



CLEAN



DATA

SUPERMARKET

market
Sales Analysis

TEAM 4:

- ☐ Ivan Anguiano
- ☐ Ligia Fascilla
- ☐ Douglas Frankish
- ☐ Aranza Orozco



Overview

Dataset consists of historical data from a supermarket that recorded sales from three different branches for three months.

Dataset: <https://www.kaggle.com/datasets/aungpyaeap/supermarket-sales>

Data Structure - Columns

- 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)

Dataset Snapshot

Invoice ID	Branch	CustomerID	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	cogs	gross margin percentage	gross income	Rating	Longitude	Latitude
750-67-8428	A	C1888	Yangon	Member	Female	Health and beauty	74.69	10	37.345	746.9	21-02-2019	13:08	Ewallet	711.3333333	4.761904762	35.56666667	9.1	96.1735	16.8409
226-31-3081	C	C1475	Naypyitaw	Normal	Female	Health and beauty	15.28	6	4.584	91.68	27-05-2019	10:29	Cash	76.4	4.761904762	15.28	10	96.0785	19.7633
631-41-3108	A	C1746	Yangon	Normal	Male	Health and beauty	46.33	7	16.2155	324.31	27-12-2019	13:23	Credit card	324.31	4.761904762	0	7.4	96.1735	16.8409
123-19-1176	A	C1896	Yangon	Member	Male	Health and beauty	58.22	11	32.021	640.42	15-11-2019	20:33	Ewallet	465.76	4.761904762	174.66	8.4	96.1735	16.8409
373-73-7910	A	C1790	Yangon	Normal	Male	Health and beauty	86.31	7	30.2085	604.17	31-03-2019	10:37	Ewallet	604.17	4.761904762	0		96.1735	16.8409
699-14-3026	C	C1423	Naypyitaw	Normal	Male	Health and beauty	85.39	8	34.156	683.12	06-01-2019	18:30	Ewallet	597.73	4.761904762	85.39		96.0785	19.7633
355-53-5943	A	C1723	Yangon	Normal	Female	Health and beauty	68.84	6	20.652	413.04	15-12-2019	14:36	Ewallet	413.04	4.761904762	0		96.1735	16.8409
315-22-5665	C	C1471	Naypyitaw	Normal	Female	Health and beauty	73.56	12	44.136	882.72	14-06-2019	11:38	Ewallet	735.6	4.761904762	147.12		96.0785	19.7633
665-32-9167	A	C1707	Yangon		Female	Health and beauty	36.26	2	3.626	72.52	25-11-2019	17:15	Credit card	72.52	4.761904762	0		96.1735	16.8409
692-92-5582	B	C1277	Mandalay	Normal	Female	Health and beauty	54.84	4	10.968	219.36	10-04-2019	13:27	Credit card	164.52	4.761904762	54.84		96.0891	21.9588
351-62-0822	B	C1300	Mandalay		Female	Health and beauty	14.48	6	4.344	86.88	23-07-2019	18:07	Ewallet	57.92	4.761904762	28.96		96.0891	21.9588
529-56-3974	B	C1264	Mandalay	Member	Male	Health and beauty	25.51	6		153.06	26-08-2019	17:03	Cash	102.04	4.761904762	51.02		96.0891	21.9588
365-64-0515	A	C1710	Yangon	Normal	Female	Electronic accessories	46.95	5		234.75	25-08-2019	10:25	Ewallet	234.75	4.761904762	0		96.1735	16.8409
252-56-2699	A	C1726	Yangon	Normal	Male	Food and beverages	43.19	10		431.9	02-10-2019	16:48	Ewallet	431.9	4.761904762	0	8.2	96.1735	16.8409
829-34-3910	A	C1702	Yangon	Normal	Female	Health and beauty	71.38	10		713.8	23-11-2019	19:21	Cash	713.8	4.761904762	0	5.7	96.1735	16.8409
299-46-1805	B	C1280	Mandalay	Member	Female	Sports and travel	93.72	11		1030.92	15-07-2019	16:19	Cash	562.32	4.761904762	468.6	5.4	96.0891	21.9588
656-95-9349	A	C1799	Yangon	Member	Female	Health and beauty	68.93	10		689.3	02-12-2019	11:03	Credit card	482.51	4.761904762	206.79	4.6	96.1735	16.8409
765-26-6951	A	C1669	Yangon	Normal	Male	Sports and travel	72.61	6		435.66	09-03-2019	10:39	Credit card	435.66	4.761904762	0	6.9	96.1735	16.8409
329-62-1586	A	C1645	Yangon	Normal	Male	Food and beverages	54.67	3		164.01	09-07-2019	18:00	Credit card	164.01	4.761904762	0	8.6	96.1735	16.8409
319-50-3348	B	C1207	Mandalay	Normal	Female	Home and lifestyle	40.3	3		120.9	08-05-2019	15:30	Ewallet	80.6	4.761904762	40.3	5.28	96.0891	21.9588
300-71-4605	C	C1462	Naypyitaw	Member	Male	Electronic accessories	86.04	8		688.32	18-10-2019	11:24	Ewallet	430.2	4.761904762	258.12	5.04	96.0785	19.7633
371-85-5789	B	C1217	Mandalay	Normal	Male	Health and beauty	87.98	4		351.92	25-12-2019	10:40	Ewallet	263.94	4.761904762	87.98	6.12	96.0891	21.9588



