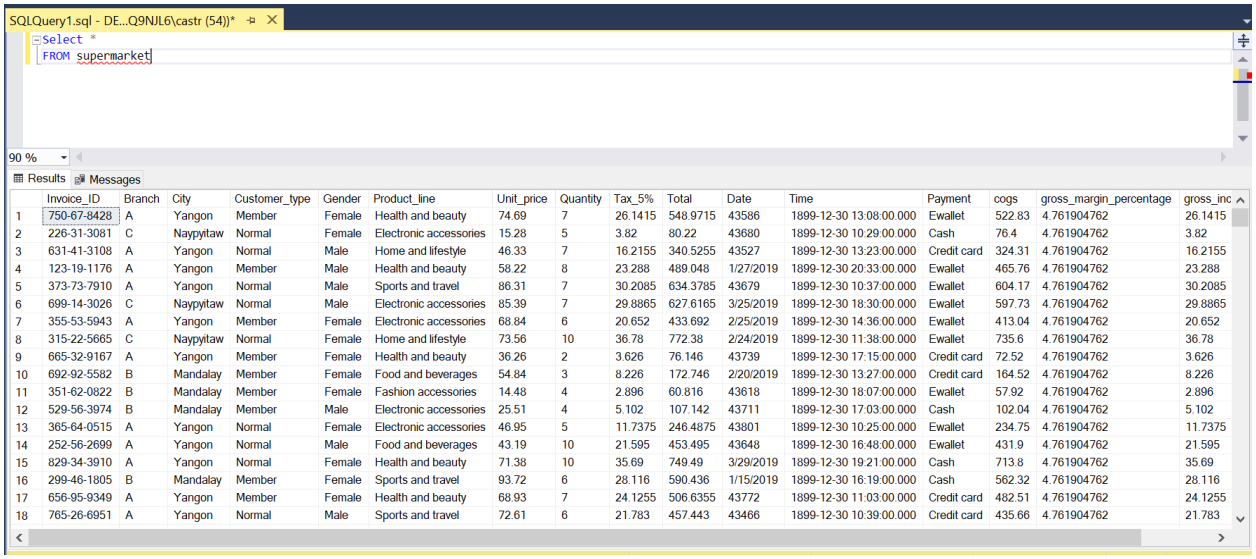
SQL Project- Supermarket Analysis

The following dataset about a three-month sale period of a supermarket has been collected from Kaggle https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales?resource=download. Analysis has been done using SQL, then insights and recommendations have been provided to improve aspects of the supermarket.

Here is a sample overview of the columns and data



	Invoice_ID	Branch	City	Customer_type	Gender	Product_line	Unit_price	Quantity	Tax_5%	Total	Date	Time	Payment	cogs	gross_margin_percentage	gross_inc
1	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	43586	1899-12-30 13:08:00.000	Ewallet	522.83	4.761904762	26.1415
2	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	43680	1899-12-30 10:29:00.000	Cash	76.4	4.761904762	3.82
3	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	43527	1899-12-30 13:23:00.000	Credit card	324.31	4.761904762	16.2155
4	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	1/27/2019	1899-12-30 20:33:00.000	Ewallet	465.76	4.761904762	23.288
5	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	43679	1899-12-30 10:37:00.000	Ewallet	604.17	4.761904762	30.2085
6	699-14-3026	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	3/25/2019	1899-12-30 18:30:00.000	Ewallet	597.73	4.761904762	29.8865
7	355-53-5943	A	Yangon	Member	Female	Electronic accessories	68.84	6	20.652	433.692	2/25/2019	1899-12-30 14:36:00.000	Ewallet	413.04	4.761904762	20.652
8	315-22-5665	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.78	772.38	2/24/2019	1899-12-30 11:38:00.000	Ewallet	735.6	4.761904762	36.78
9	665-32-9167	A	Yangon	Member	Female	Health and beauty	36.26	2	3.626	76.146	43739	1899-12-30 17:15:00.000	Credit card	72.52	4.761904762	3.626
10	692-92-5582	B	Mandalay	Member	Female	Food and beverages	54.84	3	8.226	172.746	2/20/2019	1899-12-30 13:27:00.000	Credit card	164.52	4.761904762	8.226
11	351-62-0822	B	Mandalay	Member	Female	Fashion accessories	14.48	4	2.896	60.816	43618	1899-12-30 18:07:00.000	Ewallet	57.92	4.761904762	2.896
12	529-56-3974	B	Mandalay	Member	Male	Electronic accessories	25.51	4	5.102	107.142	43711	1899-12-30 17:03:00.000	Cash	102.04	4.761904762	5.102
13	365-64-0515	A	Yangon	Normal	Female	Electronic accessories	46.95	5	11.7375	246.4875	43801	1899-12-30 10:25:00.000	Ewallet	234.75	4.761904762	11.7375
14	252-56-2699	A	Yangon	Normal	Male	Food and beverages	43.19	10	21.595	453.495	43648	1899-12-30 16:48:00.000	Ewallet	431.9	4.761904762	21.595
15	829-34-3910	A	Yangon	Normal	Female	Health and beauty	71.38	10	35.69	749.49	3/29/2019	1899-12-30 19:21:00.000	Cash	713.8	4.761904762	35.69
16	299-46-1805	B	Mandalay	Member	Female	Sports and travel	93.72	6	28.116	590.436	1/15/2019	1899-12-30 16:19:00.000	Cash	562.32	4.761904762	28.116
17	656-95-9349	A	Yangon	Member	Female	Health and beauty	68.93	7	24.1255	506.6355	43772	1899-12-30 11:03:00.000	Credit card	482.51	4.761904762	24.1255
18	765-26-6951	A	Yangon	Normal	Male	Sports and travel	72.61	6	21.783	457.443	43466	1899-12-30 10:39:00.000	Credit card	435.66	4.761904762	21.783

