# Superstore Sales Performance Analysis

## Context

The growth of supermarkets in most populated cities are increasing and market competitions are also high. The dataset is one of the historical sales of a supermarket company which has been recorded in 3 different branches for 3 months.

**Attribute information:**

**Invoice id**: Computer generated sales slip invoice identification number

**Branch:** Branch of supercenter (3 branches are available identified by A, B and C).

**City:** Location of supercenters

**Customer type**: Type of customers, recorded by Members for customers using member card and Normal for without member card.

**Gender**: Gender type of customer

**Product line**: General item categorization groups - Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel

**Unit price**: Price of each product in $

**Quantity**: Number of products purchased by customer

**Tax:** 5% tax fee for customer buying

**Total:** Total price including tax

**Date**: Date of purchase (Record available from January 2019 to March 2019)

**Time:** Purchase time (10am to 9pm)

**Payment:** Payment used by customer for purchase (3 methods are available – Cash, Credit card and E-wallet)

**COGS**: Cost of goods sold

**Gross margin percentage**: Gross margin percentage

**Gross income**: Gross income (Total - COGS)

**Rating**: Customer stratification rating on their overall shopping experience (On a scale of 1 to 10)

**Objective**

The objective is to gain a comprehensive understanding of the store's performance, identify areas for improvement, and make informed decisions to drive growth and profitability.

**Analysis Questions:**

1. How does the revenue vary across different branches?
2. Can you identify the top-selling product lines and the least popular ones?
3. What is the distribution of customer types (Members vs. Normal)?

4. What are the preferred payment methods for customers?
5. Is there a correlation between payment method and the total purchase amount?
6. How do the different branches compare in terms of customer ratings?
7. Which of the branches has more of the highest customer rating?
8. Which Gender purchases the most?

**Analysis Tool Used: MS SQL Server**

The following are the queries, results and insights to each analysis questions in order:

1. . **How does the revenue vary across different branches?**

```
-- How does the revenue vary across different branches
-- Quantity multiplied by unit price gives the revenue as COGS in the dataset

select Branch, SUM(COGS) Revenue
from skillphore..superstore
group by Branch
order by Revenue desc
```

110 %

Results | Messages

| | Branch | Revenue |
|---|---|---|
| 1 | C | 105303.53 |
| 2 | A | 101143.21 |
| 3 | B | 101140.64 |

From the above result, it can be seen that there is not much difference in how much revenue varies across the three branches.

2. **Can you identify the top-selling product lines and the least popular ones?**

```
-- Identify the top-selling product lines and the least popular ones

select ProductLine, SUM(Quantity) AS Total_Quantity
from skillphore..superstore
group by ProductLine
order by Total_Quantity DESC
```

110 %

Results | Messages

| | ProductLine | Total_Quantity |
|---|---|---|
| 1 | Electronic accessories | 971 |
| 2 | Food and beverages | 952 |
| 3 | Sports and travel | 920 |
| 4 | Home and lifestyle | 911 |
| 5 | Fashion accessories | 902 |
| 6 | Health and beauty | 854 |

The above query result shows that Electronic accessories are the top selling product lines while Health and beauty products are the least selling products.

### 3. What is the distribution of customer types (Members vs. Normal)?

```
-- What is the distribution of customer types (Members vs Normal)

select CustomerType, COUNT(*) CS_Distribution, sum(Quantity)
Quantity_Sum, round(avg(GrossIncome),2) Avg_GrossIncome
from skillphore..superstore
group by CustomerType
order by CS_Distribution DESC
```

110 %

Results | Messages

| | CustomerType | CS_Distribution | Quantity_Sum | Avg_GrossIncome |
|---|---|---|---|---|
| 1 | Member | 501 | 2785 | 15.61 |
| 2 | Normal | 499 | 2725 | 15.15 |

The distribution varies as there are 501 Members and 499 Normal (members without card) which in total makes up the 1000 rows of the dataset.

### 4. What are the preferred payment methods for customers

```
-- What are the preferred payment methods for customers

Select Payment, COUNT(*) PreferredPayment
from skillphore..superstore
group by Payment
order by PreferredPayment DESC
```

110 %

Results | Messages

| | Payment | PreferredPayment |
|---|---|---|
| 1 | E-Wallet | 345 |
| 2 | Cash | 344 |
| 3 | Credit card | 311 |

E-Wallet is the most preferred payment method for customers followed closely by Cash, with Credit card being the least preferred.

5. **Is there a correlation between payment method and total purchase amount?**

```
-- Is there a correlation between payment method and total purchase amount

select Payment, COUNT(Payment) Payment_Count, SUM(Total) Total_Purchase
from skillphore..superstore
group by Payment
order by Payment_Count
```

110 %

Results | Messages

| | Payment | Payment_Count | Total_Purchase |
|---|---|---|---|
| 1 | Credit card | 311 | 100767.072 |
| 2 | Cash | 344 | 112206.57 |
| 3 | E-Wallet | 345 | 109993.107 |

```
-- Is there a correlation between payment method and total purchase amount
WITH PaymentMethodTotal AS (
    select Payment, SUM(Total) AS total_purchase, AVG(Total) AS avg_total
    from skillphore..superstore
    group by Payment
)

SELECT
    (select SUM((total_purchase - avg_total) * (total_purchase - avg_total))
    from PaymentMethodTotal) /
    ((COUNT(Payment) - 1) * SUM(total_purchase * total_purchase)) AS correlation
FROM PaymentMethodTotal;
```

110 %

Results | Messages

| | correlation |
|---|---|
| 1 | 0.497010549322394 |

A (0.4) correlation suggests that there is a moderate positive correlation between payment method and total purchase amount. Although, the relationship between both variables is not strong to make precise predictions.