Questions

1

What is the difference in spending patterns between male and female customers? Do they tend to buy different products? What is the relationship between gender and payment type?

2

What is the average unit price and quantity of products purchased? Is there a relationship between the unit price and the quantity purchased?

3

What is the percentage of sales from members compared to non-members? Is there a difference in spending patterns between members and non-members? What is the relationship between customer type and payment type?

4

Which branch has the highest sales volume and revenue?

5

What is the correlation between customer rating and total purchase amount? Is there a relationship between purchase time of day and customer rating?

6

Which payment method is the most popular among customers? Is there a difference in spending patterns based on the payment method used?



Question

Q1

What is the difference in spending patterns between male and female customers? Do they tend to buy different products? What is the relationship between gender and payment type?



Question

01

Part 1: Total Spend by Gender and Average Spend by Gender

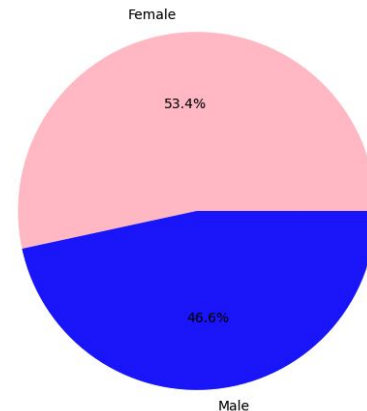
Process:

1. Data clean-up: dropped rows with no Invoice ID, Gender or Total
2. Verify data validity/representativeness: 493 females and 482 males left
3. Find total spend using `.groupby('Gender')['Total'].sum()`
4. Find average spend using same formula but with `.mean()`

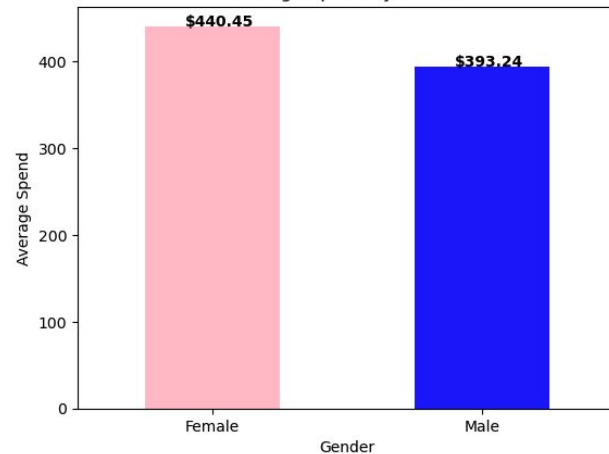
Analysis/Conclusion of Part 1:

Overall, females spent more money than males. Not only in total spend for the three months analyzed, but also on average spend per female/male. Females spent 53.4% out of the total spend. There were also more females than males, but the difference was only by 11 individuals. (Females total spend: \$217,141.20. Males total spend: \$189,539.62)

