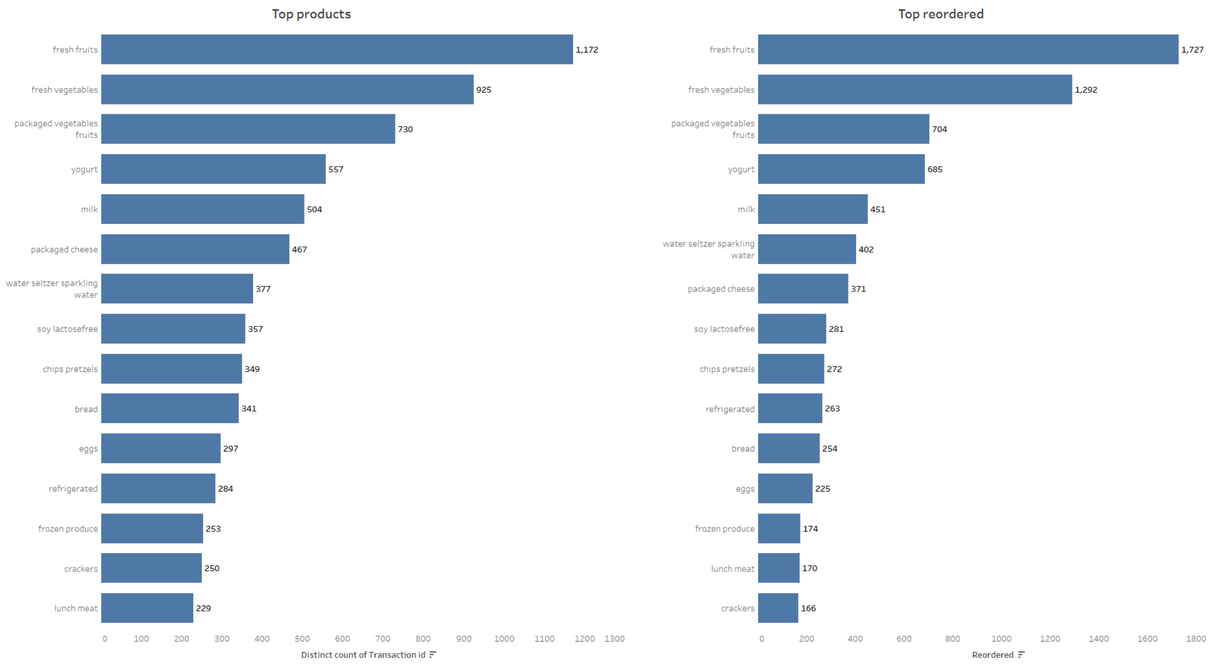


Fig. 23 Top 15 most ordered and reordered products

The analysis revealed an interesting connection between sales and customer behaviour when it comes to reordering products. The top 15 best-selling items, except for crackers, were consistently chosen for repurchase by customers. This shows that these products are highly preferred and in-demand.

Such insights could help the supermarket manage their inventory effectively, ensure these popular products are always available and develop marketing strategies that cater to customer preferences.