Objectives:

1. Which branch has the best results in the loyalty program?
2. Does the membership depend on customer rating?
3. Does gross income depend on the proportion of customers in the loyalty program? On payment method?
4. Are there any differences in indicators between men and women?
5. Which product category generates the highest income?

Column Labels

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 Ewallet)
COGS: Cost of goods sold
Gross margin percentage: Gross margin percentage
Gross income: Gross income
Rating: Customer stratification rating on their overall shopping experience (On a scale of 1 to 10)

1. Which branch has the best results in the loyalty program?

Fields required: Branch, Customer type

```
SELECT Branch, Customer_type, COUNT(Customer_type) AS Customer_count
FROM supermarket
GROUP BY Branch, Customer_type
ORDER BY 2,3 DESC;
```

90 %

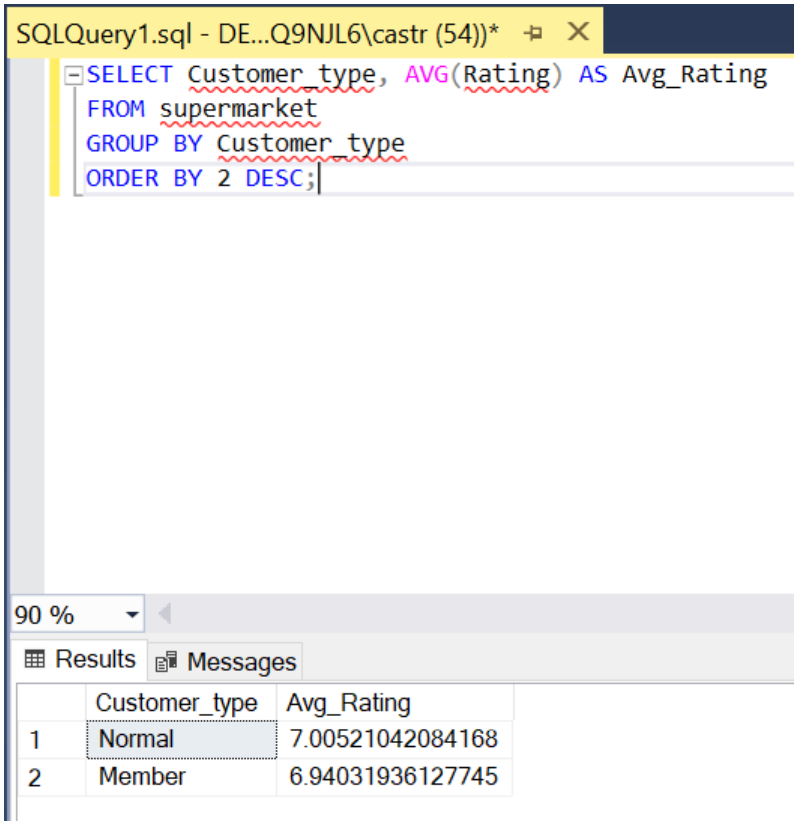
Results | Messages

	Branch	Customer_type	Customer_count
1	C	Member	169
2	A	Member	167
3	B	Member	165
4	A	Normal	173
5	B	Normal	167
6	C	Normal	159

Branch C has the most number of customers registered as a Member with 169 counts, followed by Branch A with 167 and lastly by Branch B with 165 counts. The is no significant difference among the branches in this regard. What we can see on the other hand is that there is the greatest potential from Branch A wherein we can convert Normal customers to be Members. This is followed by Branch B, and lastly by Branch C.