Percentage of Total Spend per each Gender

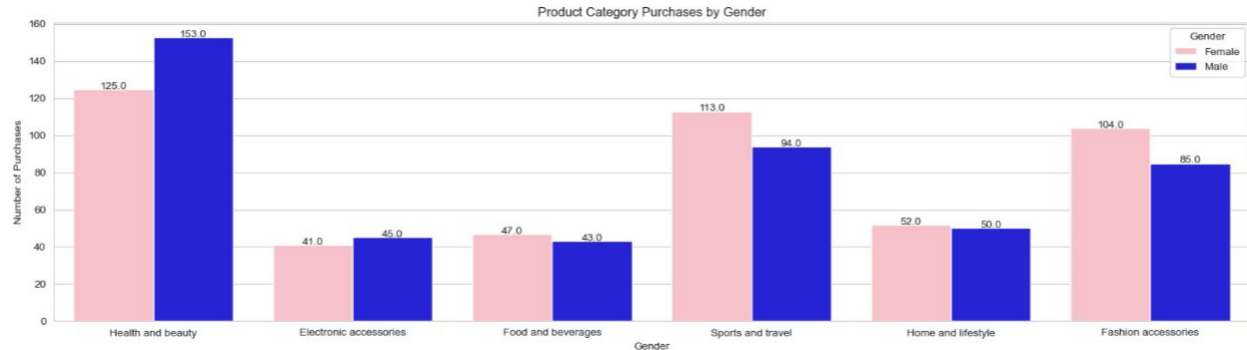


Average Spend by Gender





Question 01



Part 2: Product Line Purchases by Gender

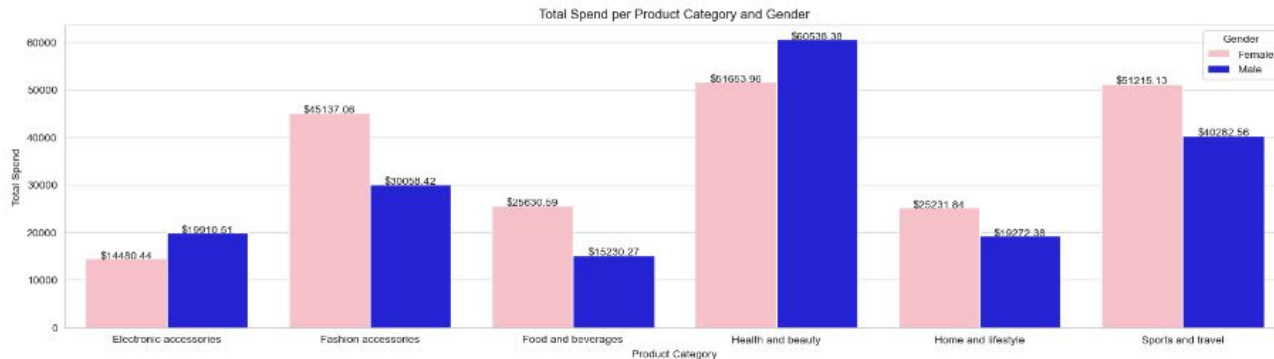
Process:

1. Data clean-up: dropped rows with no Invoice ID, Gender or Product Line
2. Verify data validity/representativeness: 482 females and 470 males left
3. Find product line purchases by gender using `.groupby('Gender')['CustomerID'].count()`

Analysis/Conclusion of Part 2:

More males than females shopped in the 'Health and Beauty' category, and in the 'Electronic Accessories' category. The 'Health and Beauty' category is where most men shopped at. Females surpassed males in shopping the 'Fashion Accessories' and 'Sports and travel' categories by the same amount (19), and did not surpass them in the 'Home and Lifestyle' and 'Food and Beverages' categories by much.

Question 01



Part 3: Total and Average Spend per Product Line by Gender

Process:

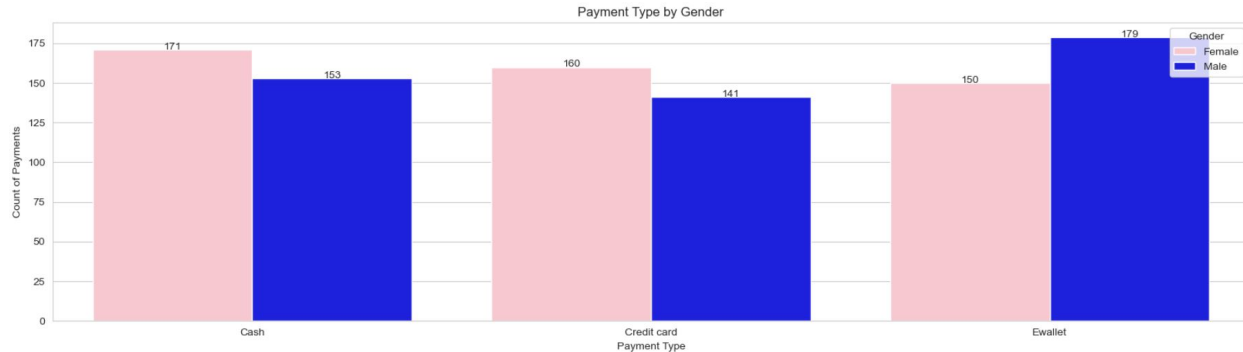
1. Data clean-up: dropped rows with no Total, Gender or Product Line
2. Find total spend per product line by gender using `.groupby(['Product line', 'Gender'])['Total'].sum().reset_index()`
3. Find average spend per product line by gender using the same code but with the mean of total spend per product line per gender

Analysis/Conclusion of Part 3:

Both genders spent the most in the line 'Health and Beauty' in total. Females spent the most on average in the 'Food and Beverages' line (\$545.33) while men spent the most on average in the 'Electronic Accessories' line (\$442.46).