Reordered vs not reordered

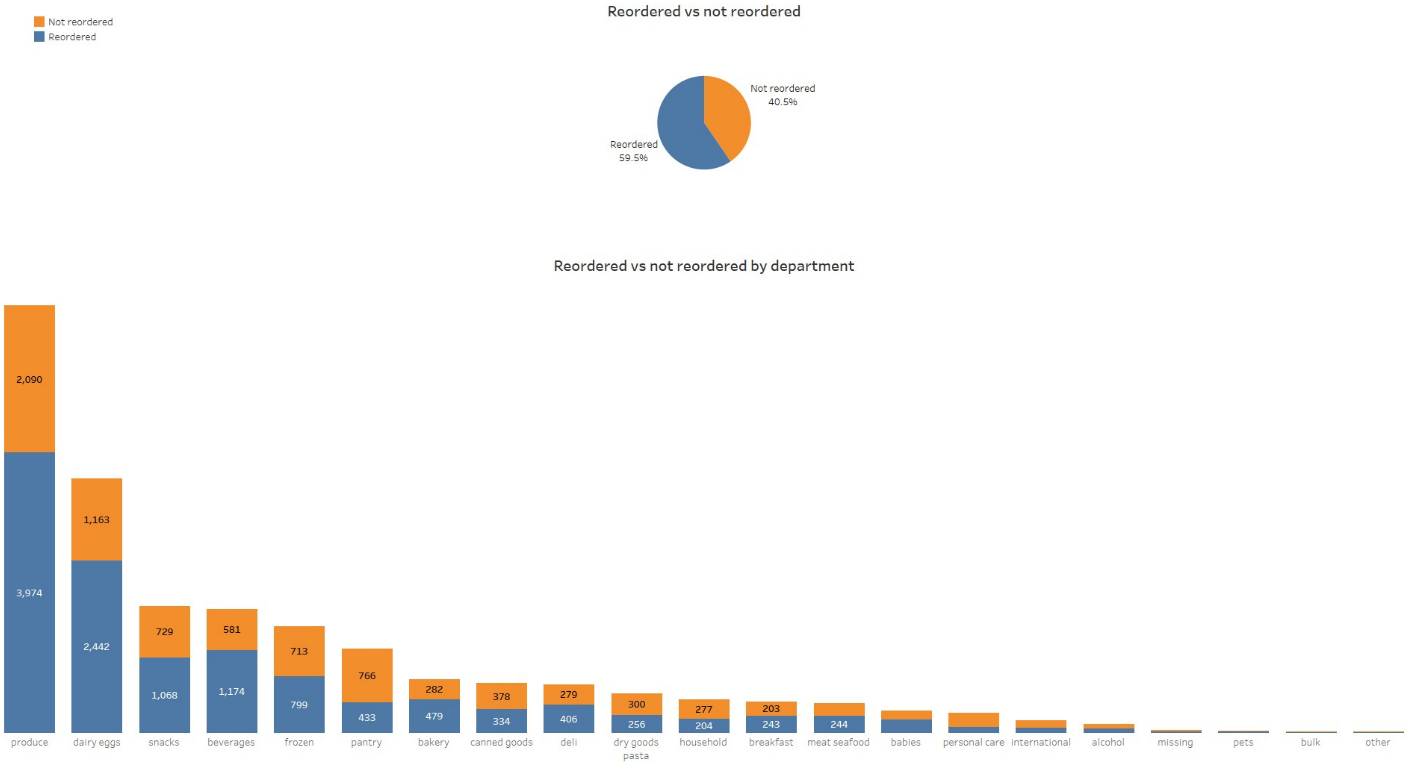


Fig. 24 Reordered vs not reordered

Approximately 59% of the products were reordered, with 49% not reordered. This breakdown suggests that a majority of customers demonstrated a preference for repurchasing items, indicating a level of satisfaction and loyalty towards those specific products.

When examining the departments, it became evident that customers frequently reordered products from departments such as produce, dairy eggs, beverages and snacks. However, it is worth noting that in 9 out of the 21 departments, there were more products that were not ordered compared to those that were ordered. This finding suggests that certain departments faced challenges in generating repeat purchases.

By understanding these patterns and trends, the supermarket can make informed decisions regarding inventory management and marketing strategies. This knowledge will enable them to better meet customer preferences and increase sales in the future.