2. Does the membership depend on customer rating?

   Fields required: Customer_type, rating



   Membership does not seem to correlate to customer rating. The average ratings of both customer types are nearly identical.

3. Does gross income depend on the proportion of customers in the loyalty program? On payment method?

   Required fields: Gross_income, Customer_type, Payment

```
SELECT Customer_type, AVG(Gross_income) AS Avg_gross_income
FROM supermarket
GROUP BY Customer_type
ORDER BY 2 DESC;
```

90 %

Results | Messages

	Customer_type	Avg_gross_income
1	Member	15.6091097804391
2	Normal	15.1487074148297

Gross income does not seem to affect loyalty significantly. Although Members have around 0.5 units higher than Normal customers for their gross income.

```
SELECT Payment, AVG(Gross_income) AS Avg_gross_income
FROM supermarket
GROUP BY Payment
ORDER BY 2 DESC;
```

90 %

Results | Messages

	Payment	Avg_gross_income
1	Cash	15.5324709302326
2	Credit card	15.4290418006431
3	Ewallet	15.1819333333333

Payment methods vary by the average gross income of the customers, but not significantly. The highest earners pay through Cash, followed by Credit Card, and the lowest earners pay via Ewallet. This may reflect the spending preferences of each customer depending on their gross income.

4. Are there any differences in indicators between men and women?

   Fields required: *, grouped by Gender

   ### Gender vs Customer_type

```
SELECT Gender, Customer_type, COUNT(Customer_type) AS Customer_count
FROM supermarket
GROUP BY Gender, Customer_type
ORDER BY 3 DESC;
```

90 %

Results | Messages

	Gender	Customer_type	Customer_count
1	Female	Member	261
2	Male	Normal	259
3	Male	Member	240
4	Female	Normal	240

   Female members outnumber male members 261:240 across branches. Male normal customers outnumber female normal customers 259:240 across branches. With a simple calculation, Female members account for a total of 501 out of the 1000 customers, while Males comprise 499 out of the 1000. Males and females regardless of membership shop equally in terms of population. It is just that females are more likely to be members than males. A suggestion is to make promotions and/or pitches to encourage more males to sign up for membership.

Gender vs Product_line

```
SELECT Gender, Product_line, COUNT(Product_line) AS Product_count
FROM supermarket
GROUP BY Gender, Product_line
ORDER BY COUNT(Product_line) DESC;
```

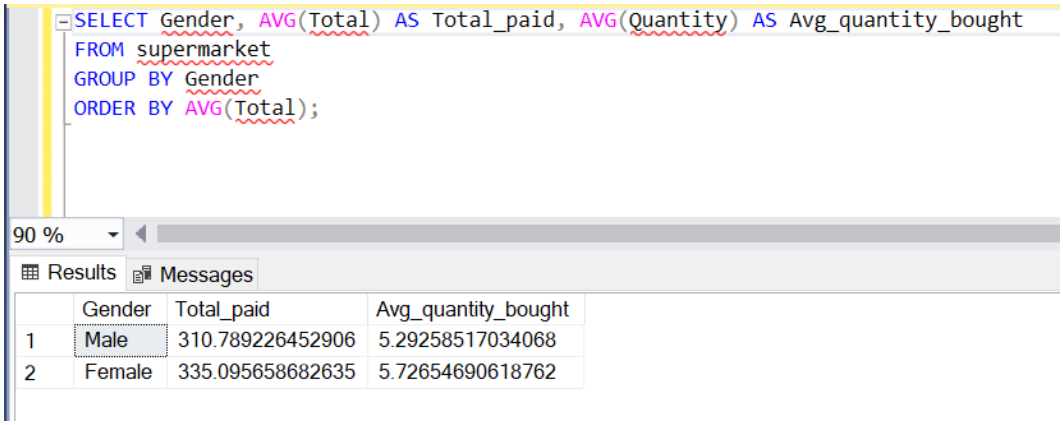
90 %

Results | Messages

	Gender	Product_line	Product_count
1	Female	Fashion accessories	96
2	Female	Food and beverages	90
3	Male	Health and beauty	88
4	Female	Sports and travel	88
5	Male	Electronic accessories	86
6	Female	Electronic accessories	84
7	Male	Food and beverages	84
8	Male	Fashion accessories	82
9	Male	Home and lifestyle	81
10	Female	Home and lifestyle	79
11	Male	Sports and travel	78
12	Female	Health and beauty	64

Females most preferred product line is Fashion accessories, least preferred is health and beauty. While Males' most preferred product line is Health and beauty, least preferred is Sports and travel. Males' preference for fashion accessories is at their bottm 3.