Question 01



Part 4: Relationship Between Payment Type and Gender

Process:

1. Data clean-up: dropped rows with no Invoice ID, Gender or Payment
2. Verify data validity: 481 females and 473 males left
3. Find total purchases per payment type by gender using `.groupby(['Gender', 'Payment'])['Invoice ID'].nunique().reset_index()`

Analysis/Conclusion of Part 3:

Males paid the most using Ewallets and the least using Credit Cards, while females paid the most using Cash and the least using Ewallets.



Question

02

What is the average unit price and quantity of products purchased? Is there a relationship between the unit price and the quantity purchased?

Question

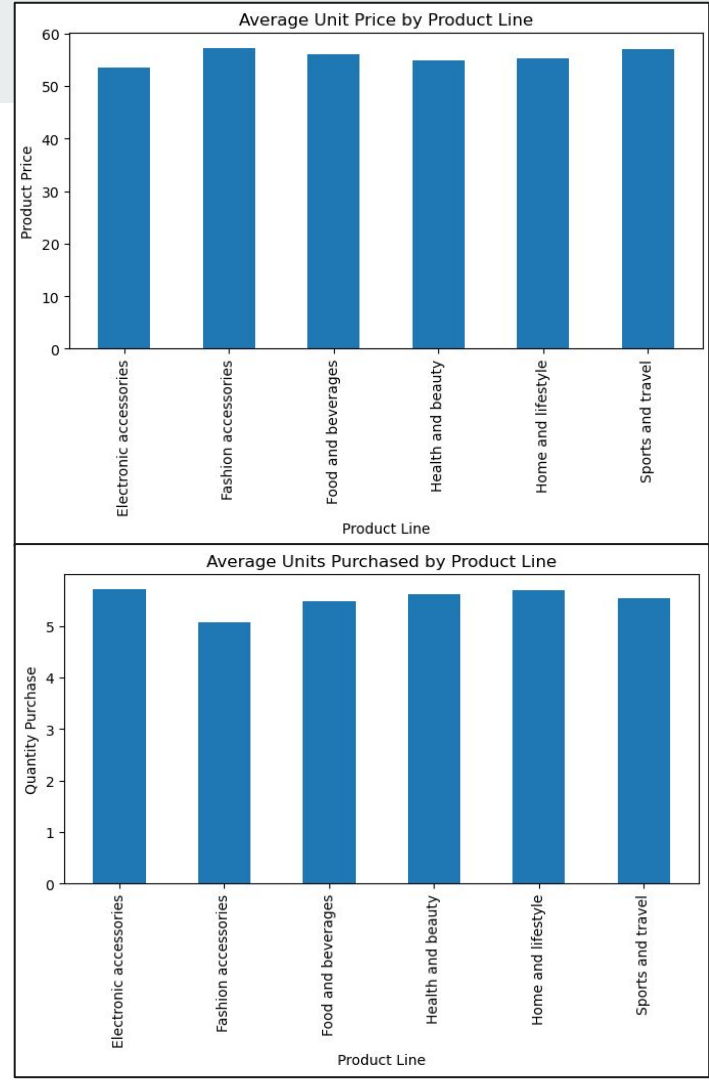
02

Process:

1. Narrowed down both sets of results using groupby ('Product line') [Unit Price/Units]. mean() for the averages.
 - **Average unit price:** \$55.67 (overall)
 - **Average quantity purchased:** 5.51

Analysis

While there appears to be some dips in the quantity prices and among the average unit prices, they are actually quite consistent, with a range of \$3.60 for product price and 0.65 for quantity purchased.