6. **How do the different branches compare in terms of customer ratings?**

```sql
-- How do the different branches compare in terms of customer ratings

select Branch, ROUND(AVG(Rating),2) Avg_Rating
from skillphore..superstore
group by Branch, City
order by Avg_Rating desc
```

110 %

Results | Messages

| | Branch | Avg_Rating |
|---|---|---|
| 1 | C | 7.07 |
| 2 | A | 7.03 |
| 3 | B | 6.82 |

From the query result above, we can see that there is not much difference in customer ratings between branches C and A as they both have an average rating of 7.07 and 7.03 respectively with branch B having the least average customer rating of 6.82.

7. **Which of the branches have more of the highest customer ratings?**

```sql
-- Which of the branches has more of the highest customer rating

select Branch, MAX(Rating) Max_Rating, COUNT(Rating) Highest_Rating_Count
from skillphore..superstore
where Rating = 10
group by Branch
order by Highest_Rating_Count DESC
```

110 %

Results | Messages

| | Branch | Max_Rating | Highest_Rating_Count |
|---|---|---|---|
| 1 | B | 10 | 3 |
| 2 | C | 10 | 1 |
| 3 | A | 10 | 1 |

With the highest customer rating being 10, it can be observed that branch B has more of the highest customer rating as it was given the highest rating of 10 by 3 customers.

8. **Which gender purchases the most?**

```sql
-- Which gender purchases the most

select Gender, sum(Quantity) Most_Purchase, sum(GrossIncome) Total_GrossIncome
from skillphore..superstore
group by Gender
order by Most_Purchase DESC
```

110 %

Results | Messages

| | Gender | Most_Purchase | Total_GrossIncome |
|---|---|---|---|
| 1 | Female | 2869 | 7994.425 |
| 2 | Male | 2641 | 7384.944 |

The female gender purchases the most as they have made a total purchase of 2869 while generating more Gross income, as compared to the male counterparts who have made a lesser purchase of 2641 and generating less gross income.

**Recommendation(s)**

- Members (with card) appear to have more purchase and generate more income than Normal (without card). The superstore should implement a loyalty program that aims at rewarding customers for repeat purchases. Membership cards could be tied to accumulating points or receiving discounts based on how frequent purchases are and the amount.
- Health and Beauty products are the least popular product items as they are the ones less purchased by customers. The superstore needs to tailor ads that are directed at the benefits of purchasing Health products which will ensure the general well-being of its customers.

**Full Query Codes**

```sql
-- Looking at the dataset

select *
from skillphore..superstore



-- How does the revenue vary across different branches
-- Quantity multiplied by unit price gives the revenue as COGS in the dataset

select Branch, SUM(COGS) Revenue
from skillphore..superstore
group by Branch
order by Revenue desc


-- Identify the top-selling product lines and the least popular ones

select ProductLine, SUM(Quantity) AS Total_Quantity
from skillphore..superstore
group by ProductLine
order by Total_Quantity DESC


-- What is the distribution of customer types (Members vs Normal)

select CustomerType, COUNT(*) CS_Distribution, sum(Quantity)
Quantity_Sum, round(avg(GrossIncome),2) Avg_GrossIncome
from skillphore..superstore
group by CustomerType
order by CS_Distribution DESC


-- What are the preferred payment methods for customers

Select Payment, COUNT(*) PreferredPayment
from skillphore..superstore
group by Payment
order by PreferredPayment DESC


-- Is there a correlation between payment method and total purchase amount

select Payment, COUNT(Payment) Payment_Count, SUM(Total) Total_Purchase
from skillphore..superstore
group by Payment
order by Payment_Count


WITH PaymentMethodTotal AS (
    select Payment, SUM(Total) AS total_purchase, AVG(Total) AS avg_total
    from skillphore..superstore
```

```
    group by Payment
)

SELECT
    (select SUM((total_purchase - avg_total) * (total_purchase - avg_total))
    from PaymentMethodTotal) /
    ((COUNT(Payment) - 1) * SUM(total_purchase * total_purchase)) AS correlation
FROM PaymentMethodTotal;


-- How do the different branches compare in terms of customer ratings

select Branch, ROUND(AVG(Rating),2) Avg_Rating
from skillphore..superstore
group by Branch, City
order by Avg_Rating desc


-- Which of the branches has more of the highest customer rating

select Branch, MAX(Rating) Max_Rating, COUNT(Rating) Highest_Rating_Count
from skillphore..superstore
where Rating = 10
group by Branch
order by Highest_Rating_Count DESC


-- Which gender purchases the most

select Gender, sum(Quantity) Most_Purchase, sum(GrossIncome) Total_GrossIncome
from skillphore..superstore
group by Gender
order by Most_Purchase DESC
```