Gender vs Avg total paid



```
SELECT Gender, AVG(Total) AS Total_paid, AVG(Quantity) AS Avg_quantity_bought
FROM supermarket
GROUP BY Gender
ORDER BY AVG(Total);
```

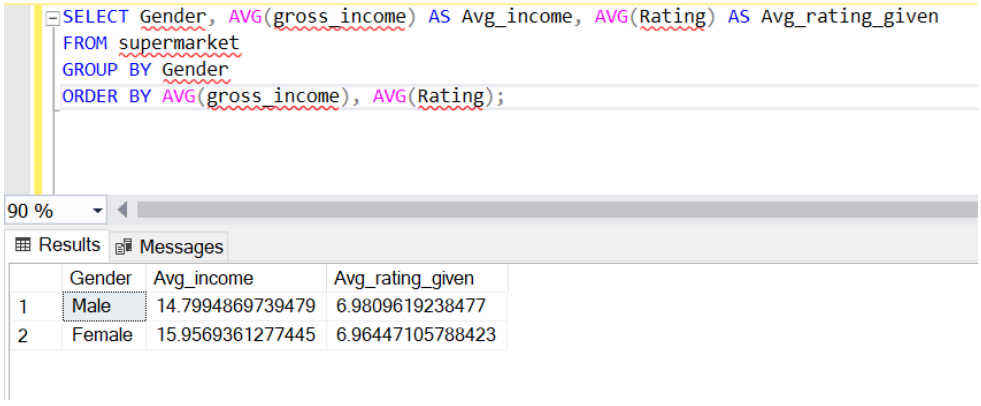
90 %

Results | Messages

	Gender	Total_paid	Avg_quantity_bought
1	Male	310.789226452906	5.29258517034068
2	Female	335.095658682635	5.72654690618762

Females on average pay $25 more than Males do. This can be associated with their preferred items to purchase (Fashion accessories) to bring up the average price. This can also be related to the average quantity of items purchased by females which is slightly more than males'.

Gender vs Avg income & Avg. Rating given



```
SELECT Gender, AVG(gross_income) AS Avg_income, AVG(Rating) AS Avg_rating_given
FROM supermarket
GROUP BY Gender
ORDER BY AVG(gross_income), AVG(Rating);
```

90 %

Results | Messages

	Gender	Avg_income	Avg_rating_given
1	Male	14.7994869739479	6.9809619238477
2	Female	15.9569361277445	6.96447105788423

Females earn 1.2 units more than males. The ratings they give are virtually similar on average.

Gender vs Mode of Payment & Count

SQLQuery1.sql - DE...Q9NJL6\castr (54))*

```
SELECT Gender, Payment, COUNT(Payment) AS Payment_count
FROM supermarket
GROUP BY Gender, Payment
ORDER BY COUNT(Payment) DESC;
```

90 %

Results | Messages

	Gender	Payment	Payment_count
1	Male	Ewallet	185
2	Female	Cash	178
3	Male	Cash	166
4	Female	Credit card	163
5	Female	Ewallet	160
6	Male	Credit card	148

Most Males prefer using Ewallet, while most Females prefer paying in Cash. This can be related to the previous finding that higher earners (majority females) prefer Cash payments while lower earners (Males) prefer Ewallets.

5. Which product category generates the highest income?

   Fields required: Product_line, Total

```
SELECT Product_line, SUM(Total) AS Revenue
FROM supermarket
GROUP BY Product_line
ORDER BY SUM(Total) DESC;;
```

90 %

Results | Messages

	Product_line	Revenue
1	Food and beverages	56144.844
2	Sports and travel	55122.8265
3	Electronic accessories	54337.5315
4	Fashion accessories	54305.895
5	Home and lifestyle	53861.913
6	Health and beauty	49193.739

Food and beverages account for the highest revenue with $56,144, followed by Sports and travel with $55122. The least income generating product line was Health and beauty, only gathering $49193. This trend can be seen due to the reason that the highest income generators are basic needs while those at the bottom can rather be categorized as auxiliary essentials only.