Question

02

Process:

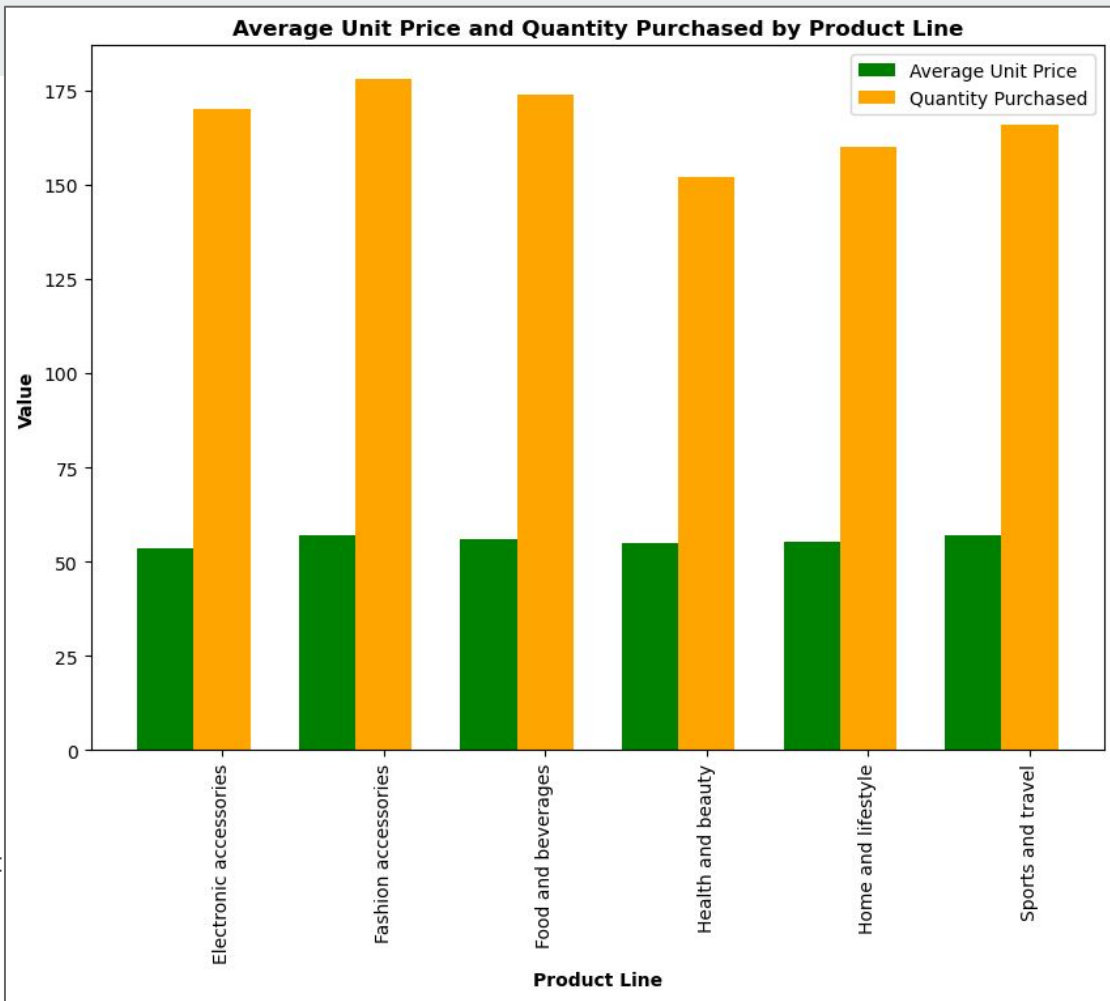
- Again using groupby('Product line')[Unit Price/Units].mean()- Compared the average unit price with the actual quantity purchased (vs the average)

Quantity Purchased Average Unit Price

•	Electronic accessories	170	55.55
•	Fashion accessories	178	57.15
•	Food and beverages	174	56
•	Health and beauty	152	54.85
•	Home and lifestyle	160	55.31
•	Sports and travel	166	56.99

Analysis:

There is a pretty consistent line across the average unit price and the quantity purchased. With a quantity range of 26 between Fashion and Health and Beauty, people appear to be buying what they want and how much of it they want.





Question

03

What is the percentage of sales from members compared to non-members? Is there a difference in spending patterns between members and non-members? What is the relationship between customer type and payment type?