Products added to cart

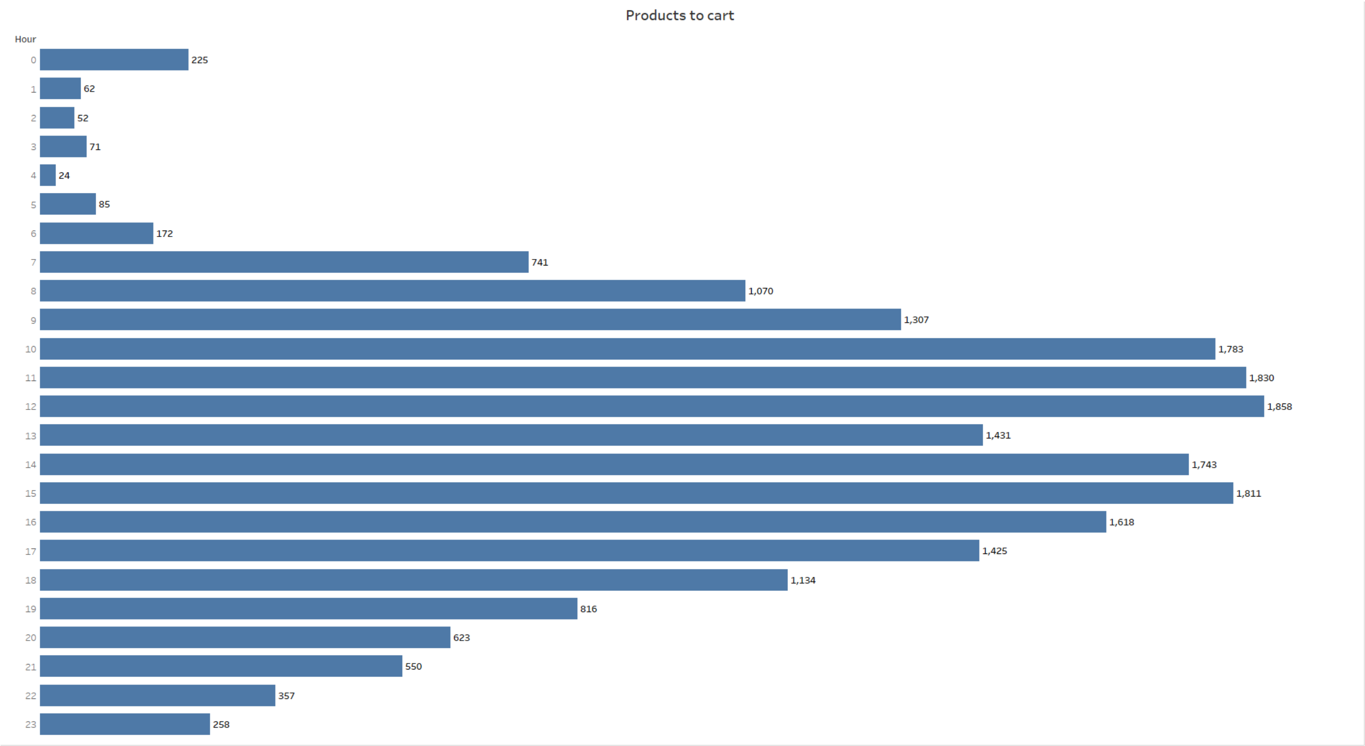


Fig. 25 Products to cart by time of the day

In the analysis, it was discovered that some customers only added one product to their cart, indicating smaller purchases. On average, most customers added around 10 products per order. However, there were also customers who added a maximum of 93 products in a single order, suggesting larger shopping sprees or bulk purchases.

The above graph highlights the peak hours when customers added more items to their carts: 12 PM, 11 AM, and 3 PM. These findings supported the previous knowledge about the busiest shopping times, reinforcing the importance of these specific hours.

This information can be valuable for improving staffing, managing inventory and enhancing overall operational efficiency during these busy periods. By understanding the peak hours, better decisions could be made to ensure that enough staff is available, the right products are stocked and the supermarket runs smoothly to meet customer demands.