Question

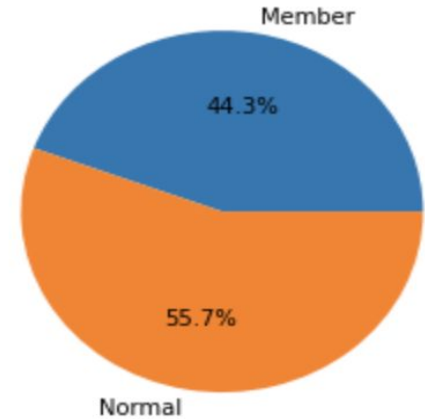
03

Part 3: Members and Non-members

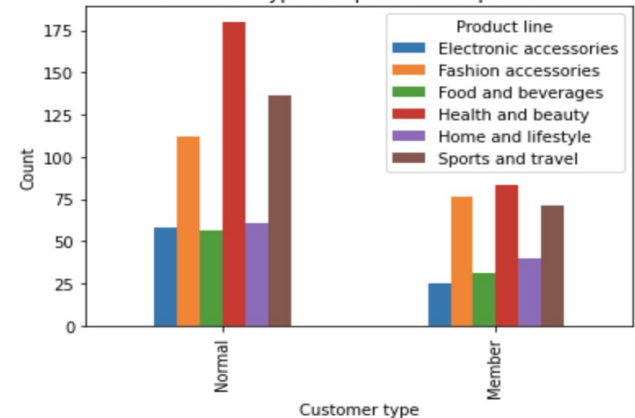
The number of sales for each customer type was added, each of the sales in every product line for each customer, and each payment method used by each customer group.

Analysis: Almost half of the sales are made by members. The spending patterns are very similar between members and non-members. Members prefer to pay with credit card while the non-members prefer to pay with Ewallet

Sales by members and non-members



Customer type and product line patterns





Question

04

Which branch has the highest sales volume and revenue?

Question

04

Process:

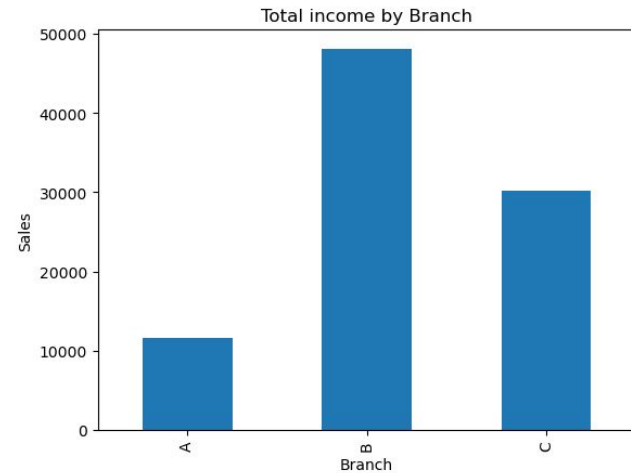
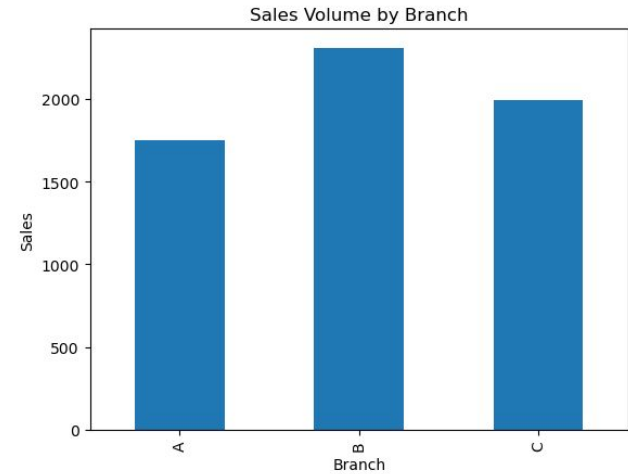
- Grouped the data in a dataframe by the values in the Branch column
- Selected the gross income and sales columns and calculated the sum of its values for each group

Which branch has the highest sales revenue?

Branch B with \$48,094

Which branch has the highest sales volume?

Branch B with 2307 sales





Question

05

What is the correlation between customer rating and total purchase amount? Is there a relationship between purchase time of day and customer rating?

Question

05

Process:

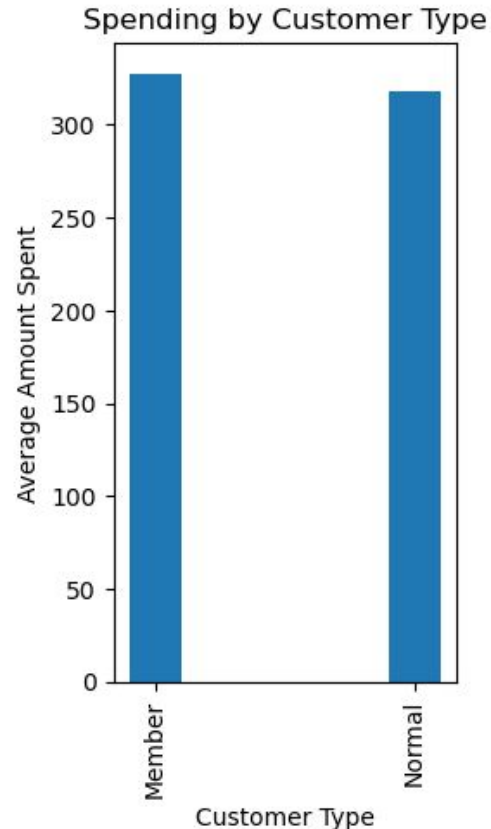
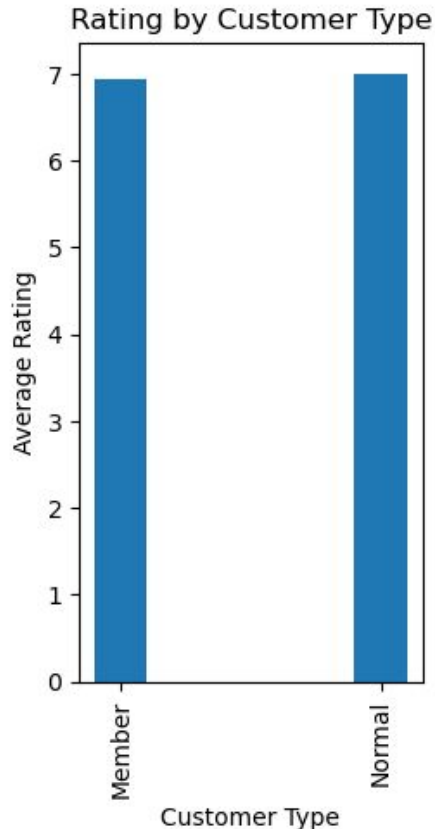
1. To make the results a little more interesting visually- rather than just showing the average rating vs amount spent, displayed is the rating/spent by customer type.
2. Information gathered using `groupby('Customer Type')['Rating/Total'].mean`

What is the correlation between customer rating and total purchase amount?

- **Member Rating:** 6.940319
- **Normal Rating:** 7.005210
- **Member Spending:** \$327.79
- **Normal Spending:** \$318.12

Analysis:

Customers like this supermarket. The ratings are nearly the same whether a member or normal customer, and with less than a \$10 difference, spending averages are also on par.



Question

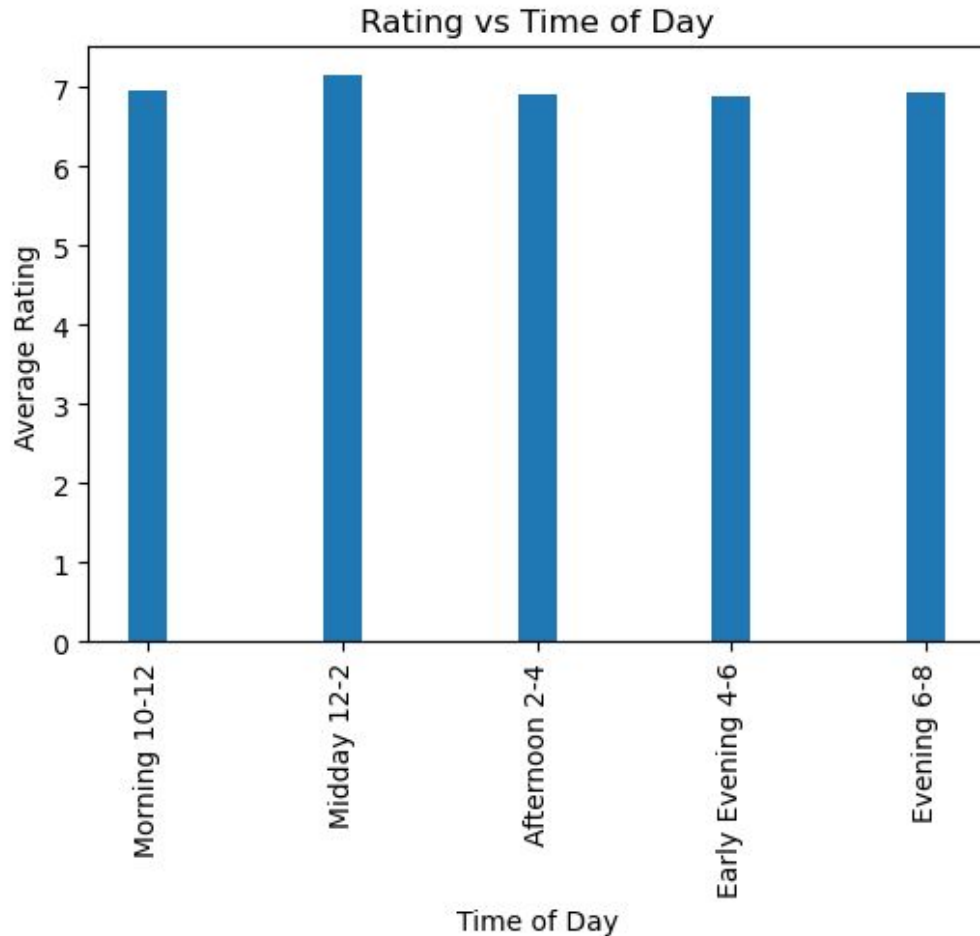
05

Process:

1. Bins and `pd.cut` were used to classify the ratings into groups based on the time of day
 - **Morning:** 6.96
 - **Midday:** 7.15
 - **Afternoon:** 6.9
 - **Early Evening:** 6.89
 - **Evening:** 6.94

Analysis:

There is not a significant difference in when customers shop. Once again, the ratings remain extremely consistent throughout the day.





Question

06

Which payment method is the most popular among customers? Is there a difference in spending patterns based on the payment method used?



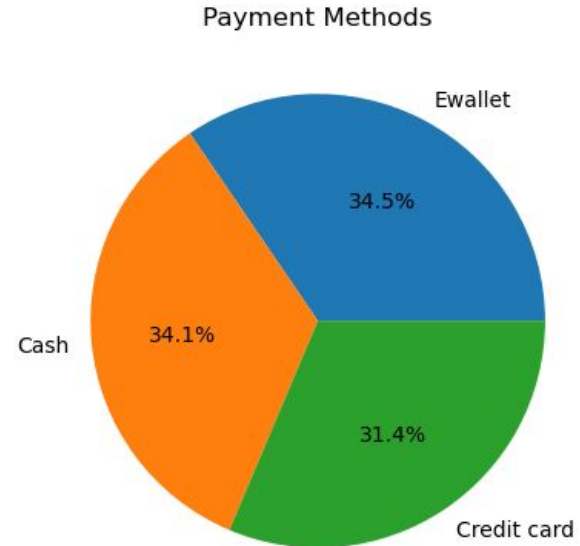
Question 06

Process:

- Used the `str.contains.sum()` function to count the number of occurrences of each payment method in the Payment column.
- For the pie chart, we used the `.value_counts()` function to count the number of occurrences of each unique word. We then used `plt.pie()`

Which payment method is most popular amongst customers?

EWallet is the most popular with 338 sales. But Cash is really close with 334 sales.



Question

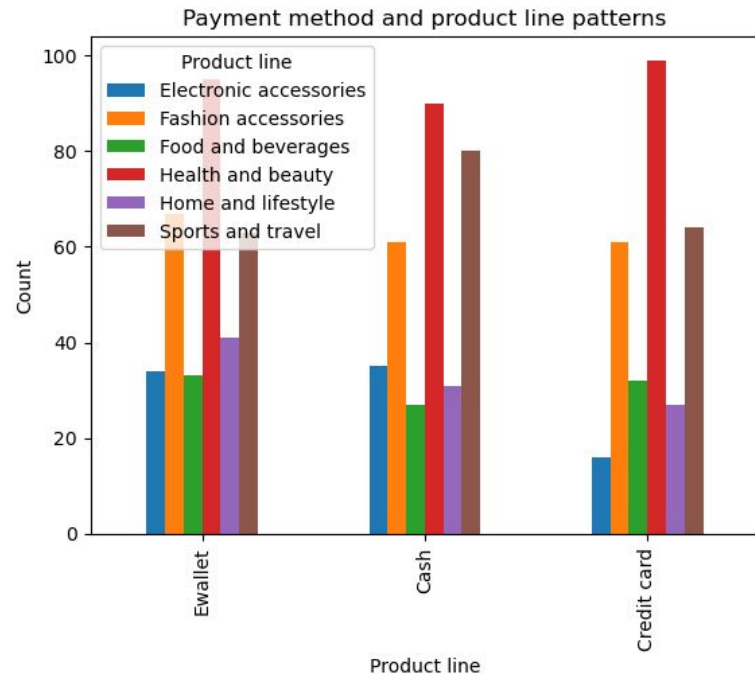
06

Process:

- Used the groupby function to count the occurrences of each payment method and product line combinations
- Sorted them by the total count of each payment method
- Plotted the sorted counts using a bar plot.

Are there any spending patterns based on the payment method used?

- Health and beauty is the most popular product line
- For Sports and travel, Cash is more frequently used
- People prefer to use Ewallet for Home and lifestyle accessories
- Credit card is by far the least popular payment method for electronic accessories





Thank you.