Customer overview

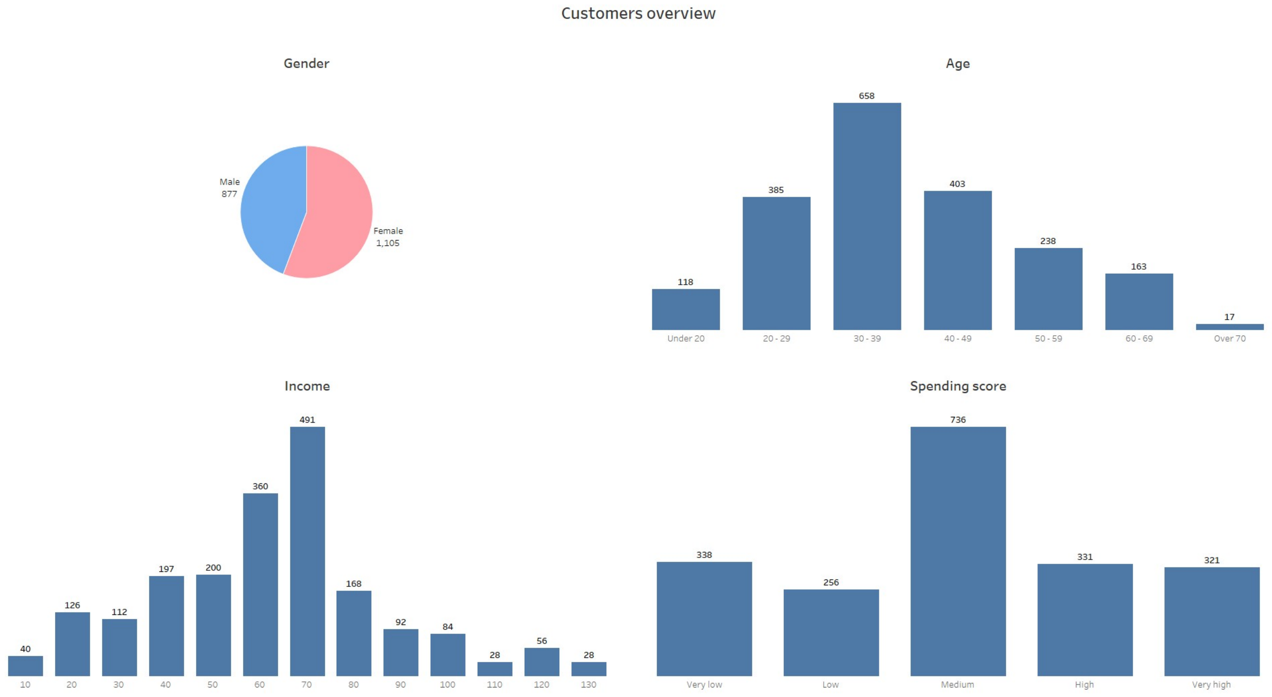
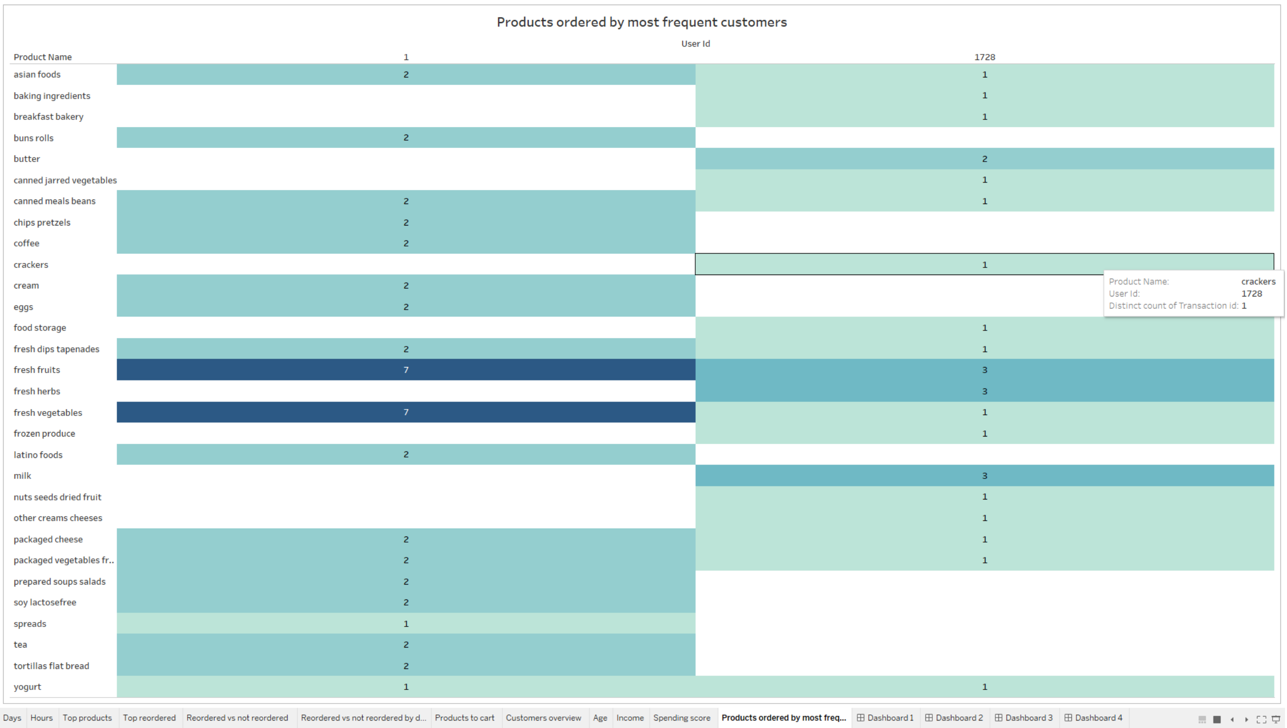


Fig. 26 Customer overview

The graph revealed that the majority of customers were female, accounting for 56% of the total customer base. The largest age group of shoppers fell between 30 and 39 years old, indicating that this age range represented a significant portion of the supermarket's customer demographic. In terms of income, the majority of customers had an income of 70k. When examining the spending score, which indicates the level of customer engagement and loyalty, the majority of customers fell into the medium range, with scores between 40 and 60.

These insights into the demographic and behavioural characteristics of the customer base could inform targeted marketing strategies and customer engagement initiatives.

Products ordered by most frequent customers



Products ordered by most frequent customers

	User Id	
Product Name	1	1728
asian foods	2	1
baking ingredients		1
breakfast bakery		1
buns rolls	2	
butter		2
canned jarred vegetables		1
canned meals beans	2	1
chips pretzels	2	
coffee	2	
crackers		1
cream	2	
eggs	2	
food storage		1
fresh dips tapenades	2	1
fresh fruits	7	3
fresh herbs		3
fresh vegetables	7	1
frozen produce		1
latino foods	2	
milk		3
nuts seeds dried fruit		1
other creams cheeses		1
packaged cheese	2	1
packaged vegetables fr..	2	1
prepared soups salads	2	
soy lactosefree	2	
spreads	1	
tea	2	
tortillas flat bread	2	
yogurt	1	1

Product Name: crackers
User Id: 1728
Distinct count of Transaction Id: 1

Fig. 27 Products ordered by most frequent customers

Upon examining the purchasing patterns of the most frequent customers, I noticed that they primarily purchased fruits, vegetables and milk. This observation aligns with the earlier identification of these products as the most popular among customers. It indicates a correlation between the preferences of the frequent customers and the overall demand for these items.

Analysing sales data is crucial for any supermarket as it provides valuable insights for growth. By examining sales data, supermarkets can make informed decisions regarding product selection, pricing strategies, and enhancing the overall shopping experience for customers. It's not just about generating profits; sales data also provides valuable insights into customer preferences and shopping behaviour. This knowledge empowers supermarkets to have better control over the sales process. In today's competitive market, understanding and utilising sales data is essential for supermarkets to thrive and succeed.

If you want to see the complete analysis, you can find the code on my GitHub page (https://github.com/mihaelakzan/Behind-the-numbers-exploring-the-sales-story-of-a-supermarket). Here you can check out the CSV files and data queries used for the supermarket sales analysis. This will give you a better understanding of the insights obtained from the dataset and the approach followed for the analysis.